Highlights

Interactions Between Soccer Teams Reveal Both Design and Emergence: Cooperation, Competition and Zipf-Mandelbrot Regularity

João Paulo Ramos,Rui J. Lopes,Duarte Araújo

- Soccer matches interactions reveal regularity properties of complex social systems

- Frequency of players' sets occurrence exhibit Zipf-Mandelbrot scaling properties

- Design is present in the exceptions to Zipf-Mandelbrot, expressing specific sets arrangement

- The two sets Goalkeeper, Goal are exceptions to Zipf-Mandelbrot regularities

- Specific sets reveal pre-defined teams' strategical constraints

Interactions Between Soccer Teams Reveal Both Design and Emergence: Cooperation, Competition and Zipf-Mandelbrot Regularity

João Paulo Ramos[a,*], Rui J. Lopes[b,c] and Duarte Araújo[d]

[a]*Universidade Europeia, Laureate International Universities, Estrada da Correia 53, 1500-210 Lisboa, Lisbon, Portugal*

[b]*Iscte - Instituto Universitário de Lisboa, Avenida das Forças Armadas, 1649-026 Lisboa, Portugal*

[c]*Instituto de Telecomunicações, Torre Norte - Piso 10, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal*

[d]*CIPER, Faculdade de Motricidade Humana, Universidade de Lisboa, Estrada da Costa, 1499-002 Cruz Quebrada - Dafundo, Portugal*

ARTICLE INFO

Keywords:
Multilevel Hypernetworks
Zipf-Mandelbrot Regularity
Design
Soccer Performance Analysis

ABSTRACT

Considering soccer matches as complex systems facilitates the identification of properties that emerge from the interactions between players. Such properties include the regularities and statistical patterns that characterize couplings and sets between players established during matches. Empirical studies on the statistical distributions of number of items (e.g., words in texts) have shown that these distributions follow scaling properties according to empirical laws known as Zipf-Mandelbrot. Here we investigate whether the (re)occurrence of pitch location of sets of players in a soccer match also obeys these empirical laws. Data collected from 10 soccer matches shows that, for most sets of players, this seems to be the case. Exceptions were found in particular types of sets, such as goalkeeper and goal, and left defender and right attacker from opposite teams. Rather than challenging the hypothesis that a Zipf-Mandelbrot law defines this system, these exceptions may be explained by the players configuration design, which is a trait of human interaction within complex systems. This design expresses match strategy, before the team enters in such dynamical processes (the game).

1. Introduction

Several approaches have been used to study complex systems, including the identification of well-known complexity features Bar-Yam (2002); Kobayashi, Kuninaka, Wakita and Matsushita (2011); Juarrero (2000); Silva, Vilar, Davids, Araújo and Garganta (2016b). For instance, in social complex systems specific behaviors emerge from self-organization Schmidt and Richardson (2008); Passos, Araújo and Davids (2013); Passos, Davids, Araújo, Paz, Minguéns and Mendes (2011); Araújo and Davids (2016). Self-organization results in most cases from the interaction between multiple parts in a system. An interesting and extensively investigated aspect of self-organization is the emerging exchange of information (e.g. verbal and non-verbal communication and their statistical properties) between people interacting in a given system Riley, Richardson, Shockley and Ramenzoni (2011). Typically these communication processes are based on cooperative interactions such as synergistic relations, but also on confrontation, which is a non-cooperative type of interaction.

Here we investigated how these processes and interactions are expressed in team sports. In particular, we asked whether soccer matches have hallmark features of other complex systems. Moreover, we assessed the influence of pre-defined design on the cooperative and competitive interactions between players. Notably, explicit inter-dependency between player-opponent behaviors is typically observed. The communication processes between players in soccer matches are often visually based and reflected in the players' moves and interpersonal spatial relationships Schmidt and Richardson (2008); Riley et al. (2011); Schmidt and Fitzpatrick (2016); Ramos, Lopes, Marques and Araújo (2017b); Silva, Chung, Carvalho, Cardoso, Davids, Araújo and Garganta (2016a). A key feature of this self-organized behavior is how players become synchronized by maintaining a perceptual link granted by their spatial proximity. The proximity-based sets thus formed may have different dimensions, both in terms of the number of players in the set and how they are organized (i.e., the team each player belongs to). According to hypernetwork theory, each player set with

---

*Corresponding author

✉ joao.ramos@universidadeeuropeia.pt (J.P. Ramos)

