Passing sequences and networks analysis in football
A study on the UEFA EURO 2020

Manuel Maria Strecht Ribeiro Hipólito Reis

Instituto Superior Técnico, Universidade de Lisboa, Portugal

December 2022

Abstract

Network Science and Graph Theory can contribute to football analysis, providing valuable tools that allow describing the interactive behaviour of teams more consistently than the traditional analysis, which is based on the individual performance of players. Few research works have studied the relationship between a team's attacking strategy (possession play or direct play) and the characteristics of the network and how these affect the team's overall performance. On the other hand, the impact of the systems of play on the network's characteristics is still unclear. Therefore, this study analysed the passing sequences and networks of the national teams that competed in the UEFA EURO 2020. Verifying that most of the teams' distributions of passes completed tend to follow the power law, an innovative way to describe the general strategy of the play was proposed, through the power law exponent,-α. Thus, teams with a possession game have a lower value of α, whereas teams with a direct play have a higher value of α. The results of statistical studies suggested that the teams that adopted a direct play, characterised by executing fewer passes and fewer passes completed, generating networks with a lower density and average clustering coefficient, were less successful, i.e., eliminated in the tournament's first stage. Finally, the clustering analysis was inconclusive in revealing how playing systems affect the networks' characteristics. In summary, this study provides relevant insights that can aid the coaching staff's work, enhancing the value of Network Science and Graph Theory in football analysis.

Keywords: Football, Passing sequences, Passing networks, Attacking strategy, Team performance, Network Science

1. Introduction

Football, also known as soccer, is the most popular sport in the world. Since the two-opponent offside rule was established in 1920, football's fundamental rules have almost not been altered [1]. However, football has been evolving and becoming more professionalised. New strategies have arisen, and due to the constant innovation in team play, the demands of match analysis have grown to the point where coaches now want to scrutinise match analysis [2].

The most appealing aspect of football is its emergent properties [3]. Goldstein [4] affirms that emergence "refers to the arising of novel and coherent structures, patterns, and properties during the process of self-organisation in complex systems". Football's complexity stems primarily from the number of interactions between teammates and opponents, but it also exists in the game context [5]. Football teams can be described as groups of individuals that interact dynamically and interdependently to achieve their common objective: score goals and prevent the opposing team from doing the same [6], [7]. Thus, a football team is a complex, dynamic, nonlinear, open, and adaptable system formed by 11 players who are also themselves systems. The nonlinearity arises from the fact that the whole is not equal to the sum of its parts, and thus a football team cannot be understood solely by examining its components, i.e. its players [8], [9]. The context and environment for creating a team system are produced by each player's relational capacity (open system), evolving capacity (dynamic system), adaptability to the environment in which he performs his tasks, and the uncertainty with which he demonstrates his competitive capacity [10].

As a result, the complexity of the play, the nearly constant flow of the ball during the match and the low scores are examples of factors that make simple statistics such as the number of goals, shots or assists insufficient as measures of player and team performance [11], [12]. In opposition, passes are the links between teammates and occur numerously in every match despite the quality of the teams [1]. Thus, by modelling the interactions based on the passes, network analysis captures teams' behaviour, organisation and performance in a way that classical analysis, based on the performance of individual players, cannot [13]–[15].

The ability to retain possession of the ball for more extended periods has been linked to success [16]–[18]. In addition, in recent years, ball possession has acquired fundamental importance in the attacking strategy of football teams [19]. However, few research works have studied the relationship between a team's attacking strategy (direct or possessive play) and how these two factors impact teams' overall performance. Alternatively, the system of play influences the team's network characteristics since it provides a reference to the players' interactions within the team [5], [20]. However, few studies have investigated the effect of the systems of play in professional football, namely the impact on the network's characteristics [21].

Therefore, three research questions are intended to analyse and answered throughout this study:

(1) Does the distribution of successful passes tend to follow a power law distribution? How can the distribution of successful passes explain the general attacking strategy (direct or possessive play)? (2) How do the general attacking strategy and network characteristics relate to each other, and how do they impact the team's overall performance? (3) How do the systems of play affect networks' characteristics? More specifically, do the same systems of play generate similar networks?

As a result, this study seeks to answer these research questions by applying statistical procedures, graph theory and network science and using as a sample the matches of UEFA EURO 2020.

2. Literature Review

This section reviews the literature on the passing sequence analysis (subsection 2.1) and the passing network analysis (subsection 2.2).

2.1. Passing sequences analysis

Over the years, few studies on passing sequences in football have been developed. Reep and his colleagues started researching this subject in the late 1960s and early 1970s. Reep & Benjamin [22] analysed statistically the passing sequences that resulted in goals from football matches, presenting them as a negative binomial distribution. Reep *et al.* [23] later expanded this work to other sports. These researchers' two primary discoveries were that a goal was scored every ten shots and that almost 80% of the goals came from a sequence of three passes or fewer [24].

These works implied that those passing sequences with few passes were more successful. Consequently, as Bate [25] deepened, it was possible to deduce that teams should adopt a "direct play" rather than a "possessive play" to be successful. However, most successful teams did not use a "direct play". So, Hughes *et al.* [26] studied the patterns of plays of the semifinalists and the national teams that were eliminated in the first stage of the 1986 World Cup and found that the most successful teams played with more passes per possession than unsuccessful teams. In this way, they determined that the conclusions made by Reep & Benjamin [22] and Bate [25] did not apply to all levels of football [24].

Years later, Hughes & Franks [24] replicated the work of Reep & Benjamin [22] and discovered that the conclusions reached by these authors could be misinterpreted. Because of this, Hughes & Franks [24] questioned whether goal-scoring or shooting was influenced by the number of passes made per possession. To assess the relative contribution of each possession from equal frequencies of occurrence, they created a new methodology in which they normalised the data by dividing the number of goals scored during each possession by the frequency of that sequence length [24].

Hughes & Franks [24] reached three conclusions when the same data were normalised. First, longer passing sequences significantly increased shots per possession compared to shorter passing sequences. Second, "direct type of play" outperformed "possession type of play" regarding the conversion rate of shots to goals. Third, although the differences between the successful and unsuccessful teams at the 1990 World Cup were not substantial, the successful teams had a better conversion ratio of possession to shots on goal [24].

2.2. Passing networks analysis

A significant contribution to the description of team interactions can be provided by network analysis. Nevertheless, despite this substantial and fascinating contribution, few studies using this methodology have been published [27]–[29]. One of the first studies that introduced the concept of football passing networks was published by Gould & Gatrell [30]. They explored the structure of a football match, specifically the Cup Final of 1977 between Liverpool and Manchester United. However, as Buldú *et al.* [31] point out, this study did not receive the attention of the scientific and sports communities. Only more than thirty years later, the research into how network science can be used to reveal vital information about the organisation and performance of football teams and players started [31].

