# NEW MEASURES OF INDIVIDUAL AND GROUP EXPERIENCE: APPLICATIONS OF SOCIAL NETWORK THEORY

*Paul Beckman, Department of Information Systems, College of Business, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, 415-338-6240, pbeckman@sfsu.edu*

## ABSTRACT

Conventional wisdom and current calculation methods state that the most experienced employee is the one who has worked in the organization the longest, that is, the one who has the longest career measured in days, months, or years. However, a new mathematical insight explained herein proposes a novel and much more subtle method of determining the experience level of an employee or groups of employees within an organization. This new method also has the ability to precisely define a much more relevant level of experience of workers or groups of workers in almost any organization.

## INTRODUCTION

This article documents an entirely new approach to measuring and analyzing individual and group experience in organizations using “social network analysis”, a subcategory of the branch mathematics called graph theory. In general, the methodology examines past workplace connections between individuals and analyzes those connections to gain insights into a logical measure of experience for those individuals or groups of individuals. The concept of social network analysis is not new and has been applied in numerous domains. However, it has never been used as a means of learning more about, or describing in more detail, individual or group “experience”. The reason we propose this new measure is because of the concept that as individuals, we become better at particular activities by interacting in those activities with others who have better tactics, techniques, and procedures than us and therefore perform better than us. As we interact with and learn from those who are better performers than us, we improve our own tactics, techniques, and procedures.

In this article, we will describe the analytical process, how it applies to one particular domain, U.S. professional sports (in the sub-domain of Major League Baseball), and provide an example of its application. The example is fully described and mathematically demonstrated and will show how our methodology can be used to learn more about an individual’s level of experience than current simple time-based measures of experience. We then propose but do not calculate a closely related process for calculating the experience level of a group of individuals.

### Social Network Analysis

Social Network Analysis (SNA) resides within a branch of mathematics more formally known as “graph theory”. Mathematically speaking, a graph is a set of nodes or points that are pairwise connected by links called “edges.” Edges can be directed (meaning that the link connecting two nodes is “from” one node “to” the other) or undirected (meaning that the link connecting two nodes has no sense of “from” or “to”). If any edge in the entire graph is directed, the graph is considered to be directed. It is also possible to measure edge “weight” or “strength”, wherein the edge that connects two nodes that are connected only one time would have an edge weight/strength of “1”. If those two nodes were connected $N$ times, the edge between the two nodes would have a weight/strength of $N$ or a value resulting from some mathematical calculation based on *N*. The weight of an edge could also be calculated based on some other process or characteristic such as the physical distance between the connected nodes if the graph were modeling a geographic system.

If an environment can be approximated by a mathematical graph, it is then possible not only to measure the number of nodes, edges, and edge weights, but also to calculate various values associated with the nodes, edges, or the graph as a whole. For example, it is possible to calculate various measures of "importance" of a single node in the larger network, depending on how the node is connected to other nodes. Graph theorists generally call such values measures of "centrality", and consider these centrality measures as representations of the importance of a node in the network. Most of these centrality measures are calculated in similar ways, and the most common centrality measures are mathematically highly correlated (Valente, et al. [11]). The four most common centrality measures are: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality, and each requires a slightly different process to calculate. Similarly, it is possible to calculate measures of the importance of an edge. For example, an edge that singularly connects two large subnetworks in a larger network is important as it is the only means by which the two subnetworks are linked. Finally, it is possible to calculate values of the network as a whole. One such measure is the "network density", which indicates how many edges actually exist in a network as compared to how many edges could possibly exist in the network. Network density is therefore a ratio that gives a sense of the overall level of connectedness of a network.

**Social Network Analysis in Organizational Theory**

One of the first thorough descriptions of the application of social network analysis to organizational theory is from Tichy at al. [9]. Their article introduces the concepts of social network analysis and their use in organizational environments. They argue that SNA is an appropriate tool for understanding static and dynamic aspects of organizations as it allows a precisely-defined way of viewing organizational objects over time. They trace the history of both network analysis and organizational theory and provide a conceptual framework for the intersection of the two fields. They go on to describe in more detail the mechanisms for describing social networks, methods of data collection, and processes for data analysis. The authors also view a previously-examined case study under the lens of SNA. They then graphically contrast the social networks of two organizations to show insights that can be derived using SNA theory and analytical processes.

