



Stochastic model for football's collective dynamicsA. Chacoma ^{*}, N. Almeida, J. I. Perotti , and O. V. Billoni*Instituto de Física Enrique Gaviola (IFEG-CONICET) and Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Córdoba 5000, Argentina*

(Received 20 April 2021; accepted 20 July 2021; published 9 August 2021)

In this paper, we study collective interaction dynamics emerging in the game of football (soccer). To do so, we surveyed a database containing body-sensor traces measured during three professional football matches, where we observed statistical patterns that we used to propose a stochastic model for the players' motion in the field. The model, which is based on linear interactions, captures to a good approximation the spatiotemporal dynamics of a football team. Our theoretical framework, therefore, can be an effective analytical tool to uncover the underlying cooperative mechanisms behind the complexity of football plays. Moreover, we showed that it can provide handy theoretical support for coaches to evaluate teams' and players' performances in both training sessions and competitive scenarios.

DOI: [10.1103/PhysRevE.104.024110](https://doi.org/10.1103/PhysRevE.104.024110)**I. INTRODUCTION**

The use of complex systems theory as an alternative paradigm for analyzing elite sport dynamics is currently arousing intense academic interest [1–4]. Fostered by the new advances in data acquisition [5,6] and artificial intelligence techniques [7–9], the use of state-of-the-art statistical tools to evaluate teams' performances is currently shaping a new profile of data-driven-based professional coaches worldwide.

One can find in the literature plentiful research works in several fields of physics that have been devoted to studying phenomena related to sports science [10–14]. The research community in statistical physics has focused on the study of sports mainly in the framework of stochastic processes, such as studying the time evolution of scores [15–17]. Alternatively, other studies propose innovative models, based on ordinary differential equations [18], stochastic agent-based game simulations [19], and network science theory [20], aiming to describe the complex dynamical behavior of teams' players.

Formally, sports teams can be thought of as complex *sociotechnical systems* [21], where a wide range of organizational factors might interact to influence athletes' performances [22–25]. Particularly in collective games such as football, cooperative interplay dynamics seems to be a key feature to be analyzed [26,27]. In principle, collective behaviors in soccer are important since they are connected to team tactics and strategies. Usually, features of these collective behaviors are described by using simple group-level metrics [28–33]. Furthermore, temporal sequences of ball and player movements in football, showing traits of complex behaviors, have been reported and studied using stochastic models and statistical analysis [19,34–36].

Recent work has focused on describing cooperative on-ball interaction in football within the framework of network science [20,37–40]. In [41], for instance, Garrido *et al.* studied

the so call *pitch passing networks* in the games of the Spanish League in the 2018/2019 season. In this outstanding work, the authors use network metrics and topological aspects to define the teams' consistency and identifiability, two highly relevant global indicators to analyze team performance during competition.

From an alternative perspective, our research group has focused on studying the dynamic interactions at the microscopic level, i.e., by modeling player-player interactions. In a previous work [19], we proposed an agent-based model that correctly reproduces key features of global statistics indicators of nearly 2000 real-life football matches. Along the same lines, the present paper aims to describe the complexity of this game, uncovering the underlying mechanisms ruling the collective dynamics. Our goal is to provide a step towards a full description of the spatiotemporal dynamics in a football match. With this purpose, we posit a simple model based on linear interactions to describe teams' cooperative evolution. To do so, we analyzed a public database containing body-sensor traces from three professional football matches of the Norwegian team *Troms IL* (see Sec. II). We will show that our model succeeds in capturing part of the cooperative dynamics among the players, and that higher-order contributions (non-linear interactions) can be carefully modeled as fluctuations. Moreover, we will show that our framework provides a useful tool to analyze and evaluate tactical aspects of the teams.

This paper is divided into three sections: In Sec. II, we describe the database, the statistical regularities found in the data, and we formally propose our theoretical framework. In Sec. III, we present our main results and relevant findings. Our conclusion and future perspectives are briefly summarized in the final section.

II. MATERIAL AND METHODS**A. The database**

In 2014, Pettersen *et al.* published a database of traces recorded with body-sensors from three professional soccer

*achacoma@famaf.unc.edu.ar

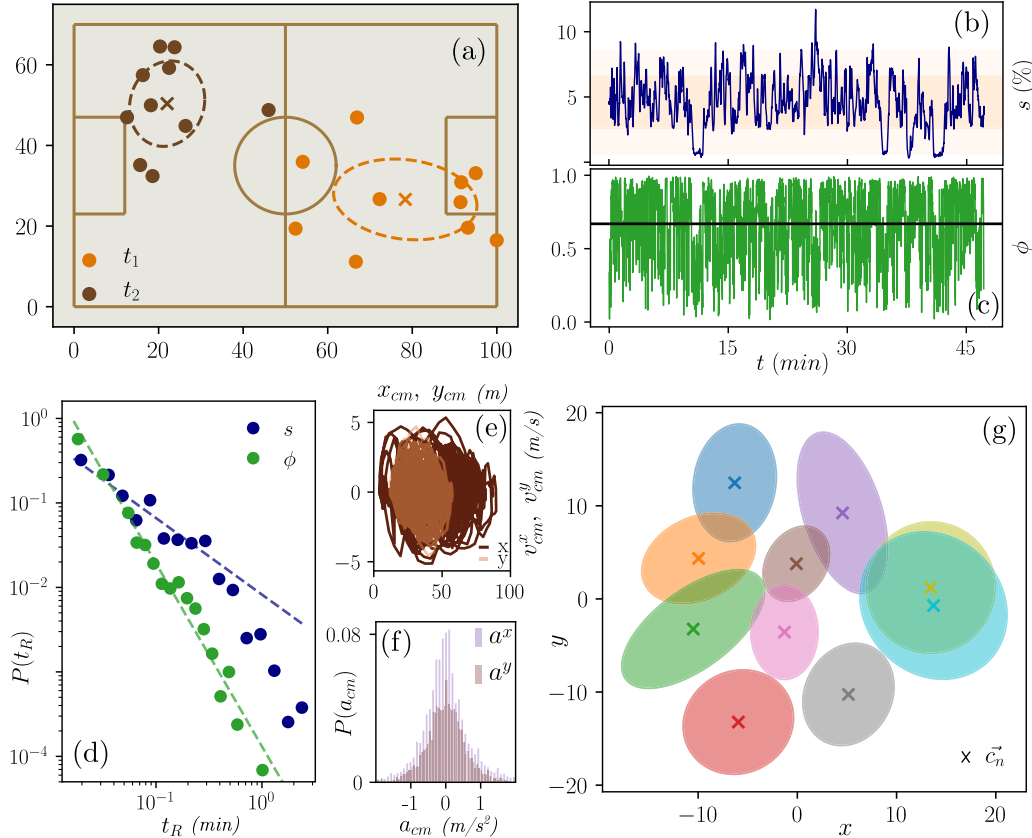


FIG. 1. Statistical regularities in *DS1*. (a) The team deployed in the field at $t_1 = 7$ min and $t_2 = 8$ min. Ellipses are a measure of the characteristic area of the team at both times (see the main text). (b) Evolution of the characteristic area s . The width of the narrow bands around the curve equals the values of $1 \times \sigma_s$ and $2 \times \sigma_s$, where σ_s is the standard deviation. (c) Evolution of the order parameter ϕ . The black solid horizontal line indicates the mean value of the series at ≈ 0.7 . (d) Time to return to the mean value calculated from series s and ϕ . Dashed lines indicate a nonlinear fit performed to measure the power-law exponent (see the main text). (e) The dynamics of the center of mass. Position vs velocity on both axes. (f) Distribution of instantaneous accelerations of the center of mass in both axes. (g) Action zones for the players (ellipses) in the center-of-mass frame of reference. \vec{c}_n indicates each player center (see the main text for further details).

games recorded in November 2013 at the Alfheim Stadium in Tromsø, Norway [6]. This database is divided into five datasets; each one contains the halves of the game Tromsø IL versus Strømsgodset IF (*DS1*, *DS2*), the halves of the game Tromsø IL versus Anzhi Makhachkala (*DS3*, *DS4*), and 40 min of the game Tromsø IL versus Tottenham Hotspurs (*DS5*). This contribution also offers video records, but they were not used in our research. We highlight that only Tromsø IL gave consent to be tracked with the body-sensors, thus the traces available in the datasets are only from this team. The goalkeeper’s position, likewise, was not tracked.

The player positions were measured at 20 Hz using the highly accurate *ZXY Sport Tracking system* by *ChyronHego* (Trondheim, Norway). However, to perform our analysis, we preprocessed the data so as to have the players’ position in the field in one-second windows. In this way, we lose resolution but it becomes simpler to analyze players’ simultaneous movements or coordination maneuvers.

B. Statistical regularities

In this section, we describe relevant aspects of the statistical observations that we used to propose our stochastic model.

In this case, we focus on the analysis of *DS1*, but similar results can be obtained for the others datasets. Let us discuss Fig. 1. In panel (a) we show the team deployed in the field at two different times: in an offensive position at $t_1 = 7$ min, and in a defensive position at $t_2 = 8$ min. The ellipses, drawn in the figure with dashed lines, give the standard deviation intervals around the mean [see the Supplemental Material (S1) for further details [42]], and they can be thought of as a measure of the team’s dispersion in the field. The ellipses’ area can be thought of as a characteristic area of the team deployed in the field. The system seems to suffer an expansion at t_1 and a contraction at t_2 . To characterize this process, we study the temporal evolution of $s := 100 \times \frac{\text{Ellipse area}}{\text{Field area}}$ (%), which we show in panel (b). We measured $\langle s(t) \rangle = 5.1$, $\sigma_s = 2.2$. The low dispersion in the time series and the symmetry around the mean indicate that the system moves around the field with a well-defined characteristic area exhibiting small variations throughout the dynamics of the game.

Let us now focus on analyzing the level of global ordering in the team. To do so, we analyze the evolution of the parameter $\phi(t) = \frac{1}{N} \sum_{n=1}^N \frac{v_n(t)}{|v_n(t)|}$ [see Eq. (1) in [43]], where v_n is the velocity of player n , and N is the total number of players. In the case $\phi \approx 1$, this parameter indicates that the

players move as a *flock*, following the same direction. On the contrary, when $\phi \approx 0$, they move in different directions. Panel (c) shows the temporal evolution of ϕ . We measured $\langle \phi(t) \rangle = 0.7$, $\sigma_\phi = 0.2$. This level of global ordering during the game shows that there are time intervals when the players tend to move as a highly coordinated *flock* [44]. Therefore, it seems there are interactions among the teammates that cause the emergence of global ordering.

