Using Network Science to Analyse Football Passing Networks: Dynamics, Space, Time, and the Multilayer Nature of the Game

Javier M. Buldú [1,2,3]*, Javier Busquets [4], Johann H. Martínez [3,5], José L. Herrera-Diestra [6], Ignacio Echegoyen [1,2,3], Javier Galeano [7] and Jordi Luque [8]

[1] Laboratory of Biological Networks, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain, [2] Complex Systems Group and GISC, Universidad Rey Juan Carlos, Móstoles, Spain, [3] Grupo Interdisciplinar de Sistemas Complejos, Madrid, Spain, [4] ESADE Business School, Barcelona, Spain, [5] INSERM-UM1127, Institute du Cerveau et de la Moelle Épinière. H. Salpêtrière, Paris, France, [6] ICTP South American Institute for Fundamental Research, IFT-UNESP, São Paulo, Brazil, [7] Grupo de Sistemas Complejos, Universidad Politécnica de Madrid, Madrid, Spain, [8] Telefónica Research, Barcelona, Spain
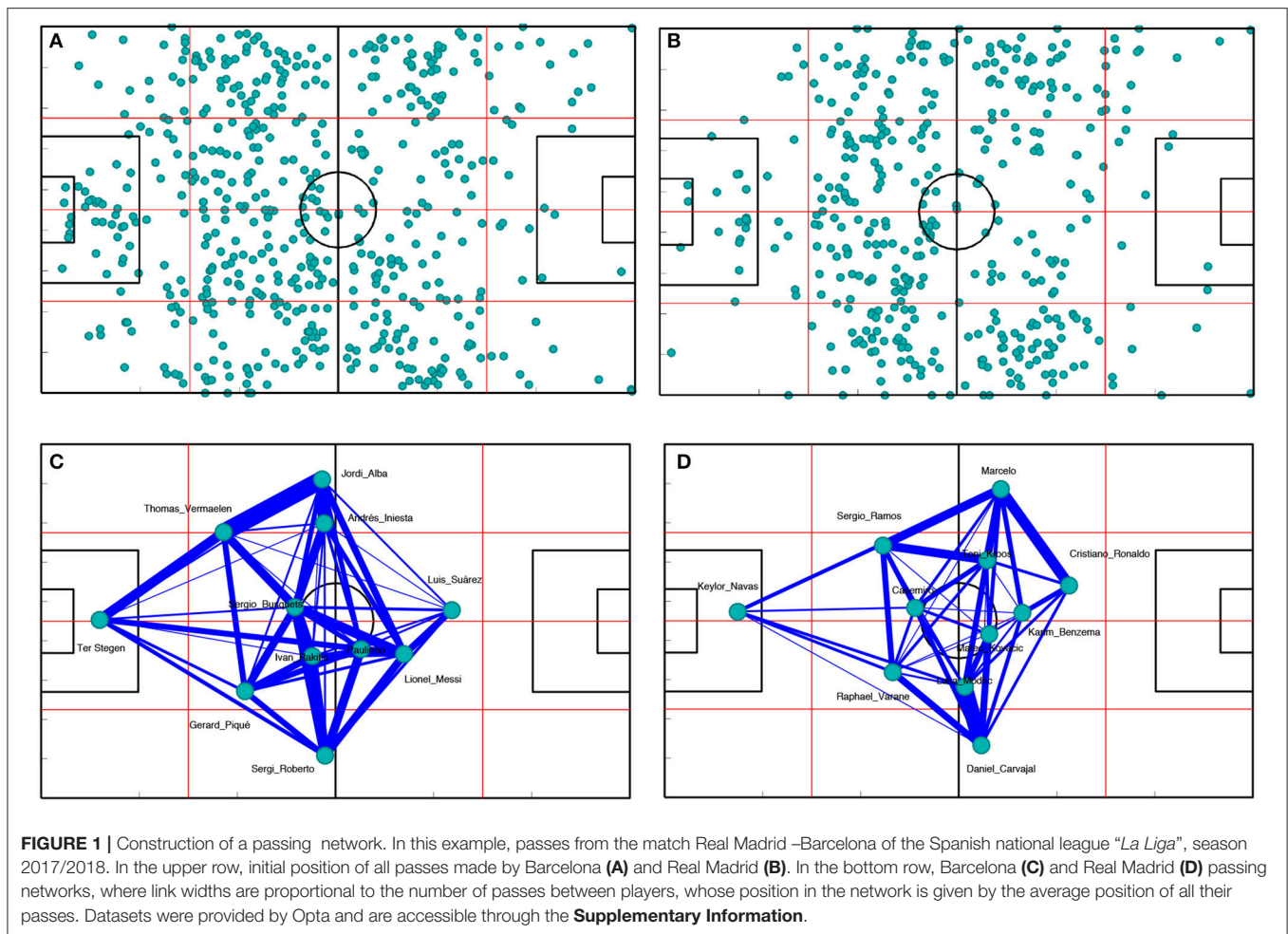
INTRODUCTION