In more recent research applying SNA to individual roles in organizations, Burt [3] found that, in an experiment of individuals in an online virtual gaming world, people who built closed subnetworks when playing one role were more likely to build closed subnetworks when playing another role. Conversely, individuals who build "open" subnetworks (those with access to structural "holes" in the larger network), were more likely to do so across roles. He found, however, that performance (therein defined as "achievement"), is determined primarily by the individual's experience in the task plus the network they decide to build for that particular role. In particular, people who build networks that bridge separate subnetworks act as "brokers" between those subnetworks, leading to more power for those individuals and hence greater success and performance than those who build closed networks.

In another experiment examining individuals, networks, and organizations, Gloor and Colladon [5] measure three dimensions of social interaction (network structure, changes in network structure over time, and degree of sharing) to learn more about the "consciousness" of groups within an organization. Using data collected about these structural, temporal, and content dimensions, they derive six "signals" that predict organizational creativity and performance. These six signals are: 1) group betweenness centrality, 2) variance in contribution, 3) rotating leadership, 4) response speed, 5) “honest sentiment”, and 6) use of innovative language. Each of these measures was quantified and then used to show that the more “emotional”, responsive, and less hierarchical is a business unit, the more successful it is.

Also investigating the relationship between individuals, networks, and performance, Fang et al. [4] completed a meta-analysis of 138 past research documents. Specifically, they examined the relationship between personality and both indegree centrality (the number of directed edges that point “to” a node) and brokerage (the measure of an individual’s ability to occupy a role in a subnetwork that exists between other subnetworks). As it has been shown that indegree centrality and brokerage have a positive impact on an individual’s performance and success, these authors wanted to find the additional moderating impact on success of “self-monitoring” (the ability of an individual to recognize and act on an examination of their own interaction with others). They found that indegree was more strongly related to performance and success than was brokerage. They also found that personality (as measured by the Big Five: extraversion, conscientiousness, openness to experience, agreeableness, and neuroticism) was a better predictor of performance and success than was network position.

Finally, Brands et al. [2] examined the role of gender and network structure (centralized vs. cohesive) on perceptions of leadership. In particular, they studied the concept that in most environments men are generally perceived to be more charismatic leaders than are women. They studied this concept by examining the structure of the networks in which those leaders operate. They found that in networks that are more centralized (wherein participants in the network are expected to go to one or a small number of individuals for advice) male leaders were seen as more charismatic than female leaders. Conversely, they found that in networks that are more cohesive (wherein individuals are expected to go to many others for advice) female leaders were seen as more charismatic than male leaders. This confirmed their supposition that network structure (centralized vs. cohesive) plays an important role in network players’ perception of the leader of the organizational unit.

These and other research projects show that there is significant interest in the role that organizational network structures play on individual and organization unit performance/success. There is also obviously interest in the role of organizational networks on such abstract concepts as “organization consciousness” and “individual charisma”. Our goal was to extend this body of knowledge by directly examining the network connections of individuals over times using data from the field of organized sports. The abundance of data in the domain of organized sports allows for precise measures of network connections that can be extrapolated to other types of organizations and organizational units. In this document, we begin this process by proposing new measures of individual and group “experience” based on measures of network connectivity of individuals in the organization.

**Social Network Analysis in Organized Sports**

Researchers have employed SNA directly to sports in a variety of applications. For example, Piette et al. [7] applied SNA to four seasons of the U.S. National Basketball Association player data to create a weighted network of players as nodes and links as connections between players who played together as a five-man unit. The authors used that network to calculate the statistical “contribution” of each player using the graph theory measure called eigenvector centrality and then determined which players under- or over-performed on offense and defense.

