Big Data Meets Team Expertise in a Dynamic Task Environment

Matthew-Donald D. Sangster <sangsm@rpi.edu>
Cognitive Science Department
Rensselaer Polytechnic Institute

David J. Mendonca <mendod@rpi.edu>
Industrial and Systems Engineering Department
Rensselaer Polytechnic Institute

Wayne D. Gray <grayw@rpi.edu>
Cognitive Science Department
Rensselaer Polytechnic Institute

Objective; This research employs large-scale data from a massively multiplayer online game to examine the links between the composition, processes and outcomes of teams operating in high tempo, data-rich environments. **Background:** Research on the performance of teams–particularly over long time scales–is often expensive and time-consuming. But Big Data from competitive, team-based games can mitigate these costs. **Methods:** Data visualization techniques are used to explore team data harvested from publicly accessible sources for the online game *League of Legends*™, one of the most popular such games in the world. **Results:** The exploratory results suggest potentially complex relationships between team composition, processes and outcomes, and in particular how team composition and process may unfold over longer time spans. **Conclusions:** The results point to the potentially substantial benefits of large-scale studies of teamwork, and–in parallel–to the need for the development of tools, techniques and measures to bring Big Data to bear in teamwork studies. **Application:** This work demonstrates the feasibility of exploring online gaming data for new insights into team and individual performance.

Introduction

A considerable obstacle to many studies of teamwork is the collection of precise and robust data on the composition, processes and outcomes of teamwork over time. Field studies in general are costly and time-consuming, while laboratory studies may rely on tasks that do not engage expertise over longer time scales. Online gaming combines field research with unobstrusive, real-time collection of massive amounts of individual, team, and environmental data. Many game companies currently collect these types of data from their games for their own purposes, and some provide public access to their databases through public Application Program Interfaces (APIs). These data, due to their inherent size, open the possibility for rich visualization of the phenomena under study and, in the longer term, for quasi-experimental research designs that provide much greater statistical power than traditional experiments (Shadish, Cook, & Campbell, 2002).

Background

Theory and observation suggest that team performance is driven, in part, by team familiarity (Mendonça, Brooks, & Grabowski, 2014; Contractor, 2013): that is, by the experience of team members in working with each other. Team familiarity can be viewed (a) as a proxy for the knowledge that team members have about each other and (b) as the knowledge a team member has about the task facing the team (Littlepage, Robison, & Reddington, 1997; Espinosa, Slaughter, Kraut, & Herbsleb, 2007).

Various researchers have argued for collecting data that attempt to capture "moment to moment interteam and intrateam interaction" on the composition, processes and outcomes of teams (Salas, Stagl, & Burke, 2004), including how these interactions unfold over longer time scales (Dickinson & McIntyre, 1997). Yet the study of teamwork "in the wild" (Hutchins, 1995) – that is, of teams of professionals performing work *in situ* – presents a number of challenges, including access to public, sufficiently detailed data and the complexity of field settings (Guzzo & Dickson, 1996; Salas, Burke, & Stagl, 2004).

Very recently, technology and culture have combined in ways that are favorable to developing new approaches to the study of teams. For example, millions of teams are currently (and continually) engaged in playing in the Multiplayer Online Battle Arena (MOBA) known as *League of Legends*™ (LoL). Just as professional athletic sports are shadowed by thousands and tens of thousands of college teams, high school teams, local teams, and groups of friends at varying levels of expertise, each of the estimated 80 million individual LoL players plays LoL in a team of three or five persons. With over 1 billion hours of LoL play logged per month in 2012 alone (Kenreck, 2012), many of these teams and players have earned the title "expert."

LoL lends itself well to the study of teamwork: (a) it is a team-based game with high demands for coordinated action across team members; (b) it is highly instrumented, with detailed records kept on many aspects of performance (in fact, detail is sufficient to enable complete post-match playback of the matches); (c) its view of performance is multi-faceted, with many explicit measures both at process and outcome levels; and (d) it enables various measures of team composition to be extracted or derived from match records, such as the working history of team members.

