Passing sequences and networks analysis in football

A study on the UEFA EURO 2020

Manuel Maria Strecht Ribeiro Hipólito Reis

Dissertation to obtain the Master of Science Degree in

Industrial Engineering and Management

Supervisors: Prof. José Rui de Matos Figueira
Prof. Francisco João Duarte Cordeiro Correia dos Santos

Examination Committee

Chairperson: Prof. Ana Isabel Cerqueira de Sousa Gouveia Carvalho

Supervisor: Prof. José Rui de Matos Figueira

Member of the committee: Prof. Andreia Sofia Teixeira

December 2022

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

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, $-\alpha$. 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, General attacking strategy, Team performance, Network Science

Resumo

A Ciência de Redes e a Teoria de Grafos podem contribuir, decisivamente, para a análise do futebol, disponibilizando ferramentas valiosas que permitem descrever o comportamento relacional das equipas, de forma mais consistente que a análise tradicional, baseada no desempenho individual dos jogadores. Até à data, poucos trabalhos de investigação estudaram a relação entre a estratégia de ataque de uma equipa (jogo de posse ou jogo direto) e as características da rede e como estas afetam o desempenho global dessa mesma equipa. Por outro lado, o impacto dos sistemas de jogo nas características da rede ainda não é claro. Com o presente estudo, pretendeu-se analisar as sequências e redes de passes das seleções nacionais que participaram no UEFA EURO 2020. Verificando-se que a maior parte das distribuições de passes completos das equipas tende a seguir a lei de potência, foi proposta uma forma inovadora de descrever a estratégia geral do jogo, através do expoente da lei de potência, $-\alpha$. Assim, as equipas com um jogo de posse têm um menor valor de α, enquanto as equipas com um jogo direto têm um maior valor de α. Os resultados dos estudos estatísticos realizados sugeriram que as equipas que adotaram uma estratégia de jogo direto, caracterizadas por executar menos passes e com menos sucesso, gerar redes menos densas e com um coeficiente de agrupamento médio mais baixo, obtiveram menos sucesso, sendo eliminadas na primeira fase do torneio. Por último, a análise de *clusters* efetuada foi inconclusiva no que se refere a revelar como os sistemas de jogo afetam as características das redes. Em suma, o presente estudo fornece vários conhecimentos relevantes que podem ser uma ferramenta útil no trabalho das equipas técnicas, reforçando a utilidade da Ciência de Redes e a Teoria de Grafos na análise do futebol.

Palavras-chave: Futebol, Sequências de passes, Redes de passes, Estrátegia geral de ataque, Desempenho da equipa, Ciência de redes

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

FIFA – *Fédération Internationale de Football Association*

IFAB – International Football Association Board

M – Mean

PC – Principal Component

SC – Silhouette Coefficient

SD – Standard Deviation

SSE – Sum of Squared Errors

UEFA – Union of European Football Associations

Chapter 1 – Introduction

This chapter introduces the dissertation and is organised into three sections. Section 1.1 describes the motivation for this work, while section 1.2 presents the objectives and research questions. Finally, section 1.3 provides the document's structure.

1.1. Motivation

Football, also known as soccer, is the most popular sport in the world. The number of players and fans has increased significantly worldwide since its inception in the 19th century in England (Garganta & Barreira, 2013). According to *Fédération Internationale de Football Association* (FIFA), there are 265 million people who play football[1], more than 130,700 active professional players[2], and a remarkable 5 billion football fans globally[3].

Since the two-opponent offside rule was established in 1920, football's fundamental rules have almost not been altered (Gyarmati *et al.*, 2014). 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 thoroughly scrutinise match analysis (Memmert & Raabe, 2018). Furthermore, the methodology of match analysis has been supported by a combination of increased computational power and new technologies, such as the global positioning system (GPS), new video recording tools, and physical data devices that allow the collection of performance data. Thus, football analysis departments have transformed into multidisciplinary panels of specialists due to the exponential growth in data availability in recent years (Caicedo-Parada *et al.*, 2020). These professionals include sports scientists, computer scientists, mathematicians, and audio-visual technicians. Their task is to extract information from the performance data and produce knowledge about their team to aid the coaching staff's decision-making (Clemente, Martins *et al.*, 2016; Diquigiovanni & Scarpa, 2019; Duarte *et al.*, 2012; Sarmento *et al.*, 2018; Vales-Vásquez, 2012).

The most appealing aspect of football is its emergent properties (Yamamoto & Yokoyama, 2011). Goldstein (1999) 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 (Bradley *et al.*, 2021). 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 (Kempe *et al.*, 2014; Ribeiro *et al.*, 2017). Hence, 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 (Hanseth & Lyytinen, 2016; Willy *et al.*, 2003). The context and environment for creating a team system

[1] Source: FIFA. (2006). FIFA Big Count 2006: 270 million people active in football. Retrieved from: https://resources.fifa.com/image/upload/big-count-stats-package-520046.pdf?cloudid=mzid0qmguixkcmruvema.
[2] Source: FIFA. (2022). Total Number of Professional Players. Retrieved from: https://landscape.fifa.com/en/landscape.
[3] Source: FIFA. (2022). Total Number of Professional Players. Retrieved from: https://landscape.fifa.com/en/landscape.

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 (Martín, 2022).

Consequently, analysing football quantitively is complicated due to its unique nature (Peña & Touchette, 2012). 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 (Duch *et al.*, 2010; Peña & Touchette, 2012). On the other hand, passes are the links between teammates and occur numerously in every match despite the quality of the teams (Gyarmati *et al.*, 2014). Therefore, the passes performed in a match provide substantial elements for applying graph and complex networks theory to football analysis (Arriaza-Ardiles *et al.*, 2018). Indeed, network analysis, by modelling the interactions based on the passes, captures teams' interactive behaviour, organisation and performance in a way that classical analysis, based on the performance of individual players, does not (Buldú *et al.*, 2018; Korte & Lames, 2019; Mclean *et al.*, 2017).

1.2. Objectives and Research Questions

The ability to retain possession of the ball for more extended periods has been linked to success (Hook & Hughes, 2001; Jones *et al.*, 2004; Lago-Peñas & Dellal, 2010). However, the difficulty in describing the possession characteristics is recognised in football performance analysis (Olsen & Larsen, 1997). Consequently, the ability to describe team possession in football must be improved. In recent years, ball possession has acquired fundamental importance in the attacking strategy of football teams (direct or possessive play) (Casal *et al.*, 2019). However, the relationship between teams’ attacking strategy and the networks’ characteristics and how these two factors impact teams’ overall performance has barely been unveiled.

Alternatively, the systems of play are the foundations of the football game, providing a reference to the team that assists players in positioning themselves and widely defining their specific roles in the attack and defence phases (Bradley *et al.*, 2021; Fernandez & Bornn, 2018). Thus, the system of play influences the team’s network characteristics. However, few studies have investigated the effect of the systems of play in professional football, making the impact on the network's characteristics unclear (Memmert *et al.*, 2019).

Therefore, this work advances with three research questions that 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 dissertation seeks to answer these research questions by applying statistical procedures, graph theory and network science and using as a sample the matches of the UEFA EURO 2020. In this way, the literature work in passing sequences and network analysis is extended.

1.3. Dissertation's Structure

This dissertation is structured into six chapters outlined below:

- **Chapter 1 – Introduction:** This first chapter introduces the dissertation, comprehending, firstly, the motivation for this work, secondly, a formulation of the research questions and resultant works' objectives and, finally, a specification of the document's structure.
- **Chapter 2 – Background:** This second chapter explains the main concepts required to understand the subsequent work. By setting a common terminology, this chapter covers the essential themes of this dissertation: football, graph theory and network science. Additionally, this chapter characterises the data studied in the analysis and the respective materials.
- **Chapter 3 – Literature Review:** This third chapter reviews the literature on passing sequences and network analysis and presents a summary enhancing the relevancy of the research questions.
- **Chapter 4 – Passing sequences analysis:** This fourth chapter exhibits the methodology, results and posterior discussion regarding the passing sequences analysis.
- **Chapter 5 – Passing networks analysis:** This fifth chapter presents the methodology, results and posterior discussion regarding the passing networks analysis.
- **Chapter 6 – Conclusions, Limitations and Future Work:** This sixth chapter summarises this dissertation's most relevant conclusions and insights, presenting the main limitations and highlighting opportunities for future work.

Chapter 2 – Background

This chapter explains the main concepts required to understand the subsequent work. By setting a common terminology, this part covers the essential themes of this work. First, in section 2.1, football's main rules are introduced (section 2.1.1), along with several notions about the game (section 2.1.2). Then, section 2.2 presents different definitions and concepts regarding graph theory and network science. Finally, section 2.3 characterises the data studied in the analyses and the used materials.

2.1. Football

This section presents football's principal rules and concepts of the game needed to comprehend the work that follows.

2.1.1. Main rules

A professional football match is played between two teams using a spherical ball on a rectangular grass field (natural or artificial), with two goals at the end of each width. Each team has eleven players, one of whom must be the goalkeeper (see section 2.1.2.2). The game's objective is to score more goals on the opposing goal than the opponent. A goal is scored when the entire ball passes over the goal line, between the goalposts, and under the crossbar (IFAB, 2021).

According to the first law of the game, the field is bounded by the touchlines (length sides) and two goal lines (width sides). The halfway line divides the field into two halves, and at its midpoint is the centre mark, which serves as the starting point for the game. Because all the opponent's players must be in their half and at least 9.15 meters from the ball until the game is started or restarted, the centre circle is marked around the centre mark and provides a reference for the kick-off (see section 2.1.2.1). Competitions can define the dimensions requirements according to the following constraints: the field's length must be between 90.00 and 120.00 meters, and the width must be between 45.00 and 90.00 meters (IFAB, 2021).

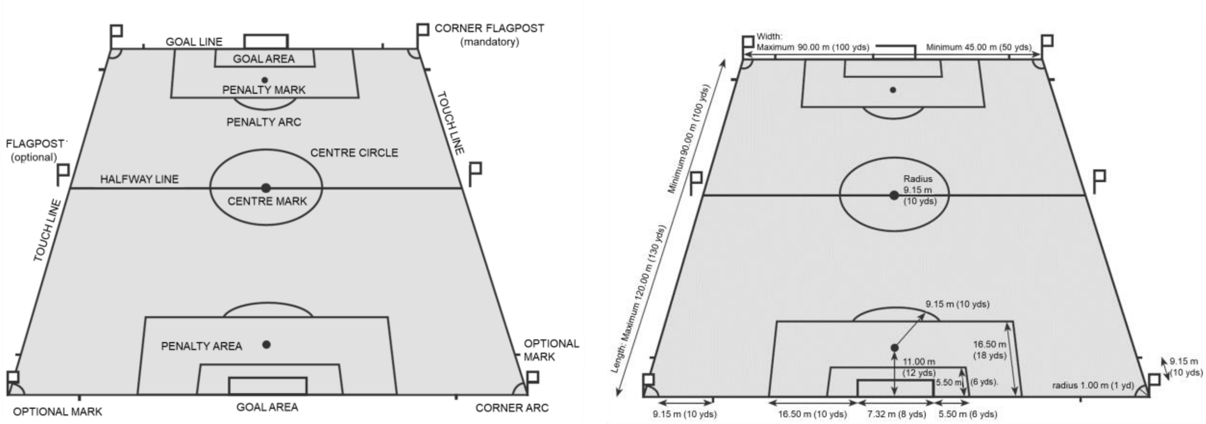


Figure 1: The field of play: components and measurements (Adapted from FIFA, (2015))

As seen in Figure 1, the penalty area is located on each half-side of the field, and inside this rectangular area, the goalkeeper can use his arms and hands to defend the ball. Inside the penalty area, a penalty mark is drawn 11 meters from the midpoint between the goalposts, and this is where the penalty kicks are taken (see section 2.2.1.). Outside the penalty area, an arc of a circle with a radius of 9.15 meters and a centre in the penalty mark is depicted as a reference for players other than the penalty kicker and goalkeeper, who must be at least 9.15 meters away from the penalty mark (IFAB, 2021). Within the penalty area is also the goal area. At 5.50 meters from the inner of each goalpost, two lines are drawn perpendicular to the goal line. A line drawn parallel to the goal line connects these lines, which extend 5.50 meters onto the playing field. The goal area is the region enclosed by these lines and the goal line and is used as a guide for goal kicks (see section 2.2.1.) since the ball must be kicked by a player inside of this area (IFAB, 2021).

A match lasts 90 minutes and is divided into two 45-minute halves. In addition, in some competition stages, if the score is tied, two equal additional periods of 15 minutes each may be played. If this extra time still ends in a draw, a penalty shootout may be held until there is a winning team.

A team can only make a certain number of substitutions during a match. Because of the COVID-19 pandemic's impact on football players, the International Football Association Board (IFAB), responsible for establishing the Laws of the Game, has approved an amendment to the third law, increasing the maximum number of substitutes from three to five. These five substitutions can be done in a maximum of three moments. Moreover, if the game goes to extra time, teams have the opportunity to use an additional substitute[4].

Finally, one of the most important rules is the offside rule. A player is in an offside position if, in the opponent's half-side, any part of his head, body, or feet is closer to the opponent's goal line than the ball and the second-last opponent. This event is penalised with a fault in favour of the opposing team (IFAB, 2021).

2.1.2. The game

The concepts of the game cycle, systems of play, and playing positions are presented in this section.

2.1.2.1. Game cycle

A football game is a whole, but it is possible to distinguish stages within it. The game can be described as a cycle that is made up of both dynamic and static phases (Castellano, 2000; Martín-Barrero & Ignacio Martínez-Cabrera, 2019). Consequently, the two phases of the game are the attacking phase and the defence phase. On the one hand, during the attacking phase of play, players attempt to move the ball toward key areas of the field to score a goal. On the other hand, in the defence phase, the team does not possess the ball and attempts to reclaim it by preventing the opponent from moving closer to the goal and scoring a goal. (Greco & Greco, 2009; Hewitt *et al.*, 2016).

Each game situation depends on the previous one and influences the next (Soriano, 2019). Therefore, despite the difficulty of precisely dividing the game into moments, four dynamic moments of play can be differentiated: organised attack, attack-defence transition, organised defence, and defence-

[4] Source: IFAB. (2021). Additional Substitutes (Covid-19). Retrieved from: https://www.theifab.com/laws/latest/temporary-amendment-covid/

attack transition (Cano, 2009). As a result, it is possible to define each dynamic moment using Cano's (2009) proposal:

- Organised attack: this is an offensive moment of the game in which the opposing team's defence is well-organised, limiting the attacking team's ability to react quickly;
- Attack-defence transition: also referred to as counterattack; this is an offensive moment in which the defending team is surprised by the attacking team because the opposing team is defensively disorganised and thus vulnerable to a quick attack;
- Organised defence: this is a defensive phase of the game in which the defensive team is well-structured and cannot be momentarily caught off guard by the opposing team;
- Defence-attack transition: this is a defensive moment in which the defending team is exposed to the opposing team's attack (Martín-Barrero & Ignacio Martínez-Cabrera, 2019).

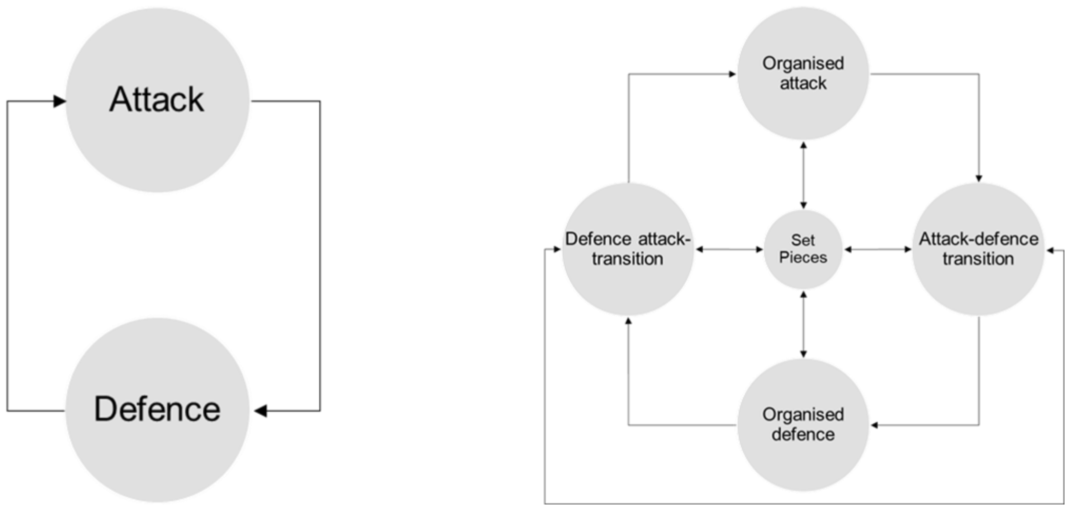


Figure 2:Phases and moments of play (Adapted from Hewitt *et al.* (2016); Martín-Barrero & Ignacio Martínez-Cabrera (2019)

The set pieces, all restarts that occur in the game, are the static phases of the game. In this phase, the game is restarted from a standing position with the foot or hand. The team in possession of the ball may begin the game whenever it wishes, as long as the time limit for restarting the game is not exceeded. Set pieces can be identified as reasonably stable conditions within the dynamic and complex football system. Are included in this phase of the game the following:

- Kick-off: starts a match's first and second halves, as well as both halves of extra time, and resumes play after a goal is scored;
- Goal-kick: is awarded when the entire ball crosses the goal line, whether on the ground or in the air, after having last touched a member of the opposing team, and no goal is scored;
- Throw-in: is awarded when the entire ball crosses over the touchline. It is awarded to the opponent team of the player who last touched the ball. A throw-in is performed with the hands, not being allowed to score a goal directly from it;
- Corner-kick: is granted when the entire ball crosses the goal line, whether on the ground or in the air, being last touched by a member of the opposing team, and no goal is scored;

- Free-kick: is granted to the opposing team when a player commits a fault. A direct free-kick is one in which the ball can be kicked directly into the opponent's goal, whereas an indirect free-kick is one in which this is not possible;
- A penalty kick: is granted to the opposing team if a player commits a direct free kick fault inside their penalty area (IFAB, 2021).

2.1.2.2. Systems of play and playing positions

Each player's position on the playing field facilitates the team's collective play development. The system of play has traditionally been the main point of reference for football players when deciding where to position on the field. This reference is organised in lines, with each player occupying a specific position within each line. The systems of play are generally defined by the number of players playing in each line (Vilar *et al.*, 2013). Each line is related to the sectors in which the field of play can be divided. The defensive, midfield, and offensive sectors are the three main sectors, and these sectors can be further subdivided. Garganta (1997), considering previous studies, presented a playing field division model that has 12 zones resulting from the combination of four sectors (a transversal division of the playing field) and three corridors (a longitudinal division of the playing field).

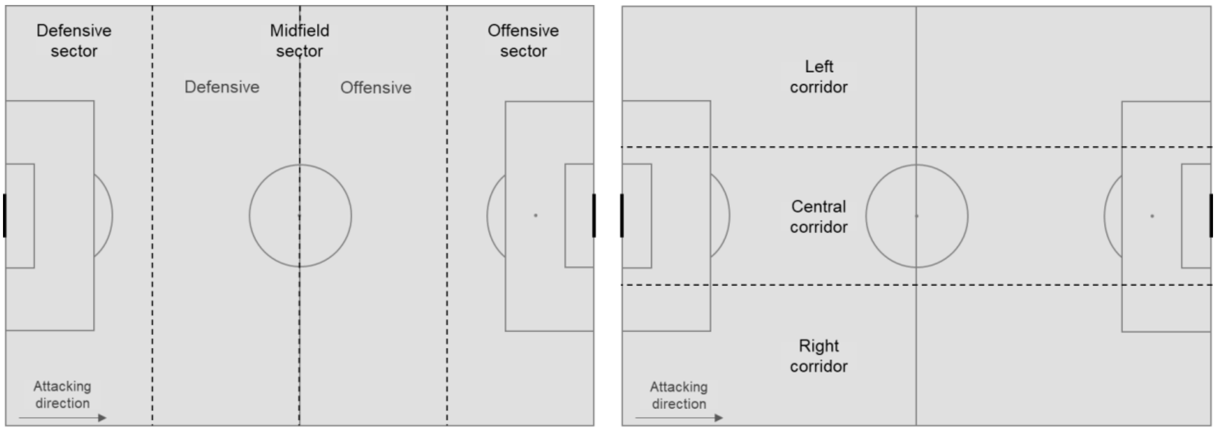


Figure 3: Division of the playing field (Adapted from Garganta, (1997))

In addition, other division models have been suggested. For instance, Diquigiovanni & Scarpa (2019) divided the playing field into nine zones, each consisting of three equal sectors and three equal corridors. Instead, Herrera-Diestra *et al.* (2020) used a thirty-zone division in their study (six equal sectors and five equal corridors).

As a result, it is possible to characterise the four main playing positions that a team has by contemplating the three main sectors:

- Goalkeeper: This player plays behind the three lines with gloves on and wearing a different colour jersey than his teammates. The goalkeeper is the only player on the team who is permitted to use his hands and arms inside the penalty area, and his primary duty is to prevent goals from the other team.
- Defender: the primary concern of this defensive line player is to contain opposing attackers and prevent the other team from scoring goals. A defender may be a centre-back if positioned in the central corridor between the full-backs or a full-back if they are placed in the outer corridors.

- Midfielder: the principal task of this player is to create the connection between the defensive and the offensive lines. The midfielder may play a more significant amount of defensive or offensive roles. First, he can serve as a defensive centre midfielder, sitting in front of the defensive line, assisting teammates with defensive responsibilities, and distributing the ball to teammates. Second, he can perform offensive and defensive tasks in various roles. Finally, this player can also be an attacking midfielder who assists the team's offensive efforts and generates opportunities for himself or the forward to score goals.
- Forward: this player's primary duty while in the attacking line is to score goals. The forward can be more versatile, helping his teammates score goals, or more of a target man, scoring goals primarily on his own.

The system of play is a method of organizing a team by creating a framework that guides the behaviours to achieve the desired interactions and relationships. Besides, the systems of play are not rigid, i.e., players play from their position rather than in their position, constantly adjusting their behaviours in a coordinated manner to achieve the performance objectives (scoring and preventing goals) (Vilar *et al.*, 2013). Hence, each team establishes an order between its players and its different lines. The systems of play are expressed in a sequence of numbers, i.e., a 4-3-3 has four defenders (defensive line), three midfielders (midfield line) and three forwards (offensive line) (Martín-Barrero & Ignacio Martínez-Cabrera, 2019; Mercé, 2004). There are several systems of play. Although they may share positions, there are differences between them. Additionally, the same systems of play can have multiple configurations. Thus, Figure 4 illustrates different examples of systems of play.

(4a) 4-3-3 (4b) 4-4-2 (4c) 4-2-3-1 (4d) 5-2-3

Figure 4: Examples of systems of play

2.2. Operational Research: Graph Theory and Network Science

Newman (2010) defines a network as “a collection of points joined together in pairs by lines” in which “the points are referred to as vertices or nodes and the lines are referred to as edges”. More specifically, social networks are networks in which nodes are people or groups of people, and the edges represent a social interaction between them, such as a pass in football (Newman, 2010).

Graph theory is a mathematics branch with technical tools to analyse networks (Newman, 2010). This section introduces a small portion of concepts of the vast field of graph theory, concentrating only on those that are relevant to this dissertation.

A network is denoted in graph theory as a graph, a set of vertices linked by edges. One or more edges can link two vertices. In addition, a vertex can be connected to itself by an edge (referred to as self-edge). A simple graph is a network with neither self-edges nor multiedges, unlike a multigraph, a network with multiedges. The number of nodes, the size of a network, represents the number of components in the network, whereas the number of edges represents the total number of interactions between the nodes (Barabási, 2016; Newman, 2010).

A network may have undirected or directed links. A network in which each edge has a direction pointing from one vertice to another is known as a directed network or directed graph (also designated as a digraph). Such edges are denoted as directed edges and can be represented by lines with arrows.