During the last decade, *Network Science* has become one of the most active fields in applied physics and mathematics (Newman, 2010). From all its possible applications, in this Opinion paper we are concerned about the analysis of one of the most extended sports, football (soccer in U.S. terminology) (Sumpter, 2016), since it allows addressing different aspects of the team organization and performance not captured by classical analyses based on the performance of individual players. The reason behind relies on the complex nature of the game, which, paraphrasing the foundational paradigm of complexity sciences *"can not be analyzed by looking at its components (i.e., players) individually but, on the contrary, considering the system as a whole"* or, in the classical words of after-match interviews *"it's not just me, it's the team."*

The recent ability of obtaining datasets of all events occurring during a match allows analysing and quantifying the behavior of a team as a whole, together with the role of each single player (Gudmundsson and Horton, 2017). Under this framework, the organization of a team can be considered as the result of the interaction between its players, creating passing *networks,* which are directed (i.e., links between players go in one direction), weighted (the weight of the links is based on the number of passes between players), spatially embedded (i.e., the Euclidean position of the ball and players is highly relevant) and time evolving (i.e., the network continuously changes its structure).

PASSING NETWORKS: INFORMATION FROM A NEW PERSPECTIVE

Passes along the match give rise to three main types of passing networks: (i) *player passing networks*, where nodes are the players of a team (Passos et al., 2011; Grund, 2012), (ii) *pitch passing networks*, where nodes are specific regions of the field connected through passes made by players occupying them (Cintia et al., 2015) or (iii) *pitch-player passing networks*, where nodes are a combination of a player and its position at the moment of the pass (Cotta et al., 2013; Narizuka et al., 2014). See **Figure 1** for an example of a player passing network.

Once the network is constructed, several *"topological scales"* can be identified inside the passing network of a football team: (i) the *microscale*, where the analysis is carried out at the level of nodes,

FIGURE 1 | Construction of a passing network. In this example, passes from the match Real Madrid –Barcelona of the Spanish national league "*La Liga*", season 2017/2018. In the upper row, initial position of all passes made by Barcelona **(A)** and Real Madrid **(B)**. In the bottom row, Barcelona **(C)** and Real Madrid **(D)** passing networks, where link widths are proportional to the number of passes between players, whose position in the network is given by the average position of all their passes. Datasets were provided by Opta and are accessible through the **Supplementary Information**.

i.e., the players and its role inside the network, (ii) the *mesoscale*, which ranges from small motifs describing the interaction between 3 or 4 players to the detection of larger groups of players that interact most frequently between them and (iii) the *macroscale*, which considers the network as a whole.

At the topological microscale, the importance of each player has been related to: its *degree*, which is the number of passes made by a player (Cotta et al., 2013); *eigenvector centrality*, a measure of importance obtained from the eigenvectors of the adjacency matrix (Cotta et al., 2013); *closeness*, measuring the minimum number of steps that the ball has to undergo from one player to reach any other in the team (López-Peña and Touchette, 2012); or *betweenness centrality*, which accounts how many times a given player is necessary for completing the routes (made by the ball) connecting any other two players of its team (Duch et al., 2010; López-Peña and Touchette, 2012). Other metrics, such as the *clustering coefficient*, which measures the number of "neighbors" of a player that also have passed the ball between them (i.e., the number of triangles around a player), has also been quantified to evaluate the contribution of a given player to the local robustness of the passing network (López-Peña and Touchette, 2012).