Some work has been done exploring traditional team constructs using Big Data from MOBAs, typically focusing on investigating team performance. For example, Leavitt, Keegan, and Clark (2016) show a link between the percentage of team members who are friends and the team's performance. While providing additional evidence for that link, Pobiedina, Neidhardt, del Carmen Calatrava Moreno, Grad-Gyenge, and Werthner (2013) also found evidence suggesting that individual decisions about player roles impact team performance. However, both of these findings treat familiarity as a binary variable and do not address varying degrees of familiarity among team members.

This paper investigates the relationship between team composition and team performance using LoL data on hundreds of teams. Team composition, referring specifically to the characteristics of the team members before the match, is assessed through the familiarity of the team members prior to the match (as a non-binary variable), while team performance is measured through both kills/assists (acting as a process variable) and whether the team won or lost (acting as an outcome variable). This paper also illustrates the potential benefits and challenges of working with highly instrumented teams in fast-paced environments.

Data Preparation

A subset of data on LoL gameplay are freely available through a public Application Program Interface (API) provided by the developer. Teams are segregated in the data set by tier: a developer-determined classification of teams that corresponds roughly to a league or division in a traditional sporting context. This study primarily uses data from the "master" and "challenger" tiers, corresponding to an upper-range league.

In LoL, two teams of equal size compete to destroy each other's *nexus*, a building which sits behind fortifications and computer-controlled defenders. To accomplish this goal, team members must attack and kill (or otherwise conquer) opponents. Each team member is controlled by a single human controller (i.e., the player). LoL play is logged extensively: all *events* (such as behaviors executed by members) are logged on execution, while *situational variables* (such as displays of the field of battle) are logged continuously at 1 min rates. Initial conditions of the game—particularly in terms of exogenous factors such as mode of play and configuration of the battle landscape—can vary. Thus, while the setting for gameplay is exceedingly rich in detail and dynamics, extensive 'instrumentation' of the setting yields data that are well suited to investigating the relationship of team performance to team familiarity.

The study sample consists of data on all teams in the sampled tiers who played at least fifteen matches. In total, 441 teams (and therefore a total of 441*15=6615 matches) are included. As each game includes two teams, for convenience these 441 teams who have played at least 15 matches, are noted the "home" teams in the discussion that follows. There are no matches in the data set where both competing teams would be considered "home" teams.

Each record in the database consists of team- and player-level identifiers along with various team-level measures. A stylized representation of the data layout of the sample is given in Table 1 for a single match between two three-member teams. The source database provides unique identifiers for the home and away teams (*HomeTeam ID* and *AwayTeam ID*, respectively), as well as identifiers (*Player ID*) for their corresponding members (represented here as a vector of *m* items). Associated with each match is a vector of *MatchAttributes*. Associated with each team are team-level attributes specific to that team (*HomeAttributes* and *AwayAttributes*), consisting of kills, deaths, by member. Match histories are then used to construct pairwise working histories for players, as well as to track the number of players on each team for each match. While many other attributes are available via the API, these are excluded for clarity of presentation as they are not included in this analysis. Match ID and HomeTeam ID are taken from the *Team* API; all other attributes are taken from the *Match* API.

	Match ID	MatchAttributes			
Match Data	100	Timestamp:	146031642		
		HomeWin:	True		
		HGameNumber:	1		
		MapId:	15		
		MatchType:	3v3 Ranked		
	HomeTeam ID	HPlayer IDs	HomeAttributes		
Home Team Data	TEAM-1		Kills:	Deaths	Assists
		1: 1000	8	3	1
		2: 1002	6	2	4
		3: 1005	9	3	2
	AwayTeam ID	APlayer IDs	AwayAttributes		
Away Team Data	TEAM-2		Kills:	Deaths	Assists
		1: 1006	4	5	1
		2: 1008	3	7	2
		3: 1010	1	11	0

Table 1
A stylized data table for a single match. The top section (Match Data) includes information at the match level; the middle section information for the Home team; and the bottom section information for the Away team. Team size and familiarity are calculated based on the Player IDs.