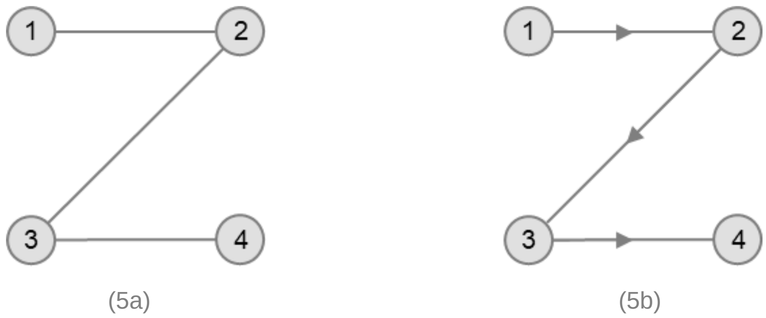


Figure 5: (5a) Simple undirected graph; (5b) Simple directed graph

2.2.1. Adjacency matrix

A network is usually represented by its adjacency matrix. The adjacency matrix A of a directed network of n nodes is a matrix that has n rows and n columns, with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1, & if\ there\ is\ an\ edge\ from\ j\ to\ i \\ 0, & otherwise \end{cases}$$

The adjacency matrix of an undirected network is symmetric $A_{ij} = A_{ji}$, thus having two entries for each link. As an example, the adjacency matrices of the networks represented in Figure 5 are:

$$A_{5a} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad A_{5b} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

In specific applications, networks might include weights reflecting the frequency of interaction between nodes. These networks in which each link (i, j) has a unique weight w_{ij} are called weighted networks and can be represented by the elements of the adjacency matrix values equal to the weight of the link, $A_{ij} = w_{ij}$ (Barabási, 2016; Newman, 2010).

2.2.2. Degree

The degree of a graph's vertex is the number of edges linked to it. On the one hand, the degree, k_i, of the vertex i for an undirected graph with n vertices can be expressed in terms of the adjacency matrix as:

$$k_i = \sum_{j=1}^{n} A_{ij}.$$

Since in an undirected graph, every edge has two ends, the total number of edges, m, can be written as:

$$m = \frac{1}{2}\sum_{j=1}^{n} k_i.$$

Moreover, the mean degree, c, of the undirected graph is:

$$c = \frac{1}{n}\sum_{j=1}^{n} k_i = \frac{2m}{n}$$

On the other hand, in a directed graph each vertex has two degrees: the in-degree and the out-degree. The in-degree is the number of ingoing edges connected to a vertex, while the out-degree is the number of outgoing edges. Thus, the in-degree, k_i^{in}, and out-degree, k_i^{out}, of the vertex i for a directed graph with n vertices can be expressed in terms of the adjacency matrix as (Barabási, 2016; Newman, 2010):

$$k_i^{in} = \sum_{j=1}^{n} A_{ij} \quad ; \quad k_i^{out} = \sum_{j=1}^{n} A_{ji}\,, \qquad with\ \ k_i = k_i^{in} + k_i^{out}.$$

Also, the total number of edges, m, in a directed graph can be written as:

$$m = \sum_{j=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_i^{out}.$$

Therefore, the mean in-degree, c_{in}, and the mean out-degree, c_{out}, are equal:

$$c_{in} = \frac{1}{n}\sum_{j=1}^{n} k_i^{in} = \frac{1}{n}\sum_{j=1}^{n} k_i^{out} = c_{out} = c = \frac{m}{n}.$$

2.2.3. Density and paths

The maximum possible number of edges in a simple undirected graph is $\frac{1}{2}n(n-1)$, whereas in a simple directed graph is $n(n-1)$. The density, ρ, is the interconnectedness of vertices of a graph and can be defined as the ratio between the number of edges and the maximum possible edges, lying in the range $0 \leq \rho \leq 1$. As a result, the density of a simple undirected graph and a simple directed graph are, respectively (Newman, 2010):

$$\rho_{undirected} = \frac{2m}{n(n-1)} \quad ; \quad \rho_{directed} = \frac{m}{n(n-1)}$$

Furthermore, a route along a network's links is referred to as a path. The length of a path is a measure of how many links are present on it. In addition, the geodesic path or shortest path, d_{ij}, between two nodes i and j is the path with the fewest number of edges (Barabási, 2016). In opposition, the diameter is the length of the longest path between any two vertices (Newman, 2010).

2.2.4. Centrality measures

The importance of the network's nodes is taken into account by the centrality measures, and each centrality measure examines a different type of importance. In this section, some measures are presented that are essential to comprehend the work that follows (Golbeck, 2015; Newman, 2010).

2.2.4.1. Degree centrality

The degree centrality is a simple centrality measure to compute, being just the degree of a vertex. It shows how many links a node has; thus, higher values mean the node is more central. In directed graphs, vertices have in-degree and out-degree centralities (Golbeck, 2015). For instance, in football, the player with the highest in-degree centrality is the one who receives more passes from teammates than the other players. In contrast, the player with the greatest out-degree centrality is the one who originated more passes than the other players (Clemente, Martins *et al.*, 2016).

2.2.4.2. Closeness centrality

The closeness centrality measure focuses on how close a node is to all other nodes in the network through the mean distance (length of the shortest path) between a vertex and other vertices (Clemente, Martins *et al.*, 2016; Newman, 2010). For example, a higher value of this measure in football indicates that one player chooses all the other players and that other players tend to primarily interact with this central player (Clemente, Martins *et al.*, 2016). Over the years, several authors have developed different closeness-based measures.

Sabidussi (1966) proposed that the sum of the geodesic distances between a vertex and all other vertices could be used to determine a vertex's centrality. However, this is a measure of inverse centrality because it increases with greater distance between a vertex and all other vertices (Freeman, 1978). Therefore, the measure of centrality vertex i for a graph with n vertices is:

$$C_c(i) = \frac{1}{\sum_j^n d_{ij}}$$

As described before, this measure depends on the number of vertices in the network from which it is computed. With the measure suggested by Sabidussi (1966), comparing graphs of different sizes is impossible. So, Beauchamp (1965) proposed a measure in which the impact of graph size was removed (Freeman, 1978):

$$C'_c(i) = \frac{n-1}{\sum_j^n d_{ij}}$$

Years later, Wasserman & Faust (1994) proposed a new closeness metric that ignores vertices that are not reachable from vertex i and focuses only on distances from vertex i to all reachable vertices. Even if the graph is not strongly connected, this measure is determined by considering the ratio of the fraction of reachable vertices to the average distance from all reachable vertices. Thus, denoting J_i as the number of vertices in the influence range of vertex i, this closeness metric can be expressed as (Wasserman & Faust, 1994):

$$C_c''(i) = \frac{\frac{J_i}{n-1}}{\frac{\sum_j^n d_{ij}}{J_i}}$$

2.2.4.3. Betweenness centrality

The betweenness centrality captures the degree to which a vertex i is on the shortest paths between other vertices (Newman, 2010). Thus, this measure is the sum of the fraction of all-pairs shortest paths that traverse the vertex i:

$$C_b(i) = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}},$$

with σ_{jk} being the number of shortest (j,k)-paths and $\sigma_{jk}(i)$ be the number of shortest (j,k)-paths passing through some vertex i other than (j,k) (Brandes, 2008). For illustration, football players with high betweenness centrality may have considerable influence within the passing network, acting as bridges between their teammates (Clemente, Martins, *et al.*, 2016).

2.2.4. Clustering coefficient

The clustering coefficient measures the degree to which the neighbours of a given vertex connect, quantifying how close a vertex and its neighbours in a graph are to forming a complete subgraph (Barabási, 2016; Clemente *et al.*, 2015). For a vertex i with a degree k_i of an undirected graph, the local clustering coefficient is defined as:

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

where T_i is the number of triangles through vertex i. Alternatively, for directed graphs, the clustering coefficient is defined as the fraction of all possible directed triangles (Fagiolo, 2007):

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

where, in this case, 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 (Pina *et al.*, 2017):

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

2.2.5. Motifs

Milo *et al.* (2002) first defined network motifs as ”patterns of interconnections occurring in complex networks in numbers that are significantly higher than those in randomized networks”, meaning that a motif is a subgraph that is statistically over-represented (Milo *et al.*, 2002; Stone *et al.*, 2019). This crucial concept, presented as a basic building block of complex networks, has been used to uncover network structural properties. Hence, as an example, Figure 6 shows all possible motifs of a three-node connected directed subgraph.

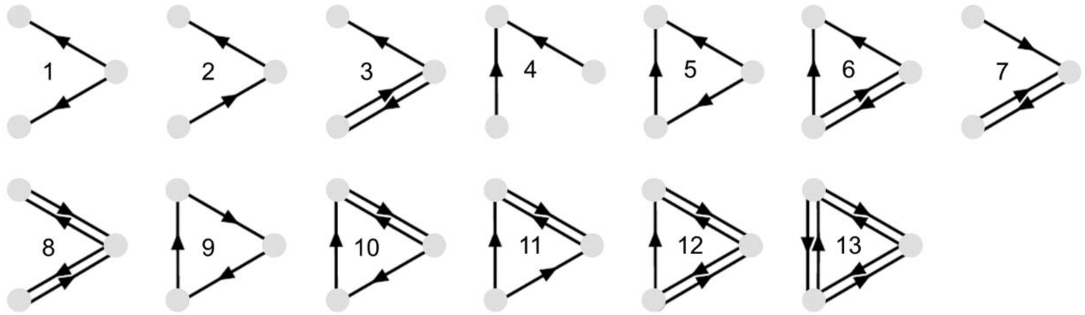


Figure 6: 13 different types of subgraphs of size 3

2.2.6. Scale-free networks

Many networks contain a small number of nodes with significantly more links than the average node. In these types of networks, termed scale-free networks, the fraction of nodes having k edges, p_k, decays according to a power law (Milo *et al.*, 2002; Yamamoto & Yokoyama, 2011):

$$p(k) \sim Ck^{-\alpha},$$

where the constant C is unimportant for the study and the exponent of the power law α usually ranges between 2 and 3, although values outside this range are feasible and sporadically observed (Newman, 2010).

2.3. Data characterisation

This section characterises the sample concerning the UEFA EURO 2020 and presents the used materials.

2.3.1. Sample

The 51 official matches from UEFA EURO 2020 were analysed (Appendix A). In this UEFA-organised tournament, the European senior men's national teams competed to crown the continental champion. The competition, held since 1960, is slated to occur every four years between FIFA World Cup competitions in the even-numbered years. However, this edition was postponed to 2021 due to the COVID-19 pandemic. The tournament was hosted in several countries to celebrate the 60th anniversary of the European Championship competition: Azerbaijan, Denmark, England, Germany, Hungary, Italy, Netherlands, Romania, Russia, Scotland, and Spain.

The tournament was competed by twenty-four teams, represented in Figure 7, 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(UEFA, 2018). The winning team of each match advanced to the next stage.

Group A	
Pos	Team
A1	Turkey
A2	Italy
A3	Wales
A4	Switzerland

Group B	
Pos	Team
B1	Denmark
B2	Finland
B3	Belgium
B4	Russia

Group C	
Pos	Team
C1	Netherlands
C2	Ukraine
C3	Austria
C4	North Macedonia

Group D	
Pos	Team
D1	England
D2	Croatia
D3	Scotland
D4	Czech Republic

Group E	
Pos	Team
E1	Spain
E2	Sweden
E3	Poland
E4	Slovakia

Group F	
Pos	Team
F1	Hungary
F2	Portugal
F3	France
F4	Germany

Figure 7: UEFA EURO 2020 participants and respective groups

2.3.2. Materials

The raw data sets were provided by StatsBomb Services Ltd, which has made the data from UEFA EURO 2020 publicly and freely available[5]. StatsBomb covers 90 different leagues worldwide, gathering data for each league match at the same granularity level. The UK-based company collects data validated in multiple layers using proprietary camera calibration, computer vision tools, and human input, guaranteeing the most accurate data for its clients[6]. In its open data, there were four different types of datasets in a JSON format[7]:

1. **StatsBomb Competition Data**: contains descriptive information about all competitions freely available;
2. **StatsBomb Match Data**: records the match information for each match, including competition and season information, home and away team information and stadium and referee information;
3. **StatsBomb Lineup Data**: reports the lineup information for the players, coaches, and referees involved in each match. The filenames correspond to the match ids.
4. **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. Once again, the filenames correspond to the match ids.

First, the StatsBomb Competition Data was accessed to get the corresponding UEFA EURO 2020 competition's id and general information about the tournament. Second, information about each tournament's match was collected from *StatsBomb Match Data*. Third, *StatsBomb Lineup Data* was not

[5] Source: StatsBomb. (2022). GitHub. Open data. Retrieved from: https://github.com/statsbomb/open-data
[6] Source: StatsBomb. (2022). Data Soccer. Retrieved from: https://statsbomb.com/what-we-do/soccer-data/
[7] Source: StatsBomb. (2022). GitHub. Open Data. StatsBomb Open Data Specification v1.1.pdf. Retrieved from: https://github.com/statsbomb/open-data/blob/master/doc/StatsBomb%20Open%20Data%20Specification%20v1.1.pdf

used. Finally, the *StatsBomb Event Data* of each tournament's match was used to perform all the studies presented in the coming chapters. 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[8].

The dissertation'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).

[8] https://github.com/statsbomb/open-data/blob/master/doc/StatsBomb%20Open%20Data%20Specification%20v1.1.pdf

Chapter 3 – Literature Review

This chapter reviews the literature on passing sequences analysis and passing network analysis in sections 3.1 and 3.2, respectively. In addition, section 3.3 presents a summary enhancing the relevancy of the research questions.

3.1. 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 (1968) statistically analysed the passing sequences that resulted in goals from football matches, presenting them as a negative binomial distribution. Reep *et al.* (1971) 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 (Hughes & Franks, 2005).

These works implied that those passing sequences with few passes were more successful. Consequently, as Bate (1988) 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.* (1988) studied the patterns of plays of the semifinalists and the national teams that were eliminated in the first round 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 (1968) and Bate (1988) did not apply to all levels of football (Hughes & Franks, 2005).

Years later, Hughes & Franks (2005) replicated the work of Reep & Benjamin (1968) and discovered that the conclusions reached by these authors could be misinterpreted. Because of this, Hughes & Franks (2005) 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 (Hughes & Franks, 2005).

Hughes & Franks (2005) 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 (Hughes & Franks, 2005).

3.2. Passing network 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 (Clemente, Martins *et al.*, 2016; Cotta *et al.*, 2013; Martins *et al.*, 2013). One of the first studies that introduced the concept of football passing networks was published by Gould & Gatrell (1979). They explored the structure of a football match, specifically the Cup Final of 1977 between Liverpool and Manchester United. However, as Buldú *et al.* (2019) 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 with the work conducted by Duch *et al.* (2010) (Buldú *et al.*, 2019).

Hence, through network analysis, Duch *et al.* (2010) evaluated players' performance in the EURO 2008 championship. The researchers identified the attacking plays that led to shots to create a directed weighted graph of the "ball flow", which included not only the players on a team but also two non-player nodes ("shots on goal" and "shots wide"). These two nodes were connected to a player's node by an arc and weighted based on the number of shots. By combining this network, denoted as the “flow network”, with passing accuracy and shooting accuracy, the probability that each network path would lead to a shot could be determined. As a result, they employed in this process a metric known as "flow centrality", i.e., the betweenness centrality of the player regarding the opponent’s goal. This metric recorded the percentage of times a player intervened in those paths that led to a shot. In addition, they defined each player's match performance as the normalised value of the logarithm of his flow centrality. Therefore, the researchers evaluated each player's influence on a game using these graph and centrality approaches, identifying the player who had the greatest influence on each team (Duch *et al.*, 2010).

They concluded that eight of the twenty players integrated into this study list were also included in the tournament's top twenty players selected by the technical panel of UEFA (the tournament organiser). Moreover, according to their study, they realised that Xavi Hernandez, the tournament's top player, was also named the tournament's best player (Clemente, Martins *et al.*, 2016; Duch *et al.*, 2010).

Similarly, Peña & Touchette (2012) constructed networks in which the players (nodes) were connected by passes (edges) using the data that was available from the FIFA World Cup 2010. The computation of centrality measures enabled researchers to examine the impact of eliminating a player from the game in addition to determining each player's relative importance (Peña & Touchette, 2012).

Another early study that used network science to investigate football was performed by Yamamoto & Yokoyama (2011). They showed how networks formed by player interactions throughout a game might characterize the team members' collective behaviours as indicated by topologies such as small-world networks and scale-free networks. Consequently, a few nodes (players) would typically exhibit more links than others in this type of network (Gama *et al.*, 2014; Yamamoto & Yokoyama, 2011).

Furthermore, the two investigators affirmed that because football teams typically have particularly dominant players who tend to dictate the game, it was reasonable to assume that the degree distribution during a game displays the power law distribution. By building networks for every five minutes of the two games analysed, Yamamoto & Yokoyama (2011) concluded that the hub's role was transferred to another node (player) as the network topology changed to follow the power of law. As a result, the authors identified “the stochastically switched dynamics of the hub player throughout the game”, a specific characteristic of football (Yamamoto & Yokoyama, 2011).

On the other hand, Passos *et al.* (2011) discussed the value of performing network analysis in team sports sciences and emphasized that small-world networks were a valuable technique for capturing dynamics in football, in line with Yamamoto & Yokoyama (2011) (Gama *et al.*, 2014; Passos *et al.*, 2011). The authors used network science to study water polo in their work, observing the passes as links between the players (nodes).

To determine how the number of intra-team interactions emerges in a game, they considered two key factors linked to successful patterns of play: the number of interactions between teammates and the probability of each player interacting with each teammate in the following phases of the attack. The results indicated that the high probability of each player interacting with other players in a team was necessary for the most successful collective system behaviours (Clemente, Martins *et al.*, 2016; Gama *et al.*, 2014; Passos *et al.*, 2011). Such evidence was also discovered in a research network analysis of twenty-three English Premier League teams, carried out in the following year by Grund (2012). Using a dataset of 760 football matches with 283,529 passes between teammates, the researcher demonstrated that high levels of interaction (density) were associated with higher team performance, which was measured by the goals scored. On the other hand, centralised interaction patterns led to lower team performance (goals scored) (Grund, 2012; Pina *et al.*, 2017).

Cotta *et al.* (2013) and Narizuka *et al.* (2014) were the only authors who applied a distinct methodology to represent the passing networks in their studies. Instead of representing the players or zones as nodes, they denoted as nodes the pairs (player, zone) to also capture the players' location. As the contributions of Narizuka *et al.* (2014) are not relevant to this dissertation, only Cotta *et al.* (2013) findings are presented. Analysing the network of passes of the Spanish national team during the FIFA World Cup 2010 tournament, the researchers made a temporal analysis of the passing networks, looking at the number of passes, length of the chain of passes, centrality measures and clustering coefficient. Studying, in particular, the last three matches of Spain, the results indicated that the clustering coefficient remained high during the game, reflecting the elaborated style of the Spanish team. Furthermore, the effectiveness of the opposing teams was shown in the change of several network measures over time, more specifically, in the decrease of not only the clustering coefficient and passing length values but also in the importance of the key players in the network.

Malta & Travassos (2014) characterised the defence-attack transition moment in football using network analysis. The two researchers considered 52 offensive play sequences from four games of one team in the Portuguese Premier League. To deal with these sequences of plays, they divided the playing field into 18 zones (6 sectors and 3 corridors) and identified the players' positions, treating these two approaches (player position as a node and zone as a node) separately. Then, for a better comprehension of the defence-attack transition moment, were computed for the two approaches different metrics, such as the betweenness centrality and the in-degree and out-degree centralities (Malta & Travassos, 2014).

Their study revealed that two types of play were preferred at this moment of the game. First, the analysed team had a possession type of play, heavily influenced by the defensive centre midfielders and in the midfield region. Second, the direct type of play was also observed, dominated in the forward region and by the centre forwards. Moreover, Malta & Travassos (2014) also found that the defensive midfielders had higher values of out-degree centrality and forwards had great levels of in-degree centralities, concluding that these player positions were the most important in this moment of play (Clemente, Martins, *et al.*, 2016; Malta & Travassos, 2014).

In another research work developed in the same year, Gyarmati *et al.* (2014) addressed whether it was possible to identify a unique style of football in the modern era by proposing a novel approach until then for quantifying teams' motif characteristics based on their passing networks. In particular, they introduced the idea of "flow motifs" of a passing network, an ordered list of players who participated in a set number of consecutive passes (in this case, three). Treating data from the top Premier Leagues (season 2012/2013), they concluded that FC Barcelona had a distinctive passing style (expressed by the unique motif characteristics) compared to the other teams. As a result, they emphasized that Barcelona's distinct *tiki-taka*[9] philosophy had a clear, structured framework rather than an uncountable number of random passes (Gyarmati *et al.*, 2014).

In a similar work, Peña & Navarro (2015) expanded the "flow motif" analysis to a player level, concentrating on researching motifs corresponding to sequences of three consecutive passes. First, they divided each of the conceivable 3-passes motifs into 15 distinct variations. Then, the frequency of each pattern occurring for each player in their dataset was determined, yielding a 15-dimensional distribution that represented the player's type of involvement with his teammates. Finally, a similarity measure was built using these feature vectors to quantify how similar any two players' playing styles were (Peña & Navarro, 2015).

Gama *et al.* (2014) sought to see if network analysis could be used to identify important players during the attacking phase of a football game. To accomplish this, they randomly selected six matches of a single team in the Portuguese Premier League and examined collective attacking actions, such as completed passes made, passes received and crosses. The investigators calculated the probability that each player would interact with any team member and used network analysis to depict the number of interactions. This work concluded that network analysis could help identify characteristics in various team strategic plans and quantify individual contributions and team interactions by studying the attacking phase actions (Caicedo-Parada *et al.*, 2020; Gama *et al.*, 2014).

One year later, part of the research group, Gama *et al.* (2015), corroborated what authors such as Yamamoto & Yokoyama (2011) and Passos *et al.* (2011) had suggested, namely that small-world networks can capture the interactions among players in a football match. In their work, they observed 30 matches of the Portuguese Premier League (season 2010/2011), analysing the same collective attacking actions as in their previous study. Based on the sectors, goalkeepers, defenders, midfielders, and forwards were the four groups into which the players were divided. According to the outcomes, defenders and midfielders interacted with their teammates to the greatest degree. Besides, it was possible to state that the key players (those who interact more) were essential for the team's process of self-organisation. The researchers concluded that network analysis could offer insights into how organizing individuals can collaborate and plan team strategies.

[9] *Tiki-taka* is a style of playing football made famous by FC Barcelona and the Spanish national team, in which a team makes a lot of short passes keeping the possession of the ball. Adapted from: Cambridge Dictionary. Retrieved from: https://dictionary.cambridge.org/pt/dicionario/ingles/tiki-taka

In 2015 and 2016, Clemente, with other colleagues, developed several works in the field. First, Clemente *et al.* (2015) examined the national team networks that competed in the FIFA World Cup 2014. Using a dataset of 37,864 passes between teammates in 64 matches of 32 different teams, the investigators studied the relationship between the characteristics of the network formed on passes among teammates and the variables of overall team performance. On the one hand, they considered the density, the centrality, and the clustering coefficient as network graph performance variables. On the other hand, they considered the maximum stage in the competition, the match result, the goals, shots, and shots on goal as team performance variables. Their most pertinent findings demonstrated a relationship between high levels of total links, network density, clustering coefficient, and high levels of goals scored. Accordingly, they evidenced that successful teams were associated with higher values of these network performance variables (Clemente *et al.*, 2015).

Second, the following year, part of the research group developed a software named Performance Analysis Tool (PATO) that enabled users to quickly identify teammate interactions and extract network data for posterior analysis. This software computed not only the total links and the density of a graph constituted by the eleven players but also various centrality metrics, such as the in-degree centrality, the out-degree centrality, and the betweenness centrality. To test the software, Clemente, Silva, *et al.* (2016) chose seven games from the FIFA World Cup 2014 involving the German national team. Using the software features, they concluded that during the attacking phase, when in organised attack moment, midfielders, followed by central defenders, were the key players, having higher values of in-degree and out-degree centralities. These findings followed the previous conclusions of the research works in the field. Moreover, the graph properties displayed high values of density and total links, demonstrating the strong ability of the team of Germany to create passes and incorporate all of the players in the attacking phase (Clemente, Silva *et al.*, 2016).