Indeed, the recent ability to obtain datasets of all events occurring during a match leveraged the investigation of how Network Science can unveil the organisation and properties of football teams [13]. Several studies have focused on football analysis in the last decade, specifically on how players interact with each other by passing the ball [31]. The sample and scope of the studies vary, ranging from pilot studies (one match from one team) and case studies (a few matches from one or more teams) to full domestic, continental, or international competitions (one or more teams in one or more competitions).

Using Network Science, the investigators construct what can be denoted as "football passing networks", which can be of three main types [13], [31]: **(1) Player/playing position passing networks**, where nodes are a team's players/playing positions [13], [32]–[34]. The majority of the research works studies this type of network; **(2) Zone passing networks**, where nodes are zones of the field of play linked through passes performed by players in those zones [13], [15], [35]–[37]. Several studies have built this type of network, whereas other studies also include this kind of analysis to complement their examination of player passing networks; (**3) Player/playing position-zone passing networks**, where nodes are the combination of a player/playing position and his location on the field of play at the moment of the pass [13], [28], [38]. Only two studies using this type of analysis were found in the literature.

According to Buldú *et al.* [13], after constructing the network, several "topological scales" can be identified: **(1) Microscale**, where analysis is performed at the node level. As presented before, most studies examine the importance of each player, considering network metrics, such as the degree, closeness, and betweenness centralities, and the clustering coefficient [13]. Some works focus their study on individual players [11], [12], while others concentrate their attention on the characteristics of the playing positions [34], [37], [39]. At this level and considering the playing positions, the research works have indicated that midfielders are usually the most influential players; **(2) Mesoscale**, where motifs depicting the interactions of three or four players are examined [13]. The analysis of motifs has revealed that most teams tend to apply a homogeneous style [1]. Also, it demonstrated how it is possible to identify the key players in the network [40], thus assisting in the scouting process [41]; **(3) Macroscale**, where the network is studied as a whole [13]. Studies have suggested that high-density and decentralised passing networks are associated with higher performance [33], [42], [43].

In addition, research works have shown that the interaction between players during a football game supports a scale-free network [3], [34]. Also, time is a dimension that is considered in a few works. Examining each game's half was one method used to investigate how the network changed over time. This method has revealed differences between the first and second halves concerning the density and centralisation of the network [13], [42]. Another technique was to build sliding windows with a specific length (between 5 to 15 minutes) [3], [13], [28]. Finally, the influence of the system of play was only found once in the literature, using only one team in two different seasons [44].

3. Methods

This section presents this study's sample (subsection 3.1), materials (subsection 3.2), and procedures (subsection 3.3).

3.1. Sample

The 51 official matches from UEFA EURO 2020 were analysed. This UEFA-organised tournament was competed by twenty-four European men's national teams and was composed of two different stages: the Group Stage and the Knockout Stage. Firstly, the twenty-four teams were divided into six groups of four in the group stage. Every team played every other team in their group once, being awarded three points for a win, one for a draw and none for a defeat. Thus, the six group winners, the six runners-up, and the four best third-placed teams qualified for the Round of 16. Secondly, the knockout stage was played in single-leg matches as follows: Round of 16, Quarter-finals, Semi-finals, and Final. In this stage, if there was no winner at the end of regular playing time, two 15-minute periods of extra time were played. Penalty kicks were required if there was still no winner after extra time. The winning team of each match advanced to the next stage.

3.2. Materials

The raw data sets were provided by StatsBomb Services Ltd, which has made the data from UEFA EURO 2020 publicly and freely available. There were different types of datasets in a JSON format in its open data. On the one hand, the *StatsBomb Match Data* records the match information for each match, including competition and season information, home and away team information, match results, stadium and referee information. On the other hand, *StatsBomb Event Data* comprises actions performed during play, concentrating on the ball. The three main characteristics of each event are (a) the timestamp, (b) the action and (c) the attributes. The timestamp registers the time in the match the event takes place; the action refers to the type of event to which it corresponds, and the attributes include general and specific information about the characteristics of the event and the entities involved in it. Thus, these types of files were used to perform all the analyses presented in the coming sections. Nevertheless, it is highly recommended to read the document *StatsBomb Data Specification v1.1*, publicly available, to get more in-depth knowledge about StatsBomb data[1].

This study's analyses were carried out using Microsoft Excel®, IBM SPSS Statistics® (version 28), Python 3.10.5 and the Python packages: *NetworkX®* (version 2.8.5), *pandas* (1.4.3), *NumPy* (1.23.1), *scipy* (1.9.0), *scikit-learn* (1.1.1), *seaborn* (0.11.2), *statsmodel* (0.13.2), *matplotlib* (3.5.2) and *powerlaw* (1.5).

3.3. Procedures

The first focus of this study was to verify if the teams' distribution of passes per possession tended to follow the power law distribution. The length of a passing sequence was used to define a team's possession. A passing sequence of length equal to one was an intended pass that a teammate received, but then the second pass either left the field of play, was contacted by the opposition, or was interrupted by a foul. On the other hand, a two-pass sequence ended when the third pass did not reach the target, and so on [24]. The sequences of passes per possession executed during the attacking phase and set pieces by each team during the regular time (90 min) of each match were examined. Note that the passes made during the extra time were

[1] Source: StatsBomb. (2022). GitHub. Open data. Retrieved from: https://github.com/statsbomb/open-data/blob/master/doc/StatsBomb%20Open%20Data%20Specification%20v1.1.pdf

excluded from the study to allow comparisons between all the matches.

The pure power law distribution, also referred to as the zeta distribution or discrete Pareto distribution, is written as follows:

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})},$$

where x is a positive integer measuring a variable of interest, $p(x)$ is the probability of observing the value x, α is the power law exponent, $\zeta(\alpha, x_{min})$ is the Riemann zeta function, defined as $\sum_{x=x_{min}}^{\infty} x^{-\alpha}$, and x_{min} is the value of x from which the power law is obeyed [45], [46].

There are several methods for fitting power law distributions. However, the maximum likelihood estimation (MLE) was used since it is one of the most robust methods for fitting the power-law distribution. It is based on finding the maximum value of the likelihood function:

$$l(\alpha \mid x) = \prod_{i=1}^{N} \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{min})},$$

$$\begin{aligned}\mathcal{L}(\alpha \mid x) &= \log l(\alpha \mid x) \\ &= \sum_{i=1}^{N} (-\alpha \log(x_i) - \log(\zeta(\alpha, x_{min}))) \\ &= -\alpha \sum_{i=1}^{N} \log(x_i) - N \log(\zeta(\alpha, x_{min})),\end{aligned}$$

where $l(\alpha \mid x)$ is the likelihood function of α given the unbinned data x and $L(\alpha \mid x)$ is the log-likelihood function.