Results

Initial exploration of the LoL data has focused on using data visualizations to explore relationships amongst study variables associated with team inputs (here, familiarity), processes (kills and assists) and outcomes (wins vs. losses). Study variables are first defined then explored in relation to each other.

Familiarity. A common measure of familiarity (Espinosa et al., 2007) takes the total number of matches (h_{ij}) in which each pair of team members (i, j) of team n has played together, and divides this by the number of possible pairs on the team. This score is computed at the onset of the match. Thus, the familiarity score for team k in match l is:

$$f_{kl} = \frac{2}{n(n-1)}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} h_{ij}\right) \tag{1}$$

The value of the measure is always zero for the first match played by the team. Assuming that the team never turns over completely, the measure increases with the number of matches played by the team. The measure may be standardized to a (0,1) scale by dividing it by the total number of matches l since the beginning of the observational period. Through standardizing the measure, familiarity is no longer monotonically increasing with the number of games played and thereby demonstrates the rate of familiarity accumulation rather than a cumulative evaluation of familiarity. This difference can be likened to evaluating the speed (distance over time) traveled at a given point in time rather than just the total distance traveled at a given point in time—representing standardized familiarity and the original familiarity measure respectively.

Figure 1, in the upper left box, depicts the values of the standardized familiarity score over 15 matches for winning vs. losing teams in the sample. A limitation of the pilot data is that familiarity is assumed to be zero at the onset of first match, as data on working histories that predate Match 1 are not included.

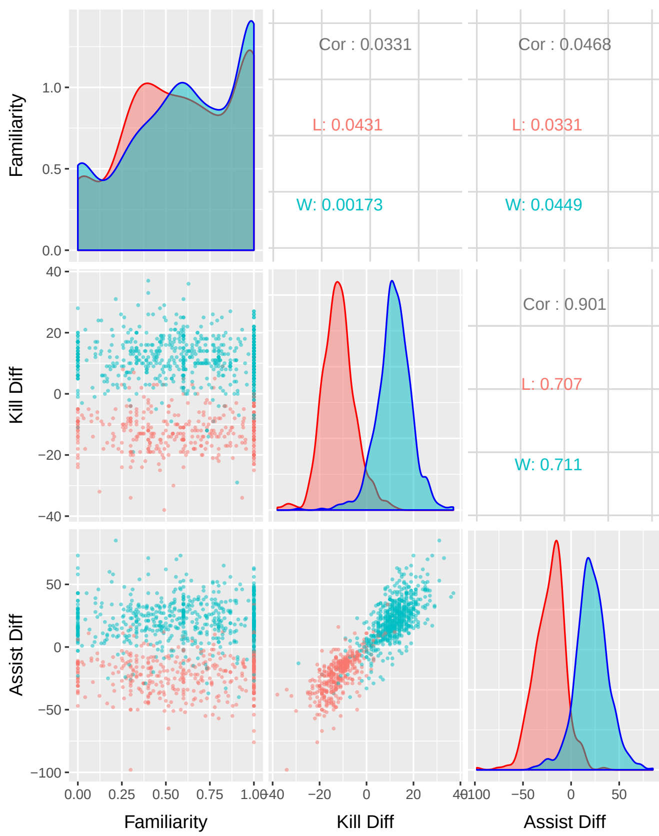

Figure 1. Correlation table for the input variable (Familiarity) and the two process variables (Kill Differential and Assist Differential). The Blue represents teams that won their game and the Red represents teams that lost their game. The scatterplots are constructed from a random sample of 1000 matches from out total sample of 6615 in order to reduce the visual clutter and make individual points more distinct. Correlation values are calculated using the complete data set, and are presented in the upper right boxes. The first value is the correlation coefficient for the two variables overall (e.g., familiarity with kill differential, using data from both winning and losing teams), the red value for losing teams, and the blue value for winning teams.