We now turn our attention to the analysis of the temporal structure of series $s(t)$ and $\phi(t)$. Let us define t_R as the time of return to the mean value. In panel (d) we show the distribution $P_s(t_R)$ and $P_\phi(t_R)$. In both cases, we can see heavy-tailed distributions. By performing a nonlinear fit, using the expression $P(t_R) = Ct_R^{-\gamma}$, for the case of the relative area, s , within the range (0, 0.4) min, we measured $\gamma_s = 0.96 \pm 0.03$. In the case of the order parameter ϕ , in the whole range, we obtained $\gamma_\phi = 2.16 \pm 0.04$. It is well known that for a random-walk process in one dimension, the probability of first return yields $\gamma = 3/2$ [45]. In our case, the measured nontrivial exponents seem to indicate the presence of a complex multiscale temporal structure in the dynamics. Notice that these two parameters, related to the team structure and order, somehow encompass the memory and complex dynamics for the team during the match.

We now focus on describing the dynamics of the center of mass (CM). In Fig. 1(e), we show the relations x_{cm} versus v_{cm}^x , and y_{cm} versus v_{cm}^y . We can see that the position variables are bounded to the field area, and the velocities in both axes seem to be bounded within the small range $(-5, 5)$ m/s. In panel (f), on the other hand, we show the distribution of accelerations. We measured $\langle a_{cm}^x \rangle = \langle a_{cm}^y \rangle = 0$, $\sigma_{a_{cm}^x} = 0.4$, and $\sigma_{a_{cm}^y} = 0.2$ m/s². Since the CM of the system is barely accelerated, we can approximate the center of mass as an inertial system. Then, to simplify our analysis, we can study the dynamical motion of the players from the center-of-mass frame of reference. In this frame, we aim to define action zones for the players. To do so, we analyze the positions in the plane that the players have explored during the match. The ellipses in panel (g) give the standard-deviation intervals around the mean, and they can be thought of as characteristic action zones for the players. Note that, from this perspective, it naturally emerges that Tromsø IL used in this half the classical tactical system 4-4-2.

Summarizing, we observed that (i) the spatial dispersion of the players follows a well-defined characteristic area, (ii) inside the team, the players' movements exhibit correlation and global ordering, (iii) the system exhibits a complex multiscale temporal structure, and (iv) the players' motion can be studied from the center-of-mass frame of reference, simplifying the analysis. In the next section, we use these insights to propose a simple stochastic model of cooperative interaction to analyze the spatiotemporal evolution of the team during the match.

C. The model

The interplay among teammates can be thought of as individuals that cooperate to run a tactical system. In this frame, we aim to define a model to describe the spatiotemporal evolution of the team. Since our goal is to define a simple theoretical framework such that we can easily interpret the results, we propose a model based on player-to-player interactions. We

proceed as follows: (i) we define the equations of evolution for the players in the team, (ii) we use the empirical data to fit the equations' parameters, and (iii) we model the error in the fitting as fluctuation in the dynamics. In the following, we present the results in this regard.

1. The equations of evolution

In the CM frame of reference, we define $\vec{r}_n(t) = (x_n(t), y_n(t))^T$ and $\vec{v}_n(t) = (v_n^x(t), v_n^y(t))^T$ as the position and velocity of player n at time t . We propose that the dynamical variables change, driven by interactions that can be thought of as springlike forces. Every player in our model is bound to (i) a place in the field related to their natural position in the team, \vec{a}_n , and (ii) the other players.

The equation of motion for a team player n can be written as follows:

$$M_n \ddot{\vec{r}}_n = -\gamma_n \vec{v}_n + k_{an}(\vec{a}_n - \vec{r}_n) + \sum'_m k_{nm}(\vec{r}_m - \vec{r}_n), \quad (1)$$

where the first term is a damping force, the second one is an "anchor" to the player's position, and the sum is the contribution of the interaction forces with the other players. We propose different interaction constants in the horizontal and vertical axis, therefore the parameters γ_n , k_{an} , and k_{nm} are 2×2 diagonal matrices such as $\gamma_n = \begin{pmatrix} \gamma_n^x & 0 \\ 0 & \gamma_n^y \end{pmatrix}$, $k_{an} = \begin{pmatrix} k_{an}^x & 0 \\ 0 & k_{an}^y \end{pmatrix}$, $k_{nm} = \begin{pmatrix} k_{nm}^x & 0 \\ 0 & k_{nm}^y \end{pmatrix}$. Moreover, since players have comparable weights, for simplicity we consider $M_n = 1$ for all the players. Notice that if equilibria exist, $\vec{r}_n(t \rightarrow \infty) = \vec{r}_n^*$ and $\vec{v}_n(t \rightarrow \infty) = 0$. Then,

$$-\left(k_{na} + \sum'_m k_{nm}\right) \vec{r}_n^* + \sum'_m k_{nm} \vec{r}_m^* + k_{an} \vec{a}_n = 0 \quad (2)$$

must hold.

Equations (1) can also be written as a first-order equation system as follows:

$$\begin{aligned} \dot{\vec{r}}_n &= \vec{v}_n, \\ \dot{\vec{v}}_n &= -\left(k_{an} + \sum'_m k_{nm}\right) \vec{r}_n + \sum'_m k_{nm} \vec{r}_m - \gamma_n \vec{v}_n + k_{an} \vec{a}_n. \end{aligned} \quad (3)$$

Furthermore, by defining $\vec{x} = (x_1, \dots, x_n, v_1^x, \dots, v_n^x)$ and $\vec{y} = (y_1, \dots, y_n, v_1^y, \dots, v_n^y)$, we can write

$$\dot{\vec{x}} = J^x(\vec{x} - \vec{x}^*), \quad \dot{\vec{y}} = J^y(\vec{y} - \vec{y}^*), \quad (4)$$

where $J^x, J^y \in \mathbb{R}^{2n \times 2n}$ are the Jacobian matrices of system (3). With Eqs. (4), we can analyze separately the system evolution along the horizontal and the vertical axis. Moreover, in Sec. III we will show that the Jacobian matrices can be used to describe the team's collective behavior.

2. Fitting the model's parameters

In this section, we show how to fit the model's parameters γ_n , k_{an} , k_{nm} , and \vec{a}_n by using the datasets, and Eqs. (3) and (2). To proceed, we have considered the following steps:

TABLE I. Model parameters inferred for dataset *DS1*.

Par.	x	y	Par.	x	y	Par.	x	y	Par.	x	y
α_1	4.0×10^{-1}	4.3×10^{-1}	k_{12}	4.2×10^{-3}	1.2×10^{-2}	k_{37}	0.0	2.5×10^{-3}	k_{79}	0.0	5.1×10^{-4}
α_2	4.1×10^{-1}	5.8×10^{-1}	k_{13}	0.0	0.0	k_{38}	0.0	0.0	k_{710}	4.5×10^{-3}	1.4×10^{-3}
α_3	3.8×10^{-1}	5.7×10^{-1}	k_{14}	6.4×10^{-3}	0.0	k_{39}	6.4×10^{-3}	0.0	k_{89}	5.6×10^{-3}	6.9×10^{-3}
α_4	4.0×10^{-1}	4.2×10^{-1}	k_{15}	3.9×10^{-3}	4.7×10^{-3}	k_{310}	0.0	2.3×10^{-3}	k_{810}	6.0×10^{-3}	3.2×10^{-3}
α_5	4.1×10^{-1}	4.1×10^{-1}	k_{16}	1.7×10^{-3}	6.7×10^{-3}	k_{45}	1.7×10^{-3}	0.0	k_{910}	4.7×10^{-3}	0.0
α_6	4.1×10^{-1}	5.4×10^{-1}	k_{17}	4.6×10^{-3}	0.0	k_{46}	0.0	1.8×10^{-3}			
α_7	4.2×10^{-1}	5.0×10^{-1}	k_{18}	1.9×10^{-3}	2.1×10^{-3}	k_{47}	0.0	5.4×10^{-3}			
α_8	4.0×10^{-1}	4.4×10^{-1}	k_{19}	3.7×10^{-3}	1.4×10^{-3}	k_{48}	1.8×10^{-3}	3.3×10^{-3}			
α_9	2.8×10^{-1}	3.3×10^{-1}	k_{110}	5.0×10^{-3}	1.6×10^{-3}	k_{49}	3.6×10^{-3}	6.5×10^{-4}			
α_{10}	2.8×10^{-1}	3.7×10^{-1}	k_{23}	6.7×10^{-3}	7.1×10^{-3}	k_{410}	0.0	1.1×10^{-3}			
k_{a1}	5.7×10^{-3}	0.0	k_{24}	0.0	8.9×10^{-4}	k_{56}	2.9×10^{-3}	6.9×10^{-4}			
k_{a2}	0.0	0.0	k_{25}	2.8×10^{-6}	0.0	k_{57}	0.0	0.0			
k_{a3}	0.0	2.2×10^{-2}	k_{26}	2.1×10^{-3}	6.7×10^{-3}	k_{58}	8.0×10^{-3}	0.0			
k_{a4}	1.0×10^{-2}	1.0×10^{-2}	k_{27}	5.5×10^{-3}	0.0	k_{59}	1.6×10^{-3}	2.7×10^{-3}			
k_{a5}	2.0×10^{-2}	0.0	k_{28}	0.0	9.9×10^{-4}	k_{510}	0.0	2.1×10^{-3}			
k_{a6}	2.7×10^{-2}	2.2×10^{-2}	k_{29}	1.9×10^{-3}	5.4×10^{-3}	k_{67}	8.6×10^{-3}	5.0×10^{-3}			
k_{a7}	3.2×10^{-2}	1.2×10^{-2}	k_{210}	1.5×10^{-3}	3.5×10^{-3}	k_{68}	2.2×10^{-4}	9.9×10^{-5}			
k_{a8}	1.9×10^{-2}	1.2×10^{-2}	k_{34}	0.0	3.3×10^{-3}	k_{69}	0.0	5.4×10^{-3}			
k_{a9}	4.2×10^{-4}	0.0	k_{35}	0.0	0.0	k_{610}	3.2×10^{-3}	1.9×10^{-3}			
k_{a10}	0.0	0.0	k_{36}	1.6×10^{-3}	4.0×10^{-3}	k_{78}	0.0	5.7×10^{-3}			

(i) For every player in the team, each dataset provides the position in the field $\vec{r}_n(t)$. The velocity is calculated as $\vec{v}_n(t) := \frac{\vec{r}_n(t+\Delta t) - \vec{r}_n(t)}{\Delta t}$ ($\Delta t = 1$ s).