ORCID(s): 0000-0002-5079-684X (J.P. Ramos); 0000-0002-8943-0415 (R.J. Lopes); 0000-0001-7932-3192 (D. Araújo)
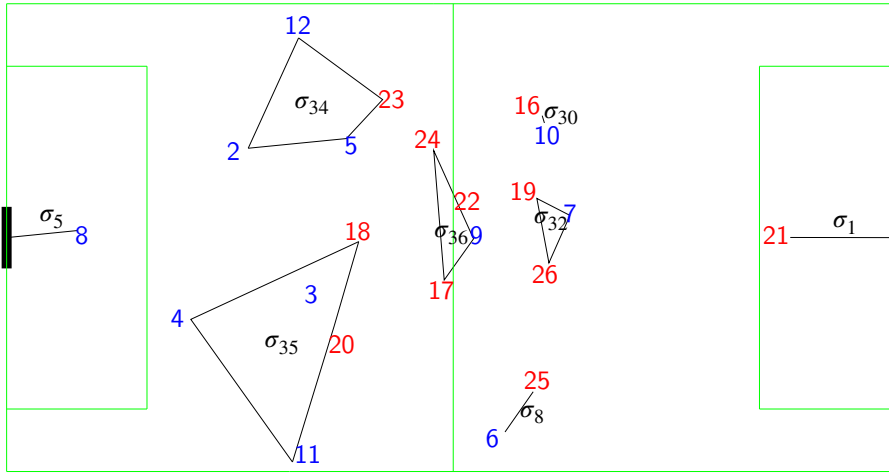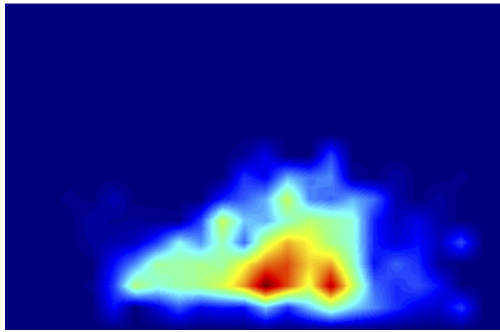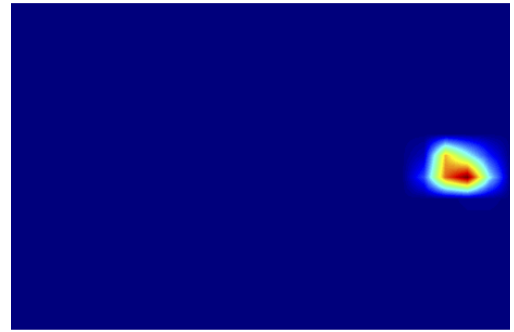
Figure 1: Players' location and simplices at frame $t = 00m : 10s$

(a) Simplex σ_{35} (3vs.2)

(b) Simplex σ_1 (B goalkeeper and goal)

Figure 2: Simplices' location heat maps

their corresponding inter-relationships form a *simplex* – plural, *simplices* – of players, representing the *n-ary* spatial interactions between at least two spatially connected players Johnson (2005b); Ramos, Lopes, Marques and Araújo (2017c). Figure 1 shows the simplices found at a particular time frame $t = 00m : 10s$ in a soccer match.

In the present study, each simplex is represented by a spatial convex hull enclosing the players in the set. For instance, simplex σ_{35}, is composed of players 3, 4, and 11 from team A (blue) and players 18 and 20 from team B (red), thus forming a 3 vs. 2 simplex Johnson (2013); Ramos et al. (2017c). By using the temporal aggregate of the geometrical center for each simplex convex hull, a spatial histogram for that simplex can be obtained. Figures 2a and 2b illustrate two of these histograms using spatial heat maps, for simplex σ_1 and σ_{35}, respectively. Simplices σ_1 and σ_5 represent the particular type of relationship, that exists between a player, the goalkeeper and the goal.

We asked whether soccer match histograms exhibit the scaling properties of other human and natural phenomena typically described by power law type models. These power laws are common signatures of chaotic processes which at one point, become self-organized, as it happens in many natural and social systems. Typical examples can be found in population distributions of big cities Izsák (2006); Li and Wang (2019); Malacarne and Mendes (2000), forest fires Malacarne and Mendes (2000), forest patch sizes Saravia, Doyle and Bond-Lamberty (2018), scientific citations Malacarne and Mendes (2000); Silagadze (1997); Komulainen (2004), WWW surfing Malacarne and Mendes (2000), ecology Malacarne and Mendes (2000); Ferrer-i Cancho and Elvevåg (2010); Izsák (2006), solar flares Malacarne and Mendes (2000), economic index Malacarne and Mendes (2000), epidemics in isolated populations Malacarne and Mendes (2000), among others. The Zipf empirical law is an example of such power laws, and its generalization by Mandelbrot. In verbal communication processes, including natural language and written texts, several studies have

shown that the frequency of word occurrence follows the Zipf power law. Indeed, the corpora of texts and languages have few words that are very frequent (e.g. "a", "the", "I", etc.) and many words which seldom occur. In Zipf's empirical law model, given the item (e.g., word) frequency, $f(r)$, and order by their assigned rank, r in decreasing order (rank 1 is for the most frequent word; rank 2 is for the second most frequent word, etc Ferrer-i Cancho and Elvevåg (2010)), occurrence frequency decays linearly as the rank increases on a double logarithmic scale, as expressed in equation 1

$$f(r) \propto \frac{1}{r^{\alpha}}$$
(1)

The generalization of this law conducted by MandelbrotPiantadosi (2014) has a better fit to empirical data 2

$$f(r) \propto \frac{1}{(\beta + r)^{\alpha}}$$
(2)

Being a generalization, the latter is also referred to as the Zipf-Mandelbrot (ZM) law.

Soccer performance features, such as goal scoring distribution, also exhibit these statistical regularities related to power laws. By computing goal distribution of several main league soccer championships such as Brazil, England, Italy and Spain, it was shown that while there are very few top-scorers, many players score only a few goals Malacarne and Mendes (2000). Rather than focusing on these performance metrics or on player individual behavior, we assessed interpersonal relationships as expressed by simplices' sets, which correspond to the meso-scale system properties. Questions addressed at this level typically concern processes Johnson (2005b); Ramos, Lopes and Araújo (2017a) and therefore aim to explain the mechanisms underlying particular simplices' set occurrence distributions Ramos et al. (2017b).