This maximum can be obtained by setting $\partial\mathcal{L}/\partial\alpha = 0$:

$$-\sum_{i=1}^{N} \log(x_i) - N \frac{1}{\zeta(\alpha, x_{min})} \frac{\partial}{\partial\alpha} \zeta(\alpha, x_{min}) = 0,$$

and, therefore, the MLE $\hat{\alpha}$ is the solution of

$$\frac{\zeta'(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} = \frac{1}{N} \sum_{i=1}^{N} \log(x_i),$$

where $\zeta'(\hat{\alpha}, x_{min})$ is the first derivate of the Riemann zeta function [45], [46].

Additionally, a test was necessary to assess the goodness-of-fit of the fitting method. Therefore, the Kolmogorov-Smirnov (KS) type test was chosen since it is one of the most simple and robust of the commonly used goodness-of-fit tests. This test is based on the following test statistic:

$$K = max_{x \geq x_{min}} |S(x) - P(x)|,$$

where $S(x)$ is the cumulative distribution function (CDF) of the data for the observations with a value of at least x_{min} and $P(x)$ is the CDF for the power-law model that best fits the data in the region $x \geq x_{min}$ [45], [46].

The passes per possession of each team in each tournament's match were fitted using the *powerlaw* Python package, which offers commands for fitting and statistical analysis of distributions. These functionalities were used to compute the fitted α parameter, i.e., the power law exponent. Thus, the discrete distribution of the passes per possession was fitted through the MLE. However, few empirical events follow a power law across the entire range of x, meaning that the optimal x_{min} for each team's distribution of passes per possession can vary from one. By fitting a power law to each distinct value in the dataset and choosing the one that minimises the KS distance between the data and the fit, the minimum value at which the power law's scaling relationship begins, x_{min}, was determined [47].

Although these tools give estimates for the parameters of α and x_{min}, they cannot determine whether the power law is a reasonable fit to the data, so it was necessary to confirm this hypothesis given the passes per possession data [45]. Hence, the methodology described by Clauset *et al.* [45] was employed. A goodness-of-fit test was used, which computes a p-value, p, that measures the plausibility of the hypothesis, given the observed data and the hypothesised power-law distribution. First, the empirical data was fitted to the power law. After that, a sizable number of power-law distributed synthetic data were created, each with parameter α and lower bound x_{min} equal to the distribution's parameters that best fit the observed data. Each synthetic data set was fitted to its power-law model, and the KS statistic was computed for each relative to its model. Then, the p-value, the percentage of the resulting statistic greater than the value of the empirical data, was calculated [45].

Therefore, to obtain an accurate estimate of the p-value, a semiparametric approach was used to produce the synthetic data that had a distribution similar to the empirical data below x_{min} but that followed the fitted power law above x_{min}. Given a data set with n observations and n_{tail} observations in which $x > x_{min}$, the new synthetic data was generated as follows: for $i = 1, \ldots, n$, with a probability of n_{tail}/n, a random number x_i was created using a power law with a scaling parameter $\hat{\alpha}$ and $x > x_{min}$. Otherwise, with a probability of $1 - n_{tail}/n$, x_i was equal to one element selected uniformly at random from among the elements of the observed data that had $x < x_{min}$ [45].

Knowing that, for the p-value to be accurate to within about ϵ of the true value, it should be created at least $0.25\epsilon^{-2}$ synthetic data sets, 2500 synthetic datasets were generated aiming to have a p-value accurate to about two decimal digits, this is $\epsilon = 0.01$. After computing the p-value, it is necessary to decide whether p is small to rule out the power-law hypothesis. Accordingly, a $p \leq 0.05$ was chosen to rule out the power-law hypothesis [45].

Considering real events, even if data are drawn from a power law, their observed distribution is unlikely to follow the power law exactly. In addition, there may be the possibility that there are samples that do not follow the power law. Nevertheless, regardless of the true data's distribution, it is always possible to fit a power law. As a result, to allow comparison of α of all teams, $x_{min} = 1$ was fixed for all samples.

There are two paradigms in the attacking strategy of play: possession play and direct play [21]. A possession play is characterised by more ball possession, expressed by more passes per possession. The teams applying this attacking strategy aim to retain the ball possession when progressing in the field of play. In contrast, direct play is characterised by trying to move the ball into a shooting position with few passes [7], [48]. Therefore, this study proposes to describe the general attacking strategy of a team (possessive or direct play) through the power law exponent, $-\alpha$. Thus, teams with a possession game have a lower value of α, while teams with a direct game have a higher value of α. As a result, the parameter α was computed for each national team in each match, aiming to study and distinguish the strategy of play of each team that competed in the tournament.

Second, this study investigated the relationship between a team's attacking strategy (possession or direct play) and the characteristics of the network and how these impact the team's overall performance. Consequently, the zone passing networks of each team in each match were built. This type of network was chosen over the player/playing position network because, in the latter, the number of nodes depends on the number of players or playing positions used throughout the game. In each *StatsBomb Event Data* file, it was only considered the "Pass" and "Ball Receipt*" types of events in the attacking phase and during all set pieces. This allowed for the collection of the following information from each completed pass: (i) the player and respective playing position who passed the ball, (ii) the player and respective playing position who received the ball, (iii) the location (coordinates (x,y)) of the sender and the receiver; (iv) the time at which the pass was made and (v) some pass attributes [31]. This way, this information enabled the construction of the networks using Python and its package NetworkX®.

However, splitting the field of play into different-sized zones leads to different networks. Consequently, the question "How many zones should the field of play be divided into?" arose. The zone networks were formed by splitting the field of play equally into Z zones, where $Z = s \times c$ is the number of nodes (zone areas), $s = \{3,4,6\}$ is the number of sectors (vertical subdivisions), and $c = \{3,5\}$ is the number of corridors (horizontal subdivisions). When a pass was made from region *i* to *j*, a link from node *i* to *j* was created. This edge had a weight that measured the total number of successful passes. As a result, different-sized zone networks were generated, where the number of nodes was the number of playing field divisions, Z={9,12,15,18,20,30}. Then, a descriptive analysis was conducted to decide the appropriate number of zones for the subsequent analysis.