At the mesoscale level, the analysis of network motifs has shown how the overabundance of certain kinds of passes between groups of three/four players can be related to both the success of a team (Gyarmati et al., 2014) and the identification of leaders in the passing network (López Peña and Sánchez Navarro, 2015). Concerning the role of communities of players playing tightly connected between them, Clemente et al. (2015), related the high heterogeneity of the number of passes between players to the existence of sub-communities, which would hinder the behavior of the team as a whole. In the same sense, Gyarmati and Anguera (2015) studied all the recurring pass sequences, relating discovered sequence patterns to teams' playing style and strategy.

Finally, at the topological macroscale, a diversity of network metrics has been shown to be informative about the style and performance of football teams. For example, the position of the network centroid has been related to the performance of the teams (the more forward, the better) and has been shown to move backwards when teams play as visitors (Bialkowski et al., 2014). Other positional variables, such as the stretch index (mean dispersion of the players around the centroid), the surface area or the team length and width have also been used as more sophisticated metrics related to team performance

(Duarte et al., 2012). On the other hand, Duch et al. (2010) designed a performance metric based on the betweenness of the players, showing how it correlated with the probability of winning a match. Other macro-scale measures such as the team average degree (i.e., average number of passes) or the variability of the players' degrees have also been proposed as proxies for evaluating team performance (Cintia et al., 2015; Pina et al., 2017). Concerning the macro-scale topology of the passing network, the small-world property (Watts and Strogatz, 1998), observed in a diversity of social, biological and technological networks, has also been reported (Narizuka et al., 2014). The average clustering coefficient of the team has also been shown to be much higher, during a match, than in equivalent random networks, unveiling the creation of triplets between players (Cotta et al., 2013). More recent studies tracking the position of the players have shown that it is better to maintain a balanced betweenness and a high closeness along the nodes of the passing network (Gonçalves et al., 2017).

Finally, it is worth mentioning that despite the micro- (i.e. player) and macro- (team) scales are perfectly identified, it is not clear how to define mesoscale units: What is the role of small motifs? Are there groups of players that can be considered as a unit? What is the role of small communities of players in the game? All these questions are of extreme importance and can be addressed using Network Science methodologies such as clustering or community identification algorithms (Fortunato, 2010), quantification of overrepresented motifs (Milo et al., 2002) or analysis of hierarchical structures (Corominas-Murtra et al., 2013).

SOME CHALLENGES: DYNAMICS, SPACE, TIME AND INTERACTION BETWEEN TEAMS

Passing networks are, in fact, dynamical system themselves, and the full identification and quantification of how variables determine the evolution of the game of a team are still open problems. Indeed, since the game cannot escape from the existence of stochastic forces combined with the high complexity of its intrinsic dynamics, modeling and forecasting a football match becomes a highly challenging task. Fortunately, distinguishing noise from determinism is an issue where Network Science can help, since it is possible to determine the *level of randomness* of the topology of the network and the dynamics occurring in it (e.g., how the ball moves along the network). We are on the way of constructing adequate null models of passing networks that are able to quantify the amount of disorder and complexity of the network topology. As explained in Sarzynska et al. (2016), the interpretation of network metrics should be referred to reference values, which can be obtained from adequate null models. However, these null models must incorporate the particular features of the system they are describing, and the Euclidean position of the nodes and temporal evolution should be taken into account (Sarzynska et al., 2016). Therefore, null models for passing networks must be as realistic as possible and include the intrinsic features of the game such as the degree

distribution, length of the passes and positions of the players in the pitch. Concerning the dynamics along the network, recent approaches using Markovian models could be a starting point to unravel hidden patterns in the passing sequences of a team (López Peña, 2014) but must include the particular features of players' movements and their ability of decision.