(ii) The discrete version of system (3) gives us the tool to estimate the states of the players at time $t + \Delta t$ by using as inputs the real states at time t and the model's parameters,

$$\begin{aligned} \vec{r}_n(t + \Delta t)' &= \vec{r}_n(t) + \vec{v}_n(t)\Delta t, \\ \vec{v}_n(t + \Delta t)' &= \vec{v}_n(t) + \left[-\gamma_n \vec{v}_n(t) - \left(k_{an} + \sum_m' k_{nm} \right) \vec{r}_n(t) \right. \\ &\quad \left. + \sum_m' k_{nm} \vec{r}_m(t) + k_{an} \vec{a}_n \right] \Delta t, \end{aligned}$$

where $\vec{r}_n(t + \Delta t)'$ and $\vec{v}_n(t + \Delta t)'$ are the model's estimations.

(iii) Note that by considering the definition of the velocity given in (i), $\vec{r}_n(t + \Delta t) = \vec{r}_n(t + \Delta t)'$. Then, at every step, the parameters are only used to predict the new velocities.

Moreover, to simplify our framework, we do the following:

(iv) Since parameters \vec{a}_n are linked to the equilibria values by Eq. (2), without loss of generality, we take $\vec{c}_n = \vec{r}_n^*$ [where \vec{c}_n is the center of the action radii for every player, empirically obtained from the datasets; see Fig. 1(d)]. By doing this, values \vec{a}_n can be calculated from the values of \vec{c}_n and the other parameters.

(v) To simplify our analysis, we normalized the players' position in the dataset such that the standard deviation (scale) of players' velocities is the unit (i.e., $\sigma_{\vec{v}} = 1$). This is useful for later assessment of the fitting performance.

In this frame, we define the error $\vec{\xi}_n(t) := \vec{v}_n(t + \Delta t) - \vec{v}_n(t + \Delta t)'$, and we fit γ_n , k_{an} , k_{nm} by minimizing the sum $\sum_t |\sum_n \vec{\xi}_n(t)|$. With this method, we obtain a unique set of parameters that govern the equations. For a more detailed description of the minimizing procedure, cf. Appendix A.

Notice that, to avoid possible large fluctuations linked to drastic tactical changes, we fitted the set of parameters to each dataset (a half match). This criterion, at the same time, let us compare the strength of the interactions among different matches halves.

The results of the optimization process for *DS1* are given in Table I. There we show the values of the fitted parameters in both coordinates, x and y . We can see $\alpha_n \approx 10^{-1}$, $k_{an} \approx 10^{-2}$, and $k_{nm} \lesssim 10^{-2}$. Particularly interesting are parameters k_{nm} , since they indicate the strength of the interactions among players. In this case, we can see a wide variety of values, from strong interactions as in the case of players 1 – 2 to negligible interactions in the case of players 3 – 8. For results on other datasets, please see the Supplemental Material S2 [42].

3. Modeling $\vec{\xi}_n(t)$ as fluctuations in the velocities

By using the optimal set of parameters calculated with the method proposed in the previous section, we can calculate for all the players at every time step the difference between the real velocity and the model's prediction. This defines N temporal series $\vec{\xi}_n(t) = \{\vec{\xi}_n(t_0), \vec{\xi}_n(t_1), \dots, \vec{\xi}_n(t_T)\}$, which can be thought of as stochastic fluctuations in the players' velocities. Note that, in the context of a football match, these fluctuations can be related to stochastic forces acting upon the players. With this idea in mind, we propose to introduce in the system (3) a noisy component linked to these fluctuations. With this aim, we focus on analyzing and describing the behavior of $\vec{\xi}_n(t)$.

Let us turn our attention to Fig. 2. Here, our goal is to characterize the fluctuations linked to *DS1*. In panel (a), we show the distributions of values related to ξ_n^x ($n = 1, \dots, 10$). We can see, in each case, that the curves approach a Gaussian shape. The dashed line indicates a nonlinear fit performed to the distribution given by the entire set of values (ξ^x); in this case, we have measured $\langle \xi^x \rangle = 0.001 \pm 0.002$ and

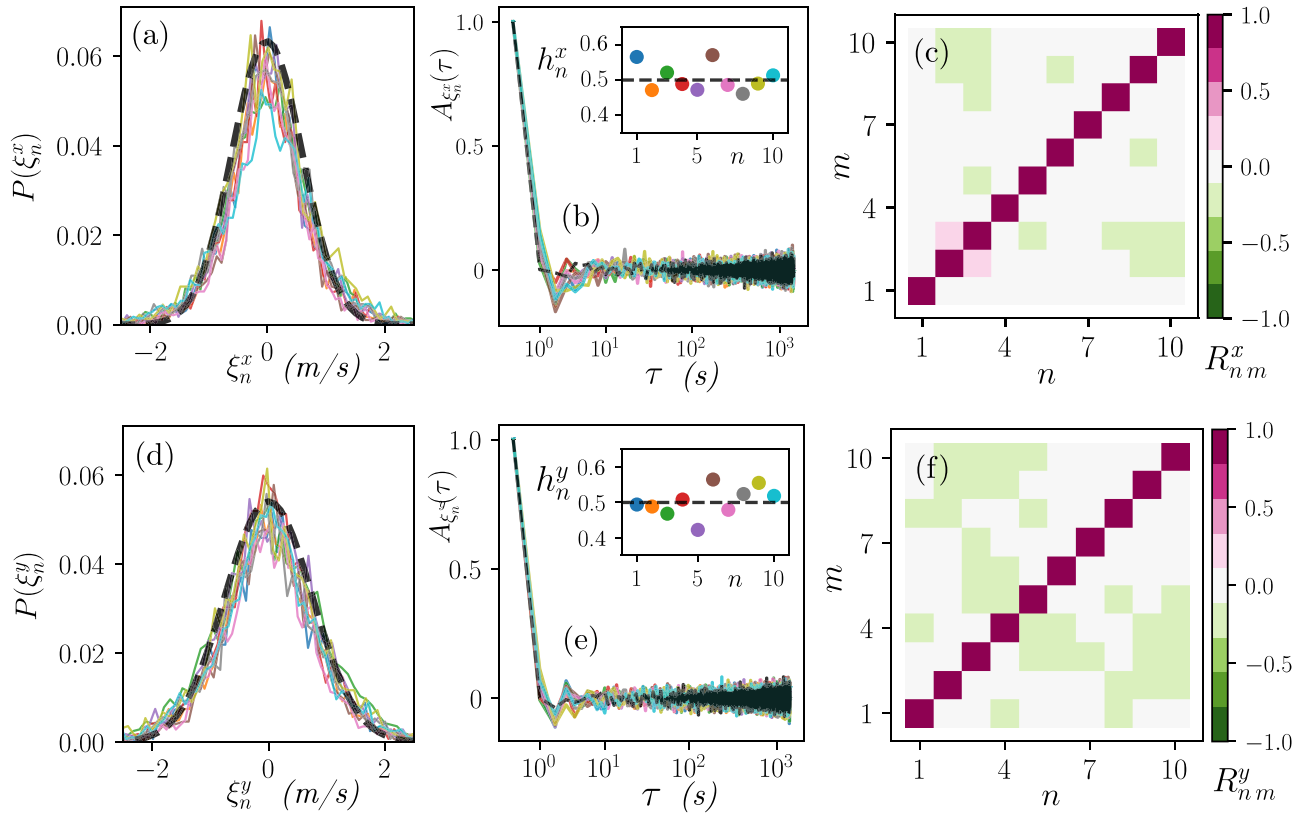


FIG. 2. Error characterization—Fluctuation analysis. For the horizontal axis, x , (a) Distributions of the error values for the ten players (colored curves). Dashed lines indicate a Gaussian fit performed to the aggregated set of values. (b) Main: autocorrelation functions for the ten cases (colored curves). The dashed line indicates the same calculation for a white noise process. Inset: Hurst exponent for the ten cases. (c) Pearson matrix indicating the value of the linear correlation among each pair of time series, including autocorrelations. For the vertical axis, y , the descriptions of panels (d), (e), and (f) are analogous to the previous case.

$\sigma_{\xi^x} = 0.60 \pm 0.02$ m/s ($R^2 = 0.97$). The fluctuations scale is lower than the velocities scale, $\sigma_{\xi^x} < \sigma_v$ [with $\sigma_v = 1$, cf. Sec. II C 2, item (v)]. Panel (b), on the other hand, shows the autocorrelation functions $A_{\xi_n^x}(\tau) = \frac{\langle ZZ' \rangle - \langle Z \rangle \langle Z' \rangle}{\sigma_Z \sigma_{Z'}}$, with $Z = \xi_n^x(t)$ and $Z' = \xi_n^x(t + \tau)$. For each case, we can see an abrupt decay at the beginning. To help the eye to visualize the behavior of the curves, the black dashed line in the plot shows the autocorrelation function for a white noise process. The inset in the panel shows the values of the Hurst exponent calculated by performing a detrended fluctuation analysis (DFA) to $\xi_n^x(t)$. We obtain values around 0.5 ± 0.06 , which is consistent with a set of memoryless processes. Panel (c), on the other hand, shows the Pearson matrix $R_{nm}^x = \frac{C_{nm}^x}{\sqrt{C_{nn}^x C_{mm}^x}}$, where C_{nm}^x is the covariance matrix of series $\xi_n^x(t)$. We can see that $R_{nm}^x < 0.25 \forall n, m$, which indicates a low level of linear correlation among the fluctuations associated with the different players. A similar description with analogous results can be done for ξ_n^y by analyzing panels (d), (e), and (f).

Based on the observations made above, we propose to model the fluctuations in both axes as noncorrelated Gaussian noise, such that $\vec{\xi}_n(t) = \vec{\sigma}_n \xi_n(t)$, with $\langle \xi_n(t) \rangle = 0$, $\langle \xi_n(t) \xi_n(t') \rangle = \delta(t - t')$, $\langle \xi_n(t) \xi_m(t) \rangle = 0$, and $\vec{\sigma}_n = (\sigma_n^x, \sigma_n^y)$ the empirical measured scales for the processes.