Several studies have asked whether Zipf's empirical law can be observed in purely random systems Ferrer-i Cancho and Elvevåg (2010); Wentian (1992) by investigating the processes that may drive these particular statistical distributions. We addressed a similar question from a different angle: by analyzing the co-design expressed in the match strategy. Despite the uncertainty of human collective behavior, and therefore of the impossibility to predict the future state of complex systems, the deliberate design of the social structures forming a system may promote specific desired behaviors Johnson (2013, 2010); Blecic and Cecchini (2008); Johnson (2008). This design is in most cases a collaborative or cooperative process Johnson (2005a).

In soccer matches, one can consider that the artificiality (i.e. the design expressed via strategic behaviors) of these social complex systems is related to specific outcomes Johnson (2013, 2008, 2010). A very relevant aspect of the coaching process in team sports is the implementation of the design Rothwell, Davids and Stone (2018). A particular challenge in studying soccer matches as social systems is that the design results from both cooperative and competitive interactions Blecic and Cecchini (2008). The most frequent simplices, which are those that seem to persist over the entire match, must therefore be a consequence of this design (i.e., the strategy of the team).

The theoretical prediction that a macro, strategic, level of organization influences the micro, local, behavior and vice -versa was already highlighted in sport sciences Araújo and Davids (2016); Ribeiro, Davids, Duarte Araújo, Silva and Garganta (2019), but without a clear empirical demonstration of its effect.

When a team distributes their players in the pitch (i.e., the team strategy; considering attacking, defending, midfielders, goalkeepers and left, right or in the center of the pitch), they naturally stand near their symmetric opponents. For instance, the right attacker from team A vs. the left defender from team B. As these positions must be maintained during most of the match, sets of opposing players from team A vs. team B (e.g., 1 vs. 1) occur frequently throughout the match. These sets may also depend on the pitch area, such as: i) the simplex set <Goalkeeper, Goal > near the goal, corresponding to players with a very specific and narrow purpose Blecic and Cecchini (2008); ii) the defending team trying to have numeric supremacy closer to its box (e.g., simplex σ_{32} in Figure 1)Blecic and Cecchini (2008).

The main questions addressed here refer to the simplices' set occurrence distribution. Specifically, we ask whether the Zipf-Mandelbrot law (ZM) fits the empirical distributions from soccer matches and if design (i.e. match strategy) has an impact on the simplices' set statistical distributions.

2. Methods

2.1. Raw data: players' coordinates

We analyzed two-dimensional displacement coordinates from 22 players provided by PROZONE (now STATS STATS (2020a)) for 10 matches (five at home and five away) of a focus team (team A) during the 2010/2011 English Premier League Season. This data was obtained via a PROZONE semi-automatic tracking system based on multiple-camera analysis. The position of the 22 players during the match was estimated based on the synchronized video files from eight cameras placed on the top of the stadium operating at a frequency of 10Hz (i.e., 10 frames per second, producing about 54000 frames per match) STATS (2020b). The player substitutions and sent-offs were also considered using ancillary descriptions of the match, e.g., commentary metadata.

2.2. Building of simplices' sets and heat maps

For each frame, the 22 players in the pitch and the two goals were organized in sets (simplices sets), according to the computational procedure adopted by Ramos and colleagues Ramos et al. (2017c). The criteria for selecting the players (or goals) for each set were based only in spatial proximity. The two goals were also considered in the simplices formation, as they act as special spatial references to the players, namely the goalkeepers. Figure 1 illustrates the player and goal positions at a particular time frame. Players and goals in the same simplex set are connected within their convex hull. Each simplex is uniquely defined by its index, i and by its element set, σ_i, such that: $\sigma_i = \sigma_j \implies i = j$. For each frame, t, Σ_t is the set of all simplices' sets that are found in that frame.

ZM model, ranking and bootstrapping the simplices sets

Zipf and Mandelbrot empirical laws relate token values and their rank using, equation 1 and 2, respectively. These laws can also be expressed by a probability density function 3.

$$p_{\theta,r} = \frac{C}{(\beta + r)^\alpha} \tag{3}$$

Where $p_{\theta,r}$ is the probability density value for token with rank r under parameter set $\theta = \{\alpha, \beta\}$. The value of C is given by equation 4.

$$\sum_{r=cut}^{s} p_{\theta,r} = 1 \tag{4}$$

The upper limit of the summation in equation 4 is s, which is the number of different simplices observed in the entire match. On the other hand, given that in this study we also investigate the impact of design in the most frequent simplices sets we also use left truncation in the generalization of these probability density functions. Correspondingly, the summation lower limit, cut, defines the rank used to left truncate the distribution. This represents a generalization to the ZM model where the most frequent simplices are not considered. For example, if $cut = 1$, all simplices sets are considered, and if $cut = 3$ the two most frequent simplices sets are not considered. This generalization allows extending the model to the description of systems where the most frequent tokens do not follow the same mechanisms or are under different constrains from other tokens. In the particular case of a soccer match this can correspond for example to particular functions or rules.

Using a counting process, computed over the entire match, we obtained the frequencies for each simplex set, n_r. These frequencies were used to rank the simplices. This scoring process is defined in equations 5 and 6

$$n_r = \sum_{t=1}^{T} I_r(t) \tag{5}$$

where T is the number of samples in the match and $I_r(t)$ is an indicator function given by:

$$I_r(t) = \begin{cases} 1, & \sigma_r \in \Sigma_t \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The total number of simplices' counts, n is given by:

$$n = \sum_{k=cut}^{s} n_r \tag{7}$$

where n_r is obtained from equation 5 and the summation upper and lower limits are the same as in equation 4. To avoid the possible artifacts resulting from using the same data set for both ranking and frequency value described by Piantadosi Piantadosi (2014) we utilized a bootstrapping procedure similar to that proposed by Piantadosi Piantadosi (2014). This process is also used for defining confidence intervals for the frequency values Babu and Bose (1988) and for assessing the ZM law fit to the empirical data.