Nonetheless, topology is only one dimension of the analysis of passing networks and at least other two must be included to have a complete picture: space and time. Concerning the former, the division of the pitch into different sub-regions has been carried out in a series of papers; however, it is not clear what is the most adequate spatial partition or, even if a unique partition (or scale) exists for a given team. From the division of the pitch into 6x3 rectangles of equal size up to a segmentation of 100 regions (Cintia et al., 2015) a diversity of field partitions has been suggested (Camerino et al., 2012; Narizuka et al., 2014; Arriaza-Ardiles et al., 2018). The translation of passes into pitch or pitch-player networks seems to be a promising complementary vision of the game, despite the information loss about the players' behavior.

Time is another dimension traditionally overlooked when constructing passing networks. Note that the analysis of passing networks must take into account its continuous evolution in time and space. For example, as shown in Duarte et al. (2012), entropy decreases as the game evolves, which is attributed to the fatigue of the players. In addition, collective behaviors decrease in complexity/irregularity during the time periods, accompanied with an increase of the deviations from the mean tendency. Network's density, heterogeneity and centralization also show differences between the 1st and 2nd part of a match (Clemente et al., 2015). However, it is common to average along the whole match (Gonçalves et al., 2017), or even along a competition (López-Peña and Touchette, 2012; Gyarmati et al., 2014), obtaining a network that can be informative about the general behavior of a team but excludes the unavoidable fluctuations during a match. Considering the networks during each of the two parts (Clemente et al., 2015) or constructing sliding windows of a certain length (between 5 and 15 min) are reasonable approaches (Yamamoto and Yokoyama, 2011; Duarte et al., 2012; Cotta et al., 2013) as suggested in Ribeiro et al. (2017). More recently, the construction of *temporal multilayer networks* has been proposed (Ramos et al., 2018). Under this framework, each temporal layer consists of a sub-network only containing the passes during a specific attacking phase, starting with the ball recovery and finishing with a shot or a ball loss. As explained in Ramos et al. (2018), aggregated or averaged networks contain unreal paths (or shortest-paths) between players, since these paths may go through passes that occurred at, for example, different parts of the match. Using temporal multilayer networks overcomes this issue, however, assuming an *a priori* temporal division of the network may introduce some bias into the network metrics and alternative approaches should focus on finding the time scales of the match from the observation of the game. This could be done by computing the network properties for any possible temporal division and, next, using community detection algorithms (Fortunato, 2010) to identify the optimal temporal partition of the match into time slots. Note that these time slots

may have different lengths and would be indicators of how (and when) the game style changed, quantifying the temporal scales associated to each team.

At the same time, a football match is the result of the competition between two teams, i.e., the interaction between two networks. Therefore, the passing network of a team must be analyzed in combination with the network of the opponent (Narizuka et al., 2014). In this way, it will be possible to draw conclusions about how a team adapts its game depending on the opponent and what kind of topological organization leads to better results. Recent studies about *networks-of-networks* in other fields have shown that when networks get connected, important properties of the ensembled systems are modified (Boccaletti et al., 2014; Kivelä et al., 2014). With this regard, a multilayer description of a match, with two interacting networks, or layers, composed of the internal passes of each team, is still missing. This multilayer approach based on networks-of-networks would be complementary to the temporal multilayer networks proposed in Ramos et al. (2018). Specifically, the intra-layer links would be composed of the passes only within each team, while the inter-layer links would consist on ball recovery/losses. The analysis of the resulting network-of-networks could be fundamental to understand the evolution and adaptability of the teams along the match, which cannot be interpreted without analysing the behavior of each team separately. In this way, the competition between the two networks, suggests the application of the network-of-networks framework proposed in Aguirre et al. (2013), where it was shown that it is possible to find the optimal strategies of interaction with other networks, or in the case of football passing networks, that each team should find the most appropriate way of organizing its passing network according to the organization of the opponent.

On the other hand, there are parallel methodologies, such as the use of hypernetworks (Johnson, 2016), which would benefit from the inclusion of the information contained in the passing networks. Ramos et al. (2017) proposed the use of hypernetworks as an alternative way of studying the spatial competition and cooperation between players, translating actions occurring between two or more players into hypernodes. As they suggest, including passes between players in the hypernetwork seems a promising step forward in the application of hypernetworks to understand the organization of the game.