III. RESULTS AND DISCUSSION

A. Simulations on the collective dynamics

As we previously stated, we first used the datasets to fit the model's parameters, and second we characterized the errors as fluctuation in the velocities. With these inputs, in the frame of our model, we can simulate the players' collective dynamics and compare the results with empirical data to assess the model performance. To do this, we modify system (3) as follows:

$$d\vec{r}_n = \vec{v}_n dt, \quad d\vec{v}_n = \left[- \left(k_{na} + \sum'_m k_{nm} \right) \vec{r}_n + \sum'_m k_{nm} \vec{r}_m - \gamma_n \vec{v}_n + k_{na} \vec{a}_n \right] dt + d\vec{W}_n, \quad (5)$$

where $d\vec{W}_n = \vec{\sigma}_n \xi_n dt$, with $\vec{\sigma}_n$ and $\xi_n dt$ as were defined in the previous section. Note that (5) is a system of stochastic differential equations (SDEs). To solve them, we use the Euler-Maruyama algorithm for Ito equations. In this section, we show the results linked to the dataset *DS1*; similar results for the other datasets can be found in Appendix B.

Let us focus on Fig. 3. Panels (a) and (b) show two heatmaps with the probability of finding a team player in

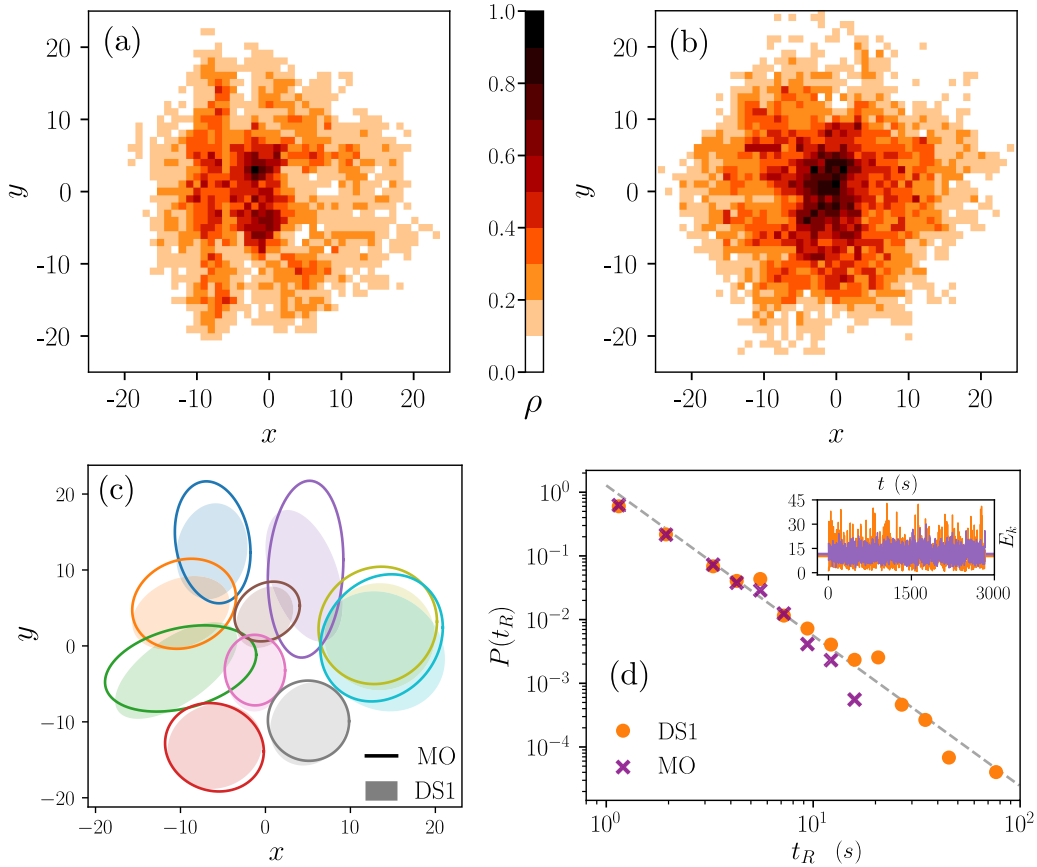


FIG. 3. Results on the collective dynamics simulations. (a),(b) Probability of finding a player in the field, in empirical data and simulations, respectively. (c) Players' action zones. Empirical data (shadow areas) compared with simulations (curves). (d) Probability distribution of the time to return to the mean value, $P(t_R)$. The dashed line indicates a nonlinear fit performed to the empirical data (orange circles). The inset shows the evolution of the kinetic energy, $E_k(t)$, from where t_R is measured.

the position (x, y) . The left panel shows the results for the empirical data, whereas the right panel shows the results for simulations. For better visualization, in both cases the probabilities were normalized to the maximum value, defining the parameter $\rho \in [0, 1]$. As we can observe, simulations reproduce reasonably well the empirical observations. To quantify the result, we calculated the Jensen-Shannon distance (*DJS*) between the distributions. We measured $DJS = 0.05$, which indicates a good similarity between distributions. In panel (c), on the other hand, we compare players' action zones. The empirical observations are the shadow ellipses, whereas the simulations are the curves. We can see, on the whole, that the model gives a good approximation, particularly for those areas with smaller dispersion. In panel (d), we analyze the kinetic energy of the system, $E_k := \sum_n \frac{1}{2} |\vec{v}_n|^2$. Our goal here is to globally describe the temporal structure of the system. In the inset, we show the temporal evolution of this quantity. Regarding the mean values of the energy, we measured $\langle E_k \rangle_{DS1} = 10.3$ for the data and $\langle E_k \rangle_{MO} = 11.5$ for simulations. We can see that the energy reaches high peaks in the empirical case that are not observed in the outcomes of the model compensating for the slightly lower mean. This effect indicates the presence of higher-order contributions, probably linked to nonlinear interactions, that our model based on linear interactions does not take into account. The main plot in the panel shows the distribution of the times to return to the mean

value for the kinetic energy, $P(t_R)$, for both empirical data and the model. We can see in both cases that $P(t_R) \propto t_R^{-\gamma_{E_k}}$, with $\gamma_{E_k} = 2.4 \pm 0.1$. Note that the value of the exponent agrees with the value measured for the time to return in the case of the variable ϕ (see Sec. II B). This seems to indicate that the temporal structure captured by the evolution of the energy may be related to the emergence of order in the system. Let us turn our attention to the distributions' tails. For the case of the empirical data, we can see the presence of extreme events that are not observed in simulations. This effect, as well as the peaks in E_k , could be linked to higher-order contributions of the interactions.

To summarize, in this section we showed that despite its simplicity, our first-order stochastic model succeeds in uncovering several aspects of the complexity of the spatiotemporal structure underlying the dynamics. In the next two sections, we show how to utilize the fitted interactions to describe the players' individual and collective behavior.

B. Describing the team behavior by analyzing the model's parameters

1. Hierarchical clustering classification on the team lineup

The interaction parameters \vec{k}_{nm} can be useful to analyze the interplay among team players, and to describe the

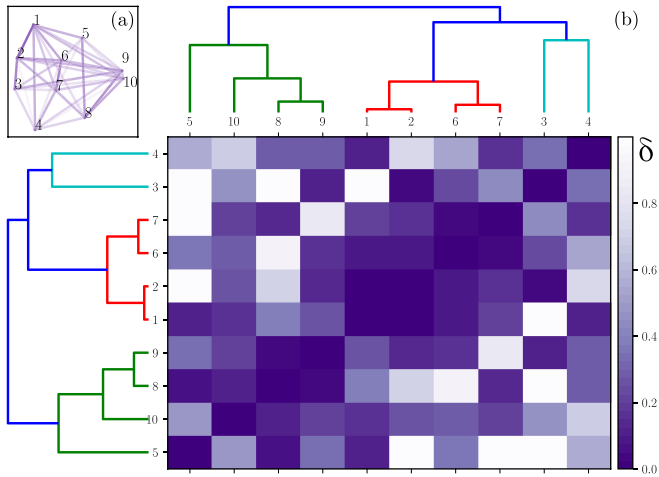


FIG. 4. Classification of the team members, based on the inferred interaction strengths. (a) Visualization of the connection strength, k_{nm} , among the players. (b) Hierarchical relations among the players. At the center: the distance matrix. At the top and the left: dendrograms to visualize the clusters of players within the team.

environment to which they are constrained. In Fig. 4(a), we show a visualization of the players' interactions magnitude. In this network of players, the links' transparency (α value) represents the connection strength between players n and m . These values are calculated as $\kappa_{nm} = \sqrt{(k^{x_{nm}})^2 + (k^{y_{nm}})^2}$. Note that the connection values can be used as a proxy of the distance among the players. Let us define the distance $\delta_{nm} := e^{-\frac{\kappa_{nm}}{\sigma_\kappa}}$ between player n and m , where σ_κ is the standard deviation of the set of values κ_{nm} . The exponential function in δ_{nm} is used to define large distances linked to small connections, and short distances related to strong connections. Then, based on this metric, we calculate the matrix of distances. With this matrix, by using a hierarchical clustering classification technique, we detect small communities of players within the team. In the following, we discuss the results in this regard. In Fig. 4(b), we show the reordered matrix of distances with two equal dendrograms, one placed at the top and the other at the left, showing the hierarchical relationship between players. To perform this calculation, we used the *ward* method [46]. With this classification, we can easily observe the presence of two main clusters of players. Those colored in green, players 5 – 10 – 8 – 9, can be related to the offensive part of the team; the others can be related to the defensive. Within the latter group, the cluster colored in red (players 1 – 2 – 6 – 7) is related to back and middle defenders at the left side of the field, whereas players 3 – 4 are related to back defenders at the right. Within the former group, we can differentiate between a central-right group of attackers, 8,9,10, and an individual group given by player 5. As we said above, this technique allows us to study groups of players with strong interactions during the match. For instance, let us focus on analyzing the group of players 6 – 7. These players cover the central zone, and they are likely in charge of covering the gaps when other defenders go to attack. One such example is the advance of the wing-backs 1 – 4 by the sides. We should expect them to behave similarly, which agrees with the result of our classification. The case of player 5 is particularly

interesting. Our results indicate that this player is less constrained than the others attackers. Our classification frame is probably detecting that 5 is a free player in the team, a classic *play builder* on the midfield, in charge of generating goal-scoring opportunities.