After building the zone networks with the number of nodes equal to the selected number of zones, the networks' density and average clustering coefficient were computed. On the one hand, density is the interconnectedness of nodes (zones) of a network (team) [49]. Since the passes have a direction, the network is represented by a digraph, so the maximum possible number of edges is $n(n-1)$. Thus, the density, ρ, is defined as the ratio between the number of edges, m, and the maximum possible edges:

$$\rho = \frac{m}{n(n-1)},$$

lying in the range $0 \leq \rho \leq 1$. On the other hand, the clustering coefficient measures the degree to which the neighbours of a given node connect, quantifying how close a node and its neighbours in a graph are to forming a complete subgraph [42], [50]. For a digraph, the clustering coefficient is defined as the fraction of all possible directed triangles [51]:

$$C_u(i) = \frac{2T_i}{k_i(k_i-1) - 2k_i^{\leftrightarrow}},$$

where, T_i is the number of triangles through node i, k_i is the sum of in-degree and out-degree and $k_i^{\leftrightarrow}$ is the reciprocal degree of i. In addition, the average local clustering coefficient can be used to measure the clustering level throughout the network [49]:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i.$$

Hence, the relationship between the parameter α, the pass statistics (number of passes, number of passes completed and percentage of passes completed), and the networks' density and the average clustering coefficient was investigated. This was performed using the Pearson Product-Moment correlation coefficient after ensuring that the assumptions of normality, linearity and homoscedasticity were not violated. When data failed these assumptions, Spearman's Rank Order Correlation was used. Thus, to classify the correlation strength, the following scale was used: very small, ($]0, 0.1[$); small, ($[0.1, 0.3[$); moderate, ($[0.3, 0.5[$); large, ($[0.5, 0.7[$); very large, ($[0.7, 0.9[$); nearly perfect ($[0.9, 1.0[$); perfect, (1.0) [42].

Then, this study aimed to relate the strategy of play with the teams' overall performance. First, the final result of the match was considered a performance variable, i.e. (i) defeat, (ii) draw or (iii) victory. Second, another overall performance of a team was determined by the stage that a team reached in the tournament, wherefore the following were the variables that determined the performance: (i) Final, (ii) Semi-finals, (iii) Quarter-finals, (iv) Round of 16 and (v) Group Stage. This way, this study sought to answer the following questions: (1) Are there any differences in the strategy of play, described by α, pass statistics and network density and average clustering coefficient between teams that achieved different match results? (2) Are there any differences in the strategy of play, described by α, pass statistics and network density and average clustering coefficient between teams that reached different stages of the tournament? After confirming the assumptions of normality and homogeneity, the influence of the match's result and the stage reached in the tournament were examined using one-way ANOVA. On the one hand, through the Kolmogorov-Smirnov tests, the assumption of normality was investigated ($p > 0.05$). Since $n \geq 30$ and considering the Central Limit Theorem, the premise of normality was assumed for any distribution that was not normal. On the other hand, Levene's test was used to investigate the homogeneity assumption. However, when this assumption was violated, the Welsh and Brown-Forsythe tests were used instead of ANOVA. When the test found significant differences between the factors, the Tukey's HSD (honestly significant difference) test or the Tukey's-Kramer test was used to determine where the differences were [42]. For measuring the effect size in ANOVA, the eta-squared, η^2, was used. The formula is:

$$\eta^2 = \frac{Sum\ of\ square\ between\ groups}{Total\ sum\ of\ squares}.$$

To interpret the strength of the eta-squared values, the guidelines from Cohen [52] 0.01=small effect; 0.06=moderate effect and 0.14=large effect.

In addition, to determine whether the article's conclusions by Hughes & Franks [24] are still observed, their research work's methodology was implemented. Initially, it was confirmed if the statement of Reep & Benjamin [22], supported by Hughes & Franks [24], that approximately 80% of the goals result from a sequence of three or fewer passes was verified or not.

Then, as Hughes & Franks [24] explains, when treating unequal frequencies of occurrences, the outcomes should be normalised by dividing the number of outcomes by the frequency of their occurrences. Consequently, the conversion rates from the different passing sequences' lengths per possession into goals were examined. The data were normalised by dividing the number of goals scored in each team's possession by the sequence length and presented as goals per 1000 possessions to avoid very small ratios. In addition, the analysis was done only to 80% of the goals to avoid biased normalisations. Finally, an independent-samples t-test was conducted to compare the goals per 1000 possessions for two groups. The eta-squared, η^2, was used as an effect size statistic for the t-test and is written as follows:

$$\eta^2 = \frac{t^2}{t^2 + df},$$

where t is the t-value and df is the degrees of freedom. Again, the guidelines from Cohen [52] were used to interpret the strength of the eta-squared values.

Finally, the differences and similarities of various systems of play were studied by conducting a clustering analysis while considering position-zone networks. However, it was not possible to draw any conclusions regarding the impact of the systems of play on the networks' characteristics. For these reasons, the procedure and results of this part are not presented. Thus, it is recommended to read the dissertation if interested in this subject.

4. Results and discussion

First, the power-law hypothesis was tested for each team's distribution of passes per possession in each match. The results indicated that approximately 70% of the 102 samples (2 teams × 51 matches) were consistent with the power-law hypothesis, while the remaining 30% were not, having a $p \leq 0.05$. Therefore, it was possible to confirm that the power law was an appropriate model for a part of the data set. Although 30% of the samples failed the power-law hypothesis, the frequency of occurrences tended to decrease as the pass sequences' length increased. Indeed, regardless of the true data's distribution, all the distribution of passes per possession were fitted to the power law and, to allow comparison of α of all teams, $x_{min} = 1$ was fixed for all samples. As a result, the parameter α was computed for each national team in each match.

At the same time, 102 networks were built for the different number of zones $Z = \{9, 12, 15, 18, 20, 30\}$. The descriptive analysis performed to select the number of zones revealed that the division of the playing field in 30 zones had a substantially greater mean of the number of edges, capturing much more information about the passes occurring in the game. In addition, in most of these networks, all zones were connected with at least another zone. For these reasons, the 102 networks with 30 nodes were used in the subsequent analysis.

Then, as the Pearson product-moment correlation and ANOVA assume that the data follow a normal distribution, the normality was assessed using the Kolmogorov-Smirnov statistic. The results of this test revealed that the distributions of the number of passes, number of passes completed and α had a non-significant result ($p > 0.05$), indicating that these data were normally distributed. In opposition, the distribution of the percentage of passes completed, density and average clustering coefficient violated the assumption of normality. However, since $n = 102$ and considering the Central Limit Theorem, the assumption of normality was assumed. In addition, the Pearson product-moment correlation's assumptions of linearity (the relationship between two variables is linear) and homoscedasticity (the variability of both variables is similar to all values) were analysed to see if there was any violation.

First, the relationship between the pass statistics and the network characteristics was not linear. Moreover, the percentage of passes completed did not have a linear relationship with the number of passes, the number of passes completed and the parameter α. Second, homoscedasticity was only violated for the relationship between the percentage of passes completed and the network's characteristics. Consequently, the Pearson product-moment was used to investigate not only the relationships between the number of passes, the number of passes completed, and α, but also between α and the network's characteristics. At the same time, Spearman's Rank Order Correlation was employed to examine the relationship between the percentage of passes completed and the remaining variables, and between the number of passes, the number of passes and the network's characteristics.