Third, Clemente, Martins, *et al.* (2016) examined the plays that resulted in goals scored and conceded by a particular team throughout an entire season in the Portuguese Premier League using network methods. Two distinct analyses were carried out: players as nodes and playing field zones as nodes. On the one hand, knowing that the team under study always adopted the same system of play, they classified each player's position on the field for the teammate's analysis. On the other hand, they chose to divide the field into 18 regions (6 sectors and 3 corridors) for the zone analysis. They treated the passes as edges in both approaches. Hence, considering the clustering coefficient and centrality measures, the findings revealed that most players who participated in the plays that led to goals were forwards in the forward regions, particularly in the penalty area. The team of researchers also discovered that most of the attacking plays that resulted in goals were started by the full-backs or midfielders, bearing in mind that the attack began at the moment that a given team recovered the ball and continued until a goal was scored (Clemente, Martins, *et al.*, 2016).

Finally, Clemente, José, *et al.* (2016) considered ten matches from the Spanish Premier League and ten matches from the English Premier League, aiming to study the variance of different competitive leagues, score status and tactical in several centrality measures. They discovered that different competitive leagues and scores did not statistically influence the centrality levels. Nevertheless, distinct centrality levels were observed in the various positions. The highest levels of in-degree and out-degree centrality were found among midfielders. The external defenders had higher values of in-degree centrality than the central defenders, but the central defenders had higher values of out-degree centrality. Additionally, the goalkeeper and the forwards had the lowest centrality level values. (Clemente, José, *et al.*, 2016).

Gama, Dias, Couceiro, Belli, *et al.* (2016) intended to study the network of contacts resulting from the collective behaviour of professional football teams. The two top teams in the 2010–2011 Portuguese Premier League were their research subjects. Their findings from an analysis of 999 attacking actions, including passes made, passes received, and crosses, highlighted the importance of passing to key players to maintain possession of the ball. (Caicedo-Parada *et al.*, 2020; Gama, Dias, Couceiro, Belli, *et al.*, 2016). This study was complemented by Gama, Dias, Couceiro, Sousa, *et al.* (2016). They used a sample of 30 matches from a single team in the Portuguese Premier League (season 2010–2011) and took into account the degree, the clustering coefficient, and a weighted function of these two metrics in their network-based approach. Thus, they confirmed, as in the earlier studies (Gama, Dias, Couceiro, Belli, *et al.* (2016) and Gama *et al.* (2015)), that teams prioritize maintaining ball possession by working with key players, as these are essential for the team's self-organisation processes. Also, they emphasized that the key players are involved in the majority of productive interactions (Gama, Dias, Couceiro, Sousa, *et al.*, 2016).

Gonçalves *et al.* (2017) evidenced how network analysis allowed the description of significant aspects of collective performance, leading to a more comprehensive understanding of team sports performance. Focusing on youth football, the authors characterize the passing network by computing the closeness and betweenness centrality. Consequently, the results indicated that less dependence on passing for a given player (lower betweenness centrality values) and greater passing relationships (high values density and closeness centrality) could improve performance and lead to better outcomes (Caicedo-Parada *et al.*, 2020; Gonçalves *et al.*, 2017). These conclusions were consistent with the work of Grund (2012).

At the same time, Pina *et al.* (2017) explored whether network density, clustering coefficient, and centralisation can predict the outcome of attacking plays. Analysing 12 matches of the group stage UEFA Champions League (season 2015/2016), the researchers, using a hierarchical logistic regression model, considered the three metrics to predict the success of the attacking plays. An offensive play was considered successful if it resulted in a shot on goal or if the team kept possession of the ball until the final sector was considered successful. Thus, the investigators showed that density was the only significant predictor of the success of attacking plays. A lower density was linked to more offensive plays, but most of them were unsuccessful. In contrast, high density was associated with less overall play and fewer ball possession losses before the attacking team entered the final area of the field, increasing the probability that the offensive plays would succeed (Pina *et al.*, 2017).

Similarly to Clemente, Martins, *et al.* (2016), Mclean *et al.* (2017) looked at networks that resulted in goals scored. However, they examined 108 passing networks from the 2016 European Championship (UEFA EURO 2016) that resulted in goals scored, intending to identify the characteristics of these networks for the entire competition. As a result, they created these networks, which consisted of the players (nodes) and passes (edges) that connected them. Each network's pass sequence was recorded, including all passes into, out of, and within the four equally sized sectors (zones) into which the playing field was divided. They used a measure known as within-degree centrality, which was defined as the total number of passes made within the attacking play's zones that resulted in a goal, in addition to the in-degree and out-degree centralities, to determine the relative contributions of the playing field zones. In addition, the competition stage and the match status were considered in the analysis (Mclean *et al.*, 2017).

Indeed, considering the two last points, they discovered that the match status significantly impacted the network metrics. These significant differences, however, were not seen between successful and unsuccessful teams or between teams in the various group stages. Regarding the field of play zones analysis, they identified differences between the four sectors when considering the degree centrality metrics. The sector closest to the opposing goal had the higher values of the chosen metrics, as would be expected when analysing attacking plays that resulted in a goal (Mclean *et al.*, 2017).

The same authors, McLean *et al.* (2018), were the first to explore the influence of the systems of play on the interaction of players through passing in another study. With this objective, they examined the passing characteristics of playing positions within an Australian professional team throughout two consecutive seasons while adopting two different systems of play: 4-2-2-2 and 4-2-3-1. Network analysis was used to determine for each playing position the centrality measures, i.e., the in-degree centrality, out-degree centrality, closeness centrality, and betweenness centrality. Consequently, it was possible to compare these measures across systems of play (McLean *et al.*, 2018).

The results showed that while the change in the system of play had little impact on the overall passing contributions, the degree of the defensive midfielders and forwards considerably changed. The defensive midfield positions had a substantially higher betweenness centrality in a 4-3-2-1 compared to the 4-4-2-2. In addition, the forward positions had a significantly higher out-degree centrality when the team played with two forwards (4-2-2-2). So, it was possible to conclude that the team's coach should switch from the 4-2-3-1 playing formation to the 4-2-2-2 if they wanted the forwards to increase passing involvement. This was one significant contribution of this work for the coaching staff.

Arriaza-Ardiles *et al.* (2018) modelled the passing networks of a single team in 32 official Spanish Premier League matches to prove that network analysis is a useful tool for the coaching staff, allowing them to characterize the play structure of a team. They used the clustering coefficient and the centrality measures (closeness and betweenness) to describe the players' contributions to the team. Additionally, they divided the field of play into 24 zones (6 sectors and 4 corridors). They recorded the number of events (passes made and received) in each zone, representing the results in a density map. Therefore, they highlighted that by capturing the game using the theory of complex systems, it was possible to analyse a player's role while also comprehending the performance of a team as a whole (Arriaza-Ardiles *et al.*, 2018; Caicedo-Parada *et al.*, 2020).

Mendes *et al.* (2018) studied the variation in general network properties at different competitive levels and periods of the season. They analysed 132 full official matches from various teams in distinct age groups (under-15, under-17, under-19 and senior) by building passing networks and computing the total links, the network density, and the in-degree, out-degree, and betweenness centrality. This study's primary outcome was a moderate-to-strong correlation between network characteristics and performance variables, namely the final score and the goals conceded. Indeed, on the one hand, the network density was positively correlated with the final score and, conversely, negatively correlated with the goals conceded. On the other hand, the elite teams (senior and under-19) had higher total links and network density (Caicedo-Parada *et al.*, 2020; Mendes *et al.*, 2018).

More recently, Buldú *et al.* (2019) used various network metrics to identify the characteristics of the FC Barcelona team coached by Pep Guardiola during the 2009/2010 season. The investigators began by evaluating various network metrics and contrasting the Barcelona team's network with its rivals in the Spanish Premier League. Next, they focused on the temporal nature of football passing networks, looking at how all network properties changed throughout the game rather than just studying the average passing networks. Creating networks with 50 consecutive passes could account for the game's temporal evolution. Thus, this study showed how different each team was, highlighting how Guardiola's FC Barcelona stood out from the competition (Buldú *et al.*, 2019). Moreover, the findings revealed that increasing the number of passes improved the passing networks' characteristics (Caicedo-Parada *et al.*, 2020).

This study was extended by Herrera-Diestra *et al.* (2020) by building the corresponding zone networks, in which the nodes of the networks are zones of the field of play. They compared FC Barcelona's network properties to their opponents' networks. They discovered significant differences in the clustering coefficient, network average shortest path, and the number of nodes occupied by a team for partitions with a large number of subdivisions of the playing field (Herrera-Diestra *et al.*, 2020).

At the same time, Korte *et al.* (2019) opted to apply a play-by-play network analysis. According to the researchers, this type of analysis was chosen to represent the actual interplay. Considering a sample of 70 matches between 35 professional football teams from Germany, they categorize a possession as successful when a team enters the final sector and unsuccessful otherwise. Also, in addition to calculating the general network metrics, they introduced a metric denoted as “flow betweenness” that measured the fraction of plays in which a player functions as an intermediate player, that is, a “player who acts as a bridge in terms of passing between any two other players” (Korte *et al.*, 2019). According to the findings, midfielders were the primary intermediaries in successful plays, while central defenders were the primary intermediaries in unsuccessful plays (Caicedo-Parada *et al.*, 2020).

Diquigiovanni & Scarpa (2019) developed a hierarchical clustering method to divide a sample of undirected weighted networks into clusters, thus, detecting different play styles. In their article, they represented the networks of the Italian Premier League teams in all matches of the season 2015/2016. In these networks, the nodes were different zones of the playing field, and the edges were the ball's movements between these areas. In addition to dividing the field into nine evenly spaced zones (3 sectors and 3 corridors), they also chose to characterize only certain degrees of connections by using the normalized weights of the edges with the selection of a threshold. If a high threshold were selected, for instance, only the communities with strong connections would be described without distinguishing between weights that were less than or equal to the threshold (Diquigiovanni & Scarpa, 2019). Their method detected six major categories identifying the main playing styles, which could still be subdivided into 15 different playing styles.

Bekkers & Dabadghao (2019) deepened the work of Gyarmati *et al.* (2014) and Peña & Navarro (2015) by analysing different motifs at the team and individual levels. Hence, they applied the network motif concept to study patterns of 155 different teams and 3532 different players in 6 top European leagues throughout four consecutive seasons (2012-2015). Along with expanding on the motif concept, they also developed an expected goals model to evaluate the efficacy of playing styles and a novel way to visualise motif data (radar charts) that made it possible to compare teams and individuals. In describing the relationships between position and playing style, they demonstrated how this analysis could aid player scouting (Bekkers & Dabadghao, 2019).

Clemente *et al.* (2020) continued their previous works by studying network centrality measures between playing positions during pass sequences and their relationships to match outcomes. Indeed, the in-degree and out-degree centralities were sensitive to changes in playing positions after researchers studied the national teams' matches at the FIFA World Cup 2018. Additionally, this study showed that midfielders, wingers, and central forwards had possibly smaller increases in degree centrality levels during won matches compared to lost matches.

3.3. Chapter considerations

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 (Buldú *et al.*, 2018). Indeed, several studies have focused on football analysis in the last decade, specifically on how players interact with each other by passing the ball (Buldú *et al.*, 2019). 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 (Buldú *et al.*, 2018, 2019):

1. **Player/playing position passing networks**, where nodes are a team's players/playing positions (Buldú *et al.*, 2018; Gama *et al.*, 2015; Grund, 2012; Passos *et al.*, 2011). 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 (Buldú *et al.*, 2018; Diquigiovanni & Scarpa, 2019; Herrera-Diestra *et al.*, 2020; Malta & Travassos, 2014; Mclean *et al.*, 2017). 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 (Buldú *et al.*, 2018; Cotta *et al.*, 2013; Narizuka & Yamazaki, 2019). Only two studies using this type of analysis were found in the literature.

According to Buldú *et al.* (2018), 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 (Buldú *et al.*, 2018). Some works focus their study on individual players (Duch *et al.*, 2010; Peña & Touchette, 2012), while others concentrate their attention on the characteristics of the playing positions (Clemente, José *et al.*, 2016; Gama *et al.*, 2015; Malta & Travassos, 2014). 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 (Buldú *et al.*, 2018). The analysis of motifs has revealed that most teams tend to apply a homogeneous style (Gyarmati *et al.*, 2014). Also, it demonstrated how it is possible to identify the key players in the network (Peña & Navarro, 2015), thus assisting in the scouting process (Bekkers & Dabadghao, 2019).
3. **Macroscale**, where the network is studied as a whole (Buldú *et al.*, 2018). Studies have suggested that high-density and decentralised passing networks are associated with higher performance (Clemente *et al.*, 2015; Gonçalves *et al.*, 2017; Grund, 2012).

The research has shown that the interaction between players during a football game supports a scale-free network (Gama *et al.*, 2015; Yamamoto & Yokoyama, 2011). Furthermore, 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 centralization of the network (Buldú *et al.*, 2018; Clemente *et al.*, 2015). Another technique was to build sliding windows with a specific length (between 5 to 15 minutes) (Buldú *et al.*, 2018; Cotta *et al.*, 2013; Yamamoto & Yokoyama, 2011). Finally, the influence of the system of play was only found once in the literature, using only one team in two different seasons (McLean *et al.*, 2018).

Regarding the analysis of passing sequences, the literature only explored the passing sequences that led to a goal (Reep & Benjamin, 1968). Additionally, they related them with the general strategy of play (direct play or possession play) and with the number of shots and conversion ratio of shots into goals (M. Hughes & Franks, 2005).

As a result, this dissertation intends to extend the work done in the passing sequence analysis to all the attacking plays, verifying if the passing distribution tends to follow the power law distribution and demonstrating how the passing distribution can translate the teams' general strategy of play. Moreover, another objective is to study the relationship between the team's overall performance variables, the general strategy of play, and the network's characteristics. Additionally, the methodology of Hughes & Franks (2005) is reproduced to verify if their outcomes are still observed. On the other hand, this dissertation aims to study the influence of the systems of play on the network's characteristics by analysing the football player/playing position-zone passing networks from a spatiotemporal perspective while considering the systems of play. The teams under study were the national teams that competed in UEFA EURO 2020 since it was the most recent professional football tournament for the men's national team, and no studies have used this sample to address these themes.

Chapter 4 – Passing sequences analysis

This chapter presents the methodology (section 4.1), results (section 4.2) and the discussion (section 4.3) regarding the passing sequences analysis.

4.1. Methodology

According to Pollard & Reep (1997), “a team possession starts when a player gains possession of the ball by any means other than from a player of the same team. The player must have enough control over the ball to be able to have a deliberate influence on its subsequent direction. The team possession may continue with a series of passes between players of the same team but ends immediately when one of the following events occurs:

a. the ball goes out of play;
b. the ball touches a player of the opposing team (e.g. by means of a tackle, an intercepted pass or a shot being saved). A momentary touch that does not significantly change the direction of the ball is excluded;
c. an infringement of the rules takes place (e.g. a player is offside or a foul is committed).”

Therefore, 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 (Hughes & Franks, 2005).

In this way, two distinct analyses were conducted. First, the passing sequences were examined to check if the distribution of passes per possession tends to follow the power law distribution. Second, the distribution of passes was considered to study the general strategy of play (possession play or direct play) of the national teams that participated in the tournament. In addition, it was also utilised to investigate the relationship between the team's overall performance variables (match result, maximum stage reached in the tournament) and the general strategy of play. Next, to determine whether the article’s conclusions by Hughes & Franks (2005) are still observed, the study was limited to those possessions that led to a shot that resulted in a goal scored. For these analyses, each match's eventing data (*StatsBomb Events Data*) was used, centring the attention on the data related to the passes performed during each team's possession.

4.1.1. Passing sequences

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 to confirm whether the power law distribution was an appropriate model for the distribution of passes per possession. The passes made during the extra time were excluded from the study to allow comparisons between all the matches. Additionally, there was no distinction between different game moments. Hence, the study included passing sequences executed during the organised attack and defence-attack transition, as well as the ones performed during set pieces.

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 (Clauset *et al.*, 2009; Goldstein *et al.*, 2004b).

There are several methods for fitting power law distributions. Many researchers make parameter estimations using linear regression. Different variations using the linear fit to the data plotted in a $log - log$ scale were suggested. First, was proposed a direct linear fit of the $log - log$ plot of the whole raw histogram of the data. However, this technique does not consider that the majority of data is collected at the first few points of the distribution, fitting all points with the same weight (Goldstein *et al.*, 2004a, 2004b). Therefore, other researchers only used the first 5 points of $log - log$ plot for the linear regression. Likewise, a linear fitting to logarithmically binned histograms was introduced. This method applies linear regression to bins with equal logarithmic sizes. With this, the tail's noise is reduced by grouping the data points into bins, so the noise reduction is determined by the bins' size (Goldstein *et al.*, 2004a). In summary, despite their easiness, due to the nonlinear nature of the data, these graphical methods tend to be biased and inaccurate (Goldstein *et al.*, 2004b).

In opposition, the maximum likelihood estimation (MLE) is a more robust method for fitting the power-law distribution. It is based on finding the maximum value of the likelihood function:

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

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

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 $\partial\mathcal{L}/\partial\alpha = 0$:

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

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 (Clauset *et al.*, 2009; M. L. Goldstein *et al.*, 2004b).

Additionally, a test is 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 $x \geq x_{min}$ (Clauset *et al.*, 2009; M. L. Goldstein *et al.*, 2004b).

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 minimizes 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 (Alstott *et al.*, 2014).

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 (Clauset *et al.*, 2009). Hence, the methodology described by Clauset *et al.* (2009) 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 hypothesized 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 (Clauset *et al.*, 2009).

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 $i = 1, \dots, n$, 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 $x < x_{min}$ (Clauset *et al.*, 2009).

Knowing that, for the p-value to be accurate to within about ϵ of the true value, should be created at least $\frac{1}{4}\epsilon^{-2}$ 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 (Clauset *et al.*, 2009).

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.

Indeed, the power law exponent $-\alpha$ is the (negative) slope of the straight line in the logarithmic plot, describing the general attacking strategy of play utilised by a team: possession play or direct play. 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 (Kempe *et al.*, 2014; Tenga *et al.*, 2010). As a result, teams with a lower value of α are teams that apply a possession-based style of play, while teams with a higher value of α are teams that favour a direct type of play. Subsequently, it was possible to explore the inherent characteristics of each strategy of play, extending, in this way, the literature's works.

Consequently, different objectives were defined considering the parameter α as well as the number of passes, the number of passes completed and the percentage of passes completed (from now on, denoted as pass statistics). First, 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, the relationship between the parameter α and the pass statistics 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) (Clemente *et al.*, 2015). Third, this study sought to relate the strategy of play with each team's overall performance. Similar to the research elaborated by Clemente *et al.* (2015), the final result of the match was considered a performance variable, i.e. (i) defeat, (ii) draw or (iii) victory. Additionally, a second overall performance of a team was determined by the stage that a team reached in the UEFA EURO 2020, 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 provide answers to the following questions:

1. Are there any differences in the strategy of play, described by α, and pass statistics between teams that achieved different match results?
2. Are there any differences in the strategy of play, described by α, and pass statistics 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 made to any distribution that was not normal. On the other hand, Levene's test was used to investigate the homogeneity assumption. 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 (Clemente *et al.*, 2015). 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 of Cohen (1988) were used: $0.01 =$ small effect; $0.06 =$ moderate effect and $0.14 =$ large effect.

4.1.2. Passing sequences that resulted in a goal scored

To determine whether the article's conclusions by M. Hughes & Franks (2005) are still observed, their research work's methodology was implemented. Initially, it was confirmed if the statement of Reep & Benjamin (1968), supported by M. Hughes & Franks (2005), that approximately 80% of the goals result from a sequence of three or fewer passes was verified or not.

Then, as M. Hughes & Franks (2005) 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 normalized by dividing the number of goals scored in each team's possession by the sequence length and presented as goals per 1000 possessions for each possession length to avoid very small ratios. On the other hand, 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 (1988) were used to interpret the strength of the eta-squared values.

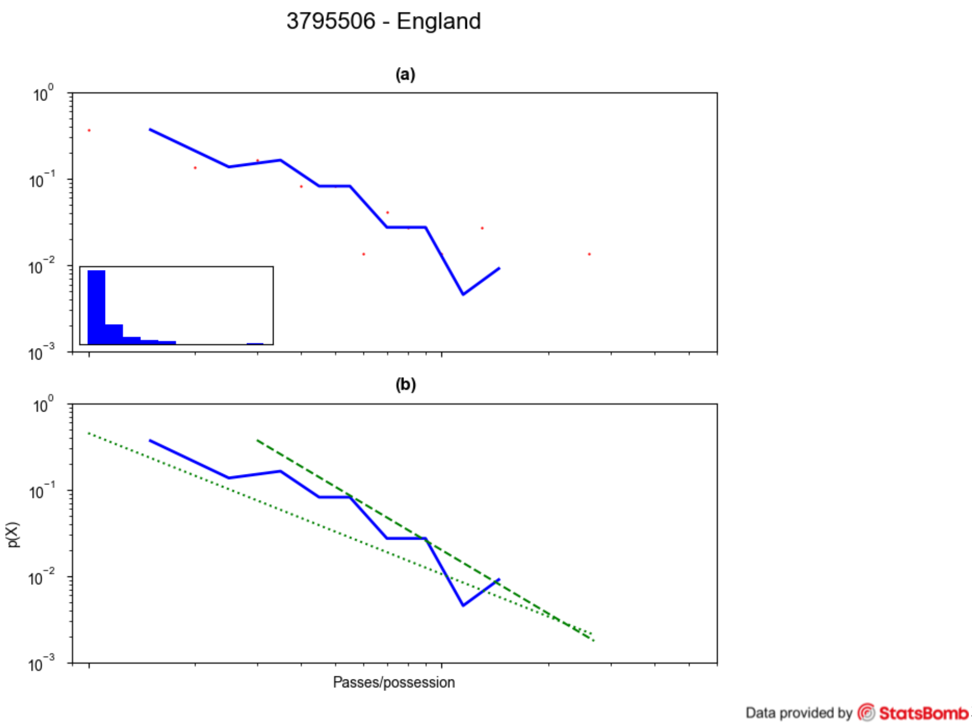
4.2. Results

This section presents the results of the analyses of the passing sequences and the passing sequences that resulted in a goal scored.

4.2.1. Passing sequences

The distribution of passes per possession, presented in Appendix B, was first examined to see if it tends to follow a power law distribution. Therefore, the power-law hypothesis was tested for each team in each match. The results, displayed in Appendix C, 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 length of the pass sequences increased distributions. 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, as illustrated in Figures 8 and 9.



[10]

Figure 8: Power law fitting of England's pass data in the regular time of the tournament's final against Italy (match id =3795506). Data visualisation with probability density functions. (a) On a log-log axis, fit using logarithmically spaced bins (blue line) of the data (red points). (b) Dotted green line: power law fit starting at $x_{min} = 1$. Dashed green line: power law fit starting from the optimal x_{min}.