The information provided by the hierarchical clustering classification allows us to characterize the players' behavior within a team and, therefore, provide useful insight into the collective organization. In light of this technique, it is possible to link the strengths and weaknesses of the team to the levels of coordination among the players. For instance, if we observe a lack in the levels of coordination among the rival players at the right side of their formation (as we can see in the case of Tromsø IL in DS1), it would be interesting to foster attacks in this sector.

We could also perform a comparative study by analyzing several games to correlate results with levels of coordination among the players. With this approach, we can detect whether there are patterns related to a winning or a losing formation, providing valuable information for coaches to use. The same idea can be applied to training sessions, to promote routines oriented to strengthen the players' connection within particular groups in the line-up, aiming to improve the team performance in competitive scenarios.

To summarize, the hierarchical clustering analysis evidences the presence of highly coordinated behavior among subgroups of players, which can be directly related to their role on the team. In this framework, coaches may find a useful tool to support the complex decision-making process involved in the analysis of the tactical aspects of the team, to assess players' performances, to propose changes, etc.

2. Collective modes

The eigenvalues and eigenvectors of the Jacobian matrices J^x and J^y [see Eqs. (4)] can be handy to describe some aspects of the team players dynamics on the field. If the system exhibits complex eigenvalues, the eigenvectors can provide information on the collective modes of the system, and, consequently, on the collective behavior of the team.

In Fig. 5(a), we plotted the system's eigenvalues $\lambda \in C$ as Re_λ versus Im_λ . We can see in most cases that $\text{Im}(\lambda) = 0$. However, around $\text{Re}(\lambda) \approx -0.2$, we can see the presence of characteristic frequencies in both coordinates. Let us focus on the case of λ_1 , the eigenvalue with smallest real part, and $\text{Im}(\lambda_1) \neq 0$. This case is particularly important, because the energy that enters the system as noise is transferred mainly to the vibration mode given by the eigenvector associated with λ_1, v_{λ_1} (first mode) [47,48]. In the frame of our model, therefore, v_{λ_1} carries information on the collective behavior of the players. We calculated $\lambda_1^x = (-0.14 \pm i 0.11)$ (1/s) and $\lambda_1^y = (-0.19 \pm i 0.04)$ (1/s) (note that complex conjugates are not shown in the plot). To describe these collective modes, let us focus on Fig. 5. Here, panel (b) is linked to the horizontal coordinate, and panel (c) is linked to the vertical one. In the plots, each circle represents the players in its field's natural positions. The circles' radii are proportional to the absolute value of the components of $v_{\lambda_1}^x$ [panel (b), blue] and $v_{\lambda_1}^y$ [panel (c), yellow]. Therefore, the size of the circles in the visualization indicates the effect of the vibration mode on the

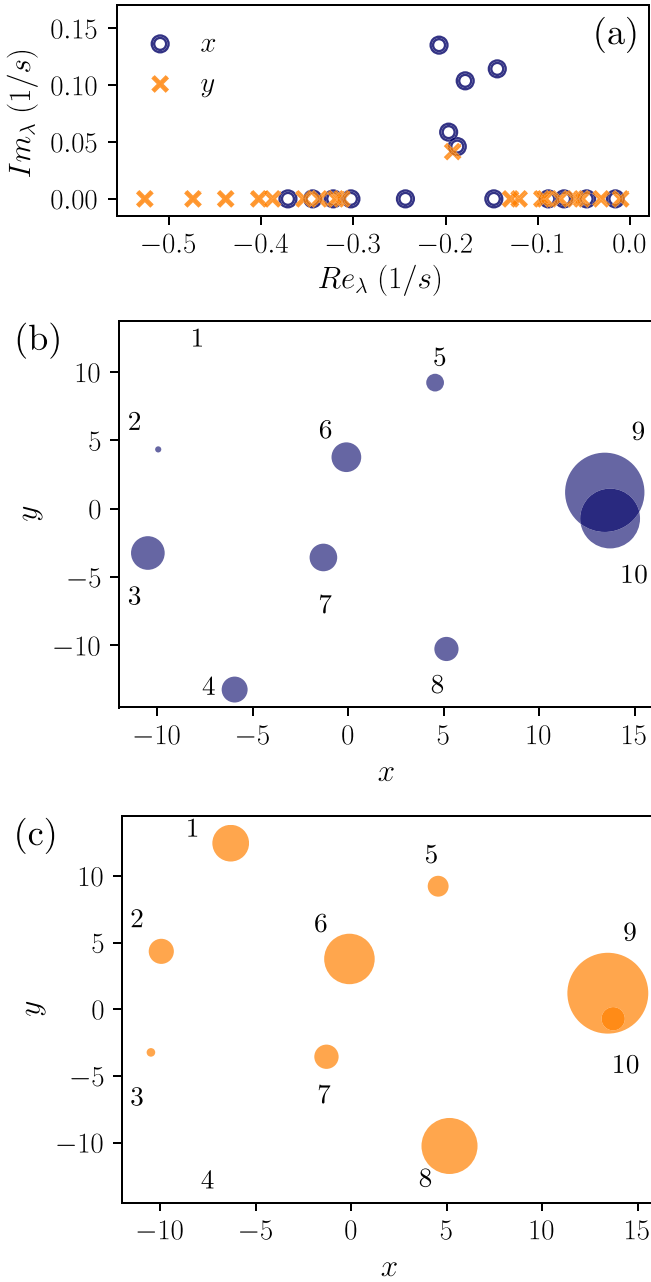


FIG. 5. Collective modes. (a) Eigenvalues plotted as the real part vs the imaginary part, for the horizontal (circles) and the vertical (crosses) axis. (b) Visualization of the first mode in the horizontal axis. (c) Visualization of the first mode in the vertical axis. See the main text for further details.

players, or, in other words, how much the player is involved in this particular collective behavior. For instance, we can see that player 1 is not affected by the mode in the horizontal coordinate, but is highly affected by the mode in the vertical coordinate. In another example, for the case of player 5, we can see that the collective modes in both coordinates have a slight effect on the motion of this player. This is because, as we have previously stated, player 5 seems to be a free player in the field, therefore his maneuvers are not constrained by other players. Conversely, player 9 is the most affected in both coordinates, which seems to indicate that the collective

behavior of the team directly affects the free movement of this particular player. A similar analysis can be performed for every team player in the field.

Let us now focus on using this information to describe the behavior of the defenders and their roles on the team. Figure 5 shows that defenders 1 and 2, at the left-back in the formation, are slightly involved in the collective mode related to the horizontal axis [panel (b)] and highly involved in the mode related to the vertical axis [panel (c)]. This indicates that these players exhibit a natural trend to coordinate with the team in the vertical direction and behave more freely when they perform movements in the horizontal direction. Conversely, players 3 and 4, at the right-back in the formation, are highly involved in the collective mode related to the horizontal axis and slightly involved in the mode related to the vertical axis. This indicates that these defenders exhibit a natural trend to follow the movements of the team in the horizontal axis (towards the goal). These observations reveal a mixed behavior in the defense, where defenders 3 and 4 are more likely to participate in attacking actions, whereas defenders 1 and 2 are more likely devoted to covering gaps. Naturally, an expert coach may easily uncover these kinds of observations while attending a game. However, our technique could be useful for the systematic analysis of hundreds or thousands of games.

The reader may note that the eigenvalues and eigenvectors of the systems provide a handy analytical tool for coaches to assess several aspects of team dynamics. In addition to the case of the hierarchical clustering analysis, in this frame it is possible to link the strengths and weaknesses of a team to collective modes uncovered by this technique, which could be useful in identifying patterns associated with a winning or a losing formation so as to act accordingly in decision-making processes.

3. Using network metrics to analyze the game Tromsø IL versus Anzhi

Datasets $DS3$ and $DS4$ are related to the first and the second half of the game Tromsø IL versus Anzhi. In this game, Anzhi scored a goal at the last minute of the second half to obtain a victory over the local team. In this context, the idea is to fit $DS3$ and $DS4$ to the model, obtain the networks of players defined by parameters \tilde{k}_{nm} , and perform a comparative study of the two cases, analyzing our results by using traditional network science metrics. Let us focus on describing Fig. 6. In panel (a), we show the largest eigenvalue, λ_1 , of the network adjacency matrix in both datasets. This parameter gives information on the network strength [49]. Higher values of λ_1 indicate that important players in the graph are connected among them. We can see that λ_1 in $DS3$ is $\approx 19\%$ higher than in $DS4$, which indicates that the network strength decreases in the second half of the game. In panel (b) we show the algebraic connectivity, $\tilde{\lambda}_2$. This value corresponds to the smallest eigenvalue of the Laplacian matrix of the players' networks [50] and carries information on the structural and dynamical properties of the networks. Small values of $\tilde{\lambda}_2$ indicate the presence of independent groups inside the network, and they are also linked to higher diffusion times, thus indicating a lack of players' connectivity. We can see that $\tilde{\lambda}_2$ decreases $\approx 31\%$ in $DS4$, which seems to indicate that in

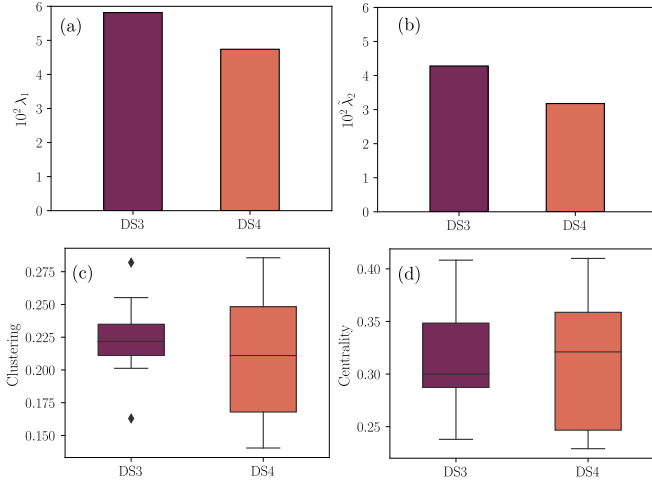


FIG. 6. Comparative analysis for the two halves of the game Tromsø IL vs Anzhi. (a) Largest eigenvalue λ_1 of the adjacency matrix A . (b) Algebraic connectivity, λ_2 , of the Laplacian matrix \tilde{L} . (c) Clustering coefficient. (d) Eigenvector centrality.

the second half of the game the team players lost cohesion. Box plots presented in panel (c) show the weighted clustering coefficient [51] of the team players. This parameter measures the local robustness of the network. We can see that in $DS3$ the clustering is slightly higher than in $DS4$, and the dispersion of the values is lower. This indicates that the network of players is more robust and homogeneous in the first half of the game. Lastly, in panel (d) we show box plots with the eigenvector centrality [52] of the players in both networks. The centrality indicates the influence of a player on the team. A higher value in a particular player is related to strong connections with the other important players on the team. The mean value of the centrality is in both cases ≈ 0.3 and the standard deviation is ≈ 0.05 . The maximum, likewise, is very similar, ≈ 0.4 , and the median, shown in the box plot, is a little lower in $DS3$. We can also observe that the network linked to $DS3$ seems to exhibit more homogeneous values among the team players than the network linked to $DS4$.