2.3. Fitting and validating the ZM distribution model

The analysis of these data structures related to the ZM distribution in real-life situations, implies an effective fitting procedure and an appropriate test for the goodness of fit Izsák (2006). The estimation of the unknown parameters, α and β for equation 3 that best fit the distribution to the empirical data set can be obtained by applying a Maximum Likelihood Estimation (MLE). In this method the estimation is performed by maximizing the likelihood of the distribution (i.e., the ZM distribution for the occurrence of each simplex) with these parameters given a particular data set (i.e., the number of occurrences of each simplex observed in the data set). How probable the distribution can be assessed via its likelihood estimator l_θ, the value to be maximized. For the ZM distribution the likelihood estimator l_θ is given by equation 8:

$$l_\theta = \frac{n!}{n_{cut}! \cdot n_{cut+1}! \dots n_s!} \cdot \prod_{r=cut}^{s} p_{\theta,r}^{n_r} \tag{8}$$

In order to ease the maximization computational process usually the logarithm of l_θ is preferred, L_θ, and given by:

$$L_\theta = \ln \left(\frac{n!}{n_{cut}! \cdot n_{cut+1}! \dots n_s!} \right) + n \ln(C) - \alpha \sum_{r=cut}^{s} \left(n_r(r + \beta) \right) \tag{9}$$

In both equations, s is the number of different simplices observed, cut is the starting index to be considered, and n_r is the observed number of occurrences for simplex with rank r.
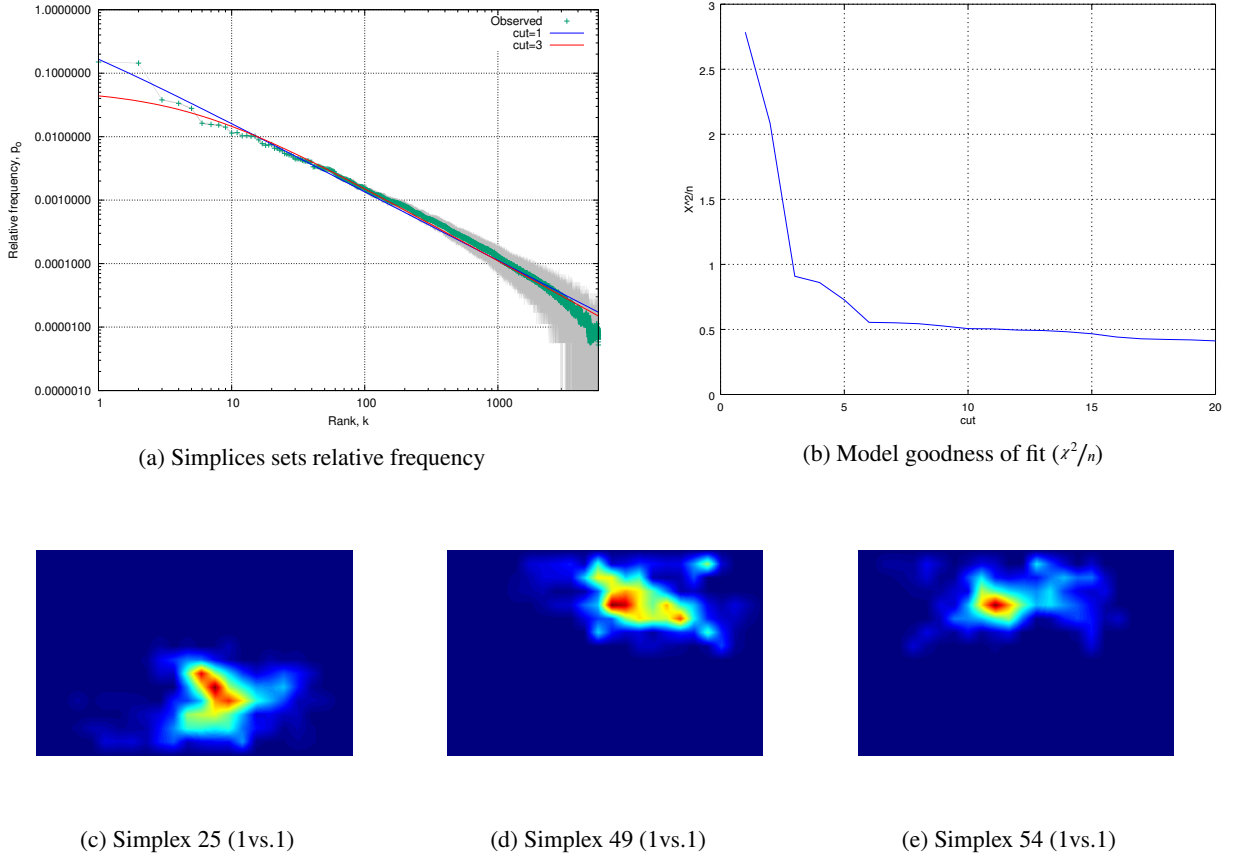
The estimation of the values for parameters α and β that minimize $-L_\theta$ is performed using the numerical minimization provided by Octave's package $optim$ function $fminsearch$ via the Nelder & Mead Simplex algorithm Nelder and Mead (1965). Parameter C is obtained from equation 4.