On the one hand, Table 1 reveals the Pearson r correlation coefficients. First, there was nearly a perfect positive correlation between the number of passes and the number of passes completed ($r = 0.994,\ n = 102,\ p < 0.01$. Such outcomes align e with the conclusions of Gama *et al.* [53]. Second, the parameter α showed a very large negative correlation with the number of passes ($r = -0.820,\ n = 102,\ p < 0.01$), the number of passes completed ($r = -0.834\ n = 102,\ p < 0.01$), the density ($r = -0.811,\ n = 102,\ p < 0.01$) and the clustering coefficient ($r = -0.781,\ n = 102,\ p < 0.01$). Finally, the average clustering coefficient showed a very large positive correlation with the density ($r = 0.894,\ n = 102,\ p < 0.01$).

Table 1: Pearson Product-Moment Correlation coefficients

Pearson Product-Moment Correlations						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
(1) Nu passes	1.000	0.994 *	——	-0.820 *	——	——
(2) Nu passes completed		1.000	——	-0.834 *	——	——
(3) % passes completed			1.000	——	——	——
(4) α				1.000	-0.811 *	-0.781 *
(5) Density					1.000	0.894 *
(6) Avg Clustering coefficient						1.000

N=102; * Correlation is significant at the 0.01 level

On the other hand, Table 4 shows the Spearman ρ correlation coefficients. The number of passes showed a very large correlation with the density ($\rho = 0.873,\ n = 102,\ p < 0.01$) and also with the average clustering coefficient ($\rho = 0.797\ n = 102,\ p < 0.01$). Secondly, the number of passes completed revealed a nearly perfect positive correlation with density ($\rho = 0.917,\ n = 102,\ p < 0.01$) and a very large positive correlation with the average clustering coefficient ($\rho = 0.825,\ n = 102,\ p < 0.01$). The nearly perfect correlation of the density with the number of passes completed can be explained by the number of edges being highly dependent on the number of successful passes since the passes are the links between nodes in these networks. In addition, the percentage of passes completed revealed a nearly perfect positive correlation with the number of passes completed ($\rho = 0.909,\ n = 102,\ p < 0.01$). This variable also indicated a very large positive correlation with the number of passes ($\rho = 0.866,\ n = 102,\ p < 0.01$), while a very large negative correlation ($\rho = -0.823,\ n = 102,\ p < 0.01$) with α. Finally, there was a large positive correlation between the percentage of passes completed and the number of edges and the density ($\rho = 0.588,\ n = 102,\ p < 0.01$), and the average clustering coefficient ($\rho = 0.521, n = 102,\ p < 0.01$).

Table 2: Spearman's Rank Order Correlation coefficients

Spearman's Rank Order Correlations						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
(1) Nu passes	1.000	——	0.866 *	——	0.873 *	0.797 *
(2) Nu passes completed		1.000	0.909 *	——	0.917 *	0.825 *
(3) % passes completed			1.000	-0.823 *	0.588 *	0.521 *
(4) α				1.000	——	——
(5) Density					1.000	——
(6) Avg Clustering coefficient						1.000

N=102; * Correlation is significant at the 0.01 level

Therefore, these results suggested that teams that adopt a possessive strategy of play perform more passes and more successfully, generating denser zone networks with a higher average clustering coefficient.

Subsequently, the one-way ANOVA or the Welch and Brown-Forsythe test were conducted to examine the differences in the variables, first, between teams that achieved different match results and, secondly, between teams that achieved different stages in the tournament.

First, the samples were divided into three groups according to the match result (Group 1: defeat; Group 2: draw; Group 3: victory). In the 51 matches played in the UEFA EURO 2020, 35 games ended in a victory for one team and 16 games resulted in a draw. After generating the descriptive statistics (Table 3), Levene's test was used to check the assumption of homogeneity of the variances. Thus, it was found that the number of passes, the number of passes completed and the density violated this assumption. For these cases, the Welsh and Brown-Forsythe tests were used instead of consulting the ANOVA. The results indicated that there were no statistical differences at the $p < 0.05$ level between the three groups concerning all variables. This result did not corroborate the research work of Clemente *et al.* [42], which found differences in network density between teams that achieved different match results. However, this disparity could be explained by differences in the type of networks, the type of competition, and the number of teams studied.

Table 3: Descriptive table (mean and standard deviation) and statistical comparison between teams that achieved different match results

	Defeat (N=35)	**Draw** (N=32)	**Victory** (N=35)
(1) Nu passes	481.110 (106.813)	519.000 (179.836)	537.830 (122.477)
(2) Nu passes completed	392.940 (104.552)	439.750 (178.550)	452.430 (129.444)
(3) % passes completed	0.809 (0.049)	0.824 (0.085)	0.828 (0.070)
(4) α	1.627 (0.101)	1.615 (0.156)	1.594 (0.109)
(5) Density	0.206 (0.027)	0.205 (0.044)	0.216 (0.097)
(6) Avg Clustering coefficient	0.376 (0.059)	0.385 (0.030)	0.388 (0.065)

Second, the analysis focused on the differences between teams that reached different stages of the tournament. Thus, the samples were divided into five groups according to the stage reached in the tournament (Group 1: Final; Group 2: Semi-finals; Group 3: Quarter-finals; Group 4: Round of 16;

Group 5: Group Stage). In this case, Levene's test verified that no variables violated the assumption of homogeneity of the variances. As a result, the analysis was accomplished using ANOVA. Table 4 shows the results of the one-way between-groups analysis of variance and the post-hoc tests conducted to explore the differences between teams that reached different stages of the tournament on the variables.

Table 4: Descriptive table (mean and standard deviation) and statistical comparison between teams that reached different stages in the tournament

	Final (N=14)		**Semi-finals** (N=12)		**Quarter-finals** (N=20)		**Round of 16** (N=32)		**Group Stage** (N=24)
(1) Nu passes	549.860 (106.923)	**	628.580 (163.71)	**	512.350 (104.276)		518.500 (145.832)		424.630 (110.683)
(2) Nu passes completed	474.360 (111.983)	**	539.670 (160.958)	**	426.050 (110.001)		435.590 (147.279)	**	336.790 (107.606)
(3) % passes completed	0.854 (0.048)	**	0.850 (0.043)	**	0.823 (0.054)		0.822 (0.082)		0.781 (0.065)
(4) α	1.567 (0.029)	**	1.565 (0.100)	**	1.600 (0.103)		1.597 (0.138)	**	1.692 (0.106)
(5) Density	0.220 (0.025)	**	0.230 (0.031)	**	0.209 (0.022)		0.212 (0.037)	**	0.187 (0.035)
(6) Avg Clustering coefficient	0.409 (0.055)	**	0.415 (0.069)	**	0.395 (0.060)	**	0.388 (0.081)	**	0.334 (0.071)