In light of the results discussed above, we can see that in the second half of the game the team decreases in connectivity, cohesion, and it becomes more heterogeneous. Considering that Tromsø IL received a goal at the end of the second half

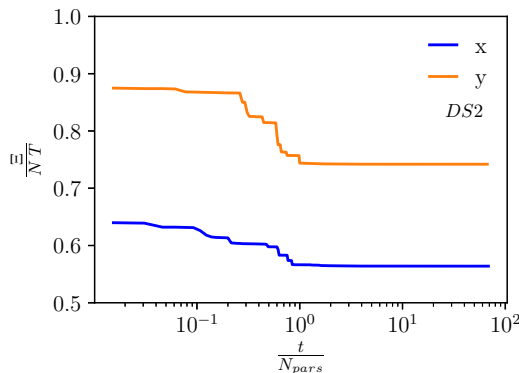


FIG. 7. Evolution of the normalized error during the minimization process.

and lost the game, the fall of these particular metrics seems to be related to a decrease in the team's performance. Previous works devoted to the analysis of passing networks have also reported a relationship between the magnitudes of these particular metrics in these networks and team performances [39]. In this regard, the previous analysis suggests that there is consistency between the results reached through our methods and those previously reported in the literature.

IV. SUMMARY AND CONCLUSIONS

In this work, we studied the spatiotemporal dynamics of a professional football team. Based on empirical observations, we proposed to model the player cooperative interactions to describe the global behavior of the group. In this section, we summarize our main results.

First, we surveyed a database containing body-sensor traces from one team on three professional soccer games. We observed statistic regularities in the dynamics of the games that reveal the presence of a strong correlation in the players' movements. With this insight, we proposed a model for the team's dynamic consisting of a fully connected system where the players interact with each other following linear-springlike forces. In this frame, we performed a minimization process to obtain the parameters that fit the model to the datasets. Furthermore, we showed that is possible to treat the higher-order contributions as stochastic forces in the players' velocities, which we modeled as Gaussian fluctuations.

Second, once we defined the model, we carried out numerical simulations and evaluated the model performance by comparing the outcomes with the empirical data. We showed that the model generates spatiotemporal dynamics that give a good approximation to the real observations. In particular, we analyzed (i) the probability of finding a player in a position (x, y) , (ii) the action zones of the players, and (iii) the temporal structure of the system by studying the time to return to the mean value in the temporal series of the kinetic energy of the system. Despite its simplicity, in all the cases the model exhibited a good performance.

Third, we described the system at the local level by using the parameters we obtained from the minimization process. In this regard, we proposed to use two analytical tools, namely a hierarchical cluster classification and an eigenvalues-eigenvectors-based analysis. We found that it is possible to describe the team behavior at several organization levels and to uncover nontrivial collective interactions.

Lastly, we used network science metrics to carry out a comparative analysis on the two halves of the game Tromsø IL versus Anzhi. We observed that a decrease in connectivity and cohesion, and an increase in the heterogeneity of the network of players, seem to be related to a decrease in the team performance.

We consider this contribution to be a new step towards a better understanding of the game of football as a complex system. The proposed stochastic model, based on linear interactions, is simple and can be easily understood in the frame of standard dynamical variables. Moreover, our framework provides a handy analytical tool to analyze and evaluate tactical aspects of teams, something helpful to support the decision-making processes that coaches face in their working

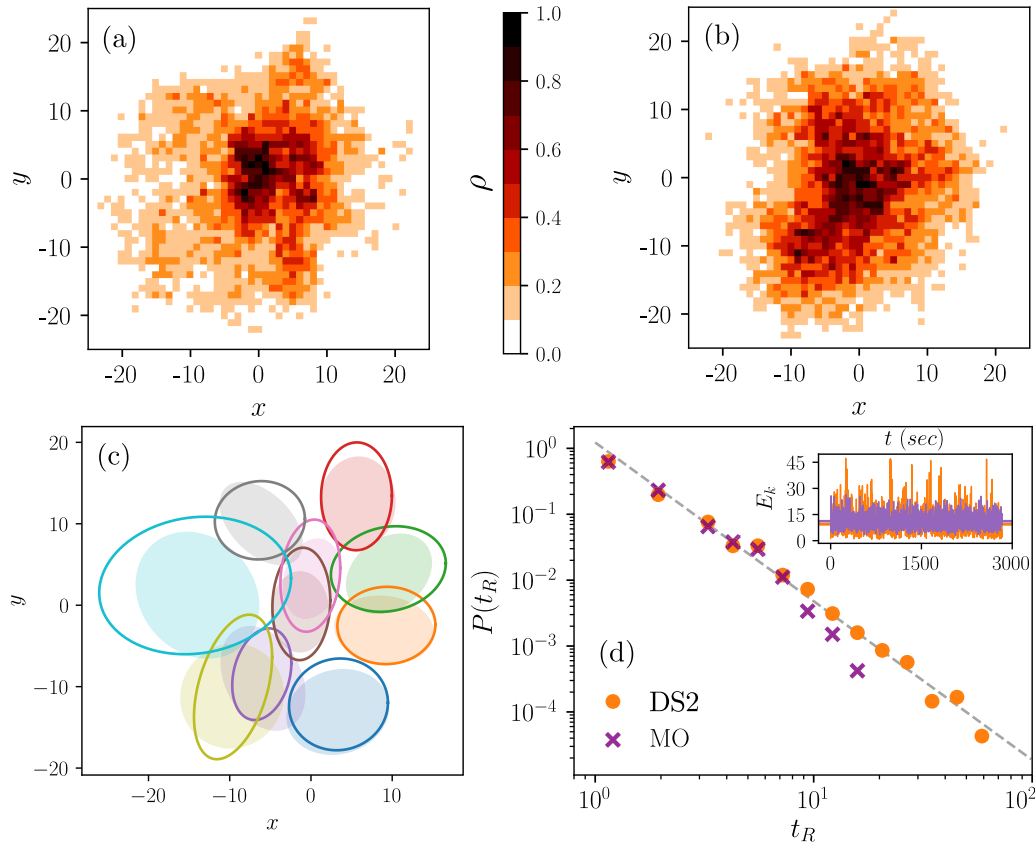


FIG. 8. Collective dynamics simulations corresponding to *DS2*. (a),(b) Probability of finding a player in the field, in empirical data and simulations, respectively. (c) Players’ action zones. Empirical data (shadow areas) compared with simulations (curves). (d) Probability distribution of the time to return to the mean value, $P(t_R)$. The dashed line indicates a nonlinear fit performed to the empirical data (yellow circles). The inset shows the evolution of the kinetic energy, $E_k(t)$, from where t_R is measured.

activities. It is important to highlight that our framework is not limited to being used only in the analysis of football games. A similar approach can be performed to study others sport disciplines, mainly when the evaluation of players’ interactions is key to understanding the game results. In [53], for instance, the authors use players-tracking data in basketball games to estimate the expected number of points obtained by the end of a possession. In this case, a complementary analysis within our framework could also unveil collective behavior patterns linked to players’ coordination interactions that can be correlated to the upshot at the end of the possession intervals.

However, we point out that the full dynamics of a football match (and other sports) cannot be addressed by only analyzing cooperative aspects within a particular team. To describe the full dynamics, we should also model the competitive interactions among the players in both teams. To do so, we need to measure the interplay among rivals, which in our framework implies having body-sensor traces for both teams, something that football clubs reject because of competing interests. Moreover, it could be useful to have a record of the ball position to improve our analysis. In this context, the use of alternative measurement techniques based on artificial intelligence and visual recognition becomes relevant [54–56].

To summarize, our model provides a simple approach to describe the collective dynamics of a football team untan-

gling interactions among players, and stochastic inputs. The structure of interactions that results from this approach can be considered a new metric for this sport. In this sense, our analysis complements recent contributions in the framework of network science [20,37–41]. Note that there is a major difference between our approach and current network-science-oriented methods: The latter analyze interaction based on players’ passes, whereas our approach analyzes interaction based on players’ movements. In this regard, the problem with studying a football team from only passing/pitch networks is that this approach is entirely based on on-ball action, which completely neglects how players behave when they are far from the center of the plays (off-ball actions). Our approach, instead, integrates the information of the entire team to calculate every single link between players, therefore in our model we also consider off-ball actions, which is key to correctly evaluating a team’s performance [57]. In addition, many of the metrics of current use in this sport can be tested through our model [28–33]. On the other hand, to perform an analysis based on passing/pitch networks, one must have a record of the ball position and be able to characterize events (passes) during the match. Our model, instead, employs just the data of players’ positions, something that nowadays can be easily measured, both in training sessions and competitive scenarios, with the simplest GPS trackers available on the market.

Finally, we consider that it is possible to improve our model by incorporating nonlinear interactions in the equations of motion. An analysis of the most commonly observed players' maneuvers may help to find new statistical patterns to be used as an insight to propose a nonlinear approach. In this regard, we leave the door open to future research projects in the field.

ACKNOWLEDGMENTS

This work was partially supported by grants from CONICET (PIP 112 20150 10028), FonCyT (PICT-2017-0973), SeCyTUNC (Argentina), and MinCyT Córdoba (Grant No. PID PGC 2018).

APPENDIX A: THE MINIMIZATION PROCESS

As we explained in Sec. II C 2, to obtain the model's parameters, we define the error $\xi_n(t) := \vec{v}_n(t + \Delta t) - \vec{v}_n(t + \Delta t)'$, and we fit γ_n , k_{an} , and k_{nm} by minimizing the sum $\Xi = \sum_t \sum_n |\xi_n(t)|$. To do so, we coded a python script based on the use of the *scipy.optimize.minimize* routine (see the online documentation [58]). In this script, we proceed as follows:

- (i) At the beginning, we set the parameters equal to zero.
- (ii) We randomly choose one of the parameters and minimize Ξ via only this parameter. If the minimization succeeds, we keep it.