To test the validity of the model we used the χ^2 metric for assessing its goodness of fit Baker and Cousins (1984); Bentler and Bonett (1980); Spiess and Neumeyer (2010). Although the $p - value$ obtained from the χ^2 statistic was used to decide whether the hypothesis should be rejected, we nevertheless show the χ^2/n value, as it does not depend on the sample size and to which "rule of thumb" $\chi^2/n < 1$ can be applied.

3. Results and Discussion

The figures below show the simplices formation results from three matches (selected out of 10 soccer matches) opposing team A against team B (Figure 3), team A against team C (Figure 4) and team A against team H (Figure 5). In figures: 3a, 4a, and 5a, we plotted the simplices' set relative frequencies versus the rank from the observed data. The gray area in these figures is obtained via a bootstrapping process where the limits correspond to the 10% and 90% percentiles (as described above). The red and blue lines in sub-figures 3a, 4a and 5a correspond to the values obtained from the ZM model, for $cut = 1$ and $cut = 3$, with parameters C, α and β estimated via Maximum Likelihood Estimation (MLE) Izsák (2006).

In figures 3b, 4b and 5b we plotted the χ^2/n metric for assessing the goodness of fit Baker and Cousins (1984); Bentler and Bonett (1980); Spiess and Neumeyer (2010) of the ZM model. We opted for plotting the χ^2/n instead of χ^2 as it is easier to identify the $\chi^2/n < 1$ rule of thumb criteria for not rejecting the hypothesis.) This metric was computed according to expressions 8 and 9 and plotted against the cut value.

(a) Simplices sets relative frequency

(b) Model goodness of fit (χ^2/n)

(c) Simplex 25 (1vs.1)

(d) Simplex 49 (1vs.1)

(e) Simplex 54 (1vs.1)

Figure 3: Team B vs. Team A

Sub-figures c), d) and e) for all three cases show the heat maps of the simplices' position for the 3rd to 5th most frequent occurring simplices (i.e., we do not show the <Goalkeeper, Goal > simplices sets, as further explained in the Discussion), for the corresponding matches. Finally, in Table 1 we present the parameter values (β, α and χ^2/n) of the Mandelbrot generalization for all the 10 matches considered. These results were obtained considering two different conditions: considering all the existing simplices (*cut*=1), and removing the two most occurring simplices (*cut*=3).

The results in Figure 3a suggest that the frequency versus rank follows a power law (we used as hypothesis the ZM model). However, the results shown in Figure 3b, where χ^2 is used to assess the goodness of fit of the ZM model, lead to different conclusions depending on how many simplices sets are considered. When all simplices sets are considered (*cut* = 1, depicted in Figure 3a as a blue line) the ZM model hypothesis must be rejected. The high value of χ^2 results mostly from the most frequent simplices sets, which clearly do not follow a power law, as they form groups with very similar and high frequencies. Figure 3b shows that χ^2 decreases with the *cut* value, and that for *cut* ≥ 3 the ZM hypothesis should not be rejected. The red line in Figure 3a represents the ZM model resulting for this threshold (*cut* = 3). A notable difference between the two thresholds (*cut* = 1) and (*cut* = 3) is the shift of the β parameter from almost Zipf-like ($\beta = 0.16978$) to clearly Mandelbrot ($\beta = 4.6029$). Figures 3c to 3e show the spatial position heat maps for the simplices ranked 3rd to 5th in Match B vs. A. They correspond to simplices sets formed by one player of team A and one player of team B (i.e., 1vs.1) located in particular zones of the pitch, along the side lines.

Figures 4a to 4e show similar results for Match Team C vs. Team A. Figure 4b reveals the same threshold value, *cut* = 3, does not reject the ZM model hypothesis. This *cut* value is more significant for this match, as there is no substantial change in the χ^2 value above this threshold. Moreover, Figure 4a shows the same trend, with the red line exhibiting a much better fit to the observed data after the 2nd most frequent simplex set. Again, a notable difference is found in the β parameter between the two thresholds (*cut* = 1) and (*cut* = 3) when using the ZM Model, shifting from almost Zipf-like ($\beta = -0.20276$) to clearly Mandelbrot ($\beta = 43.603$). Figures 4c to 4e show the spatial position heat
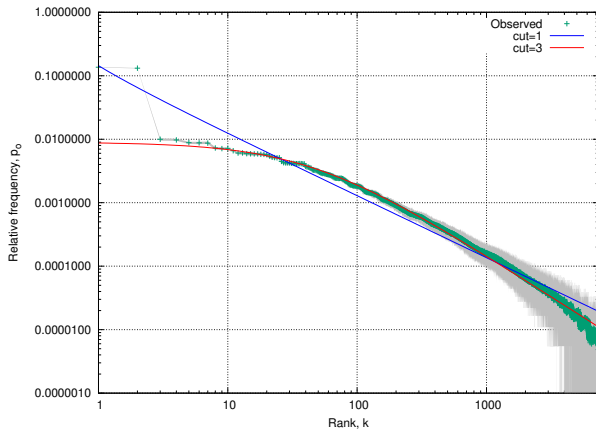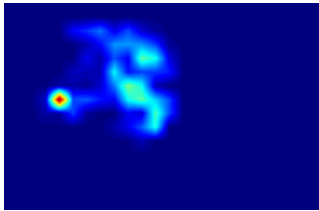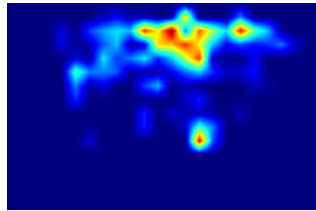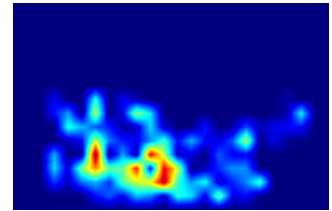
(a) Simplices sets relative frequency
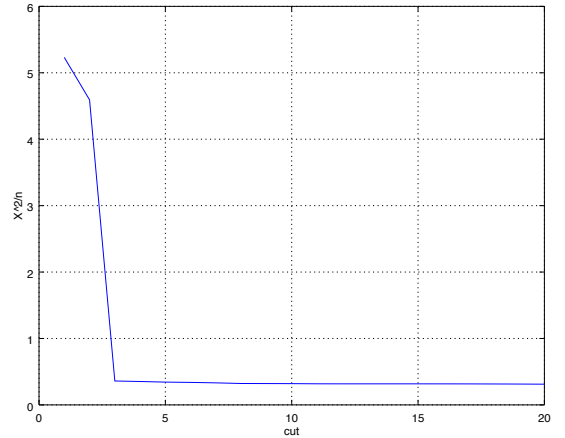
(b) Model goodness of fit (x^2/n)