Beckman and Chi [1] used SNA to map into a graph U.S. Major League Baseball players and their connections as team-mates. They used the resulting graph to calculate measures of player centrality and found positive correlations (varying from +0.25 to +0.48) between those centrality measures and offensive performance (BA, HR, RBI, SLG). They did not find a correlation between player centrality and defensive performance, which they hypothesized was due to the ability of MLB players to learn offensive batting tactics from all other players but to learn defensive fielding tactics only from similar position players.

Reifman [8] examined videotapes of a subset of the NCAA men's and women's 2003 basketball season and mapped ball-passing (edges) between players (nodes) into a graph. He then proposed analyzing the graph to determine if there was a difference between teams passing distribution, believing that more complex offenses would be represented by more complex sets of edges.

Grund [6] analyzed almost 300,000 passes between soccer players over 760 English Premier League matches to investigate whether more passing between more players was associated with higher team performance as measured by "goals scored." The author found that higher passing rates were associated with more goals scored and that more centralized passing (passing between fewer players) was associated with fewer goals scored.

Finally, Skinner [9] also used graph theory in organized sports by mapping basketball players as nodes and passes as edges and applying network traffic theory to come to an interesting conclusion. He analogized basketball possessions as vehicular routes on roads to suggest that it is not always the "best" play to choose the highest probability path to a shot, given the complexity of actions that others might be taking (either on other roads in the traffic domain or on the court in the basketball domain).

**Calculating Individual Experience**

Given these descriptions of past applications of SNA to sports, the focus on baseball in the case of this research project will view U.S. Major League Baseball players as nodes in the graph while edges are created when two players appear together on the same team roster in the same year. For example, Rod Carew and Harmon Killebrew both first played together for the Minnesota Twins in 1967 when Carew first signed with the Twins as a rookie (Killebrew had been playing the major leagues since 1954 and had joined the Twins in 1961). Both players appear in the network of all MLB players because they each signed professional MLB contracts. An edge appears in the network of all connected MLB players because both Carew and Killebrew played together on the same team in the same year.

It is possible in this application of SNA to measure the strength of a link, as it is possible for two players to play on the same team together over many years (in fact, the players could play together on different teams across or even within years). The Carew:Killebrew edge strength is therefore "8", using a value of "1" for each year on which they both appeared together on the same team roster. This is so because both players played for the Minnesota Twins together from 1967 through 1974, after which Killebrew was traded to the Kansas City Royals. Carew continued playing with the Twins through 1978 after which he was traded to the California Angels where he ended his career in 1985. This is one example of the millions of pairwise player dyads that were used to construct a complete node-and-edge graph of all MLB players, their team-mates, and their teams.

We examined team rosters for each year of each team's existence and counted as a connection each (player1, player2) dyad for all players on the roster. We did not count as links those connections from a player to themselves, so-called "self-links," because these are spurious links in our process of measuring experience as "connections to others". We also removed so-called "reverse links" which are those links for "player1" of the form (player2, player1) where we counted as a link the player dyad (player1, player2). Simple adjustments to a database query can avoid counting these "self-links" and "reverse links."

Given this description of the research project, the connection-based experience level of an individual player at a specific point in time is calculated as the sum of the number of connections that player has made to other players (teammates) up through that point in time. The underlying premise here is that an individual will increase their experience level (and therefore their value to the organization) by associating with a greater number of other individuals. Their value to the organization will increase because, as they increase their number of connections to other individuals, they increase the probability that they will connect with another individual who has a tactic, technique, or procedure that they do not yet know or have. They will then be able to incorporate that tactic, technique, or procedure into their own skillset and improve their performance thus increasing their value to the organization. In our example of baseball players and their teammates, all other things being equal, a player will have increasing value to their team as they increase the number of connections they make to other players.

### Calculating Team Experience

In this section, we describe the simple process by which SNA can be used to measure the experience level of a team. Extending the calculation of connection-based experience from players to teams is straightforward, as the connection-based experience level of a team at some point in time is just the sum, up to that point in time, of the current cumulative connection-based experience levels of each of its players. Therefore, the total player-contributed experience of a team in a particular year is calculated by summing across all players on the roster of that team, the cumulative connection-based experience of each player up to that point in time. This is only slightly more complicated of a computation than that of the connection-based experience level of a player.