Finally, it is possible to translate and generalize the results of network theory in football to organizational studies (Orlikowski, 1996; Padgett and Powell, 2012). When two teams are competing in the field, they need to develop strategies to create new options and "entrepreneurial actions" that generate "surprises" to the opponent (Dew, 2009). Furthermore, the interaction between teams goes beyond the notion of "adaptation" challenging the concept of the "interface" or dynamic limit between the two teams. We believe that teams (beyond football and sports) need to generate new competencies such as systematic creativity and organizational learning that allow them to anticipate to the competition, promote their superiority and to create order and optimal organizational structures but, at the same time, to generate "disorder" in the opponent with the aim of creating situations of superiority.

AUTHOR CONTRIBUTIONS

All authors participated in the conception of the article. JMB wrote the initial draft. All authors revised the manuscript together; all authors gave final approval for publication. Datasets were provided by Opta and are available through the **Supplementary Information**.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01900/full#supplementary-material

REFERENCES

Aguirre, J., Papo, D., and Buldú, J. M. (2013). Successful strategies for competing networks. *Nat. Phys.* 9, 230. doi: 10.1038/nphys2556

Arriaza-Ardiles, E., Martín-González, J. M., Zuniga, M. D., Sánchez-Flores, J., de Saa, Y., and García-Manso, J. M. (2018). Applying graphs and complex networks to football metric interpretation. *Hum. Mov. Sci.* 57, 236–243. doi: 10.1016/j.humov.2017.08.022

Bialkowski, A., Lucey, P., Peter, G., Carr, K., Yue, Y., and Matthews, I. (2014). "Win at home and draw away: Automatic formation analysis highlighting the differences in home and away team behaviors," in *Proceedings of the 8th Annual MIT Sloan Sports Analytics Conference* (Boston, MA: MIT), 1–7.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544, 1–122. doi: 10.1016/j.physrep.2014.07.001

Camerino, O. F., Chaverri, J., Anguera, M. T., and Jonsson, G. K. (2012). Dynamics of the game in soccer: detection of t-patterns. *Eur. J. Sport Sci.* 12, 216–224. doi: 10.1080/17461391.2011.566362

Cintia, P., Rinzivillo, S., and Pappalardo, L. (2015). "A network-based approach to evaluate the performance of football teams," in *Machine Learning and Data Mining for Sports Analytics Workshop* (Porto).

Clemente, F. M., Santos Couceiro, M., Lourenço Martins, F. M., and Sousa Mendes, R. (2015). Using network metrics in soccer: a macro-analysis. *J. Hum. Kinet.* 45, 123–134. doi: 10.1515/hukin-2015-0013

Corominas-Murtra, B., Goñi, J., Solé, R. V., and Rodríguez-Caso, C. (2013). On the origins of hierarchy in complex networks. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13316–13321. doi: 10.1073/pnas.1300832110

Cotta, C., Mora, A. M., Merelo, J. J., and Merelo-Molina, C. (2013). A network analysis of the 2010 FIFA world cup champion team play. *J Syst Sci Complex* 26, 21. doi: 10.1007/s11424-013-2291-2

Dew, N. (2009). Serendipity in Entrepreneurship, *Organizational* 30, 735–753. doi: 10.1177/0170840609104815

Duarte, R., Araújo, D., Folgado, H., Esteves, P., Marques, P., and Davids, K. (2012). Capturing complex, non-linear team behaviours during competitive football performance. *J Syst. Sci. Complex.* 26, 62–72. doi: 10.1007/s11424-013-2290-3

Duch, J., Waitzman, J. S., and Amaral, L. A. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE* 5:e10937. doi: 10.1371/journal.pone.0010937

Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.* 486, 75–174. doi: 10.1016/j.physrep.2009.11.002

Gonçalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jimnéz, S., and Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. *PLoS ONE* 12:e0171156. doi: 10.1371/journal.pone.0171156

Grund, T. U. (2012). Network structure and team performance: the case of English Premier League soccer teams. *Soc. Netw.* 34, 682–690. doi: 10.1016/j.socnet.2012.08.004

Gudmundsson, J., and Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Comput. Surv.* 50, 22. doi: 10.1145/3054132

Gyarmati, L., and Anguera, X. (2015). Automatic extraction of the passing strategies of soccer teams. *arXiv:1508.02171 [Preprint]*.