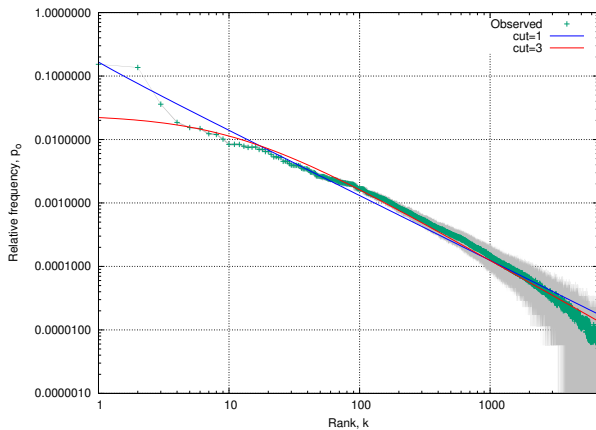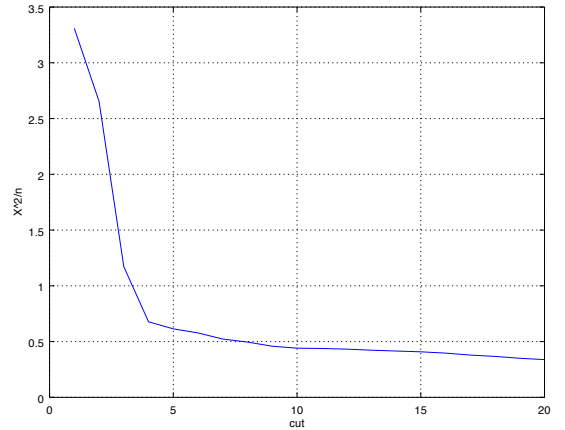(c) Simplex 38 (2vs.1)

(d) Simplex 352 (1vs.1)

(e) Simplex 1371 (1vs.1)

Figure 4: Team C vs. Team A

(a) Simplices sets relative frequency

(b) Model goodness of fit (x^2/n)

maps for the simplices ranked 3rd to 5th in Match C vs. A. Two of these heat maps (4d and 4e) correspond to 1vs.1 simplices along the side lines. On the other hand, the 3rd most frequent simplex set (4e) corresponds to an unbalanced set (two players from Team C and one player from Team A) and the spatial position of the heat map is more intense in the central zone of the pitch and close to Team's C goal.

Figures 5a to 5e present the results for Match Team A vs. Team H. Figure 5b shows that the threshold value for not rejecting the ZM model hypothesis in this match is $cut = 4$, which is also clear in Figure 4a where the red line reveals a
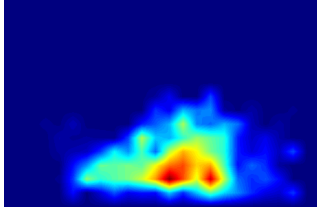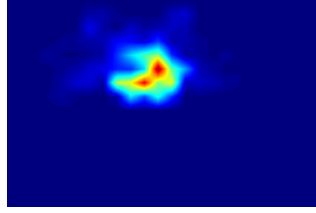
(c) Simplex 8 (1vs.1)   (d) Simplex 28 (1vs.1)   (e) Simplex 32 (1vs.2)

Figure 5: Team A vs. Team H

Table 1
Mandelbrot parameter of β and α

(a) $cut = 1$

Match	β	α	χ^2/n
B × A	0.16978	1.0808	2.7847
C × A	-0.20276	0.9763	5.2315
D × A	0.09027	1.0601	3.3324
E × A	0.09536	1.0442	3.4538
F × A	0.49587	1.1060	2.3015
A × G	0.00658	1.0536	2.6180
A × H	-0.12456	1.0217	3.3092
A × I	-0.02822	1.0419	4.0488
A × J	0.10487	1.0427	2.0488
A × K	0.07262	1.0579	3.2862