** Significantly different compared with the Group Stage

Statistically significant differences were found between groups (stage reached in the competition) in all variables: number of passes ($F_{4,97} = 5.605$, $p < 0.001, \eta^2 = 0.188,$ large effect); the number of passes completed ($F_{4,97} = 5.719$, $p < 0.001, \eta^2 = 0.191,$ large effect); the percentage of passes ($F_{4,97} = 3.770$, $p = 0.007, \eta^2 = 0.134,$ moderate effect); the parameter α ($F_{4,97} = 4.048$, $p = 0.004, \eta^2 = 0.143,$ large effect); the density ($F_{4,97} = 4.648$, $p = 0.002, \eta^2 = 0.162,$ large effect), and the average clustering coefficient ($F_{4,97} = 4.218$, $p = 0.003, \eta^2 = 0.148,$ large effect). Consequently, regarding, firstly, the number of passes and the percentage, the post-hoc tests indicated that the mean for Group 5 (Group Stage) was significantly different at the $p < 0.05$ from Group 2 (Semi-finals) and Group 1 (Final). Secondly, concerning the number of passes completed, the parameter α and the density, the test revealed that the mean number of passes completed for Group 5 (Group Stage) was significantly different at the $p < 0.05$ from Group 4 (Round of 16), from Group 2 (Semi-finals) and, from Group 1 (Final). Finally, post-hoc tests showed that the average clustering coefficient for Group 5 (Group Stage) was significantly different from all the other groups, i.e., from Group 4 (Round of 16), Group 3 (Quarter-finals), Group 2 (Semi-finals) and Group 1 (Final).

These results revealed that unsuccessful teams, i.e., teams eliminated in the first stage of the tournament (Group Stage), adopted a more direct type of play and were characterised by performing fewer passes and fewer passes completed. These findings contradict Bate (1988) and extend the findings of Hughes *et al.* (1988). On the one hand, the idea of Bate (1988) that teams should adopt direct play with fewer passes per possession rather than a possessive type of play to be successful was refuted by this study's findings. On the other hand, the findings of Hughes *et al.* [26], in which it was suggested that most successful teams played with more passes per possession than unsuccessful teams, were extended with the introduction of the parameter α and the discovered relationships of it with the pass statistics. Furthermore, the unsuccessful teams had lower values of density and average clustering coefficient. Thus, these findings are consistent with the conclusions of Grund [33], Clemente *et al.* [42] and Gonçalves *et al.* [43]. They found that successful teams are associated with high levels and distribution of interactions. Clemente *et al.* [42] also concluded that high cooperation and interconnectivity could lead to better performance outcomes, as also suggested by the previous results.

Additionally, the research work's methodology of Hughes & Franks [24] was reproduced. All goals scored from a sequence of one or more passes during regular time and extra time were considered in this analysis. The 14 goals that came from a possession without any passes (such as penalty kicks, direct free kicks, and ball recoveries immediately following a goal) and the 11 own goals were, thus, excluded from the analysis of the 142 goals scored during the tournament. The UEFA EURO 2020 data revealed that 80% of the goals resulted from 12 passes or less. Indeed, approximately 50% of the goals resulted from possessions of five or fewer passes.

As seen in Figures 1 and 2, these results can be explained by the tail's elongation of the goal-scoring possessions' distribution which in turn is explained by the tail's elongation of the passing sequences' distribution.

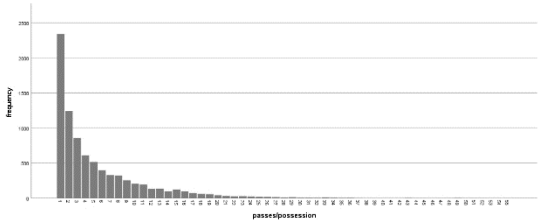

Figure 1: Frequency of each sequence length in the tournament

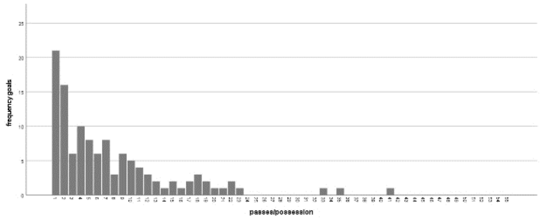

Figure 2: Frequency of goals concerning the length of the possession in the tournament

However, as the frequencies of occurrences are unequal, the results were normalised by dividing the number of goals scored in each team's possession by the sequence length. Therefore, a profile of the relative importance of the different passing sequence lengths was obtained. Figure 3 shows that the longer passing sequence lengths have a higher conversion ratio of goals per 1000 possessions. These results indicate that teams that have the capacity to sustain long passing sequences tend to score more goals [24]. Note that the low value of the goals/1000 possession that resulted from an eight-pass sequence can be classified as an outlier of the dataset.

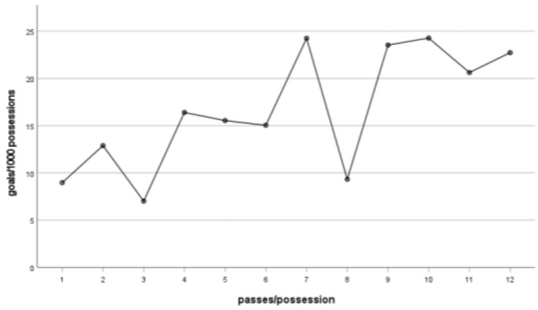

Figure 3: Analysis of the number of goals scored per 1000 possession for the tournament

Finally, the sample was divided into two groups: the goals per 1000 possessions that resulted from a sequence of 6 or fewer passes and the goals per 1000 possessions that resulted from a sequence of 7 or more passes. The means of goals per 1000 possessions for sequence lengths 0–6 and 7-12 were compared using a t-test after performing a descriptive group statistics analysis (Table 5). There was a significant difference between the two groups ($t_{10} = -2.878; p = 0.016; \eta^2 = 0.45,$ large effect).

Table 5: Group Statistics for each group (goals per 1000 possessions for sequence lengths 0–6 and 7-12)

	Length	N	Mean	Std. Deviation	Std. Error
Goals/1000 possessions	1-6	6	12.637	3.834	1.565
	7-12	6	20.784	5.779	2.359

These results indicated that, nowadays, teams score more goals from longer passing sequences compared to data from the last century. Moreover, this reveals how professional football has evolved in the last decades, with teams exchanging and sustaining the ball longer in their possessions. The increase in this threshold demonstrates how football has become more organised, being necessary to exchange the ball more, creating unbalances and disassembling the opposing team's structure to score goals.