[10] Note that, according to the StatsBomb Public Data User Agreement, is required to accredit any publication of analysis formed from StatsBomb Data with the StatsBomb brand logo. So, it is informed that all the subsequent work was performed using StatsBomb publicly and freely available data. Retrieved from: https://github.com/statsbomb/open-data/blob/master/LICENSE.pdf

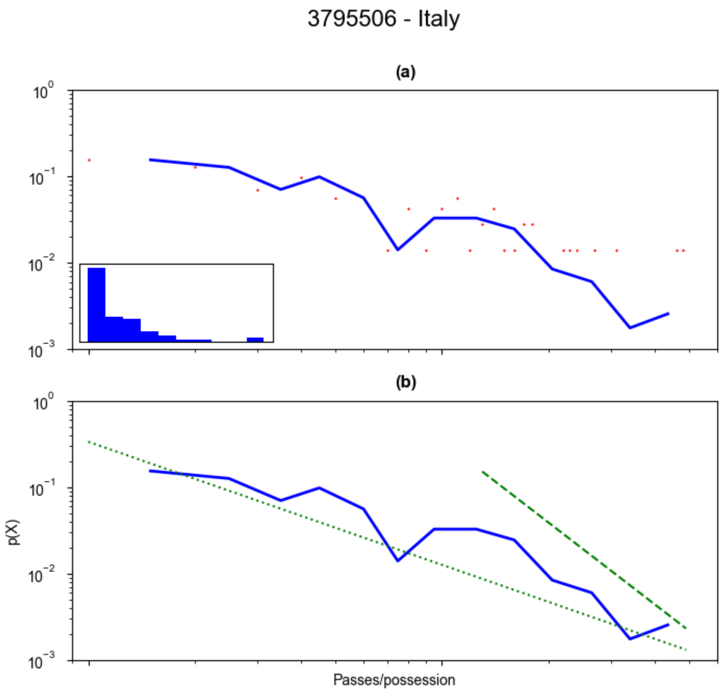


Figure 9: Power law fitting of Italy's pass data in the regular time of the tournament's final against Italy (match id =3795506). Data visualisation with probability density functions. (a) On a log-log axis, fit using logarithmically spaced bins (blue line) of the data (red points). (b) Dotted green line: power law fit starting at $x_{min} = 1$. Dashed green line: power law fit starting from the optimal x_{min}.

After computing α for all teams in all matches, the distributions of the pass statistics and the parameter α were studied using descriptive statistics. Table 1 shows, firstly, that the mean values (and the standard deviation) of the number of passes and number of passes completed are, respectively, 512.460 ($\pm$139.262) and 428.040 ($\pm$140.647). The percentage of passes completed had a mean (and a standard deviation) equal to 0.820 ($\pm$0.069). Furthermore, the parameter α had a mean (and a standard deviation) of 1.612 ($\pm$0.123). Additionally, it is essential to highlight the range of the number of passes (773) and the number of passes completed (747). In particular, the minimum and maximum values of the number of passes completed were 94 and 841. Lastly, the descriptive statistics also provided some information concerning the distribution of the variables. The number of passes and the number of passes completed both had skewness values close to 0, representing the symmetry of the distribution. In contrast, the number of passes completed and α had negative and positive skewness values, indicating that the values clustered to the right and left-hand sides of the distribution, respectively. Furthermore, Kurtosis, which provides information about the 'peakedness' of the distribution, revealed a high positive value for the percentage of passes completed, indicating that the distribution is peaked, with long thin tails (Pallant, 2005).

Table 1: Descriptive table of the pass statistics and the parameter α

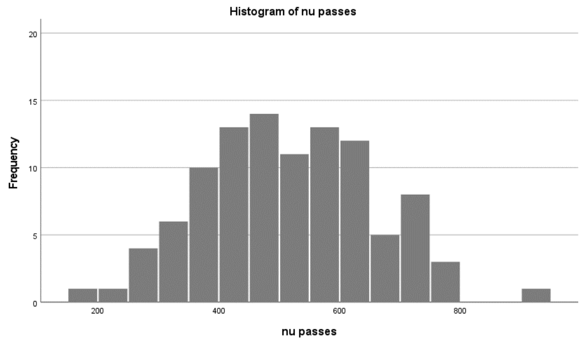
Descriptive Statistics

	Mean	Std. Error (Mean)	95% Confidence Interval for Mean Lower Bound	Upper Bound	5% Trimmed Mean	Median	Variance	Std Deviation	Minimum	Maximum	Range	Interquartil Range	Skewness	Std. Error (Skewness)	Kurtosis	Std. Error (Kurtosis)
Nu passes	512.460	13.789	485.110	539.810	511.600	505.000	19393.974	139.262	173.000	946.000	773.000	188.000	0.185	0.239	0.015	0.474
Nu passes completed	428.040	13.926	400.410	455.660	427.700	407.000	19781.662	140.647	94.000	841.000	747.000	200.000	0.149	0.239	-0.204	0.474
% passes completed	0.820	0.007	0.807	0.834	0.827	0.830	0.005	0.069	0.540	0.920	0.380	0.080	-1.485	0.239	3.311	0.474
α	1.612	0.012	1.588	1.636	1.603	1.601	0.015	0.123	1.387	2.025	0.638	0.162	1.028	0.239	1.915	1.381

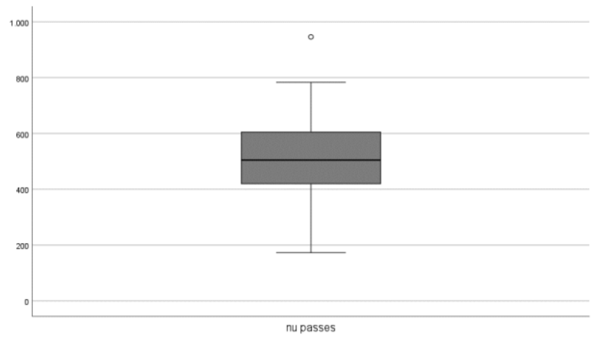
The descriptive study was then complemented with the inspection of the histograms (shape of the distribution) and the box plots, which simultaneously display several features of the data and allow the identification of outliers (Montgomery & Runger, 2003). According to the histograms in Figures 10 (a1) and 10 (b1), the distribution of the number of passes and the number of passes completed appeared to follow a normal distribution. In contrast, the visual inspection of the histograms of the percentage of passes completed and of α along with the skewness and kurtosis values seemed not to reveal the same. Furthermore, one outlier was visible in the boxplots for the number of passes and completed passes. This outlier referred to Spain's match against Sweeden during the group stage, in which the Spanish team performed 946 passes, of which 841 were successful.

By analysing the boxplot regarding the percentage of passes completed (Figure 10 (c2)), four outliers were identified, one very similar to the minimum value of the box plot. The two lower values belonged to Sweeden. In contrast to Spain, the Swedish team only completed 94 of the 174 passes it attempted against Spain, resulting in a percentage of passes completed equal to 54%. In the match against Poland, the Swedish team completed 59% (163/278) of the passes. On the other hand, Poland's match versus Spain was the other outlier in the match against Spain. In this match, the Polish team completed 60% (176/289) of the passes, while the Spanish team performed 754 passes, of which 658 were successful.

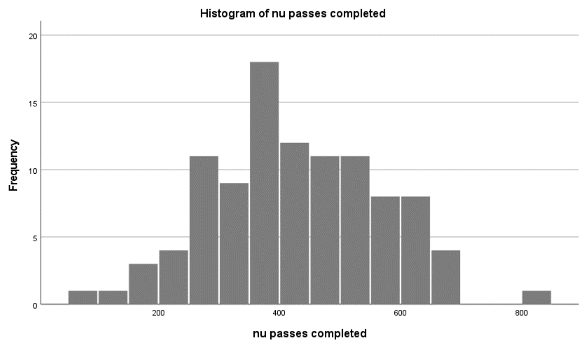
Additionally, the outcomes of the box plot of α (Figure 10 (d2)) agreed with the outcomes of the box plot of the total number of passes completed. In the match versus Spain, Sweden had the higher outlier (2.025). Then, Hungary had an α of 1.977 against Germany. Following that, Poland, in their match versus Spain, had an alpha equal to 1.953, and, finally, against Poland, Sweden had an alpha of 1.922.



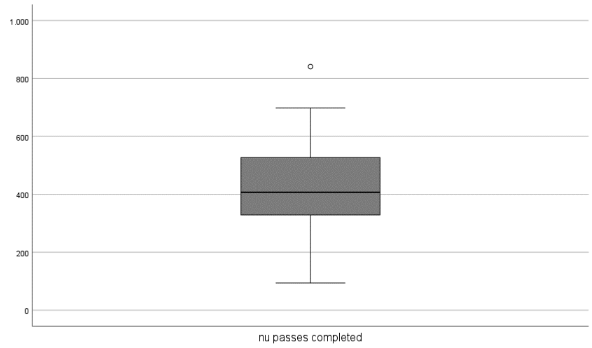
(a1)



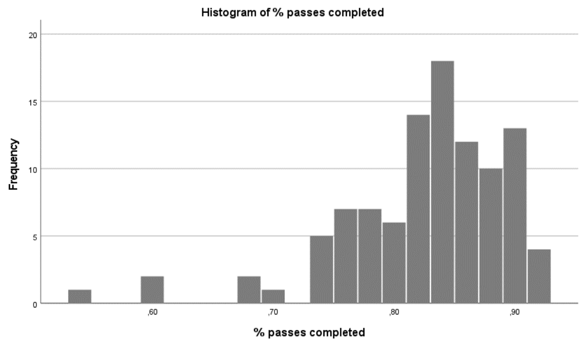
(a2)

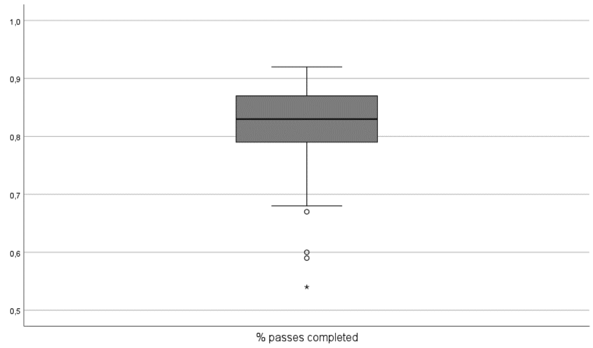


(b1)

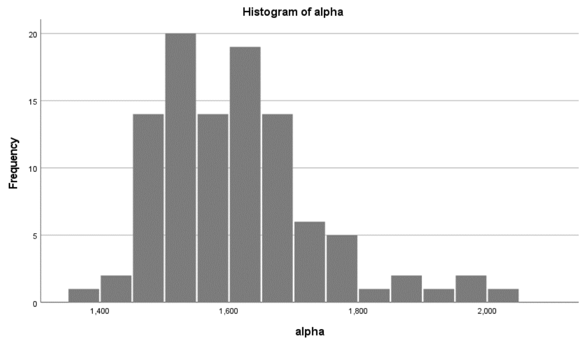


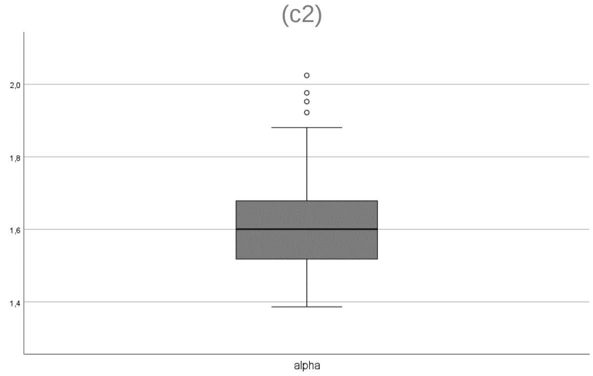
(b2)





(c1) (c2)





(d1) (d2)

Figure 10: (1) Histograms for the (a) number of passes, (b) the number of passes completed, (c) the percentage of passes completed, and (d) parameter α. (2) Box plots for the (a) number of passes, (b) the number of passes completed, (c) the percentage of passes completed, and (d) parameter α.

The Pearson product-moment correlation and ANOVA assume that the data follow a normal distribution. The normality could be somewhat evaluated with descriptive statistics, specifically the skewness and kurtosis values (Pallant, 2005). However, the normality was assessed using the Kolmogorov-Smirnov statistic, a more reliable procedure. Table 2 shows the results of this test statistics that revealed that the distribution of the number of passes, number of passes completed and α had a non-significant result (Sig.>0.05), indicating that these data were normally distributed. In opposition, in the case of the percentage of passes completed, the Sig. was less than 0.001, implying a violation of the assumption of normality.

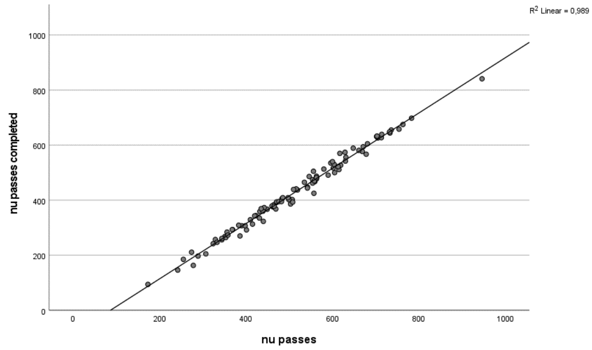
Table 2: Test of Normality for the number of passes, the number of passes completed, the percentage of passes completed, and parameter α.

Test of Normality

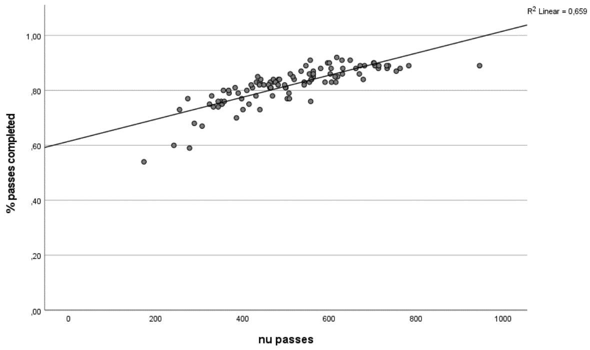
	Kolmogorov-Smirnov		
	Statistic	**df**	**Sig.**
nu passes	,045	102.000	0.200
nu passes completed	,073	102.000	0.200
% passes completed	,137	102.000	<0.001
α	,083	102.000	0.077

However, the Central Limit Theorem states in its most basic formulation that the sum of n independently distributed random variables will tend to be normally distributed as n becomes larger and $n \geq 30$, the normal approximation is satisfactory regardless of the shape of the population (Montgomery & Runger, 2003). Consequently, although the distribution of the percentage of passes completed was not normal, since $n = 102$ and considering the Central Limit Theorem, the assumption of normality was assumed (Clemente *et al.*, 2015).

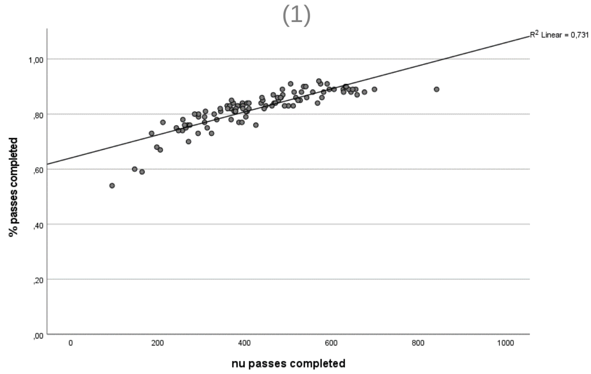
Next, the Pearson product-moment correlation's assumptions of linearity (the relationship between two variables is linear) and homoscedasticity (the variability of both variables is similar to all values) were analysed to see if there was any violation. The linearity was assessed by generating scatterplots between each pair of variables. Figure 11 shows that only the percentage of passes completed did not have a linear relationship with the other variables since a straight-line relationship between them was not present. In addition, the homoscedasticity assumption was not violated, as seen in Appendix D. Consequently, the Pearson Product-Moment was used to investigate the relationships between the number of passes, the number of passes completed, and α. 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.



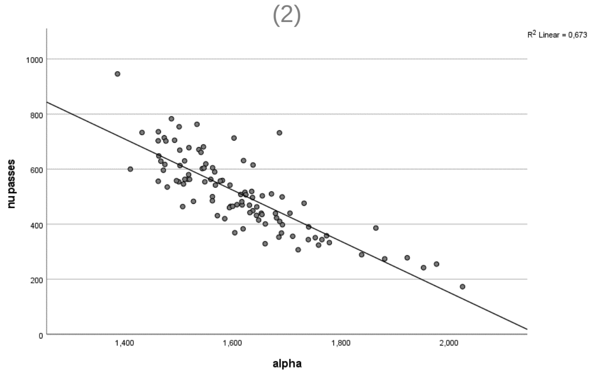
(1)



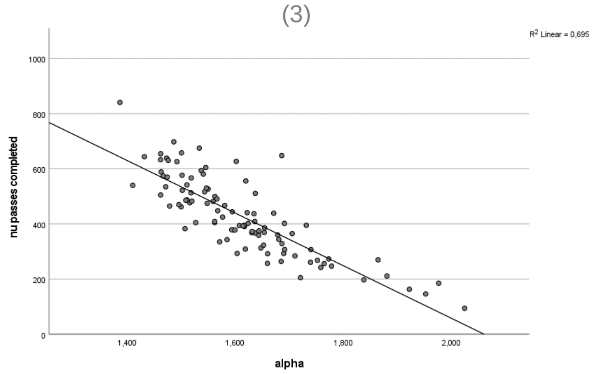
(2)



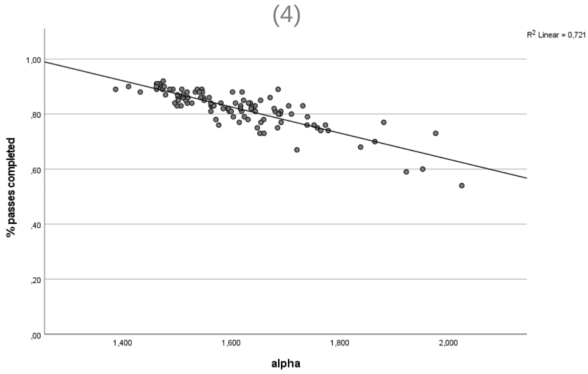
(3)



(4)



(5)



(6)

Figure 11: (1) Plot of the number of passes vs the number of passes completed; (2) Plot of the number of passes vs the percentage of passes completed; (3) Plot of the number completed vs the percentage of passes completed; (4) Plot of the parameter α vs the number of passes; (5) Plot of the parameter α vs the number of passes completed; (6) Plot of the parameter α vs the percentage of passes completed.

Table 3 reveals the Pearson r correlation coefficients between each pair of variables except for the percentage of passes completed. There was nearly a perfect positive correlation between the number of passes and the number of passes completed ($r = 0.994,\ n = 102,\ p < 0.01$), with high levels of the number of passes completed associated with high levels of the number of passes. The parameter α showed a very large negative correlation with the number of passes ($r = -0.820,\ n = 102,\ p < 0.01$) and the number of passes completed ($r = -0.834\ n = 102,\ p < 0.01$). Thus, lower levels of α were associated with high levels of the number of passes and the number of passes completed, suggesting that teams that adopt a possessive type of play tend to perform not only more passes but more successful passes.

Table 3: Pearson Product-Moment Correlation values between the number of passes, the number of passes completed, and parameter α.

Pearson Product-Moment Correlations

Measures	1	2
(1) Nu passes		
(2) Nu passes completed	0.994 **	
(3) α	-0.820 **	-0.834 **

N=102

** Correlation is significant at the 0.01 level

Table 4 shows the Spearman ρ correlation coefficients between the percentage of passes completed and the remaining variables. The percentage of passes completed revealed a nearly perfect positive correlation with the number of passes completed ($\rho = 0.909,\ n = 102,\ p < 0.01$), with high levels of the percentage of passes completed being associated with high levels of the number of passes completed. Additionally, this variable 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. This means that high levels of the percentage of passes completed were associated with high (low) levels of the number of passes completed (the parameter α).

Table 4: Spearman Rank's Order Correlation values between the percentage of passes completed and the number of passes, the number of passes completed, and parameter α, respectevely.

Spearman's Rank Order Correlations

Measures	1 (Nu passes)	2 (Nu passes completed)	3 (α)
% passes completed	0.866 **	0.909 **	-0.823 **

N=102

** Correlation is significant at the 0.01 level

Indeed, this methodology made it possible to describe the general playing strategy by studying the differences between national teams throughout the tournament. Table 5 shows the mean and standard deviation of the number of passes, number of passes completed, percentage of passes completed, and the parameter α. On the one hand, Spain was the team with the highest mean values for passes made (755.197), passes completed (669.000), and percentage of passes completed (89%), followed by Germany. On the other hand, the Spanish team had $\alpha = 1.509$, while the German team had $\alpha = 1.471$, switching places in terms of the national teams with the lowest mean value of α. In addition, Germany was the national team with the lowest standard deviation regarding the mean α, denoting its loyalty to the possessive strategy of play. Besides, Spain was the team with the lowest standard deviation concerning the percentage of passes completed, demonstrating its ability to sustain the ball while moving it. In opposition, Hungary was the team that exhibited not only the lowest values in the mean of the number of passes, the number of passes completed, and the percentage of passes completed but also the highest mean value of α ($\alpha = 1.779$). Furthermore, Sweeden had the lowest mean percentage of passes completed (70%) and the highest standard deviation of the mean α, followed by Poland and Hungary, suggesting that these national teams sometimes used a less direct strategy even though they had higher mean values of α.

Table 5: Descriptive statistics (mean and standard deviation) of the number of passes, the number of passes completed, and parameter α.

	Nu passes		**Nu passes completed**		**% passes completed**		**α**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Austria	537.750	48.808	439.750	44.977	82%	0.019	1.597	0.038
Belgium	606.400	132.221	524.000	131.924	86%	0.029	1.555	0.059
Croatia	493.750	130.811	415.250	129.962	84%	0.040	1.593	0.084
Czech Republic	441.800	47.620	335.800	47.851	76%	0.034	1.666	0.115
Denmark	502.000	103.454	410.333	94.580	81%	0.027	1.621	0.065
England	519.286	98.655	441.857	105.056	84%	0.048	1.590	0.111
Finland	396.667	79.827	309.333	92.034	77%	0.084	1.665	0.065
France	574.000	95.114	503.750	104.433	87%	0.046	1.536	0.079
Germany	667.500	107.600	586.750	103.219	88%	0.026	1.471	0.028
Hungary	316.000	57.420	240.000	54.028	76%	0.031	1.779	0.187
Italy	580.429	113.389	506.857	116.915	87%	0.050	1.544	0.107
Netherlands	566.000	118.830	470.250	133.440	82%	0.068	1.564	0.066
North Macedonia	410.000	51.098	331.000	54.617	80%	0.038	1.712	0.037
Poland	463.000	192.102	354.000	183.131	73%	0.118	1.698	0.220
Portugal	585.250	115.034	511.750	105.664	87%	0.030	1.551	0.079
Russia	435.000	165.638	333.333	161.029	75%	0.076	1.712	0.110
Scotland	413.000	79.373	319.667	61.695	78%	0.031	1.653	0.033
Slovakia	477.000	137.153	402.667	136.830	84%	0.040	1.651	0.036
Spain	755.167	99.012	669.000	88.916	89%	0.010	1.509	0.101
Sweden	382.000	195.433	290.500	197.499	70%	0.160	1.739	0.274
Switzerland	482.000	98.346	403.400	98.503	83%	0.050	1.618	0.108
Turkey	486.333	86.950	404.333	85.448	83%	0.036	1.664	0.080
Ukraine	519.200	61.141	441.000	70.601	85%	0.037	1.562	0.109
Wales	341.750	60.224	266.750	55.175	78%	0.033	1.722	0.129

Figure 12 shows the plots generated to explore the relationships between the mean α and the pass statistics. For each plot, the average value of each variable was computed, thus forming four quadrants that helped analyse the data. Figure 12.1. shows the mean of the parameter α versus its standard deviation. This plot reveals which teams were loyal to a general strategy of play and which ones did not. As mentioned before, Germany was the national team with the lowest mean value of α and the lowest standard deviation of α, indicating loyalty to the possessive strategy of play. In opposition, some teams opted to play a more direct type of play, such as North Macedonia, Slovakia, and Scotland, as described by the higher mean α and lower standard deviation values. On the other hand, the tournament's finalists, Italy and England, were below the average value of the mean value but above the average of the standard deviation.

Moreover, Figures 12.2. and 12.3. displays the mean α versus, respectively, the number of passes and the number of passes completed. Thus, Spain stands out from the simple linear regression that considers the mean α and, respectively, the number of passes and the number of passes completed. These plots demonstrate the capacity of Spain to exchange the ball and, consequently, to have more passes and more passes completed than its opponents.