(b) $cut = 3$

Match	β	α	χ^2/n
B × A	4.6029	1.1532	0.9089
C × A	43.603	1.3058	0.3598
D × A	8.7968	1.1876	0.6676
E × A	12.5346	1.2009	0.5386
F × A	7.7196	1.2231	0.4570
A × G	5.5558	1.1347	0.9276
A × H	10.466	1.1576	1.1718
A × I	12.634	1.2019	0.7568
A × J	6.7537	1.1448	0.8013
A × K	8.7551	1.1851	1.0925

much better fit to the observed data after the 3rd most frequent simplex set. Interestingly, the 3rd most frequent simplex set also stands out from all the others. Moreover, as for the other matches, the β parameter shifts from almost Zipf-like ($\beta = -0.12456$) to clearly Mandelbrot ($\beta = 10.466$) when using the ZM model for the two thresholds ($cut = 1$) and ($cut = 3$). Figures 5c to 5e show the spatia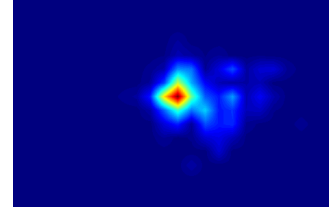l position heat maps for the simplices ranked 3rd to 5th in Match A vs. H. Two of these heat maps (5c and 5d) correspond to 1vs.1 simplices along the side lines. The results for the 3rd most frequent simplex set (Figure 5c) further reveal the particularity of this set as observed above. Figure 5e corresponds to an unbalanced set (one player from Team A and two players from Team H) and the spatial position of the heat map is more intense in the central zone of the pitch and close to Team's H goal.

In the cut=1 table, the results for the β values are closer to 0, which approximates to a Zipf's like distribution and the distribution begins to approximate a Mandelbrot distribution when the two most frequently occurring simplices are removed and the β values are significantly higher.

In these three matches, removing the two most frequent simplices improves the goodness of fit, as shown in panels a) and b) of all figures. Moreover, the two most frequent simplices stand out from all the other simplices sets, not only because of their high frequency values, but also because when removed from the data set, the χ^2 values on the goodness of fit tests Baker and Cousins (1984); Bentler and Bonett (1980) are significantly reduced. It is also interesting to note that when these simplices are considered, the distribution is approximately Zipf (i.e., $\beta \approx 0$), however, in the opposite scenario the Mandelbrot generalization better describe the results (i.e., $\beta > 4.5$).

Collectively our results reveal that, in the 10 soccer matches analyzed, the frequency of the overwhelming majority of the simplices that emerge follows the typical distribution of a complex system. The goodness of fit tests supports these findings and allows us to validate the null hypothesis postulating that the simplices frequencies follow a Zipf-Mandelbrot (ZM) like distribution Bentler and Bonett (1980).

4. Conclusions

In the present study we show that most of the simplices observed in the 10 soccer matches follow a statistical distribution of occurrence typical of complex systems. This result is supported by goodness of fit tests on the hypothesis of Zipf-Mandelbrot (ZM) like distribution Bentler and Bonett (1980), which correspond to hallmarks of complex and self-organized systems Schmidt and Fitzpatrick (2016). Moreover we found that the two most frequent simplices stand out from other simplices sets in frequency values and in their impact on the ZM distribution *beta* parameter (from $\beta \approx 0$, Zipf, to $\beta > 4.5$, Mandelbrot).

The players involved in the two most frequent simplices, the goalkeepers, have a very distinctive purpose (defending the goal) and specific rules, when compared with the other players. First, these simplices sets are of the type <Goalkeeper, Goal> and the design of the competition field is established with specific delimited areas in the pitch, as goalkeepers can touch the ball with the hands in this specific area. Second, the specific role of these players anchors the goalkeepers to their goals, to prevent the opposing team players from scoring a goal. Moreover, we can observe another typical feature of social complexity, namely, intentionality in the behavior of the actors Johnson (2013, 2010); Blecic and Cecchini (2008); Johnson (2008).

Notably, the results from some of the matches (e.g. A against teams B and H), reveal that simplices that seem to be designed, preplanned or conceived before the match, to behave differently from the others, i.e., where subsets of players are more frequently close to each other than the others (Figures 3a and 5a). For instance, in Match B vs. A which has mainly 1vs.1 simplices and a typical positioning in the field (figures c), d) and e)), during the entire match, it is clear that the players remained connected in a very specific area of the field, and a similar scenario can be observed in Match A vs. H (for the six first more frequent simplices). However, in Match C vs. A no simplices stand out from the rank distribution, with the exception of <Goalkeeper, Goal> simplices, which is also revealed in Figure b), where even after removing the <Goalkeeper, Goal> simplices the χ^2 values remain low (less than 0.5) and stable.

In conclusion, we found several 1vs.1 simplices, as well as close combinations 1vs.2 or 2vs.1 and also 2vs.2, that might reflect a preformed design and strategy of the teams. In addiction, we also found large number of sets of simplices that appear less frequently, revealing that many interactions between players are self-organized. The frequency distribution of simplices sets is well modeled by the ZM model, a hallmark of complex systems, with α parameter in the range of other systems (e.g., written text, population size). However, large deviations from this model occurs for the most common simplices sets, revealing design - a well identified means to deal with complexity. This aspect is particularly relevant as it results not only from the traditional cooperative design Johnson (2005a), but in this case from both cooperative and competitive processes.

Data Availability

The raw data used in this study, i.e., the players' coordinates on the pitch, are available from City Football Services, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Secondary data, i.e., simplices' constitution at each time frame, are however available from the authors upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

References

Araújo, D., Davids, K., 2016. Team synergies in sport: theory and measures. Frontiers in psychology 7, 1449.
Babu, G., Bose, A., 1988. Bootstrap confidence intervals. Statistics & Probability Letters 7, 151–160.