Process Variables

The remainder of Figure 1 refers to team processes and outcomes, and are followed by a presentation of one process-level measure (kills) as a function of familiarity. As mentioned, the primary objective of LoL is to destroy the enemy's Nexus, thus winning the match. A player on one team is awarded a 'kill' after dealing fatal damage to an opponent. Killing a player removes that player from combat for a period of time, and produces a reward for the player who dealt the death blow. Any teammate who assisted with the kill in the preceding ten seconds is awarded an 'assist.'. In practice, the rewards for kills and assists vary based on a number of circumstantial factors (e.g., time elapsed in the game, the number of assisting players). Here, all kills are treated equally, as are all assists.

Except where explicitly stated, the analyses are expressed in terms of the home team, h: that is, the set of 441 teams that have played at least 15 matches.

Kill Differential. Kill differential (that is, the difference between the number of kills for the home versus the away team) serves as one measure of performance effectiveness. For a given match m, the kill differential k_m is the difference between the sum of kills across all players ($i = 1, 2, \dots, n_h$) for team h and the sum of kills across all players ($j = 1, 2, \dots, n_a$) for team a. The total count of kills made by one team can exceed the number of players on the opposing team for the reasons described previously. Figure 1, center box, shows the density of Kill Differential for the home teams across the 15 matches.

Assist Differential. Assist differential is calculated using the same approach as with Kill Differential, using assists instead of kills. Figure 1, lower right box, shows the density of Assist Differential for home teams in the sample.

Relation Among Study Variables

Figures 1 and Figure 2 are used in this section to explore the relationship among the input, process and output variables described previously.

Input-Process. The distribution of familiarity is decidedly non-normal, with more higher (>0.50) than lower (<0.50) familiarity teams in the sample. There is some suggestion of a difference in the distribution of familiarity for winning teams (blue) vs. losing team (red) towards the middle of the distribution. The raw correlations–calculated using the complete data set–between familiarity and kills (a process measure) are low, whether overall (0.0331) or considering only winning (0.0431) or losing (0.00173) teams. Similar statements can be made with respect to assist differential. Overall, none of the correlations between the input variable and either of the process variables approaches 0.10.

Process-Output. The process variables kill differential and assist differential are at least symmetrical, though with some overlap, suggesting that these measures may be only imperfect predictors of performance. The measures themselves are fairly highly correlated, both overall and for winning and losing teams.

It should be noted that this lack of perfect agreement between team process quality (here, kills) and outcomes (wins vs. losses) is not exclusive to LoL (Brannick & Prince, 1997). Teams can win and play poorly, or lose after playing a nearly perfect game. Further, it is a peculiarity of this and other game data sets that the methods for calculating a team's so-called *true score* (i.e., its overall rating) are not publicly available. Thus, work such as that presented here must, for the time being, rely on proxy measures for assessing performance through both processes and outcomes.

Input-Output. The data may also be used to examine the relationship between team familiarity and performance over time (putting aside the role of process-level variables). Figure 2 depicts the relationship between composition (expressed through the standardized version of Espinosa's familiarity measure) and outcome (win/loss) for home teams. The plots demonstrate that as a team plays more games, they experience more turnover in team members. Further, and as suggested by the previous discussion, there appears to be little effect of familiarity on winning vs. losing, as the distribution of familiarity is nearly the same for winners and losers in a given match. A caveat is that for the first two plots, a team with no turnover (standardized familiarity score of 1) appears to have a slightly higher chance of winning.