Note that we should not count as experience for a team, the final career total of a player's experience. This is so because a player can only contribute to a team the experience they have accumulated up to that point in time. For example, to calculate the connection-based experience level of the 1969 "Amazin' Mets" as a team requires that we examine the roster of that team, calculate the connection-based experience level (up through 1968) of each player on that team, and sum those experience counts across the roster players.

## METHODOLOGY

The process by which this new SNA measure of individual and team experience was implemented was to 1) collect data for all MLB teams, players, and rosters, and 2) run database queries to calculate the running total of the connection-based experience through each year of a player's career.

### Experience Calculation 1: By Career Years

The current standard measure of a professional athlete's experience is to total the number of years in which the player had a signed contract with a team that is a member of the league(s) in which that player has played. A slightly more precise measure, and one almost as easy to calculate, but following the past framework of "experience = time", is the total number of games in which a player appeared on a team roster. For the purposes of this project, we calculated "time experience" only as "career years." While the "career length in years" value was calculated for all players since the inception of U.S. professional baseball in 1871, we only present (in Table 1, below) the top 9 players with the longest careers (and not the top 10 because of the many players tied at position #10 with 24 years of experience), measured as "years in which the player had a signed contract with a MLB team". In Table 2 (below), we calculate player experience using our new SNA connection-based process, and display the top 10 most-experienced players.

**TABLE 1.**

| Rank | Name | Career Length (In Years) |
|---|---|---|
| 1 (tied) | Cap Anson | 27 |
| 1 (tied) | Nolan Ryan | 27 |
| 3 (tied) | Tommy John | 26 |
| 3 (tied) | Deacon McGuire | 26 |
| 5 (tied) | Eddie Collins | 25 |
| 5 (tied) | Rickey Henderson | 25 |
| 5 (tied) | Charlie Hough | 25 |
| 5 (tied) | Jim Kaat | 25 |
| 5 (tied) | Bobby Wallace | 25 |

**Top 9 Most Experienced MLB Players by "Career Years" (Descending)**

**Experience Calculation 2: By Connections to Other Players**

The result of our database manipulations shows in Table 2 the top 10 most-experienced players by the total number of connections to other players throughout each player's career. As described above, we did count the possible multiple links from one player to another over time. That is, if two players played together on the same (or different) teams across more than one year, as in the case described above of Rod Carew and Harmon Killebrew, each player was awarded one link for each year on which the two appeared together on a team's roster. In this case, Mr. Carew and Mr. Killebrew each were awarded 8 links from each other because they appeared on the same MLB team roster in 8 different years. The result of this process shows in Table 2 (below) the top 10 most-experienced baseball players of all time where experience is calculated as "all connections to all other players over all years."

**TABLE 2.**

| Rank | Player | All Connections to All Other Players |
|---|---|---|
| 1 | Rickey Henderson | 1197 |
| 2 | David Weathers | 1167 |
| 3 | Terry Mulholland | 1146 |
| 4 | Harold Baines | 1109 |
| 5 | Ruben Sierra | 1091 |
| 6 | Tommy John | 1089 |
| 7 | Jamie Moyer | 1081 |
| 8 | Jesse Orosco | 1068 |
| 9 | Mike Stanton | 1064 |
| 10 | Ken Griffey Jr. | 1048 |

**Top 10 Most Experienced MLB Players by "Connections to Other Players" (Descending)**

## CONCLUSIONS

This research project was performed to show that SNA can be applied to professional sports (or to any situation wherein individuals work together to complete a series of tasks) in a way that proposes a completely new method of calculating the experience of an individual or of a workgroup. The application of SNA to Major League Baseball was described, the calculation process explained, and analytical results provided for a precise measure of connection-based player experience. The process to extend player-based experience to team-based experience was also explained although not calculated.