(iii) We randomly choose another parameter and repeat the previous step. Note that now we are using the previously fitted parameter to fit the new one.

(iv) Repeat the previous steps M times, with $M \gg N_{\text{pars}}$, where $N_{\text{pars}} = 65$ is the number of parameters.

(v) After M steps, we fit the entire set of parameters simultaneously using as seeds the values obtained in the previous steps.

In Fig. 7, we show the evolution of the normalized total error as a function of the normalized number of parameters fitted in the process described above. Here we focus on the case of DS2. Notice that N is the number of players in the team, T is the total match time, and N_{pars} is the number of model parameters. We can see that after fitting the total set of parameters by performing steps (i)–(iv), the error enters a plateau indicating that the system reaches a stable set of optimized values. After this process, step (v) is performed to achieve a joint optimization.

APPENDIX B: COLLECTIVE DYNAMICS SIMULATION RELATED TO DS2, DS3, AND DS4

For DS2, DS3, and DS4, we simulated the players collective dynamics and compared the results with empirical data in order to assess the model performance. The results are shown in Figs. 8, 9, and 10. The panels follow all the same pattern:

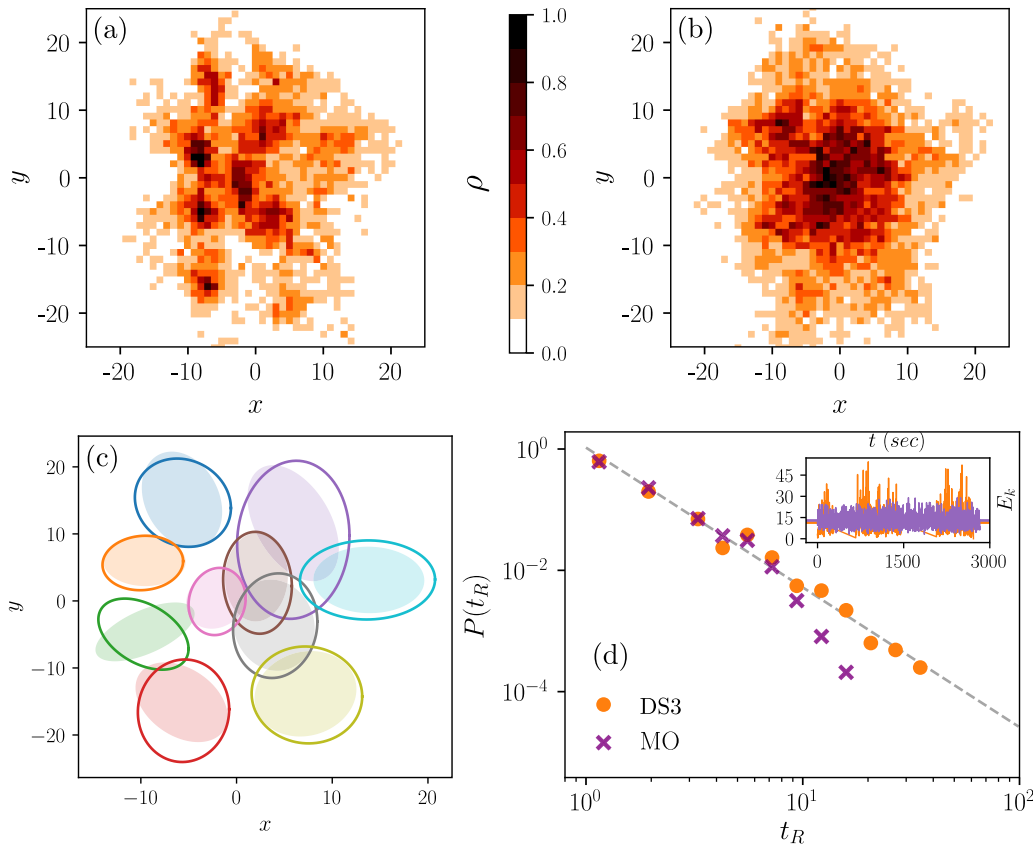


FIG. 9. Collective dynamics simulations corresponding to DS3. (a),(b) Probability of finding a player in the field, in empirical data and simulations, respectively. (c) Players' action zones. Empirical data (shadow areas) compared with simulations (curves). (d) Probability distribution of the time to return to the mean value, $P(t_R)$. The dashed line indicates a nonlinear fit performed to the empirical data (yellow circles). The inset shows the evolution of the kinetic energy, $E_k(t)$, from where t_R is measured.

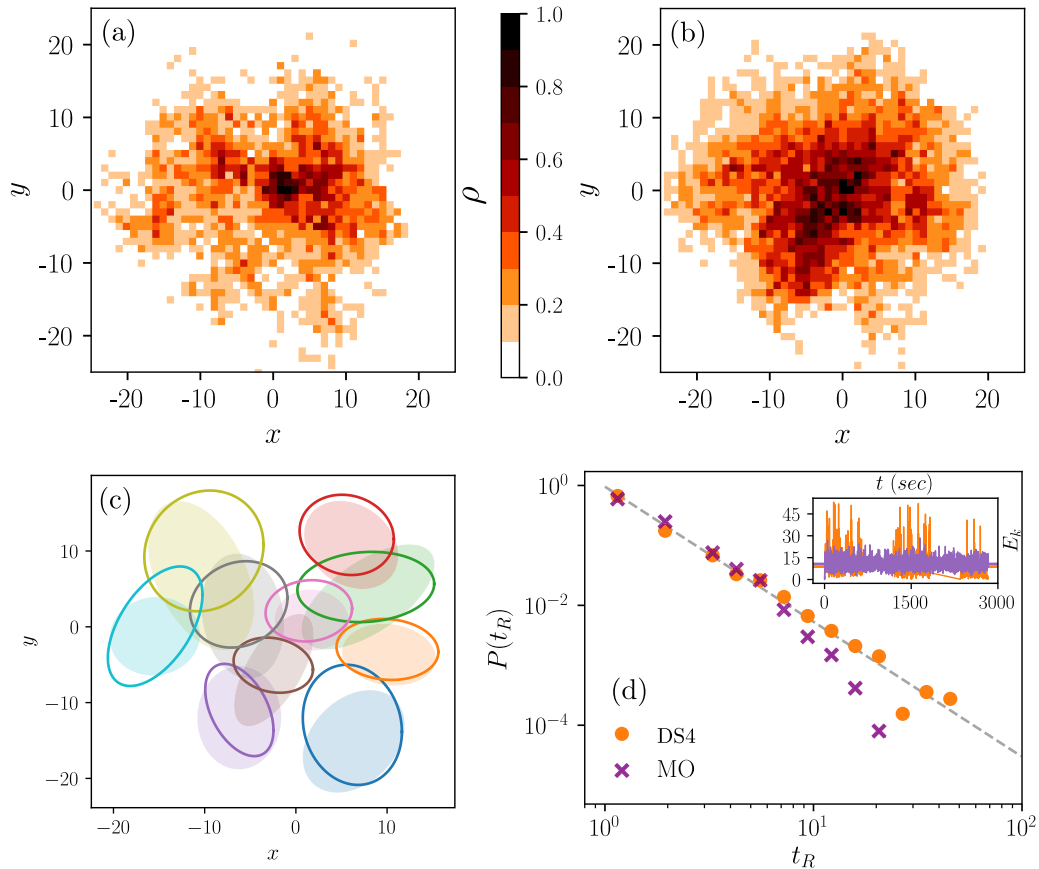


FIG. 10. Collective dynamics simulations corresponding to DS4. (a),(b) Probability of finding a player in the field, in empirical data and simulations, respectively. (c) Players’ action zones. Empirical data (shadow areas) compared with simulations (curves). (d) Probability distribution of the time to return to the mean value, $P(t_R)$. The dashed line indicates a nonlinear fit performed to the empirical data (yellow circles). The inset shows the evolution of the kinetic energy, $E_k(t)$, from where t_R is measured.

Panels (a) and (b) show two heatmaps with the probability of finding a team player in the position (x, y) . The left panel shows the results for the empirical data, while the right panel shows the results for simulations. For a better visualization, in both cases the probabilities were normalized to the maximum value, defining the parameter $\rho \in (0, 1)$. In panel (c), we compare players’ action zones. The empirical observations are the shadow ellipses, whereas the simulations are the curves. In panel (d), we analyze the kinetic energy of the system, $E_k := \sum_n \frac{1}{2} |v_n|^2$. Our goal here is to globally describe the temporal structure of the system. In the inset, we show the temporal evolution. We measured $\langle E_k \rangle_{DS1} = 10.3$ for the data, and $\langle E_k \rangle_{MO} = 11.5$ for simulations. The main plot in the panel

shows the distribution of the time to return to the mean value, $P(t_R)$. We can see that in both cases, $P(t_R) \propto t_R^{-\gamma_{E_k}}$. The values of the exponents for the three cases are $\gamma_{E_k}^{DS2} = 2.41 \pm 0.07$, $\gamma_{E_k}^{DS3} = 2.3 \pm 0.1$, and $\gamma_{E_k}^{DS4} = 2.2 \pm 0.1$. Note that the values are similar to the value obtained for DS1 (see Sec. III A).

The results in all the cases agree with the result for DS1. Here, it is important to point out that in the datasets analyzed in this section, there are periods during the match when the body-sensor traces were not recorded. This lack of information might have a direct effect on the process we used to obtain the parameters. However, despite the simplicity of the model and the lack of available data to fit the model, we can conclude that it reproduces the empirical data reasonably well.

[1] N. Almeida, A. L. Schaigorodsky, J. I. Perotti, and O. V. Billoni, Structure constrained by metadata in networks of chess players, *Sci. Rep.* **7**, 1 (2017).
 [2] A. L. Schaigorodsky, J. I. Perotti, and O. V. Billoni, Memory and long-range correlations in chess games, *Physica A* **394**, 304 (2014).
 [3] J. I. Perotti, H.-H. Jo, A. L. Schaigorodsky, and O. V. Billoni, Innovation and nested preferential growth in chess playing behavior, *Europhys. Lett.* **104**, 48005 (2013).