Additional work will be needed in further exploring these data. Figure 2 could be an artifact of the selection criteria, as teams that eventually play 15 games together may form as a result of a previous team disbanding. This is highighted by a relatively high win rate (60.68%) for teams that play 15 matches together. In games 3 and 6 in Figure 2, for example, there is a difference in maximum familiarity between winning vs. losing teams. It may be that teams with higher familiarity are more likely to win than lose, but only for games earlier in the team's history (the matchmaking system for League of Legends tends to match teams having equivalent or nearly equivalent number of matches played). Segmentation of the data may be used in the future to identiify matches where familiarity differs, and thus to explore the role of this factor more comprehensively.

These data, overall, suggest that all three aspects of teamwork (composition, processes, and outcomes) may be closely related and possibly mutually dependent (e.g., greater familiarity may yield either higher or lower kill differentials, and changes in kill differential–perhaps particularly negative ones–may precipitate decisions to change team members, thus reducing familiarity).

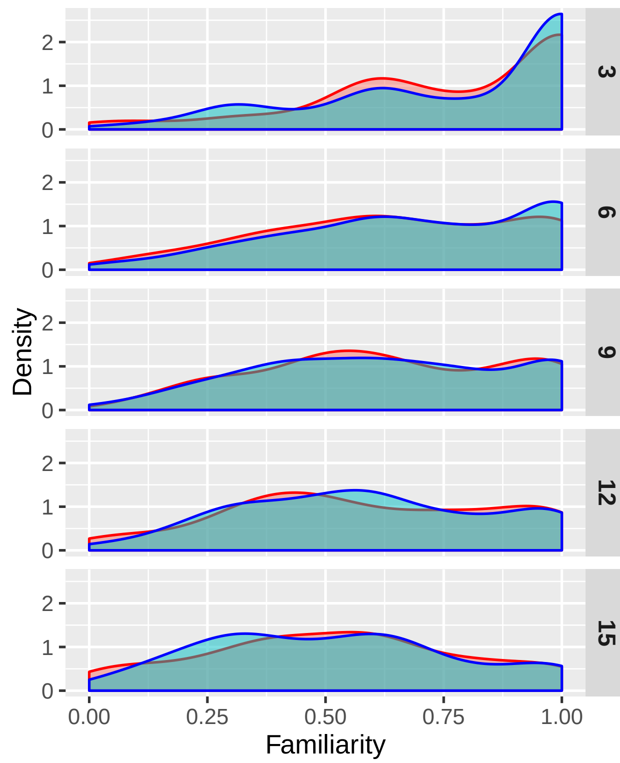

Figure 2. Density of Familiarity for matches 3, 6, 9, 12, and 15 for home teams. Y-axis values represent kernel density estimates whereby the area under the curve between any two points along the x-axis (x_1 and x_2) is an esitmate for the probability of getting a value for x between x_1 and x_2. The green curves represent matches where the home team won while the pink curves represent matches where the home team lost.

Conclusions and Discussion

This exploratory analysis yields several insights into various interrelated aspects of teamwork as evinced by rich and objective data captured from LoL gameplay. It also suggests a number of directions for future research with this data set.

First, this work highlights opportunities for extending and sharpening approaches to measuring team composition, processes and outcomes in the era of Big Data. For example, the major difference between Espinosa's original familiarity measure and the standardized one used here is that standardized familiarity accounts for all instances of pairwise working history among players, irrespective of current team, whereas relative familiarity, in order to be scaled by games played, can only account for instances of pairwise working history within a given team. Each measure of familiarity provides its own benefits for use as a possible predictor variable in future analysis, though due to their inherent assumptions neither is perfect. Moreover, typical approaches (including those used here) to measuring familiarity assume that only working history within the team is salient, while the case may be that working history outside the team (such as in past teams) also effects performance. Future data analyses may be able to address this limitation through inclusion of extra-team working history data. Of course, both approaches to measuring familiarity used here depend on access to data over the lifetime of teams, while the postulated alternative measures will depend on access to data over the lifetime of pairwise working histories–a considerably higher barrier.