Baker, S., Cousins, R.D., 1984. Clarification of the use of chi-square and likelihood functions in fits to histograms. Nuclear Instruments and Methods in Physics Research 221, 437–442.

Bar-Yam, Y., 2002. General features of complex systems, in: Encyclopedia of Life Support Systems (EOLSS). UNESCO, EOLSS Publishers, Oxford, UK. volume 1.

Bentler, P., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 88, 588–606.

Blecic, I., Cecchini, A.B., 2008. Design beyond complexity: Possible futures—prediction or design?(and techniques and tools to make it possible). Futures 40, 537–551.

Ferrer-i Cancho, R., Elvevåg, B., 2010. Random texts do not exhibit the real Zipf's law-like rank distribution. PLoS One 5, e9411.

Izsák, F., 2006. Maximum likelihood estimation for constrained parameters of multinomial distributions—application to Zipf–Mandelbrot models. CMStatistics 51, 1575–1583.

Johnson, J., 2005a. Complexity science in collaborative design. CoDesign 1, 223–242.

Johnson, J., 2005b. Multilevel multidimensional networks for designing and managing complex human systems, in: Conference on Complexity Science and Society, pp. 11–14.

Johnson, J., 2008. Science and policy in designing complex futures. Futures 40, 520 – 536.

Johnson, J., 2010. Embracing Complexity in Design. Routledge. chapter Embracing design in complexity. pp. 193–203.

Johnson, J., 2013. Hypernetworks In The Science Of Complex Systems. volume 3. World Scientific.

Juarrero, A., 2000. Dynamics in action: Intentional behavior as a complex system. Emergence 2, 24–57.

Kobayashi, N., Kuninaka, H., Wakita, J.i., Matsushita, M., 2011. Statistical features of complex systems–toward establishing sociological physics–. Journal of the Physical Society of Japan 80, 072001.

Komulainen, T., 2004. Self-similarity and power laws, in: Complex systems-science on the edge of chaos (Report 145). chapter 10, pp. 109–122.

Li, G., Wang, X., 2019. Prediction accuracy measures for a nonlinear model and for right-censored time-to-event data. Journal of the American Statistical Association 114, 1815–1825.

Malacarne, L., Mendes, R., 2000. Regularities in football goal distributions. Physica A: Statistical Mechanics and Its Applications 286, 391–395.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The computer journal 7, 308–313.

Passos, P., Araújo, D., Davids, K., 2013. Self-organization processes in field-invasion team sports. Sports Medicine 43, 1–7.

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., Mendes, J., 2011. Networks as a novel tool for studying team ball sports as complex social systems. Journal of Science and Medicine in Sport 14, 170–176.

Piantadosi, S.T., 2014. Zipf's word frequency law in natural language: A critical review and future directions. Psychon Bull Rev 21, 1112–1130.

Ramos, J., Lopes, R.J., Araújo, D., 2017a. What's next in complex networks? capturing the concept of attacking play in invasive team sports. Sports Medicine 48, 17–28.

Ramos, J., Lopes, R.J., Marques, P., Araújo, D., 2017b. Hypernetworks: capturing the multilayers of cooperative and competitive interactions in soccer, in: International Congress Complex Systems in Sport, Frontiers. pp. 150–153.

Ramos, J., Lopes, R.J., Marques, P., Araújo, D., 2017c. Hypernetworks reveal compound variables that capture cooperative and competitive interactions in a soccer match. Frontiers in Psychology 8, 1379.

Ribeiro, J., Davids, K., Duarte Araújo, J.G., Silva, P., Garganta, J., 2019. Exploiting bi-directional self-organizing tendencies in team sports: The role of the game model and tactical principles of play. Frontiers in psychology 10, 2213.

Riley, M.A., Richardson, M., Shockley, K., Ramenzoni, V.C., 2011. Interpersonal synergies. Frontiers in psychology 2, 38.

Rothwell, M., Davids, K., Stone, J., 2018. Harnessing socio-cultural constraints on athlete development to create a form of life. Journal of Expertise 1, 94–102.

Saravia, L.A., Doyle, S.R., Bond-Lamberty, B., 2018. Power laws and critical fragmentation in global forests. Scientific Reports 8, 17766.

Schmidt, R.C., Fitzpatrick, P., 2016. The origin of the ideas of interpersonal synchrony and synergies. Interpersonal coordination and performance in social systems. London: Routledge .

Schmidt, R.C., Richardson, M.J., 2008. Dynamics of interpersonal coordination, in: Coordination: Neural, behavioral and social dynamics. Springer, pp. 281–308.

Silagadze, Z., 1997. Citations and the Zipf-Mandelbrot's law. Complex Systems 11, 487–499.

Silva, P., Chung, D., Carvalho, T., Cardoso, T., Davids, K., Araújo, D., Garganta, J., 2016a. Practice effects on intra-team synergies in football teams. Human movement science 46, 39–51.

Silva, P., Vilar, L., Davids, K., Araújo, D., Garganta, J., 2016b. Sports teams as complex adaptive systems: manipulating player numbers shapes behaviours during football small-sided games. SpringerPlus 5, 191.

Spiess, A.N., Neumeyer, N., 2010. An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. BMC Pharmacology 10.

STATS, 2020a. About STATS Perform. URL: https://www.statsperform.com/about/. last accessed 18 March 2020.

STATS, 2020b. Football performance analysis solutions. URL: https://www.stats.com/football/. last accessed 18 March 2020.

Wentian, L., 1992. Random texts exhibit Zipf's-law-like word frequency distribution. IEEE Transactions on Information Theory 38, 1842–1845.