There are numerous interesting conclusions that can be drawn simply by examining the difference in the two "Top *N* Most Experienced Players" lists provided in this article. For example, it is interesting to note that using the current time-based measure of player experience, players with the most experience are spread across the history of baseball, from Cap Anson, who began playing MLB in 1871, to Rickey Henderson, who began playing MLB in 1979. When we use our connection-based measure of experience, however, we note that the "Top 10 Most Experienced Players" list is populated by relatively recent players. This implies that, while the length of players' careers may be comparable across the history of baseball, a higher rate of player movement among teams, causing an increased level of player connections, is concentrated in the recent history of baseball.

This result by itself would be interesting to know for many organizations. That is, almost all current organizations of any size track and therefore could know something about the distribution across time of the time-based experience level of their employees. However, they are very unlikely to know, much less track across time, our connection-based measure of experience. Therefore, those organizations would not know if there is a difference between the levels of their employees' experience using these two different measures. Those organizations would then not know if there was a difference between their employees' experience level measured by time versus their experience level measured by connections. Such an organization may think that the level of its "time-based" most-experienced workers has been very constant over time, while the level of its "connection-based" most experienced workers has changed drastically over time. If workers are staying in the organization for generally the same number of years, but are making many more (or less) connections in their careers, this would very likely be a cause of concern. Without measuring its workers connections over time, this result is unknowable.

There are obviously many different directions that can be examined using connection-based experience that are not possible with the current simple time-based measure of experience. As the methodology explained in this article is completely novel, our intent was not to enumerate these directions. Our goal was to suggest that our connection-based process might yield insights into new ways to examine and measure work experience and ultimately its relationship to performance in Major League Baseball as well as any other domain in which individuals come together to complete group-based tasks in which participants might learn from each other and add to their own level of experience. It should be obvious that the application of our connection-based analytical methodology to the corporate world will yield insights into worker experience that are more complex and insightful than simple time-based measures.

## REFERENCES


[1] Beckman, P. and Chi, J. "More Highly Connected Baseball Players Have Better Offensive Performance", *Baseball Research Journal*, Fall, 2011.

[2] Brands, R., Menges, J., and Kilduff, M. "The Leader-in-Social-Network Schema: Perceptions of Network Structure Affect Gendered Attributions of Charisma", *Organization Science*, 2015.

[3] Burt, R. "Network-related personality and the agency question: Multi-role evidence from a virtual world", *American Journal of Sociology*, Vol. 118, pp. 543–591, 2012.

[4] Fang, R., Landis, B., Zhang, Z., Anderson, M., Shaw, J., and Kilduff, M. "Integrating Personality and Social Networks: A Meta-Analysis of Personality, Network Position, and Work Outcomes in Organizations", *Organization Science*, 26(4):1243-1260, 2015.

[5] Gloor, P., and Colladon, A. "Measuring Organizational Consciousness Through E-Mail Based Social Network Analysis", *arXiv preprint* arXiv:1502.04997, 2015.

[6] Grund, T. "Network Structure and Team Performance: The Case of English Premier League Soccer Teams." *Social Networks* 34.4 (2012): 682-690.

[7] Piette, J., Pham, L., and Anand, S. "Evaluating Basketball Player Performance via Statistical Network Modeling", *MIT Sloan Sports Analytics Conference*, 2011, March 4-5, 2011, Boston, MA, USA.

[8] Reifman, A. "Network Analysis of Basketball Passing Patterns II", *NetSci2006*, May 22-25, 2006, Bloomington, IN, USA.

[9] Skinner, B. "The Price of Anarchy in Basketball." *Journal of Quantitative Analysis in Sports*, 6.1 (2010).

[10] Tichy, N., Tushman, M., and Fombrun, C. "Social network analysis for organizations", *Academy of Management Review 4*, no. 4, pp. 507-519, 1979.

[11] Valente, Thomas W., et al. "How Correlated are Network Centrality Measures?" *Connections* (Toronto, Ont.) 28.1 (2008): 16.