Second, the role of *task* familiarity (Espinosa et al., 2007) has thus far been unexplored: that is, the competencies of players either individually or as whole in undertaking the task at hand. This concern touches upon questions of repertoires of team skills, established patterns in games, and the portfolio of skills and knowledge in the team. The inclusion of data on individual skill and experience may enable us to differentiate between a team made up of task novices and a team made up of task experts, for example. Similarly, because the game is competitive, it is also imperative to include opposing team information (e.g., on its inferred rank) to assess the extent to which opponents are evenly matched.

Third and relatedly, there is a clear theoretical argument for examining performance over time, and in particular to discover mechanisms that lead to earlier games impacting performance on later ones, especially if the time frame between the two games is sufficiently small. A host of phenomena that are relatively unexplored in teamwork could be addressed here, including the phenomena of 'winning streaks' and 'losing streaks' and their role in driving team turnover.

Finally, future work is also expected to explore possible reciprocal effects, in which the outcomes of matches drive subsequent decisions about team composition, and thereby subsequent match outcomes. In the extreme, there may be factors from the result of a team's first game that produce a decision not to play a second game. Essentially, team performance for a given game may have an impact on the likelihood of a team remaining together.

This research provides a first step in applying an individual-focused assessment of team performance within the world of big data analytics for MOBAs. League of Legends, and MOBAs, in general offer an untapped wealth of opportunity for team studies. Through using a Big Data approach to a field that is typically muddled down by high cost experiments, we can begin to make substantial leaps in exploring the field.

Acknowledgments

Matthew-Donald Sangster's work on this project was supported by a Rensselaer Presidential Graduate Research Fellowship.

References

Brannick, M. T. & Prince, C. (1997). Team performance assessment and measurement: theory, methods, and applications. In M. T. Brannick, E. Salas, & C. Prince (Eds.), (Chap. An overview of team performance measurement, pp. 3–18). Mahwah, NJ: Erlbaum.

Contractor, N. (2013). Some assembly required: leveraging web science to understand and enable team assembly. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *371*(1987).

Dickinson, T. L. & McIntyre, R. M. (1997). A conceptual framework of teamwork measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (pp. 331–356). Lawrence Erlbaum.

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., & Herbsleb, J. D. (2007). Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science*, *18*(4), 613–630.

Guzzo, R. A. & Dickson, M. W. (1996). Teams in organizations: recent research on performance and effectiveness. *47*, 307–338.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Kenreck, T. (2012). 'league of legends' players log 1 billion hours a month. [Online; accessed 10-April-2016]. Retrieved from http://www.nbcnews.com/tech/video-games/league-legends-players-log-1-billion-hours-month-f1C6423906

Leavitt, A., Keegan, B. C., & Clark, J. (2016). Ping to win? non-verbal communication and team performance in competitive online multiplayer games. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems*. CHI '16.

Littlepage, G., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. *Organizational Behavior and Human Decision Processes*, *69*(2), 133–147.

Mendonça, D., Brooks, J. D., & Grabowski, M. (2014). Linking team composition to team performance: an application to postdisaster debris removal operations. *Transactions on Human- Machine Systems*, *44*(3), 315–325.

Pobiedina, N., Neidhardt, J., del Carmen Calatrava Moreno, M., Grad-Gyenge, L., & Werthner, H. (2013). On successful team formation: statistical analysis of a multiplayer online game. In *2013 IEEE International Conference on Business Informatics* (pp. 55–62). Vienna: IEEE.

Salas, E., Burke, C. S., & Stagl, K. C. (2004). Developing teams and team leaders: strategies and principles. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds.), *Leader development for transforming organizations: Growing leaders for tomorrow* (pp. 325–355). Mahwah, NJ: Erbaum.

Salas, E., Stagl, K. C., & Burke, C. S. (2004). 25 years of team effectiveness in organizations: research themes and emerging needs. In *International Review of Industrial and Organizational Psychology* (Vol. 19, pp. 47–91).

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd). Chicago: Wadsworth.