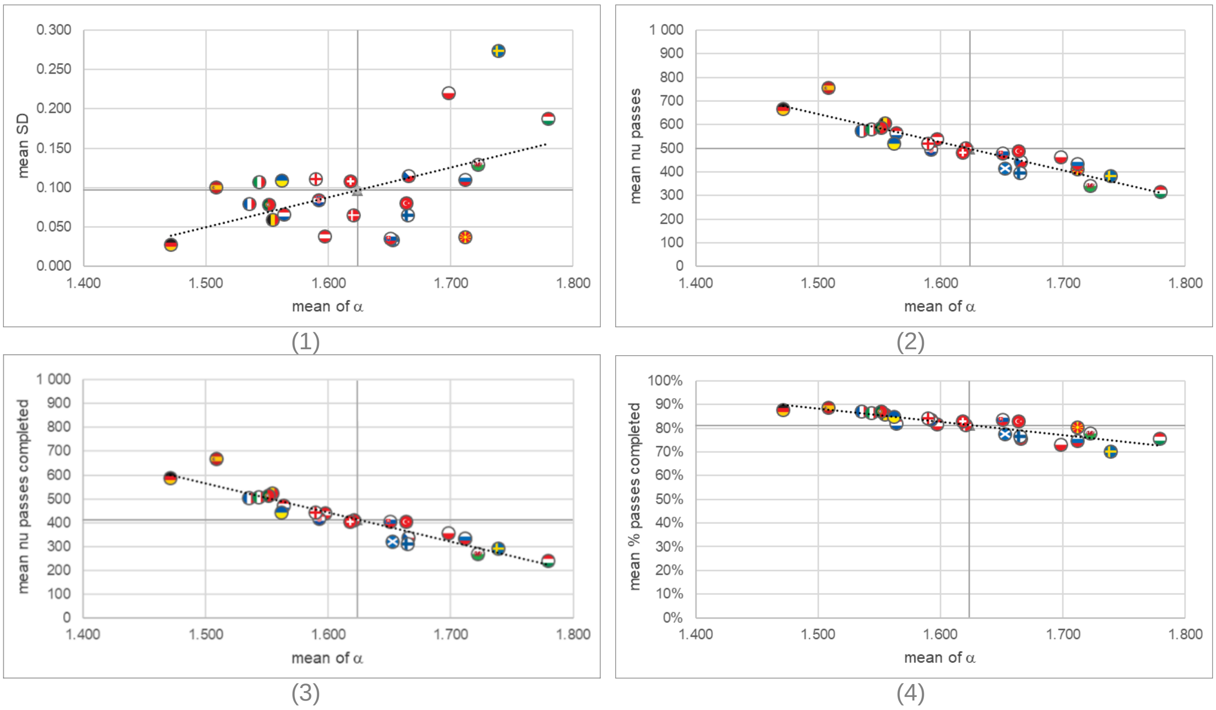


Figure 12: (1) Plot of the mean of parameter α vs the standard deviation of the parameter α; (2) Plot of the mean of parameter α vs mean of the number of passes; (3) Plot of the mean of parameter α vs mean of the number of passes completed; (4) Plot of the mean of parameter α vs mean of the percentage of passes completed.

Although it was possible to differentiate teams and to take several conclusions by examining Table 5 and Figure 12, the one-way ANOVA or the Welch and Brown-Forsythe tests were conducted to answer the two questions mentioned earlier. First, the differences in the parameter α and the pass statistics between teams that achieved different match results (defeat, draw or victory) were analysed. Thus, 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, as can be understood from Table 6.

Table 6: Descriptive table and statistical comparison between groups (match results), considering the pass statistics and parameter α.

Descriptive Statistics

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Nu passes	Defeat	35	481.110	106.813	18.055	444.420	517.810	274.000	733.000
	Draw	32	519.000	179.836	31.791	454.160	583.840	173.000	946.000
	Victory	35	537.830	122.477	20.702	495.760	579.900	278.000	763.000
	Total	**102**	**512.460**	**139.262**	**13.789**	**485.110**	**539.810**	**173.000**	**946.000**
Nu passes completed	Defeat	35	392.940	104.552	17.673	357.030	428.860	197.000	644.000
	Draw	32	439.750	178.550	31.563	375.380	504.120	94.000	841.000
	Victory	35	452.430	129.444	21.880	407.960	496.890	163.000	675.000
	Total	**102**	**428.040**	**140.647**	**13.926**	**400.410**	**455.660**	**94.000**	**841.000**
% passes completed	Defeat	35	0.809	0.049	0.008	0.793	0.826	0.680	0.900
	Draw	32	0.824	0.085	0.015	0.793	0.854	0.540	0.920
	Victory	35	0.828	0.070	0.012	0.804	0.852	0.590	0.910
	Total	**102**	**0.820**	**0.069**	**0.007**	**0.807**	**0.834**	**0.540**	**0.920**
α	Defeat	35	1.627	0.101	0.017	1.592	1.662	1.410	1.881
	Draw	32	1.615	0.156	0.028	1.559	1.671	1.387	2.025
	Victory	35	1.594	0.109	0.018	1.557	1.631	1.463	1.922
	Total	**102**	**1.612**	**0.123**	**0.012**	**1.588**	**1.636**	**1.387**	**2.025**

After generating the descriptive statistics, Levene's test tested the assumption of homogeneity of the variances. This assumption was not violated if the significance value, Sig., was greater than 0.05. However, assessing Levene's test in Table 7, it was found that the number of passes and the number of passes completed violated this assumption. For these cases, the Welsh and Brown-Forsythe were used instead of consulting the ANOVA because they are preferable when this assumption is violated (Pallant, 2005).

Table 7: Test of Homogeneity of variances between groups (match results), considering the pass statistics and parameter α.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Nu passes	Based on Mean	5.946	2.000	99.000	0.004
Nu passes completed	Based on Mean	6.677	2.000	99.000	0.002
% passes completed	Based on Mean	2.363	2.000	99.000	0.099
α	Based on Mean	2.940	2.000	99.000	0.057

Therefore, a one-way between-groups analysis of variance (Table 8) was conducted to explore the impact of the percentage of passes and the parameter α on the match result. There was not a statistically significant difference at the $p < 0.05$. In addition, the Welch and Brown-Forsythe tests (Table 9) were conducted to investigate the impact of the number of passes and the number of passes completed. As previously mentioned, the samples were divided into three groups, and there was not a statistically significant difference at the $p < 0.05$.

Table 8: One-way between-groups analysis of variance (match results), considering the pass statistics and parameter α.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Nu passes	Between Groups	58282.829	2.000	29141.414	1.518	0.224
	Within Groups	1900508.514	99.000	19197.056		
	Total	**1958791.343**	**101.000**			
Nu passes completed	Between Groups	68319.386	2.000	34159.693	1.753	0.179
	Within Groups	1929628.457	99.000	19491.197		
	Total	**1997947.843**	**101.000**			
% passes completed	Between Groups	0.006	2.000	0.003	0.675	0.511
	Within Groups	0.472	99.000	0.005		
	Total	**0.478**	**101.000**			
α	Between Groups	0.019	2.000	0.010	0.639	0.530
	Within Groups	1.508	99.000	0.015		
	Total	**1.527**	**101.000**			

Table 9: Welch and Brown-Forsythe tests (match results), considering the pass statistics and parameter α.

Robust Tests of Equality of Means

		Statistic	df1	df2	Sig.
Nu passes	Welch	2.183	2.000	62.234	0.121
	Brown-Forsythe	1.474	2.000	76.654	0.235
Nu passes completed	Welch	2.440	2.000	61.987	0.096
	Brown-Forsythe	1.705	2.000	78.226	0.188
% passes competed	Welch	0.923	2.000	61.117	0.403
	Brown-Forsythe	0.660	2.000	80.954	0.519
α	Welch	0.857	2.000	62.923	0.430
	Brown-Forsythe	0.622	2.000	80.483	0.539

Second, the analysis focused on the differences between teams that reached different stages of the tournament. As before, descriptive statistics, represented in Table 10, were initially produced, and then the assumption of homogeneity of the variances was again tested using Levene’s test with the same significance value. Table 11 demonstrates that this assumption was not violated, so the ANOVA test was executed.

Table 10: Descriptive table and statistical comparison between groups (stage reached in the tournament), considering the pass statistics and parameter α.

Descriptive Statistics

						95% Confidence Interval for Mean			
		N	**Mean**	**Std. Deviation**	**Std. Error**	**Lower Bound**	**Upper Bound**	**Minimum**	**Maximum**
Nu passes	Final	14	549.860	106.923	28.576	488.120	611.590	344.000	703.000
	Semi-finals	12	628.580	163.710	47.259	524.570	732.600	398.000	946.000
	Quarter-finals	20	512.350	104.276	23.317	463.550	561.150	333.000	763.000
	Round of 16	32	518.500	145.832	25.780	465.920	571.080	173.000	783.000
	Group Stage	24	424.630	110.683	22.593	377.890	471.360	242.000	631.000
	Total	**102**	**512.460**	**139.262**	**13.789**	**485.110**	**539.810**	**173.000**	**946.000**
Nu passes completed	Final	14	474.360	111.983	29.929	409.700	539.010	261.000	633.000
	Semi-finals	12	539.670	160.958	46.465	437.400	641.930	307.000	841.000
	Quarter-finals	20	426.050	110.001	24.597	374.570	477.530	247.000	675.000
	Round of 16	32	435.590	147.279	26.036	382.490	488.690	94.000	698.000
	Group Stage	24	336.790	107.606	21.965	291.350	382.230	146.000	556.000
	Total	**102**	**428.040**	**140.647**	**13.926**	**400.410**	**455.660**	**94.000**	**841.000**
% passes completed	Final	14	0.854	0.048	0.013	0.826	0.882	0.760	0.910
	Semi-finals	12	0.850	0.043	0.012	0.823	0.877	0.770	0.900
	Quarter-finals	20	0.823	0.054	0.012	0.798	0.848	0.700	0.900
	Round of 16	32	0.822	0.082	0.015	0.792	0.852	0.540	0.920
	Group Stage	24	0.781	0.065	0.013	0.753	0.808	0.600	0.880
	Total	**102**	**0.820**	**0.069**	**0.007**	**0.807**	**0.834**	**0.540**	**0.920**
α	Final	14	1.567	0.107	0.029	1.505	1.629	1.462	1.773
	Semi-finals	12	1.565	0.100	0.029	1.501	1.628	1.387	1.692
	Quarter-finals	20	1.600	0.103	0.023	1.552	1.648	1.410	1.864
	Round of 16	32	1.597	0.138	0.024	1.547	1.646	1.432	2.025
	Group Stage	24	1.692	0.106	0.022	1.647	1.737	1.566	1.977
	Total	**102**	**1.612**	**0.123**	**0.012**	**1.588**	**1.636**	**1.387**	**2.025**

Table 11: Test of Homogeneity of variances between groups (stage reached in the tournament), considering the pass statistics and parameter α.

Test of Homogeneity of Variances

		Levene Statistic	**df1**	**df2**	**Sig.**
Nu passes	Based on Mean	1.855	4.000	97.000	0.124
Nu passes completed	Based on Mean	1.744	4.000	97.000	0.147
% passes completed	Based on Mean	0.663	4.000	97.000	0.619
α	Based on Mean	0.451	4.000	97.000	0.772

Therefore, a one-way between-groups analysis of variance (Table 12) was conducted to explore the impact of the percentage of passes and the parameter α on the stage reached in the tournament. 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). There were statistically significant differences at the $p < 0.05$ between the different groups (stage reached in the tournament) in the 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) and the parameter α ($F_{4,97} = 4.048,\ p = 0.004, \eta^2 = 0.143,$ large effect).

Table 12: One-way between-groups analysis of variance (stage reached in the tournament), considering the pass statistics and parameter α.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Nu passes	Between Groups	367722.537	4.000	91930.634	5.605	<0.001
	Within Groups	1591068.806	97.000	16402.771		
	Total	**1958791.343**	**101.000**			
Nu passes completed	Between Groups	381295.335	4.000	95323.834	5.719	<0.001
	Within Groups	1616652.508	97.000	16666.521		
	Total	**1997947.843**	**101.000**			
% passes completed	Between Groups	0.064	4.000	0.016	3.770	0.007
	Within Groups	0.414	97.000	0.004		
	Total	**0.478**	**101.000**			
α	Between Groups	0.218	4.000	0.055	4.048	0.004
	Within Groups	1.309	97.000	0.013		
	Total	**1.527**	**101.000**			

As ANOVA detected significant statistical differences, the Tukey-Kramer modification of Tukey's HSD test was implemented as the sample sizes were unequal. First, regarding the number of passes, the post-hoc comparisons indicated that the mean number of passes for Group 5 (Group Stage) $[M = 424.630, SD = 110.683]$ was significantly different at the $p < 0.05$ from Group 2 (Semi-finals) $[M = 628.580, SD = 163.710]$ and from Group 1 (Final) $[M = 549.860, SD = 106.923]$, as discriminated in Table 13.

Table 13: Post-hoc test for the number of passes

Multiple Comparisons

Nu passes

Tukey HSD

(I) competition stage	(J) competition stage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Final	Quarter-finals	-78.726	50.384	0.525	-218.780	61.330
	Semi-finals	37.507	44.629	0.917	-86.550	161.570
	Round of 16	31.357	41.039	0.940	-82.720	145.440
	Group Stage	125.232 *	43.071	0.036	5.510	244.960
Semi-finals	Final	78.726	50.384	0.525	-61.330	218.780
	Quarter-finals	116.233	46.766	0.102	-13.760	246.230
	Round of 16	110.083	43.353	0.090	-10.430	230.590
	Group Stage	203.958 *	45.281	<.001	78.090	329.830
Quarter-finals	Final	-37.507	44.629	0.917	-161.570	86.550
	Semi-finals	-116.233	46.766	0.102	-246.230	13.760
	Round of 16	-6.150	36.507	1.000	-107.630	95.330
	Group Stage	87.725	38.776	0.166	-20.060	195.510
Round of 16	Final	-31.357	41.039	0.940	-145.440	82.720
	Semi-finals	-110.083	43.353	0.090	-230.590	10.430
	Quarter-finals	6.150	36.507	1.000	-95.330	107.630
	Group Stage	93.875	34.584	0.059	-2.260	190.010
Group Stage	Final	-125.232 *	43.071	0.036	-244.960	-5.510
	Semi-finals	-203.958 *	45.281	<.001	-329.830	-78.090
	Quarter-finals	-87.725	38.776	0.166	-195.510	20.060
	Round of 16	-93.875	34.584	0.059	-190.010	2.260

* The mean difference is significant at the 0.05 level.

Second, Table 14 displays the post-hoc comparisons for the number of passes completed. The test's result revealed that the mean number of passes completed for Group 5 (Group Stage) $[M = 336.790, SD = 107.606]$ was significantly different at the $p < 0.05$ from, firstly, Group 4 (Round of 16) $[M = 435.590, SD = 147.279]$, secondly from Group 2 (Semi-finals) $[M = 539.670, SD = 160.958]$, and, finally, from Group 1 (Final) $[M = 474.360, SD = 111.983]$.

Table 14: Post-hoc test for the number of passes completed

Multiple Comparisons

Nu passes completed

Tukey HSD

(I) competition stage	(J) competition stage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Final	Quarter-finals	-65.310	50.787	0.700	-206.490	75.870
	Semi-finals	48.307	44.987	0.820	-76.740	173.360
	Round of 16	38.763	41.368	0.882	-76.230	153.760
	Group Stage	137.565 *	43.415	0.017	16.880	258.250
Semi-finals	Final	65.310	50.787	0.700	-75.870	206.490
	Quarter-finals	113.617	47.140	0.121	-17.420	244.660
	Round of 16	104.073	43.700	0.129	-17.400	225.550
	Group Stage	202.875 *	45.643	<0.001	76.000	329.750
Quarter-finals	Final	-48.307	44.987	0.820	-173.360	76.740
	Semi-finals	-113.617	47.140	0.121	-244.660	17.420
	Round of 16	-9.544	36.799	0.999	-111.840	92.750
	Group Stage	89.258	39.087	0.159	-19.390	197.910
Round of 16	Final	-38.763	41.368	0.882	-153.760	76.230
	Semi-finals	-104.073	43.700	0.129	-225.550	17.400
	Quarter-finals	9.544	36.799	0.999	-92.750	111.840
	Group Stage	98.802 *	34.861	0.043	1.900	195.710
Group Stage	Final	-137.565 *	43.415	0.017	-258.250	-16.880
	Semi-finals	-202.875 *	45.643	<0.001	-329.750	-76.000
	Quarter-finals	-89.258	39.087	0.159	-197.910	19.390
	Round of 16	-98.802 *	34.861	0.043	-195.710	-1.900

* The mean difference is significant at the 0.05 level.

Third, the post-hoc comparisons point out that the mean percentage of passes completed for Group 5 (Group Stage) $[M = 0.781, SD = 0.065]$ was significantly different from Group 2 (Semi-finals) $[M = 0.850, SD = 0.043]$ and from Group 1 (Final) $[M = 0.854, SD = 0.048]$, as seen in Table 15.

Table 15: Post-hoc test for the percentage of passes completed

Multiple Comparisons

% passes completed

Tukey HSD

(I) competition stage	(J) competition stage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Final	Quarter-finals	0.004	0.026	1.000	-0.067	0.076
	Semi-finals	0.031	0.023	0.646	-0.032	0.095
	Round of 16	0.032	0.021	0.534	-0.026	0.091
	Group Stage	0.073 *	0.022	0.010	0.012	0.135
Semi-finals	Final	-0.004	0.026	1.000	-0.076	0.067
	Quarter-finals	0.027	0.024	0.789	-0.039	0.093
	Round of 16	0.028	0.022	0.709	-0.033	0.090
	Group Stage	0.069 *	0.023	0.028	0.005	0.133
Quarter-finals	Final	-0.031	0.023	0.646	-0.095	0.032
	Semi-finals	-0.027	0.024	0.789	-0.093	0.039
	Round of 16	0.001	0.019	1.000	-0.051	0.053
	Group Stage	0.042	0.020	0.215	-0.013	0.097
Round of 16	Final	-0.032	0.021	0.534	-0.091	0.026
	Semi-finals	-0.028	0.022	0.709	-0.090	0.033
	Quarter-finals	-0.001	0.019	1.000	-0.053	0.051
	Group Stage	0.041	0.018	0.145	-0.008	0.090
Group Stage	Final	-0.073 *	0.022	0.010	-0.135	-0.012
	Semi-finals	-0.069 *	0.023	0.028	-0.133	-0.005
	Quarter-finals	-0.042	0.020	0.215	-0.097	0.013
	Round of 16	-0.041	0.018	0.145	-0.090	0.008

* The mean difference is significant at the 0.05 level.

Finally, the post-hoc comparisons, presented in Table 16, showed that the mean parameter for Group 5 (Group Stage) $[M = 1.692, SD = 0.106]$ was significantly different from Group 4 (Round of 16) $[M = 1.597, SD = 0.138]$, from Group 2 (Semi-finals) $[M = 1.565, SD = 0.100]$ and from Group 1 (Final) $[M = 1.567, SD = 0.107]$.

Table 16: Post-hoc test for the parameter α

Multiple Comparisons

α

Tukey HSD

(I) competition stage	(J) competition stage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Final	Quarter-finals	0.002	0.046	1.000	-0.125	0.129
	Semi-finals	-0.033	0.040	0.922	-0.146	0.079
	Round of 16	-0.030	0.037	0.928	-0.133	0.073
	Group Stage	-0.125 *	0.039	0.016	-0.234	-0.016
Semi-finals	Final	-0.002	0.046	1.000	-0.129	0.125
	Quarter-finals	-0.035	0.042	0.919	-0.153	0.083
	Round of 16	-0.032	0.039	0.926	-0.141	0.077
	Group Stage	-0.127 *	0.041	0.021	-0.241	-0.013
Quarter-finals	Final	0.033	0.040	0.922	-0.079	0.146
	Semi-finals	0.035	0.042	0.919	-0.083	0.153
	Round of 16	0.003	0.033	1.000	-0.089	0.095
	Group Stage	-0.092	0.035	0.077	-0.189	0.006
Round of 16	Final	0.030	0.037	0.928	-0.073	0.133
	Semi-finals	0.032	0.039	0.926	-0.077	0.141
	Quarter-finals	-0.003	0.033	1.000	-0.095	0.089
	Group Stage	-0.095 *	0.031	0.025	-0.182	-0.008
Group Stage	Final	0.125 *	0.039	0.016	0.016	0.234
	Semi-finals	0.127 *	0.041	0.021	0.013	0.241
	Quarter-finals	0.092	0.035	0.077	-0.006	0.189
	Round of 16	0.095 *	0.031	0.025	0.008	0.182

4.2.2. Passing sequences that resulted in a goal scored

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.

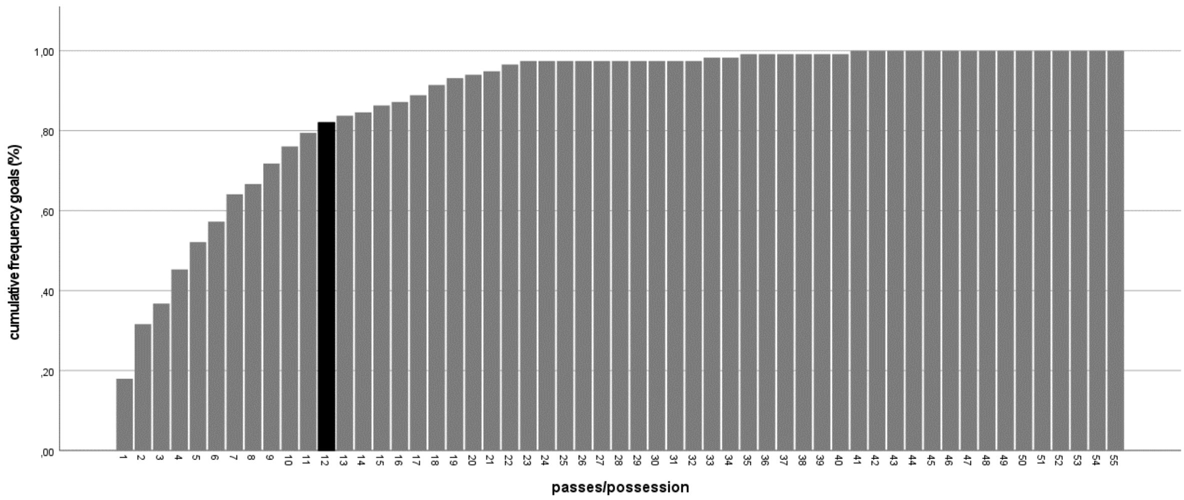


Figure 13: Cumulative frequency of goals

The UEFA EURO 2020 data on the passing sequence revealed that 80% of the goals resulted from 12 passes or less, as represented in Figure 13. Indeed, approximately 50% of the goals resulted from possessions of five or fewer passes. In addition, Figures 14 and 15 show the frequency of each sequence length and the frequency of goals concerning the possession length. As a result, the previous 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.

