Experts' Commentary: The Soccer Team Problem

Brian Macdonald
Director of Sports Analytics
Stats & Information Group
ESPN
bmacnhlfin1@gmail.com

Alexander J. Gates
Network Science Institute
Northeastern University
Boston, MA
a.gates@northeastern.edu

Introduction

Data have always been a part of sports. Talk to any die-hard fan, and she'll provide an endless list of team and player statistics. Unfortunately, while box scores summarize bottom-line outcomes in a game, they don't provide enough information or context to explain why those outcomes happened; underdogs or teams with poorer statistics win all of the time. This is especially true for free-flowing team sports such as basketball, soccer, and hockey, which have fewer isolated one-on-one matchups relative to a sport such as baseball. The success of teams in these sports is much more than the sum of the abilities of individual players. Rather, team success is based on the team composition, how well the teammates play together, how leaders and coaches inspire performance, decisions made by referees, and even the weather.

Modern sports analytics takes advantage of recent advances in big data, interdisciplinary modeling, and machine learning [Morgulev 2018]. Sports models not only draw insight into the collective success of teams during games and matches, they provide valuable insight for player drafts and trades, and help to keep players healthy and injury free.

The availability of play-by-play data that include time-stamped events and player substitutions was a big step forward for sports analytics. It al-

The UMAP Journal 41 (3) (2020) 257–260.

lowed for more-advanced analysis of the game by capturing all of the players actively playing at that particular time in the game, as well as player roles in the most important game events. Combined with new statistical tools, play-by-play data permit analysts to find appropriate explanations of what happened in a game or season and make more accurate predictions of future outcomes. The data even allow us to start tackling challenging problems related to player interactions, team chemistry, and collective behaviors of a team and their opponent.

The possibilities for sports analysts are only going to grow. Because of improvements in technology and an increase in the acceptance of analytics, many leagues are gathering data on the locations of all of the players and the ball (or puck) when important events occur, and in some cases, many times per second throughout the entire game. This means that spatial relationships among the players and the ball are known at all times. To take full advantage of these new data, and to fully analyze cooperation among teammates or matchups against opponents in a team sport, new modeling tools and techniques will be required that account for the spatial relationships.

From a mathematical standpoint, network science provides a unique set of tools for sports analytics. Network analysis allows modelers to quantify interactions within a team, extract regular motifs and hierarchies, and characterize the role of individual players in the collective context. Extended over time, networks become a powerful record of how the game has evolved.

In soccer, passing networks have been particularly useful to capture common patterns in the way players from the same team move the ball among each other. For example, F.C. Barcelona's infamous tiki-taka style of play reveals itself through short, rapid passes between closely-spaced players, resulting in dense network structures [Buldú et al. 2019].

Formulation and Intent of the Problem

This year's ICM problem challenged student teams to quantitatively capture the collective interactions among soccer players and model how changes in team behavior translated towards soccer team success. This problem was further complicated by asking students to clean and interpret a real-world dataset, taken from the 2017 Champions League.

Student modelers were expected to build player passing networks and characterize their structure. Outstanding student teams found additional variations of this setup: building passing networks between different regions of the field, changing the time-scales over which passes were aggregated, and even including other related player events in the network. The resulting statistical models integrated individual player characteristics with collective team performance measures to provide policy recommendations to enhance the team's play.

Student modeling solutions were evaluated based on the team's understanding of the nature of the problem, their creativity in quantifying collective behavior, and their approach for communicating their proposed solution. While knowledge of soccer team formations and professional strategies helped to guide some teams, many creative solutions did not rely on such familiarity with the game.

Solving the Problem

Many ICM teams moved beyond an understanding of game outcome as win / loss / tie, and proposed graduated measures of team success such as goal differential and ball possession time. The very best solutions leveraged information over multiple time scales, finding coherent patterns over the course of minutes, game halves, games, and the entire season. We were particularly impressed by ICM solutions that captured strategies of both the Huskies and their opponents.

Sports analytics and computational social science are continuously evolving and producing interdisciplinary problems that need modeling and analysis. Every team took on a highly-complex and impactful modeling problem with no "right answer." Combined with the unexpected difficulties posed by a global health crisis, this year's ICM participants are to be commended! With the expanding nature of sports modeling, it is very likely some of this year's contestants will model sports teams and events in the future.

References

Buldú, J.M., J. Busquets, I. Echegoyen, and F. Seirul·lo. 2019. Defining a historic football team: Using network science to analyze Guardiola's F.C. Barcelona. *Scientific Reports* 9: article 13602. https://www.nature.com/articles/s41598-019-49969-2?sf221670398=1.

Morgulev, Elia, Ofer H. Azar, and Ronnie Lidor. 2018. Sports analytics and the big-data era. *International Journal of Data Science and Analytics* 5: 213–222. https://doi.org/10.1007/s41060-017-0093-7.

Santos-Fernandez, Edgar, Paul Wu, and Kerrie L. Mengersen. 2019. Bayesian statistics meets sports: A comprehensive review. *Journal of Quantitative Analysis in Sports* 15 (4): 289–312. https://doi.org/10.1515/jqas-2018-0106.

About the Authors

Brian Macdonald is Director of Sports Analytics in the Stats & Information Group at ESPN. He was previously the Director of Hockey Analytics with the Florida Panthers Hockey Club, and an Associate Professor in the Dept. of Mathematical Sciences at the US Military Academy at West Point. He received a B.S. in Electrical Engineering from Lafayette College, Easton, PA, and a Master of Arts and a Ph.D. in Mathematics from Johns Hopkins University, Baltimore, MD.

Alexander J. Gates is an associate research scientist in the Network Science Institute at Northeastern University. His research explores the science of success in human behavior. Alex holds a joint Ph.D. in Informatics (Complex Systems) and Cognitive Science from Indiana University, Bloomington, and an M.Sc. in Complex Systems Modeling from King's College, London.