Gyarmati, L., Kwak, H., and Rodríguez, P. (2014). "Searching for a unique style in soccer," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD) Workshop on Large-Scale Sports Analytics* (New York, NY).

Johnson, J. H. (2016). Hypernetworks: multidimensional relationships in multilevel systems. *Eur. Phys. J. Special Top.* 225, 1037–1052. doi: 10.1140/epjst/e2016-02653-4

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *J. Cplx. Netw.* 2, 203–271.

López Peña, J. (2014). A Markovian model for association football possession and its outcomes. *arXiv:1403.7993. [Preprint]*

López Peña, J., and Sánchez Navarro, R. (2015). Who can replace Xavi? A passing motif analysis of football players. *arXiv:1506.07768 [Preprint]*.

López-Peña, J., and Touchette, H. (2012). "A network theory analysis of football strategies," in *Sports Physics: Proc. 2012 Euromech Physics of Sports Conference*, ed C. Clanet (Palaiseau: Éditions de l'École Polytechnique). (ISBN 978-2-7302-1615-9) 517–528.

Milo, R., Shen-Orr, S. S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827. doi: 10.1126/science.298.5594.824

Narizuka, T., Yamamoto, K., and Yamazaki, Y. (2014). Statistical properties of position-dependent ball-passing networks in football games. *Physica A* 412, 157–168. doi: 10.1016/j.physa.2014.06.037

Newman, M. E. J. (2010). *Networks: An introduction*. New York, NY: Oxford University Press.

Orlikowski, W. J. (1996). Improvising organizational transformation over time: a situated change perspective. *Inf. Syst. Res.*, 7, 63–92. doi: 10.1287/isre.7.1.63

Padgett, J. F., and Powell, W. (2012). *The Emergence of Organizations and Markets*. Princetown, NJ: Princeton University Press.

Passos, P., Davids, K., Araújo, D., Paz, N., Mingueéns, J., and Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *J. Sci. Med. Sport* 14, 170–176. doi: 10.1016/j.jsams.2010.10.459

Pina, T. J., Paulo, A., and Araújo, D. (2017). Network characteristics of successful performance in association football. a study on the UEFA champions league. *Front. Psychol.* 8:1173. doi: 10.3389/fpsyg.2017.01173

Ramos, J., Lopes, R. J., and Araújo, D. (2018). What's next in complex networks? capturing the concept of attacking play in invasive team sports. *Sports Med.* 48, 17–28. doi: 10.1007/s40279-017-0786-z

Ramos, J., Lopes, R. J., Marques, P., and Araújo, D. (2017). Hypernetworks reveal compound variables that capture cooperative and competitive interactions in a soccer match. *Front. Psychol.* 8:1379. doi: 10.3389/fpsyg.2017.01379

Ribeiro, J., Silva, P., Duarte, R., Davids, K., and Garganta, J. (2017). Team sports performance analysed through the lens of social network theory: implications for research and practice. *Sports Med.* 47, 1689–1696. doi: 10.1007/s40279-017-0695-1

Sarzynska, M., Leicht, E. A., Chowell, G., and Porter, M. A. (2016). Null models for community detection in spatially embedded, temporal networks. *J. Comp. Net.* 4, 363–406. doi: 10.1093/comnet/cnv027

Sumpter, D. (2016). *Soccermatics. Mathematical Adventures in the Beautiful Game, Bloomsbury Sigma*. London: Bloomsbury Publishing.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature* 393, 440–442. doi: 10.1038/30918

Yamamoto, Y., and Yokoyama, K. (2011). Common and unique network dynamics in football games. *PLoS ONE* 6:e29638. doi: 10.1371/journal.pone.0029638