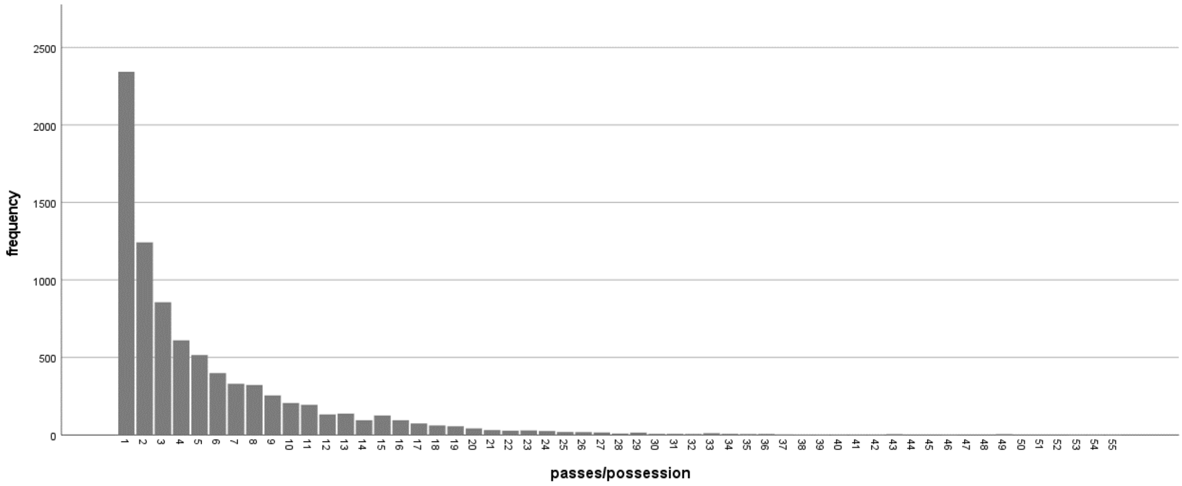


Figure 14: Frequency of each sequence length in the UEFA EURO 2020 tournament

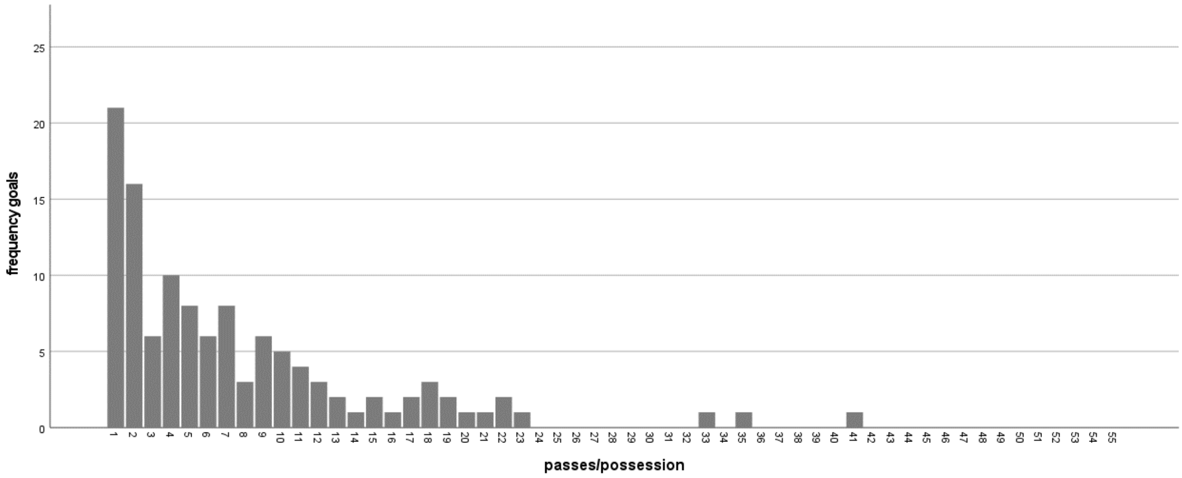


Figure 15: Frequency of goals concerning the length of the possession in the UEFA EURO 2020 tournament

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 16 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 (Hughes & Franks, 2005). 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.

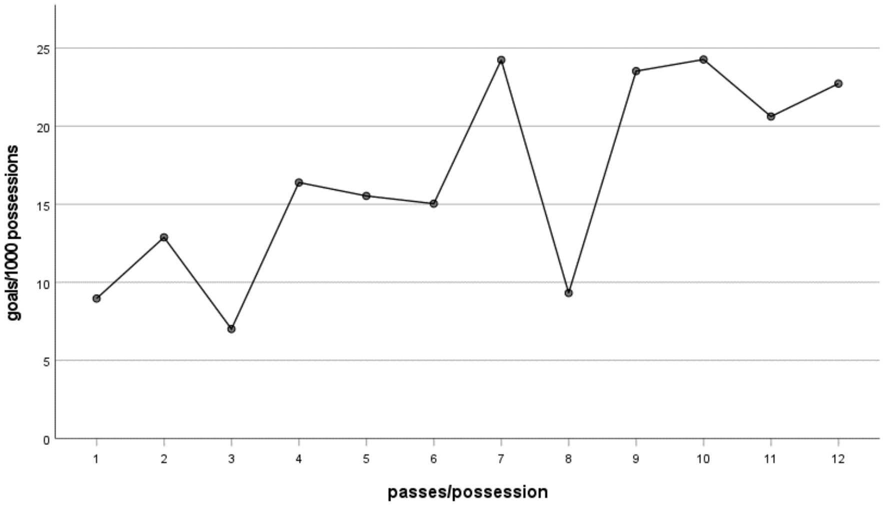


Figure 16: Analysis of the number of goals scored per 1000 possession for the UEFA EURO 2020

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 1–6 and 7-12 were compared using a t-test after performing a descriptive group statistics analysis (Table 17). There was a significant difference between the two groups ($t_{10} = -2.878; p = 0.016; \eta^2 = 0.45,$ large effect), as seen in Table 18.

Table 17: Group Statistics for each group (goals per 1000 possessions for sequence lengths 0–6 and 7-12)

Group Statistics

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

Table 18: Independent-samples t-test for comparing the two groups (goals per 1000 possessions for sequence lengths 0–6 and 7-12)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				Mean	Std. Error	95% Confidence Interval of the	
		F	Sig.	t	df	One-Sided p	Two-sided p	Difference	Difference	Lower	Upper
Goals/1000 possessions	Equal variances assumed	0.196	0.667	-2.878	10.000	0.008	0.016	-8.148	2.831	-14.456	-1.840
	Equal variances not assumed			-2.878	8.687	0.009	0.019	-8.148	2.831	-14.588	-1.708

4.3. Discussion

This section discusses the results of the analyses of the passing sequences and the passing sequences that resulted in a goal scored.

4.3.1. Passing sequences

Previous works have reported that the general attacking strategy, depicted by the possession characteristics, influences teams' performance in a football match (Garganta, 1997; M. Hughes & Franks, 2005; Tenga & Sigmundstad, 2011). Therefore, the work developed in section 4.2.1 aimed to study the passing sequences, specifically, the distribution of passes per possession, which describes the general attacking strategy. First, it was found that roughly 70% of the 102 samples were consistent with the power-law hypothesis by fitting the passing distribution to the power law and testing it at $p \leq 0.05$. More precisely, it was possible to confirm that, for those samples consistent with the power-law hypothesis, the power-law distribution was an appropriate model for a portion of the sample of the passes per possession distributions (Clauset *et al.*, 2009). However, as few real events follow power laws for all range of x, the passing distributions were fitted to the power law regardless of their true distribution, and $x_{min} = 1$ was fixed for all samples to allow comparison of the power law exponent, $-\alpha$, of the different teams.

As a result, a novel way of describing the general attacking strategy of football teams was introduced. Hence, this dissertation proposed to use α to describe the general attacking strategy of football teams. Indeed, teams with a lower value of α are teams that employ a possession-based strategy of play, while teams with a higher value of α are teams that adopt a direct strategy of play. Next, the relationship between α and pass statistics, such as the number of passes, the number of passes completed and the percentage of passes completed, was examined. The findings revealed that teams that performed more passes during the regular time of the match also executed more successful passes, as suggested by the nearly perfect positive correlation between the number of passes and the number of passes completed. Such outcomes align with the work of Gama, Dias, Couceiro, Sousa, *et al.* (2016). Moreover, the results demonstrated that the percentage of passes completed was very large and nearly perfect positively correlated with the number of passes and the number of passes completed, respectively. These findings indicated that teams that execute more passes and, so, more passes completed exchange the ball more successfully. Additionally, the parameter α showed a very large negative correlation with all the pass statistics. Thus, these results suggested that teams adopting a possessive play perform more passes and more successfully, losing the ball less when exchanging it during the attacking phase and set pieces.

Then, the mean values of α for each national team combined with the mean values of the passes statistics unveiled that Germany was the team with the lowest mean values of α and its standard deviation. This indicated that Germany was the national team that played a more possessive type of play and was loyal to its strategy. This result is in line with the findings of Clemente, Silva, *et al.* (2016). These researchers studied Germany in the FIFA World Cup 2014. They highlighted the capacity of the German team to have ball possession and create passes, as was found to be a feature of teams with a lower α, i.e., teams that adopt a possession-based strategy of play. As a result, it is possible to understand that possessive play is an inherent characteristic of the German team in different tournaments. In contrast, North Macedonia, Slovakia, and Scotland opted to play a direct type of play, as suggested by the higher mean α and lower standard deviation values.

Furthermore, Spain stood out from its rivals concerning the number of passes and the number of passes completed, also presenting the second lowest mean α. Cotta *et al.* (2013) studied Spain during the FIFA World Cup 2010 tournament, highlighting the Spanish team's elaborated style. This elaborated style, reflected in the high number of passes completed, is a characteristic of *tiki-taka*, a style of playing football implemented by the Spanish national team, in which teams execute a lot of short passes, keeping the possession of the ball. Lastly, England and Italy, the tournament's finalists, presented a standard deviation of α above the average, which raises the question of whether adjusting the strategy of play for each match leads to success in a tournament. This question was not answered and is presented for future work.

The relationship of the overall performance variables (match result and stage reached in the tournament) with the pass statistics and the general strategy of play was examined. First, no statistical differences were found between teams that achieved different match results (defeat, draw or victory) concerning the variables: pass statistics and α. In opposition, statistical differences were found between the stage reached in the competition and all the variables. The results indicated that teams that achieved the highest stages of the tournament, namely the Semi-finals and Final, were significantly different from the teams that were eliminated in the first stage of the tournament (Group Stage) concerning the number of passes and the percentage of passes completed. Thus, this showed that the most unsuccessful teams performed a lower number of passes in the matches played and did so less successfully. In the same way, regarding the number of passes completed and the parameter α, teams that achieved the Round of 16, Semi-finals and Final were significantly different from the teams that were eliminated in the Group Stage. These results revealed that unsuccessful teams adopted a more direct type of play while executing 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 dissertation's findings. On the other hand, Hughes *et al.* (1988) findings in which 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.

4.3.2. Passing sequences that resulted in a goal scored

In section 4.2.2, the study was limited to those possessions that resulted in a goal scored. Thus, it was revealed that 80% of the goals resulted from 12 passes or less. Consequently, the outcomes did not agree with Reep & Benjamin (1968) and M. Hughes & Franks (2005), whose results showed that 80% of the goals came from three/four passes or fewer. This finding and the t-test 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.

Chapter 5 – Passing network analysis

This chapter presents the methodology (section 5.1), results (section 5.2) and posterior discussion (section 5.3) regarding the passing network analysis.

5.1. Methodology

The second part of this dissertation had two main objectives. First, the analysis aimed to study the impact of macro network properties on performance variables, namely the match result and the stage reached in the tournament. Second, the goal was to analyse player/playing position-zone passing networks, and study the differences and similarities of distinct systems of play, while capturing the spatial-temporal components of the passing network.

The eventing data sets (*StatsBomb Event Data*) were again used to accomplish these objectives. For each team in each match, it was only considered the “Pass” and “Ball Receipt*” types of events in the attacking phase and during all set pieces. This allowed 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 (see Table 19 as an example). Therefore, this information enabled the construction of the different types of networks using Python and its package *NetworkX*®.

Table 19: Example of the dataset structure, with a sequence of passes of Portugal in the match against Belgium.

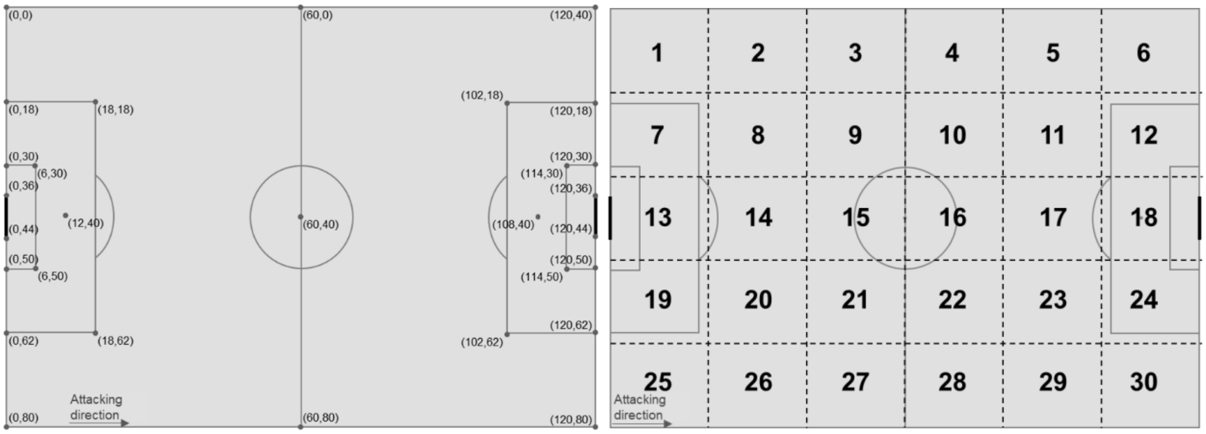
index	timestamp	type_name	play_pattern_name	team_name	location	player_id	position_id	pass_recipient_id	pass_length	pass_height_name	pass_end_location	pass_body_part_name	pass_outcome_name
1451	00:31:04	Pass	Regular Play	Portugal	[75.1, 65.6]	16028.0	2.0	5207.0	24.446268	Ground Pass	[79.2, 41.5]	Right Foot	
1452	00:31:06	Ball Receipt*	Regular Play	Portugal	[79.2, 41.5]	5207.0	23.0						
1453	00:31:06	Pass	Regular Play	Portugal	[79.2, 41.5]	5207.0	23.0	9929.0	18.681005	Ground Pass	[65.5, 28.8]	Right Foot	
1454	00:31:09	Ball Receipt*	Regular Play	Portugal	[65.5, 28.8]	9929.0	21.0						
1456	00:31:12	Pass	Regular Play	Portugal	[76.8, 23.1]	9929.0	21.0	5209.0	14.676852	Ground Pass	[75.3, 8.5]	Right Foot	
1457	00:31:14	Ball Receipt*	Regular Play	Portugal	[75.3, 8.5]	5209.0	6.0						
1458	00:31:14	Pass	Regular Play	Portugal	[75.3, 7.2]	5209.0	6.0	3168.0	9.552486	Ground Pass	[70.4, 15.4]	Left Foot	
1459	00:31:16	Ball Receipt*	Regular Play	Portugal	[70.4, 15.4]	3168.0	13.0						
1461	00:31:20	Pass	Regular Play	Portugal	[72.8, 23.3]	3168.0	13.0	20016.0	28.255442	Low Pass	[58.9, 47.9]	Right Foot	
1462	00:31:22	Ball Receipt*	Regular Play	Portugal	[58.9, 47.9]	20016.0	3.0						
1464	00:31:23	Pass	Regular Play	Portugal	[59.3, 49.8]	20016.0	3.0	3593.0	10.606602	Ground Pass	[69.8, 48.3]	Right Foot	
1465	00:31:24	Ball Receipt*	Regular Play	Portugal	[69.8, 48.3]	3593.0	15.0						
1466	00:31:24	Pass	Regular Play	Portugal	[70.0, 49.6]	3593.0	15.0	20016.0	14.045996	Ground Pass	[58.0, 56.9]	Right Foot	
1467	00:31:26	Ball Receipt*	Regular Play	Portugal	[58.0, 56.9]	20016.0	3.0						
1469	00:31:31	Pass	Regular Play	Portugal	[57.4, 55.1]	20016.0	3.0	5206.0	23.11731	Ground Pass	[50.3, 33.1]	Right Foot	
1470	00:31:33	Ball Receipt*	Regular Play	Portugal	[50.3, 33.1]	5206.0	5.0						

Data provided by StatsBomb

5.1.1. Zone passing networks analysis

Initially, the analysis was carried out by building zone passing networks of the passes performed during the regular time (90 min). 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. 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. Therefore, a preliminary analysis was conducted to answer this question.

The zone networks were formed by splitting the field of play (Figure 17.1) equally into Z zones, as illustrated in Figure 17.2; 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 with a weight that measured the total number of successful passes. As a result, as discriminated in Table 20, 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.



(1) (2)

Figure 17: (1) Field of play coordinates (x,y) in yards; (2) Example of the field of play's division into 30 zones (6 sectors x 5 corridors)

Table 20: Zone passing networks analysed

#	Type of passing network	Node	Nu of sectors	Nu of corridors	Nu of zones
1	Zone passing network	Zone	3	3	9
2	Zone passing network	Zone	4	3	12
3	Zone passing network	Zone	3	5	15
4	Zone passing network	Zone	6	3	18
5	Zone passing network	Zone	4	5	20
6	Zone passing network	Zone	6	5	30

After selecting the number of zones, the analysis examined the relationship between the networks' density and average clustering coefficient and the variables covered in Chapter 4, namely the parameter α and the pass statistics. The same procedure of the previous chapter was adopted. After confirming that the assumptions of normality, linearity and homoscedasticity were not violated, the Pearson Product-Moment correlation coefficient was used to study the relationship between these variables. Spearman's Rank Order Correlation was employed when the data did not respect the assumptions. An identical scale was applied to classify the correlation strength: 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).

Additionally, the differences in the networks' density and average clustering coefficient were explored between teams that achieved different match results ((i) defeat, (ii) draw or (iii) victory) and teams that reached different stages of the tournament ((i) Final, (ii) Semi-finals, (iii) Quarter-finals, (iv) Round of 16 and (v) Group Stage). These investigations were performed using one-way ANOVA or the Welsh and Brown-Forsythe tests after verifying the assumptions of normality and homogeneity. Firstly, the assumption of normality was evaluated using Kolmogorov-Smirnov tests $(p > 0.05)$. However, the Central Limit Theorem was evoked since $n \geq 30$ to assume the assumption of normality. Secondly, the homogeneity assumption was investigated using Levene's test. ANOVA was substituted with the Welsh and Brown-Forsythe tests when this assumption was violated. The Tukey's HSD (honestly significant difference) test or the Tukey's-Kramer test was used to identify the differences when the test found significant differences between the factors (Clemente *et al.*, 2015). Finally, the eta-squared, η^2, was used for measuring the effect size in ANOVA, and the guidelines from Cohen (1988) were followed to translate the strength of the eta-squared: 0.01=small effect; 0.06=moderate effect and 0.14=large effect.

5.1.2. Clustering analysis

Clustering analysis was conducted to study the zone networks and the differences and similarities of various systems of play in the playing position-zone networks. Clustering analysis can be a valuable tool for discovering and exploring data characteristics by organising them into subgroups or clusters (Everitt *et al.*, 2011). Han & Kamber (2006) designate clustering as "the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another but are very dissimilar to objects in other clusters". The methodology for this section's work was an adaptation of the work of Milligan (1996), which includes a framework with seven steps that typically constitute the clustering analysis (Everitt *et al.*, 2011). As a result, the main steps are outlined below (Milligan, 1996):

1. Clustering objects: the objects to be clustered must be selected and chosen in such a way as to be representative of the cluster structure believed to be present in the data;
2. Clustering variables: the variables used in the clustering analysis must be chosen and contain sufficient information to permit the clustering of the objects;
3. Variable standardisation: a choice must be made regarding the standardization of each variable used in the cluster analysis. However, variable standardisation is not always a requirement that must be fulfilled and can sometimes be misleading (Everitt *et al.*, 2011);
4. Proximity measure: a similarity or dissimilarity measure must be selected to reflect the degree of closeness or separation of the objects to be clustered;
5. Clustering method: a method suitable for the kind of clustering that is expected to be present in the data must be selected.
6. Number of clusters: the number of clusters must be determined with the help of different techniques;
7. Interpretation: the results must be interpreted within their context with the auxiliary of graphical representation and descriptive statistics (Everitt *et al.*, 2011).

5.1.2.1. Clustering objects

Initially, a preliminary clustering analysis was performed on the 102 zone networks examined in section 5.1.1 to determine if any general differences were observed. Then, instead of using the player/playing position or the zone of the playing field as a node, the combination of both was considered. Thus, the playing position-zone networks were constructed by representing each pair $(playing\ position,\ zone)$ as a node. In this type of network, the size of this type of network is determined by multiplying the number of playing positions by the number of zones. A pass from the pair $i = (playing\ position\ _a,\ zone_b)$ to $j =$

$(playing\ position_c,\ \ zone_d)$ results in the creation of a link from node i to j, and this link has a weight that quantifies the total number of complete passes.

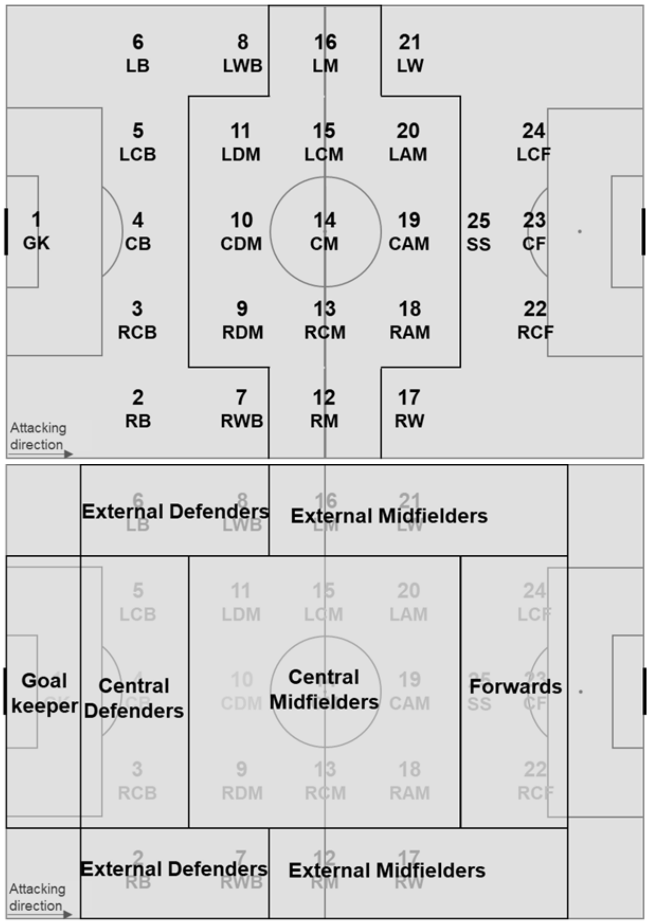
Several decisions were made to achieve this section's objectives. First, to allow the comparison between teams and systems of play, rather than analysing the players individually, the study concentrated on analysing the playing positions, which various players throughout the match could carry out. In addition, the number of zones chosen for this analysis was equal to the number of zones selected in section 4.1.1. after the descriptive statistics.

Second, a team can adopt more than one system of play throughout the match. For this reason, a sliding window technique was applied. As Cotta *et al.* (2013) and Clemente *et al.* (2015) suggested, a sliding window's size equal to 15 minutes and a step of 5 minutes were chosen. This window's size was selected because it is "long enough to capture the state of the game" (Cotta *et al.*, 2013). Furthermore, regular and extra time were studied, but the distinct parts of the match were treated separately. The last sliding window of each part contained the additional time for that part. Therefore, 14 playing position-zone networks were constructed for each team in each match's regular time. Two additional networks were constructed for teams in matches that went to extra time. However, when a team changed its system of play within a sliding window, an extra network was built if the respective team made ten or more passes. In addition, this value was used as a threshold, i.e., teams that performed less than ten passes during the sliding window length did not enter the clustering analysis. To sum up, by generating distinct networks for the different systems of play, building networks with the same number of nodes was possible since the number of playing positions and zones remained constant. Thus, the networks' size was equal to the multiplication of the eleven playing positions by the number of zones.

Third, although the different systems of play may share playing positions, they always have differences. For illustration purposes, considering the possible positions shown in Figure 18, a 4-3-3 system of play, $SP_{4-3-3} = \{GK, RB, RCB, RLB, LB, \boldsymbol{CDM}, RCM, LCM, \boldsymbol{RW}, \boldsymbol{LW}, \boldsymbol{CF}\}$, and a 4-4-2 system of play, $SP_{4-4-2} = \{GK, RB, RCB, RLB, LB, \boldsymbol{RM}, RCM, LCM, \boldsymbol{LM}, \boldsymbol{RCF}, \boldsymbol{LCF}\}$, share seven playing positions and have four different ones. Moreover, the same system of play can have different configurations. Therefore, an adaptation of the general classification of the playing positions proposed by Clemente, José, *et al.* (2016) was adopted. As represented in Figure 18, the 25 positions categorized by StatsBomb have been reduced to 6 common positions:

- Goalkeeper = $\{GK\}$;
- Central Defenders = $\{RCB, CB, LCB\}$;
- External Defenders = $\{RB, LB, RWB, LWB\}$;
- Central Midfielders = $\{RDM, CDM, LDM, RCM, CM, LCM, RAM, CAM, LAM\}$;
- External Midfielders = $\{RM, LM, RW, LW\}$;
- Forwards = $\{RCF, CF, LCF, SS\}$.