5. Conclusions

This study concluded that the power law was an appropriate model for a part of the distributions of passes per possession. Consequently, this work contributed with an innovative way to describe the general attacking strategy of football teams through the power law exponent, $-\alpha$. Thus, teams that use a possession-based strategy of play have a lower value of α, whereas teams that use a direct strategy of play have a higher value of α. The main findings of this study suggest that unsuccessful teams adopt a direct play, which is characterised by lower values of all variables, i.e., number of passes, number of passes completed, percentage of passes completed, network's density and average clustering coefficient.

This study faces some limitations that should be addressed. On the one hand, as this work was time constrained, the scope of the analysis was limited since much time was consumed in designing and developing the Python scripts that made the multiple analysis from StatsBomb's raw data possible. However, this limitation can be considered an advantage because, with the code developed, the study can be replicated for other tournaments provided by StatsBomb. On the other hand, although 30% of the samples were not consistent with the power-law hypothesis, all the samples were fitted to the power-law distribution.

As a result, this study provides future research proposals that complement the present work and the literature in general. First, further investigations should replicate this study's methodology with other data sets to validate and corroborate this work's findings. Second, the matter of how the systems of play affect the characteristics of the networks should be a subject of future studies using different methodologies. Different clustering methods should be experienced to verify if they can unveil the differences between the different systems of play. The motifs between playing positions of the same sector and between playing positions of different sectors should also be studied in addition to the micro and macro levels of networks. Third, future studies should continuously focus on the study of network metrics at a macro, but more particularly at a micro level that can reflect the teams' general attacking strategies. Fourth, how adapting the general strategy of play to the opponent can lead to winning the match should be investigated. Fifth, further studies should consider the spatiotemporal evolution of the football passing networks, namely the player/playing position-zone networks, to enhance the knowledge of how teams organise and evolve during a match and how it relates to their performance. Finally, future research should address one significant gap in the literature: the need to consider how players and teams adapt to the ball's location in the field of play. This will provide pertinent and detailed information on how players interact within the game's dynamics.

References

[1] L. Gyarmati, H. Kwak, and P. Rodriguez, "Searching for a Unique Style in Soccer," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.0308

[2] D. Memmert and D. Raabe, "Data analytics in football: Positional data collection, modelling and analysis. Routledge.," 2018.

[3] Y. Yamamoto and K. Yokoyama, "Common and unique network dynamics in football games," *PLoS One*, vol. 6, no. 12, Dec. 2011, doi: 10.1371/journal.pone.0029638.

[4] J. Goldstein, "Emergence as a Construct: History and Issues," *Emergence*, vol. 1, no. 1, pp. 49–72, Mar. 1999, doi: 10.1207/s15327000em0101_4.

[5] P. Bradley *et al.*, *Football Analytics 2021*. 2021.

[6] J. Ribeiro, P. Silva, R. Duarte, K. Davids, and J. Garganta, "Team Sports Performance Analysed Through the Lens of Social Network Theory: Implications for Research and Practice," *Sports Medicine*, vol. 47, no. 9, pp. 1689–1696, Sep. 2017, doi: 10.1007/s40279-017-0695-1.

[7] M. Kempe, M. Vogelbein, D. Memmert, and S. Nopp, "Possession vs. Direct Play: Evaluating Tactical Behavior in Elite Soccer," *International Journal of Sports Science*, vol. 2014, no. 6A, pp. 35–41, 2014, doi: 10.5923/s.sports.201401.05.

[8] C. Willy, E. A. M. Neugebauer, and H. Gerngroß, "The concept of nonlinearity in complex systems: An additional approach to understand the pathophysiology of severe trauma and shock," *European Journal of Trauma*, vol. 29, no. 1. pp. 11–22, Feb. 2003. doi: 10.1007/s00068-003-1248-x.

[9] O. Hanseth and K. Lyytinen, "Design theory for dynamic complexity in information infrastructures: The case of