[4] Ş. Erkol and F. Radicchi, Who is the best coach of all time? a network-based assessment of the career performance of professional sports coaches, *J. Complex Netw.* **9**, cnab012 (2021).
 [5] L. Pappalardo, P. Cintia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi, and F. Giannotti, A public data set of spatio-temporal match events in soccer competitions, *Sci. Data* **6**, 1 (2019).
 [6] S. A. Pettersen, D. Johansen, H. Johansen, V. Berg-Johansen, V. R. Gaddam, A. Mortensen, R. Langseth, C. Griwodz, H. K. Stensland, and P. Halvorsen, Soccer video and player position

- dataset, in *Proceedings of the 5th ACM Multimedia Systems Conference* (ACM, New York, 2014), pp. 18–23.
- [7] I. Fister Jr., K. Ljubič, P. N. Suganthan, M. Perc, and I. Fister, Computational intelligence in sports: Challenges and opportunities within a new research domain, *Appl. Math. Comput.* **262**, 178 (2015).
 - [8] T. Neiman and Y. Loewenstein, Reinforcement learning in professional basketball players, *Nat. Commun.* **2**, 1 (2011).
 - [9] S. Mukherjee, Y. Huang, J. Neidhardt, B. Uzzi, and N. Contractor, Prior shared success predicts victory in team competitions, *Nat. Human Behav.* **3**, 74 (2019).
 - [10] K. Laksari, M. Kurt, H. Babae, S. Kleiven, and D. Camarillo, Mechanistic Insights into Human Brain Impact Dynamics Through Modal Analysis, *Phys. Rev. Lett.* **120**, 138101 (2018).
 - [11] F. Schade and A. Arampatzis, Influence of pole plant time on the performance of a special jump and plant exercise in the pole vault, *J. Biomech.* **45**, 1625 (2012).
 - [12] M. Le Berre and Y. Pomeau, Theory of ice-skating, *Int. J. Non-Lin. Mech.* **75**, 77 (2015).
 - [13] S. S. Yu, S. Zhang, Z. W. Xia, S. Liu, H. J. Lu, and X. T. Zeng, Textured hybrid nanocomposite coatings for surface wear protection of sports equipment, *Surf. Coat. Technol.* **287**, 76 (2016).
 - [14] H. Trenchard, A. Richardson, E. Ratamero, and M. Perc, Collective behavior and the identification of phases in bicycle pelotons, *Physica A* **405**, 92 (2014).
 - [15] A. Clauset, M. Kogan, and S. Redner, Safe leads and lead changes in competitive team sports, *Phys. Rev. E* **91**, 062815 (2015).
 - [16] H. V. Ribeiro, S. Mukherjee, and X. H. T. Zeng, Anomalous diffusion and long-range correlations in the score evolution of the game of cricket, *Phys. Rev. E* **86**, 022102 (2012).
 - [17] D. P. Kiley, A. J. Reagan, L. Mitchell, C. M. Danforth, and P. S. Dodds, Game story space of professional sports: Australian rules football, *Phys. Rev. E* **93**, 052314 (2016).
 - [18] P. E. Ruth and J. G. Restrepo, Dodge and survive: Modeling the predatory nature of dodgeball, *Phys. Rev. E* **102**, 062302 (2020).
 - [19] A. Chacoma, N. Almeida, J. I. Perotti, and O. V. Billoni, Modeling ball possession dynamics in the game of football, *Phys. Rev. E* **102**, 042120 (2020).
 - [20] K. Yamamoto and T. Narizuka, Preferential model for the evolution of pass networks in ball sports, *Phys. Rev. E* **103**, 032302 (2021).
 - [21] K. Davids, R. Hristovski, D. Araújo, N. B. Serre, C. Button, and P. Passos, *Complex Systems in Sport* (Routledge, England, UK, 2013).
 - [22] A. Hulme, S. McLean, G. J. M. Read, C. Dallat, A. Bedford, and P. M. Salmon, Sports organisations as complex systems: Using cognitive work analysis to identify the factors influencing performance in an elite netball organisation, *Front. Sports Active Liv.* **1**, 56 (2019).
 - [23] A. M. Petersen and O. Penner, Renormalizing individual performance metrics for cultural heritage management of sports records, *Chaos Solitons Fractals* **136**, 109821 (2020).
 - [24] S. McLean, P. M. Salmon, A. D. Gorman, G. J. M. Read, and C. Solomon, What's in a game? a systems approach to enhancing performance analysis in football, *PLoS One* **12**, e0172565 (2017).
 - [25] S. Soltanzadeh and M. Mooney, Systems thinking and team performance analysis, *Int. Sport Coaching J.* **3**, 184 (2016).
 - [26] J. Gudmundsson and M. Horton, Spatio-temporal analysis of team sports, *ACM Comput. Surv.* **50**, 1 (2017).
 - [27] B. Gonçalves, R. Marcelino, L. Torres-Ronda, C. Torrents, and J. Sampaio, Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games, *J. Sports Sci.* **34**, 1346 (2016).
 - [28] M. F. Clemente, S. M. Couceiro, F. M. L. Martins, R. Mendes, and A. J. Figueiredo, Measuring collective behaviour in football teams: Inspecting the impact of each half of the match on ball possession, *Int. J. Perf. Anal. Sport* **13**, 678 (2013).
 - [29] W. Frencken, K. Lemmink, N. Delleman, and C. Visscher, Oscillations of centroid position and surface area of soccer teams in small-sided games, *Eur. J. Sport Sci.* **11**, 215 (2011).
 - [30] F. A. Moura, L. E. B. Martins, R. O. Anido, P. R. C. Ruffino, R. M. L. Barros, and S. A. Cunha, A spectral analysis of team dynamics and tactics in brazilian football, *J. Sports Sci.* **31**, 1568 (2013).
 - [31] Z. Yue, H. Broich, F. Seifriz, and J. Mester, Mathematical analysis of a soccer game. part i: Individual and collective behaviors, *Stud. Appl. Math.* **121**, 223 (2008).
 - [32] Z. Yue, H. Broich, F. Seifriz, and J. Mester, Mathematical analysis of a soccer game. part ii: Energy, spectral, and correlation analyses, *Stud. Appl. Math.* **121**, 245 (2008).
 - [33] T. Narizuka and Y. Yamazaki, Clustering algorithm for formations in football games, *Sci. Rep.* **9**, 1 (2019).
 - [34] R. S. Mendes, L. C. Malacarne, and C. Antenodo, Statistics of football dynamics, *Eur. Phys. J. B* **57**, 357 (2007).
 - [35] A. Kijima, K. Yokoyama, H. Shima, and Y. Yamamoto, Emergence of self-similarity in football dynamics, *Eur. Phys. J. B* **87**, 41 (2014).
 - [36] T. Narizuka and Y. Yamazaki, Statistical properties for directional alignment and chasing of players in football games, *Europhys. Lett.* **116**, 68001 (2017).
 - [37] J. H. Martínez, D. Garrido, J. L. Herrera-Diestra, J. Busquets, R. Sevilla-Escoboza, and J. M. Buldú, Spatial and temporal entropies in the spanish football league: A network science perspective, *Entropy* **22**, 172 (2020).
 - [38] J. L. Herrera-Diestra, I. Echegoyen, J. H. Martínez, D. Garrido, J. Busquets, F. S. Io, and J. M. Buldú, Pitch networks reveal organizational and spatial patterns of guardiolas fc barcelona, *Chaos Solitons Fractals* **138**, 109934 (2020).
 - [39] J. M. Buldu, J. Busquets, I. Echegoyen *et al.*, Defining a historic football team: Using network science to analyze guardiolas fc barcelona, *Sci. Rep.* **9**, 1 (2019).
 - [40] B. Gonçalves, D. Coutinho, S. Santos, C. Lago-Penas, S. Jiménez, and J. Sampaio, Exploring team passing networks and player movement dynamics in youth association football, *PLoS One* **12**, e0171156 (2017).
 - [41] D. Garrido, D. R. Antequera, J. Busquets, R. L. D. Campo, R. R. Serra, S. J. Vielcazat, and J. M. Buldú, Consistency and identifiability of football teams: A network science perspective, *Sci. Rep.* **10**, 19735 (2020).
 - [42] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.104.024110> for further details on the calculation of the standard deviation intervals around the mean and to access the model parameters for $DS2$, $DS3$, and $DS4$.

- [43] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Scale-free correlations in starling flocks, *Proc. Natl. Acad. Sci. (USA)* **107**, 11865 (2010).
- [44] M. Welch, T. M. Schaerf, and A. Murphy, Collective states and their transitions in football, *PLoS One* **16**, e0251970 (2021).
- [45] S. Redner, *A Guide to First-passage Processes* (Cambridge University Press, Cambridge, 2001).
- [46] J. H. Ward Jr., Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.* **58**, 236 (1963).
- [47] S. H. Strogatz, *Nonlinear Dynamics and Chaos with Student Solutions Manual: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC, Boca Raton, FL, 2018).
- [48] A. H. Nayfeh and B. Balachandran, *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods* (Wiley, New Jersey, 2008).
- [49] J. Aguirre, D. Papo, and J. M. Buldú, Successful strategies for competing networks, *Nat. Phys.* **9**, 230 (2013).
- [50] M. Newman, *Networks* (Oxford University Press, Oxford, 2018).
- [51] S. E. Ahnert, D. Garlaschelli, T. M. A. Fink, and G. Caldarelli, Ensemble approach to the analysis of weighted networks, *Phys. Rev. E* **76**, 016101 (2007).
- [52] M. E. J. Newman, The mathematics of networks, *New Palgrave Encycl. Econ.* **2**, 1 (2008).
- [53] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, A multiresolution stochastic process model for predicting basketball possession outcomes, *J. Am. Stat. Assoc.* **111**, 585 (2016).
- [54] R. Sanford, S. Gorji, L. G. Hafemann, B. Pourbabae, and M. Javan, Group activity detection from trajectory and video data in soccer, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, Piscataway, NJ, 2020), pp. 898–899.
- [55] Y. Ganesh, A. S. Teja, S. K. Munnangi, and G. R. Murthy, A novel framework for fine grained action recognition in soccer, in *International Work-Conference on Artificial Neural Networks* (Springer, Heidelberg, 2019), pp. 137–150.
- [56] A. Khan, B. Lazzerini, G. Calabrese, and L. Serafini, Soccer event detection, in *4th International Conference on Image Processing and Pattern Recognition (IPPR 2018)* (AIRCC, India, 2018), pp. 119–129.
- [57] C. A. Casal, R. Maneiro, T. Ardá, F. J. Marí, and J. L. Losada, Possession zone as a performance indicator in football. the game of the best teams, *Front. Psych.* **8**, 1176 (2017).
- [58] SciPy v1.6.2 Reference Guide, <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>.