As a result, this process established a common classification for the various play systems, enabling comparisons between them.



Legend:

Position ID	Position Abbreviation	Position Name
1	GK	Goalkeeper
2	RB	Right Back
3	RCB	Right Centre Back
4	CB	Centre Back
5	LCB	Left Centre Back
6	LB	Left Back
7	RWB	Right Wing Back
8	LWB	Left Wing Back
9	RDM	Right Defensive Midfield
10	CDM	Centre Defensive Midfield
11	LDM	Left Defensive Midfield
12	RM	Right Midfield
13	RCM	Right Centre Midfield
14	CM	Centre Midfield
15	LCM	Left Centre Midfield
16	LM	Left Midfield
17	RW	Right Wing
18	RAM	Right Attacking Midfield
19	CAM	Centre Attacking Midfield
20	LAM	Left Attacking Midfield
21	LW	Left Wing
22	RCF	Right Centre Forward
23	CF	Centre Forward
24	LCF	Left Centre Forward
25	SS	Secondary Striker

Data provided by StatsBomb

Figure 18: Statsbomb's playing positions and the respective six common positions

5.1.2.2. Clustering variables

Clustering analysis was proposed to divide the generated networks into groups according to specific criteria: the local clustering coefficient and the degree (Diquigiovanni & Scarpa, 2019). These two metrics were selected because they have been reported as good descriptors to capture how teams play (Pina *et al.*, 2017). Thus, the clustering analysis was conducted using the specific criteria independently. On the one hand, for the zone networks' analysis, the clustering analysis was first carried out using the nodes' clustering coefficient as variables (attributes) of the objects and then using the nodes' degree as variables. On the other hand, instead of using the nodes' metric values, the average clustering coefficient and the sum of the degree per common position per zone were used as variables for the playing position-zone networks' analysis. For example, in a team playing in a 4-4-2, the clustering coefficient and degree of the right and left centre forwards in zone z were, respectively, averaged ($\bar{C}_{(Forwards,z)} = \frac{1}{2}\left[C_{(RCF,z)} + C_{(LCF,z)}\right]$) and summed ($k_{(Forwards,z)} = k_{(RCF,z)} + k_{(LCF,z)}$). Consequently, each object was described by n variables, where n is equal to the multiplication of the number of 6 common positions by the selected number of zones.

The clustering analysis requires particular attention to a specific case: clustering high-dimensional data, and this is challenging due to the curse of dimensionality. According to the curse of dimensionality, initially formulated by Bellman (1961), the number of samples required to estimate any function with a given level of accuracy increases exponentially concerning the number of input variables (i.e., dimensionality) of the function (Chen, 2009). The data in clustering become increasingly sparse as the number of dimensions rises, rendering the distance between pairs of points irrelevant and making it likely that any one point's average density will be low (Han & Kamber, 2006). As a result, a feature transformation technique was applied to reduce dimensionality (Chen, 2009; Han & Kamber, 2006). The Principal Component Analysis (PCA) was the method selected for the reduction of dimensionality.

The PCA, introduced by Pearson (1901) and later developed by (Pearson, 1901), seeks to represent the data using k n-dimensional orthogonal vectors, where $k \leq n$ (Han & Kamber, 2006). The dimensionality reduction is achieved by projecting the data onto a smaller space. The original variables are transformed into a new set of variables, named principal components (PCs), uncorrelated and ordered so that the first PCs retain the most variation in the data (Roessner *et al.*, 2011). The PCs represent a selection of a new coordinate system obtained by rotating the original axis to a set of new axis. Thus, the first PC is a linear combination of all the original variables, representing the direction of maximum variability (Roessner *et al.*, 2011). The second PC represents the direction of maximum variability orthogonal to the first. Accordingly, the last PC represents the direction of maximum variability and is orthogonal to all the others. Because the reduction of dimensionality is an objective of PCA, several criteria have been proposed for determining how many PCs should be used in the clustering analysis. However, the criterion used was to include all those PCs up to a predetermined total percentage variance explained equally to 90% (Holland, 2019).

5.1.2.2. Variable standardisation

Each variable's measurement unit affects the clustering analysis, so the data needs to be standardized to give each variable the same weight. Converting the original measurements to unitless variables is one way to standardize measurements (Han & Kamber, 2006). However, since the clustering analysis was conducted using the specific criteria separately, the objects clustered were measured in the same units, so the standardization process was not performed.

6.1.2.4. Proximity measure

The dissimilarity between objects was computed based on the distance between objects. The distance measure chosen was the Euclidian distance, which is written as:

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + \left(x_{in} - x_{jn}\right)^2},$$

where $i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $j = (x_{j1}, x_{j2}, \ldots, x_{jn})$ are two n-dimensional data objects (Han & Kamber, 2006).

6.1.2.5. Clustering method

Numerous clustering algorithms have been proposed in the literature; however, *K-means* was chosen. This method is one of the most well-known and widely used partitioning techniques and has various advantages, such as simple mathematical principles and easy implementation (Han & Kamber, 2006; Jain, 2010; Li *et al.*, 2017; Yuan & Yang, 2019). *K-means* is a partitioning algorithm that uses a dissimilarity function based on distance as a partitioning criterion (Hartigan & Wong, 1979). This algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high and the intercluster similarity is low. Cluster similarity is measured regarding the mean value of the objects in a cluster, which can be viewed as the cluster's centroid or centre of gravity (Han & Kamber, 2006).

K-means first initializes the cluster means by randomly selecting k points. Each iteration consists of two steps: clustering assignment and centroid update. Thus, first, each remaining point is assigned to the most similar cluster, i.e., with the closest mean. Second, the new means for each cluster are updated. This process is done until the scoring function converges. Usually, the sum of squared errors (SSE) is used as the scoring function and is defined as:

$$SSE(C) = \sum_{i=1}^{k} \sum_{x_j \in C_i} \left\| x_j - \mu_i \right\|^2 ,$$

where x_j is the point in space representing a given object and μ_i is the mean of the cluster C_i. *K-means* has converged if the centroids do not change from one iteration to the next or if $\sum_{i=1}^{k} \|\mu_i^t - \mu_i^{t-1}\|^2 \leq \epsilon$, where $\epsilon > 0$ denotes the convergence threshold, t the current iteration and μ_i^t the mean for the cluster C_i in iteration t (Zaki *et al.*, 2014).

Usually, *K-means* is performed multiple times and for different values of k. The method starts with a random choice for the initial centroids. Hence, the K-means pseudo-code is presented in Algorithm 1 (Zaki *et al.*, 2014).

Algorithm 1: *K-means* algorithm

K-means (D, k, ϵ)

$t = 0$

Initialize randomly k centroids: $\mu_1^t, \mu_2^t, \ldots, \mu_k^t \in \mathbb{R}^d$

repeat

  $t \longleftarrow t + 1$

  $C_j \longleftarrow \emptyset$ for all $j = 1, \ldots, k$

  foreach $x_j \in D$ **do**

    $j^* \longleftarrow \arg min_i \left\{ \left\| x_j - \mu_i^t \right\|^2 \right\}$

    $C_{j^*} \longleftarrow C_{j^*} \cup \{x_j\}$

  foreach $i = 1 \; to \; k$ **do**

    $\mu_i^t \longleftarrow \frac{1}{C_i} \sum_{x_j \in C_i} x_j$

until $\sum_{i=1}^{k} \left\| \mu_i^t - \mu_i^{t-1} \right\|^2 \leq \epsilon$

6.1.2.6. Number of clusters

Selecting the number of clusters is one of the most challenging decisions (Everitt *et al.*, 2011). Hence, the elbow method was employed to select the number of clusters, k, to analyse. This approach uses the square of the distance between the objects in each cluster and the cluster's centroid. So, the SSE is computed and used as a performance indicator, so smaller values indicate that each cluster is more convergent. The k value can be determined by observing the plot of the k-SSE curve and finding the "elbow" (Yuan & Yang, 2019).

It is imperative and crucial the validating the results of the clustering algorithm. Consequently, the Silhouette analysis, proposed by Rousseeuw (1987), was done. This technique compares each cluster's tightness and separation. Each cluster is represented by one silhouette, revealing which objects are well within their cluster and which are not. To compare the quality of the clusters, the entire clustering can be viewed by plotting all the silhouettes into a single diagram (Kaufman & Rousseeuw, 1990). The silhouette width, $s(i)$, compares the within-cluster cohesion, based on the distance to all entities in the same cluster, to the cluster separation and is written as follows (de Amorim & Hennig, 2015; Kaufman & Rousseeuw, 1990):

$$s(i) = \frac{b(i) - a(i)}{max\{a(i),\ b(i)\}},$$

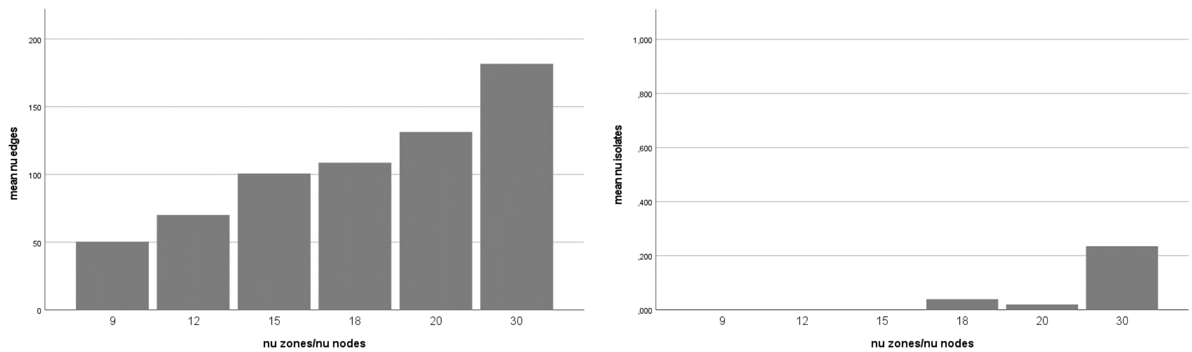
where $a(i)$ is the average dissimilarity of $i \in C_k$ to all other $j \in C_k$, $b(i)$ is the minimum dissimilarity over all clusters C_l, to which i is not assigned, of the average dissimilarities to $j \in C_l, l \neq k$. Therefore, -$-1 \leq s(i) \leq 1$ (de Amorim & Hennig, 2015). A silhouette width near 1 implies that the object is far away from the neighbouring clusters. On the other hand, a value of 0 suggests that the object should be assigned to its or a neighbour cluster. Finally, a negative value indicates that those objects might have been assigned to the wrong cluster. The average silhouette width, or silhouette coefficient (SC), can be used not only to assess the validity of the clustering but also might be used to select the number of clusters (Rousseeuw, 1987).

5.2. Results

This section presents the results of zone passing networks analysis and the clustering analyses.

5.2.1. Zone passing networks analysis

First, 102 networks were built for the different number of zones $Z = \{9, 12, 15, 18, 20, 30\}$. Next, a descriptive analysis, shown in detail in Appendix E, was performed to select the number of zones, i.e., the number of nodes to be used in this section's analysis. Therefore, as seen in Figure 19, 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 the majority of these networks, all zones were connected with at least another zone. For these reasons, the 102 networks with 30 nodes, presented in Appendix F, were selected to analyse the impact of macro network properties on performance variables, such as the match result and the maximum stage reached in the tournament.



(1) (2)

Figure 19: (1) Mean number of edges of the different-sized networks; (2) Mean number of isolated nodes of the different-sized networks

As in Chapter 4, a descriptive study of the networks' macro properties, namely the density and the average clustering coefficient, was initially done, as presented in Table 21. This study was then accompanied by an analysis of the histograms and boxplot. The main finding from the descriptive analysis was the negative skewness of both variables that suggested the cluster of values in the right at high distribution levels, as confirmed by the inspection of the histograms in Figure 22.

Table 21: Descriptive table of the networks' density and average clustering coefficient

Descriptive Statistics

	Mean	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	5% Trimmed Mean	Median	Variance	Std Deviation	Minimum	Maximum	Range	Interquartil Range	Skewness	Kurtosis
Density	0.209	0.003	0.202	0.216	0.211	0.213	0.001	0.034	0.087	0.274	0.186	0.045	-0.916	1.068
Average clustering coefficient	0.383	0.007	0.368	0.397	0.387	0.394	0.006	0.074	0.086	0.554	0.468	0.100	-1.006	1.915

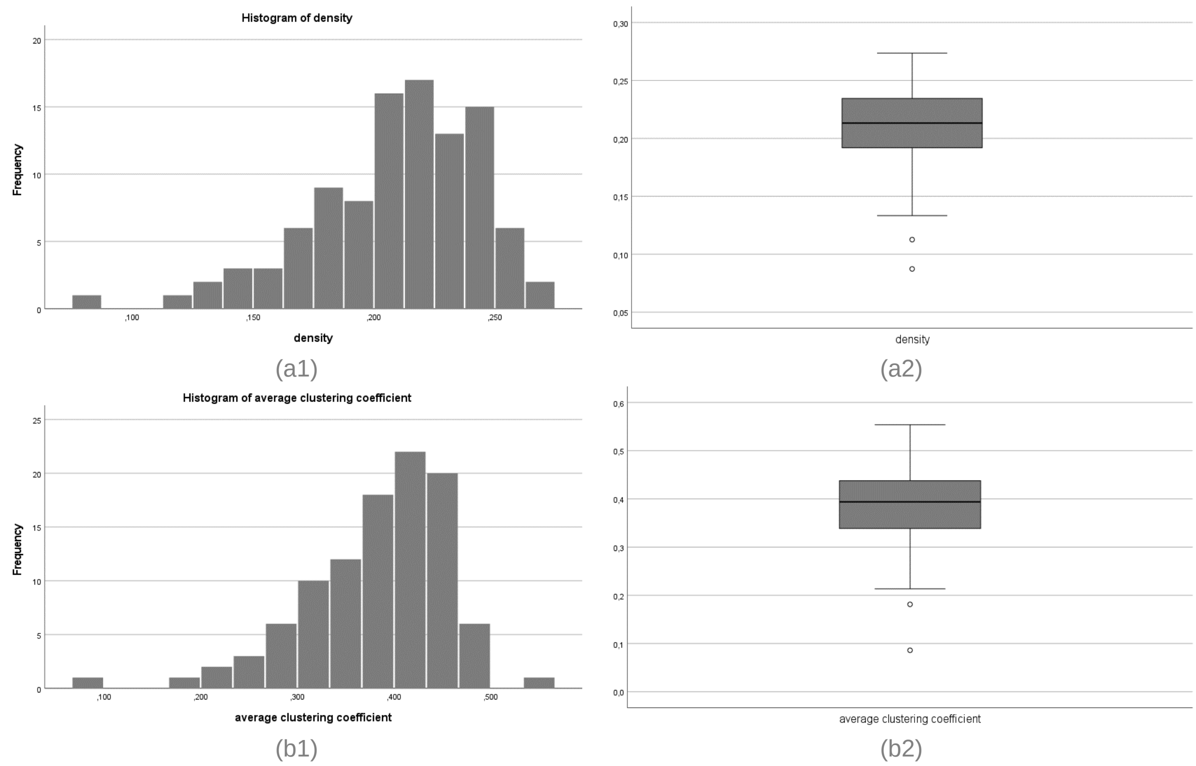


Figure 20: (1) Histograms for the (a) density and (b) the average clustering coefficient. (2) Box plots for the (a) density and (b) the average clustering coefficient

Moreover, the box plot's examination identified two outliers for all variables. These two outliers belong to Sweeden and Poland in their group stage matches against Spain. On the one hand, the Swedish team had the lowest values in the density (0.087) and average clustering coefficient (0.086) of the tournament. On the other hand, the Polish team had the second lowest values in the density (0.113) and average clustering coefficient (0.181) of the tournament.

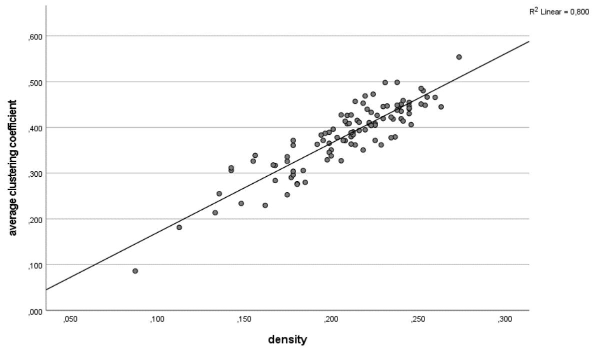
Again, the normality assumption of Pearson product-moment and ANOVA was evaluated using the Kolmogorov-Smirnov statistic. Table 22 displays the results of this test statistics that showed that the distribution of the number of edges, density and average clustering coefficient had a significant result (Sig.<0.05), indicating a violation of the assumption of normality. However, again, since $n \geq 30$ and considering the Central Limit Theorem, the assumption was assumed.

Table 22: Test of Normality for the density and the average clustering coefficient

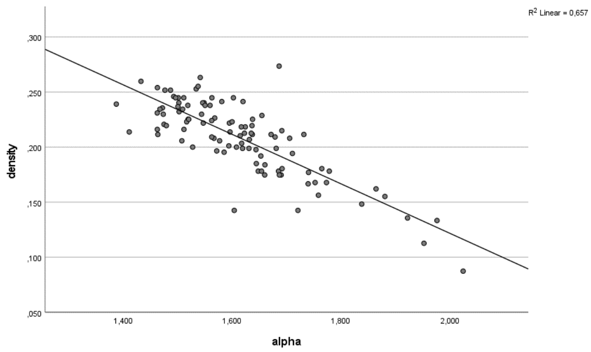
Test of Normality

	Kolmogorov-Smirnov		
	Statistic	**df**	**Sig.**
Density	0.101	102.000	0.012
Average clustering coefficient	0.100	102.000	0.014

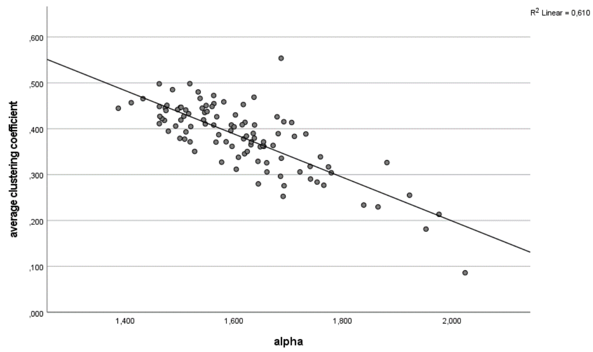
In addition to examining the relationships between density and clustering coefficient, the relationships between these variables and the variables presented in Chapter 4, namely the pass statistics and the parameter α, were also investigated. Thus, the assumptions of linearity and homoscedasticity of the Pearson product-moment were assessed to see if there was any violation. Figure 21 shows that the relationship between the network's macro properties and pass statistics was not linear. In addition, the homoscedasticity was only violated for the relationship between the percentage of passes completed and the network's macro properties, as seen in Appendix D.



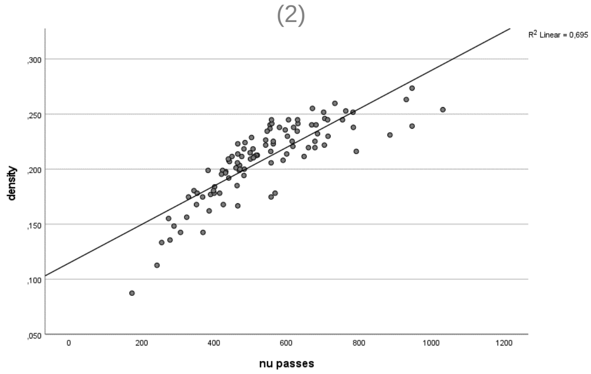
(1)



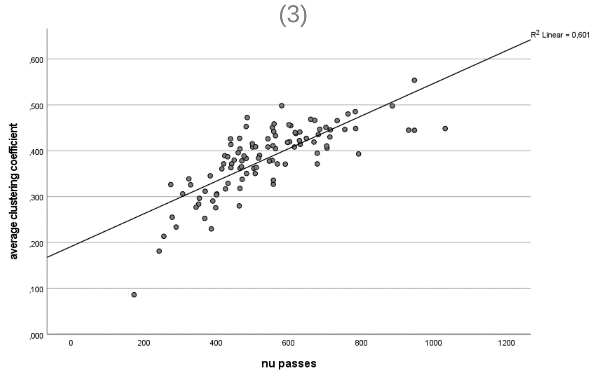
(2)



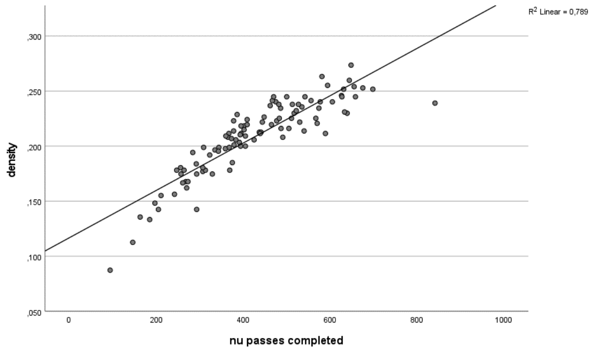
(3)



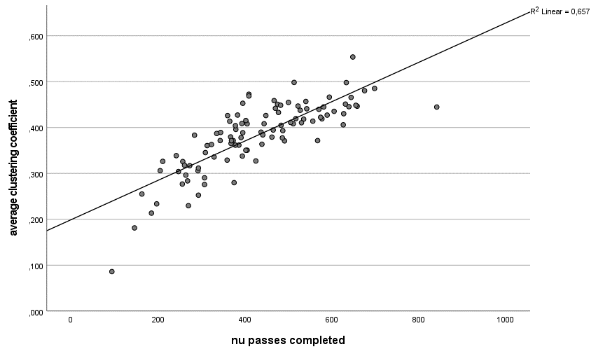
(4)



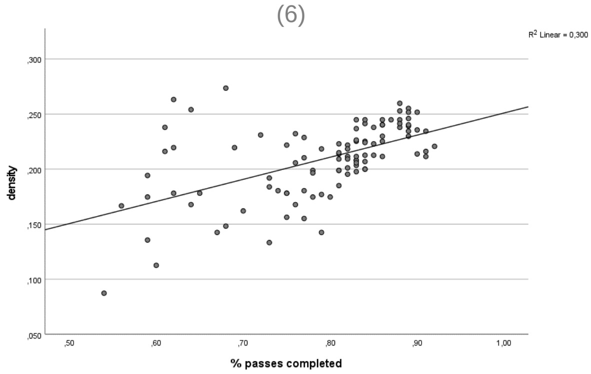
(5)



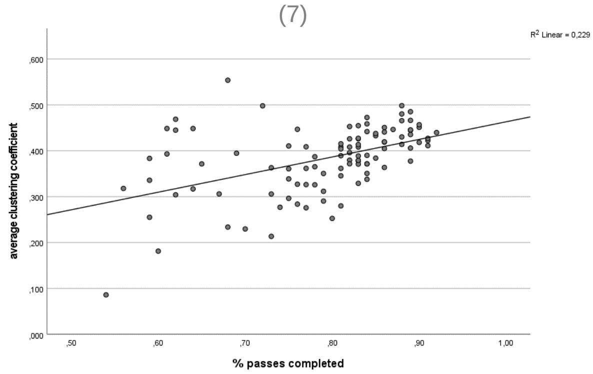
(6)



(7)



(8)



(9)

Figure 21: (1) Plot of the density vs the average clustering coefficient; (2) Plot of the parameter α vs the density; (3) Plot of the parameter α vs the average clustering coefficient; (4) Plot of the number passes vs the density; (5) Plot of the number passes α vs the average clustering coefficient; (6) Plot of the number passes completed vs the density; (7) Plot of the number passes completed vs the average clustering coefficient; (8) Plot of the percentage passes vs the density; (9) Plot of the number passes vs the average clustering coefficient;

As a result, Pearson Product-Moment was applied to investigate the relationships between the networks' macro-properties and α. Simultaneously, Spearman's Rank Order Correlation was used to study the relationship between the networks' macro-properties and pass statistics.