building internet," in *Enacting Research Methods in Information Systems: Volume 3*, Springer International Publishing, 2016, pp. 104–142. doi: 10.1007/978-3-319-29272-4_4.
[10] A. Martín, “Match Analysis: Final Assessement.” Barça Innovation Hub, 2022.
[11] J. L. Peña and H. Touchette, "A network theory analysis of football strategies," Jun. 2012, [Online]. Available: http://arxiv.org/abs/1206.6904
[12] J. Duch, J. S. Waitzman, and L. A. Nunes Amaral, "Quantifying the performance of individual players in a team activity," *PLoS One*, vol. 5, no. 6, 2010, doi: 10.1371/journal.pone.0010937.
[13] J. M. Buldú *et al.*, "Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game," *Frontiers in Psychology*, vol. 9, no. OCT. Frontiers Media S.A., Oct. 08, 2018. doi: 10.3389/fpsyg.2018.01900.
[14] F. Korte and M. Lames, "Passing Network Analysis of Positional Attack Formations in Handball," *J Hum Kinet*, vol. 70, no. 1, pp. 209–221, Nov. 2019, doi: 10.2478/hukin-2019-0044.
[15] S. Mclean, P. M. Salmon, A. D. Gorman, N. J. Stevens, and C. Solomon, "A social network analysis of the goal scoring passing networks of the 2016 European Football Championships," *Hum Mov Sci*, vol. 57, pp. 400–408, Feb. 2017, doi: 10.1016/j.humov.2017.10.001.
[16] C. Hook and M. D. Hughes, "Patterns of play leading to shots in 'Euro 2000,'" *Pass.com. Cardiff: UWIC, pp. 295-302*, 2001.
[17] C. Lago-Peñas and A. Dellal, "Ball possession strategies in elite soccer according to the evolution of the match-score: The influence of situational variables," *J Hum Kinet*, vol. 25, no. 1, pp. 93–100, Sep. 2010, doi: 10.2478/v10078-010-0036-z.
[18] P. D. Jones, N. James, and S. D. Mellalieu, "Possession as a performance indicator in soccer.," *Int J Perform Anal Sport*, vol. 4, no. 1, pp. 98–102, Aug. 2004, doi: 10.1080/24748668.2004.11868295.
[19] C. A. Casal, M. T. Anguera, R. Maneiro, and J. L. Losada, "Possession in football: More than a quantitative aspect - A mixed method study," *Front Psychol*, vol. 10, no. MAR, 2019, doi: 10.3389/fpsyg.2019.00501.
[20] J. Fernandez and L. Bornn, "Wide Open Spaces: A Statistical Technique for Measuring Space Creation In Professional Soccer," *Sloan Sports Analytics Conference, retrieved from http://www.sloansportsconference. com/wp-content/uuploads/ 2018/03/1003. pdf*, 2018.
[21] D. Memmert, D. Raabe, S. Schwab, and R. Rein, "A tactical comparison of the 4-2-3-1 and 3-5-2 formation in soccer: A theory-oriented, experimental approach based on positional data in an 11 vs. 11 game set-up," *PLoS One*, vol. 14, no. 1, Jan. 2019, doi: 10.1371/journal.pone.0210191.
[22] C. Reep and B. Benjamin, "Skill and Chance in Association Football," 1968.
[23] C. Reep, R. Pollard, and B. Benjamin, "Skill and chance inball games.," *Journal of the Royal Statistical Society*, vol. 134, pp. 623–329, 1971.
[24] M. Hughes and I. Franks, "Analysis of passing sequences, shots and goals in soccer," *J Sports Sci*, vol. 23, no. 5, pp. 509–514, May 2005, doi: 10.1080/02640410410001716779.
[25] R. Bate, *Football chance: tactics and strategy.* . 1988.
[26] M. D. Hughes, K. Robertson, and A. Nicholson, "An analysis of the 1984 World Cup of Association Football.," *In Science and Football*, 1988.
[27] F. M. Clemente, F. M. L. Martins, and R. S. Mendes, "Analysis of scored and conceded goals by a football team throughout a season: A network analysis," *Kinesiology*, vol. 48, no. 1, pp. 103–114, Jun. 2016, doi: 10.26582/k.48.1.5.
[28] C. Cotta, A. M. Mora, J. J. Merelo, and C. Merelo-Molina, "A network analysis of the 2010 FIFA world cup champion team play," *J Syst Sci Complex*, vol. 26, no. 1, pp. 21–42, Feb. 2013, doi: 10.1007/s11424-013-2291-2.
[29] F. M. L. Martins, F. M. Clemente, and M. S. Couceiro, "From the individual to the collective analysis at the football game.," *Paper presented at the Proceedings Mathematical Methods in Engineering International Conference, Porto*, 2013, doi: 10.5890/JAND.2012.02.001.
[30] P. Gould and A. Gatrell, "A Structural Analysis of a Game: The Liverpool v Manchester United Cup Final of 1977," 1979.
[31] J. M. Buldú, J. Busquets, I. Echegoyen, and F. Seirul.lo, "Defining a historic football team: Using Network Science to analyse Guardiola's F.C. Barcelona," *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-49969-2.
[32] P. Passos, K. Davids, D. Araújo, N. Paz, J. Minguéns, and J. Mendes, "Networks as a novel tool for studying team ball sports as complex social systems," *J Sci Med Sport*, vol. 14, no. 2, pp. 170–176, Mar. 2011, doi: 10.1016/j.jsams.2010.10.459.
[33] T. U. Grund, "Network structure and team performance: The case of English Premier League soccer teams," *Soc Networks*, vol. 34, no. 4, pp. 682–690, Oct. 2012, doi: 10.1016/j.socnet.2012.08.004.
[34] J. Gama, M. Couceiro, ; Gonçalo Dias, and ; Vasco Vaz, "SMALL-WORLD NETWORKS IN PROFESSIONAL FOOTBALL: CONCEPTUAL MODEL AND DATA," 2015.
[35] J. L. Herrera-Diestra *et al.*, "Pitch networks reveal organisational and spatial patterns of Guardiola's F.C. Barcelona," *Chaos Solitons Fractals*, vol. 138, Sep. 2020, doi: 10.1016/j.chaos.2020.109934.
[36] J. Diquigiovanni and B. Scarpa, "Analysis of association football playing styles: An innovative method to cluster networks," *Stat Modelling*, vol. 19, no. 1, pp. 28–54, Feb. 2019, doi: 10.1177/1471082X18808628.
[37] P. Malta and B. Travassos, “Caraterização da transição defesa-ataque de uma equipa de Futebol,” *Motricidade*, vol. 10, no. 1, pp. 27–37, 2014, doi: 10.6063/motricidade.10(1).1544.
[38] T. Narizuka and Y. Yamazaki, "Clustering algorithm for formations in football games," *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-48623-1.
[39] F. M. Clemente *et al.*, "Network structure and centralisation tendencies in professional football teams from Spanish La Liga and english premier leagues," *Journal of Human Sport and Exercise*, vol. 11, no. 3, pp. 376–389, 2016, doi: 10.14198/jhse.2016.113.06.
[40] J. L. Peña and R. S. Navarro, “Who can replace Xavi? A passing motif analysis of football players," Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.07768
[41] J. Bekkers and S. Dabadghao, "Flow motifs in soccer: What can passing behavior tell us?," *Journal of Sports Analytics*, vol. 5, no. 4, pp. 299–311, Aug. 2019, doi: 10.3233/jsa-190290.
[42] F. M. Clemente, F. M. L. Martins, D. Kalamaras, D. P. Wong, and R. S. Mendes, "General network analysis of national soccer teams in Fifa World Cup 2014," *Int J Perform Anal Sport*, vol. 15, no. 1, pp. 80–96, Mar. 2015, doi: 10.1080/24748668.2015.11868778.
[43] 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*, vol. 12, no. 1, Jan. 2017, doi: 10.1371/journal.pone.0171156.
[44] S. McLean, P. M. Salmon, A. D. Gorman, J. Wickham, E. Berber, and C. Solomon, "The effect of playing formation on the passing network characteristics of a professional football team," *Human Movement*, vol. 2018, no. 5, pp. 14–22, 2018, doi: 10.5114/hm.2018.79416.
[45] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4. pp. 661–703, 2009. doi: 10.1137/070710111.
[46] M. L. Goldstein, S. A. Morris, and G. G. Yen, "Problems with fitting to the power-law distribution," *European Physical Journal B*, vol. 41, no. 2, pp. 255–258, Sep. 2004, doi: 10.1140/epjb/e2004-00316-5.
[47] J. Alstott, E. Bullmore, and D. Plenz, "Powerlaw: A python package for analysis of heavy-tailed distributions," *PLoS One*, vol. 9, no. 1, Jan. 2014, doi: 10.1371/journal.pone.0085777.
[48] A. Tenga, L. T. Ronglan, and R. Bahr, "Measuring the effectiveness of offensive match-play in professional soccer," *Eur J Sport Sci*, vol. 10, no. 4, pp. 269–277, Jul. 2010, doi: 10.1080/17461390903515170.
[49] T. J. Pina, A. Paulo, and D. Araújo, "Network characteristics of successful performance in association football. A study on the UEFA champions league," *Front Psychol*, vol. 8, no. JUL, Jul. 2017, doi: 10.3389/fpsyg.2017.01173.
[50] A. L. Barabási, *Network Science*. Cambridge University Press, Cambridge., 2016.
[51] G. Fagiolo, "Clustering in Complex Directed Networks," 2007.
[52] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences Second Edition," 1988.
[53] J. Gama, G. Dias, M. Couceiro, T. Sousa, and V. Vaz, "Networks metrics and ball possession in professional football," *Complexity*, vol. 21, pp. 342–354, Nov. 2016, doi: 10.1002/cplx.21813.