The Pearson r correlation coefficients between each pair of networks' macro properties and α are exhibited in Table 23. The average clustering coefficient showed, on the one hand, a very large positive correlation with the density ($r = 0.894,\ n = 102,\ p < 0.01$) and, on the other hand, a very large negative correlation with the parameter α ($r = -0.781,\ n = 102,\ p < 0.01$). In addition, there was a very large negative correlation between the parameter α and the density ($r = -0.811,\ n = 102,\ p < 0.01$). So, low values of the parameter α were associated with high values of these networks' macro properties.

Table 23: Pearson Product-Moment Correlation values between the density, the average clustering coefficient and parameter α

Pearson Product-Moment Correlations

Measures	1	2
(1) Density		
(2) Average clustering coefficient	0.894 **	
(3) α	-0.811 **	-0.781 **

N=102

** Correlation is significant at the 0.01 level

Table 24 exposes the Spearman ρ correlation coefficients between pass statistics and the networks' macro properties. Firstly, 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 the density ($\rho = 0.917,\ n = 102,\ p < 0.01$) and a very large positive correlation with the average clustering coefficient ($\rho = 0.825,\ n = 102,\ p < 0.01$). 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 ($\rho = 0.521, n = 102,\ p < 0.01$). Thus, high values of the pass statistics were associated with high values of the network's macro properties.

Table 24: Spearman Rank's Order Correlation values between pass statistics and the density and the average clustering coefficient, respectively.

Spearman's Rank Order Correlations

Measures	1 (Density)	2 (Avg. Clustering coefficient)
Nu passes	0.873 **	0.797 **
Nu passes completed	0.917 **	0.825 **
% passes completed	0.588 **	0.521 **

N=102

** Correlation is significant at the 0.01 level

Then, similarly to the procedure in Chapter 4, the analysis focused on the differences between teams that achieved different match results. As before, descriptive statistics were initially produced, and then the assumption of homogeneity of the variances was again tested using Levene's test with the same significance value. Table 25 demonstrates that this assumption was not violated, so the ANOVA test was executed.

Table 25: Descriptive table and statistical comparison between groups (match results), considering the density and the average clustering coefficient

Descriptive Statistics

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Density	Defeat	35	0.206	0.027	0.004	0.196	0.215	0.148	0.260
	Draw	32	0.205	0.044	0.008	0.189	0.221	0.087	0.274
	Victory	35	0.216	0.030	0.005	0.205	0.226	0.136	0.255
	Total	**102**	**0.209**	**0.034**	**0.003**	**0.202**	**0.216**	**0.087**	**0.274**
Average Clustering Coefficient	Defeat	35	0.376	0.059	0.010	0.355	0.396	0.234	0.466
	Draw	32	0.385	0.097	0.017	0.350	0.420	0.086	0.554
	Victory	35	0.388	0.065	0.011	0.365	0.410	0.230	0.480
	Total	**102**	**0.383**	**0.074**	**0.007**	**0.368**	**0.397**	**0.086**	**0.554**

After generating the descriptive statistics, Levene's test (Table 26) was again used to assess the assumption of homogeneity of the variances. This assumption was violated for the density since the significance value, Sig., was lower than 0.05. Consequently, on the one hand, the ANOVA test was performed for the average clustering coefficient. On the other hand, the Welsh and Brown-Forsythe tests were used for the density because they are preferable when this assumption is violated (Pallant, 2005).

Table 26: Test of Homogeneity of variances between groups (match results), considering the density and the average clustering coefficient.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Density	Based on Mean	4.660	2.000	99.000	0.012
Average clustering coefficient	Based on Mean	3.069	2.000	99.000	0.051

Firstly, a one-way between-groups analysis of variance (Table 27) was conducted to explore the impact of the average clustering coefficient on the match result. Like before, the samples were divided into three groups (Group 1: defeat; Group 2: draw; Group 3: victory), but there were no statistically significant differences at the $p < 0.05$ level. Secondly, the Welch and Brown-Forsythe tests (Table 28) investigated the impact of the density on the match result – however, no statistically significant differences at the $p < 0.05$ level were found as well.

Table 27: One-way between-groups analysis of variance (match results), considering the density and the average clustering coefficient.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Density	Between Groups	0.002	2.000	0.001	1.030	0.361
	Within Groups	0.115	99.000	0.001		
	Total	**0.117**	**101.000**			
Average Clustering Coefficient	Between Groups	0.003	2.000	0.001	0.246	0.782
	Within Groups	0.557	99.000	0.006		
	Total	**0.559**	**101.000**			

Table 28: Welch and Brown-Forsythe tests (match results), considering the density and the average clustering coefficient.

Robust Tests of Equality of Means

		Statistic	df1	df2	Sig.
Density	Welch	1.222	2.000	62.391	0.302
	Brown-Forsythe	1.001	2.000	77.746	0.372
Average Clustering Coefficient	Welch	0.344	2.000	62.400	0.711
	Brown-Forsythe	0.239	2.000	76.881	0.788

Alternatively, the differences between the networks' macro properties of teams that reached different stages of the tournament were studied. As before, descriptive statistics were initially made, as displayed in Table 29. Levene's test was used to evaluate the assumption of homogeneity of the variances and was again tested using the same significance value. Table 30 demonstrates that this assumption was not violated since the significance value, Sig., of both networks' macro properties was greater than 0.05. As a result, the analysis was accomplished using ANOVA.

Table 29: Descriptive table and statistical comparison between groups (stage reached in the tournament), considering the density and the average clustering coefficient

Descriptive Statistics

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Density	Final	14	0.220	0.025	0.007	0.205	0.234	0.167	0.240
	Semi-finals	12	0.230	0.031	0.009	0.211	0.250	0.175	0.274
	Quarter-finals	20	0.209	0.022	0.005	0.198	0.219	0.162	0.253
	Round of 16	32	0.212	0.037	0.007	0.199	0.226	0.087	0.260
	Group Stage	24	0.187	0.035	0.007	0.173	0.202	0.113	0.241
	Total	**102**	**0.209**	**0.034**	**0.003**	**0.202**	**0.216**	**0.087**	**0.274**
Average Clustering Coefficient	Final	14	0.409	0.055	0.015	0.378	0.441	0.317	0.498
	Semi-finals	12	0.415	0.069	0.020	0.371	0.459	0.276	0.554
	Quarter-finals	20	0.395	0.060	0.013	0.367	0.423	0.230	0.480
	Round of 16	32	0.388	0.081	0.014	0.359	0.417	0.086	0.485
	Group Stage	24	0.334	0.071	0.014	0.304	0.364	0.181	0.459
	Total	**102**	**0.383**	**0.074**	**0.007**	**0.368**	**0.397**	**0.086**	**0.554**

Table 30: Test of Homogeneity of variances between groups (stage reached in the tournament), considering the density and the average clustering coefficient.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Density	Based on Mean	1.404	4.000	97.000	0.238
Average clustering coefficient	Based on Mean	3.069	4.000	97.000	0.697

Table 31 shows the results of the one-way between-groups analysis of variance conducted to explore the impact of the percentage of passes and the parameter α on the stage reached in the tournament. 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). There were statistically significant differences at the $p < 0.05$ between the different groups (stage reached in the tournament) in 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).

Table 31: One-way between-groups analysis of variance (stage reached in the tournament), considering the density and the average clustering coefficient.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Density	Between Groups	0.019	4.000	0.005	4.648	0.002
	Within Groups	0.098	97.000	0.001		
	Total	**0.117**	**101.000**			
Average Clustering Coefficient	Between Groups	0.083	4.000	0.021	4.218	0.003
	Within Groups	0.476	97.000	0.005		
	Total	**0.559**	**101.000**			

The Tukey-Kramer modification of Tukey's HSD post-hoc test was executed since ANOVA detected statistical differences. First, Table 32 shows the post-hoc comparisons for the number of edges. The findings indicated that the mean number of passes for Group 5 (Group Stage) $[M = 162.880, SD = 30.115]$ was significantly different from Group 4 (Round of 16) $[M = 184.720, SD = 32.401]$, from Group 2 (Semi-finals) $[M = 200.420, SD = 26.623]$ and from Group 1 (Final) $[M = 191.140, SD = 21.947]$.

Table 32: Post-hoc test for the density

Multiple Comparisons

Density

Tukey HSD

(I) competition stage	(J) competition stage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Final	Quarter-finals	-0.011	0.013	0.914	-0.045	0.024
	Semi-finals	0.011	0.011	0.860	-0.020	0.042
	Round of 16	0.007	0.010	0.950	-0.021	0.036
	Group Stage	0.032 *	0.011	0.025	0.003	0.062
Semi-finals	Final	0.011	0.013	0.914	-0.024	0.045
	Quarter-finals	0.022	0.012	0.345	-0.011	0.054
	Round of 16	0.018	0.011	0.454	-0.012	0.048
	Group Stage	0.043 *	0.011	0.002	0.012	0.074
Quarter-finals	Final	-0.011	0.011	0.860	-0.042	0.020
	Semi-finals	-0.022	0.012	0.345	-0.054	0.011
	Round of 16	-0.004	0.009	0.995	-0.029	0.022
	Group Stage	0.022	0.010	0.176	-0.005	0.048
Round of 16	Final	-0.007	0.010	0.950	-0.036	0.021
	Semi-finals	-0.018	0.011	0.454	-0.048	0.012
	Quarter-finals	0.004	0.009	0.995	-0.022	0.029
	Group Stage	0.025 *	0.009	0.034	0.001	0.049
Group Stage	Final	-0.032 *	0.011	0.025	-0.062	-0.003
	Semi-finals	-0.043 *	0.011	0.002	-0.074	-0.012
	Quarter-finals	-0.022	0.010	0.176	-0.048	0.005
	Round of 16	-0.025 *	0.009	0.034	-0.049	-0.001

Second, the post-hoc comparisons, exhibited in Table 33, showed that the average clustering coefficient for Group 5 (Group Stage) $[M = 0.334, SD = 0.071]$ was significantly different from all the other groups, i.e., from Group 4 (Round of 16) $[M = 0.388, SD = 0.081]$, from Group 3 (Quarter-finals) $[M = 0.395, SD = 0.060]$, from Group 2 (Semi-finals) $[M = 0.415, SD = 0.069]$ and from Group 1 (Final) $[M = 0.409, SD = 0.055]$.

Table 33: Post-hoc test for the clustering coefficient

Multiple Comparisons

Average clustering coeffiicient

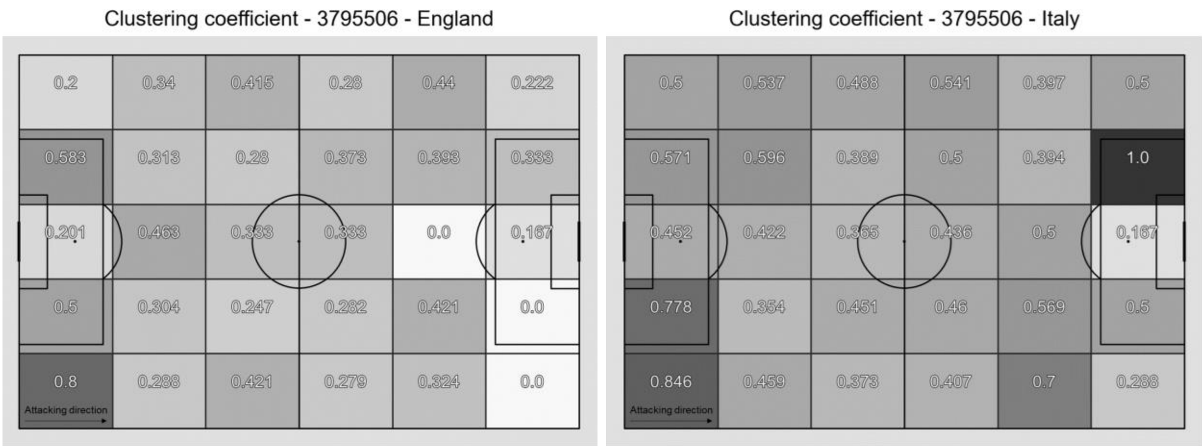
Tukey HSD

(I) competition stage	(J) competition stage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Final	Quarter-finals	-0.005	0.028	1.000	-0.082	0.072
	Semi-finals	0.014	0.024	0.976	-0.054	0.082
	Round of 16	0.022	0.022	0.869	-0.041	0.084
	Group Stage	0.075 *	0.024	0.016	0.010	0.141
Semi-finals	Final	0.005	0.028	1.000	-0.072	0.082
	Quarter-finals	0.019	0.026	0.941	-0.052	0.091
	Round of 16	0.027	0.024	0.789	-0.039	0.093
	Group Stage	0.081 *	0.025	0.013	0.012	0.149
Quarter-finals	Final	-0.014	0.024	0.976	-0.082	0.054
	Semi-finals	-0.019	0.026	0.941	-0.091	0.052
	Round of 16	0.007	0.020	0.996	-0.048	0.063
	Group Stage	0.061 *	0.021	0.039	0.002	0.120
Round of 16	Final	-0.022	0.022	0.869	-0.084	0.041
	Semi-finals	-0.027	0.024	0.789	-0.093	0.039
	Quarter-finals	-0.007	0.020	0.996	-0.063	0.048
	Group Stage	0.054 *	0.019	0.043	0.001	0.106
Group Stage	Final	-0.075 *	0.024	0.016	-0.141	-0.010
	Semi-finals	-0.081 *	0.025	0.013	-0.149	-0.012
	Quarter-finals	-0.061 *	0.021	0.039	-0.120	-0.002
	Round of 16	-0.054 *	0.019	0.043	-0.106	-0.001

5.2.2. Clustering analysis

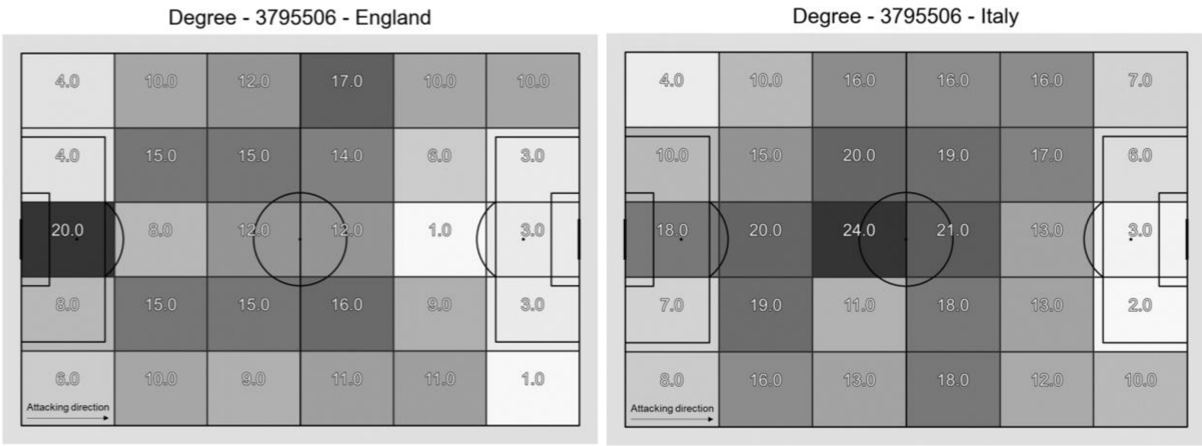
5.2.2.1. Clustering analysis on the zone passing networks

Initially, the local clustering coefficient and degree were computed for all zone networks, as exemplified in Figures 22 and 23. Consequently, two clustering analyses were performed using the 30 nodes' clustering coefficient and degree as clustering variables.



(1) (2)

Figure 22: Local clustering coefficient of each node in the zone network of size 30 of (1) England and (2) Italy in the tournament's final match.



(b1) (b2)

Figure 23: (1) Degree of each node in the zone network of size 30 of (1) England and (2) Italy in the tournament's final match

Figures 24 and 25 show the selection process of the number of PC using the PCA technique for the clustering analysis using the local clustering coefficient and the degree, respectively. The criterion that states to include all those PCs up to the predetermined 90% total percentage variance explained was applied. On the one hand, 13 PCs were selected for the clustering analysis using the local clustering coefficient, and the two first PC explains 37.42% of the variance in this instance. On the other hand, for the clustering analysis using degree, 17 PCs were chosen. In this case, the two first PC explains 51.75% of the variance (Holland, 2019).

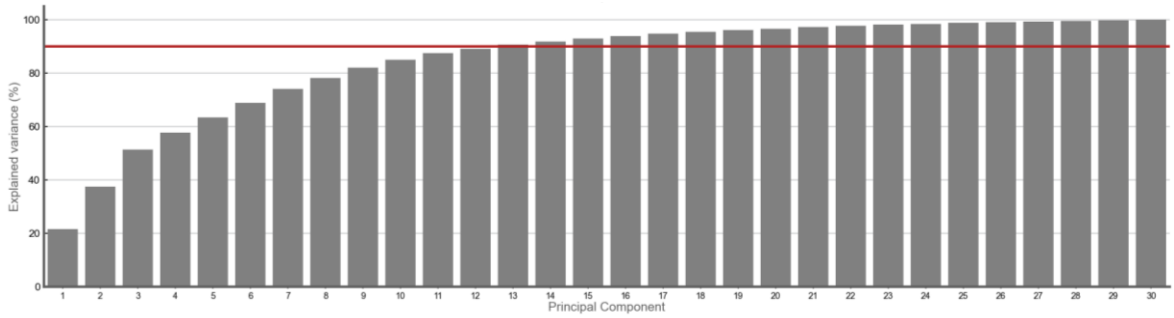


Figure 24: Cumulative explained variance by components for the clustering analysis on the zone passing networks using the local clustering coefficient

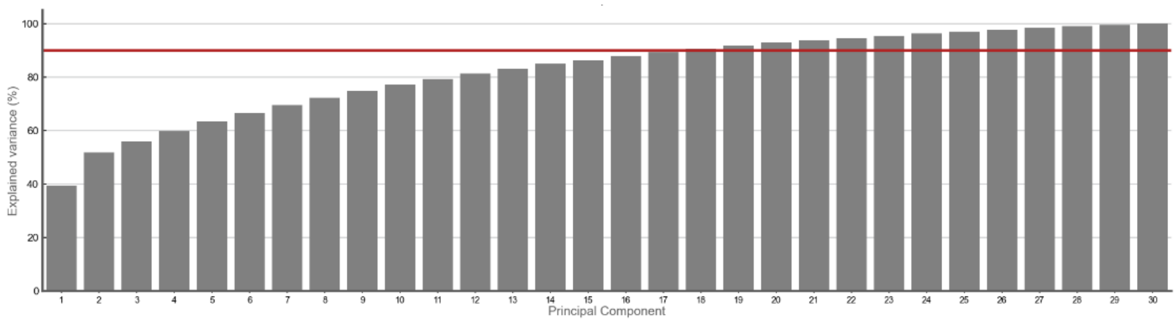
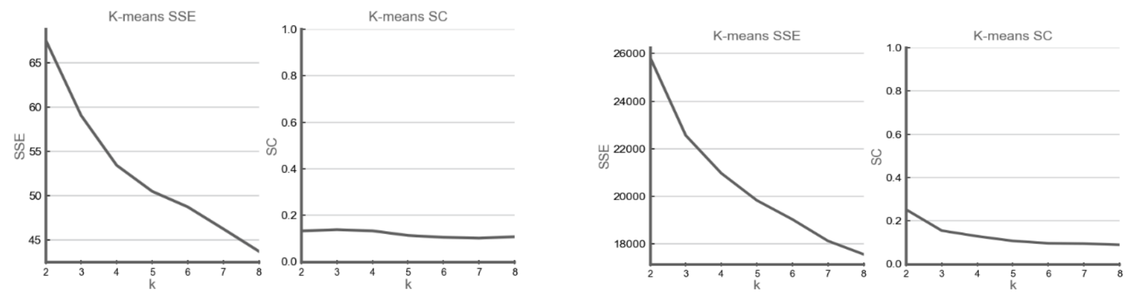


Figure 25: Cumulative explained variance by components for the clustering analysis on the zone passing networks using the degree

The choice of the number of clusters is one of the hardest decisions. Indeed, this decision was made by the joint use of the elbow method and the silhouette analysis. Figure 26.1. shows that there is no evident elbow in both k-SSE plots. Thus, the decision was made by mainly looking at the silhouette coefficient. On the one hand, $k = 3$ was selected for the clustering analysis using the clustering coefficient since $k = 3$ had the highest silhouette coefficient ($SC_{k=3} = 0.138$), as seen in the k-SC plot in Figure 26.2. On the other hand, the k-SC plot for the clustering analysis using the degree reveals that $k = 2$ ($SC_{k=2} = 0.252$) had the highest silhouette coefficient, followed by $k = 3$ ($SC_{k=3} = 0.133$). Therefore, as $k = 3$ has a lower SSE than $k = 2$, $k = 3$ was chosen as the number of clusters to use in the clustering analysis using the degree. Figures 27 and 28 show the silhouette analysis and the visualisation of the clustered data for the clustering analysis using the clustering coefficient and the degree. Note that the clustering analysis results must be interpreted with caution since there were some objects with a silhouette width near 0, which means that they should have been assigned to their or another cluster.



(1) (2)

Figure 26:(1) k-SSE and k-SC plots using the for the clustering analysis on the zone passing networks clustering coefficient. (2) k -SSE and k-SC plots for the clustering analysis on the zone passing networks using the degree.

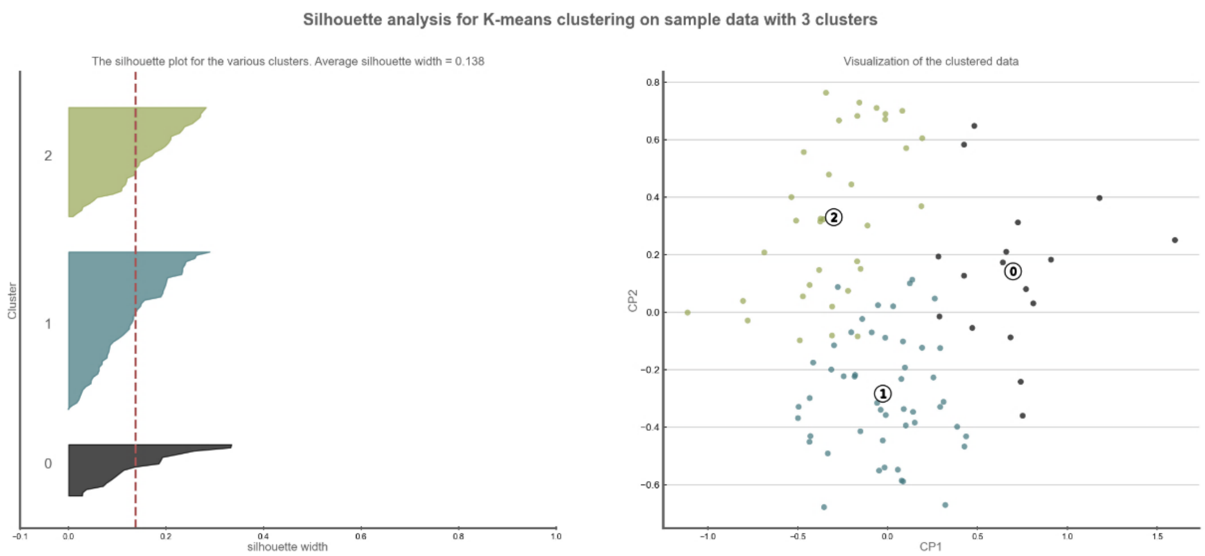


Figure 27: Silhouette analysis and visualisation of the clustered data for the clustering analysis with 3 clusters on the zone passing networks using the clustering coefficient.

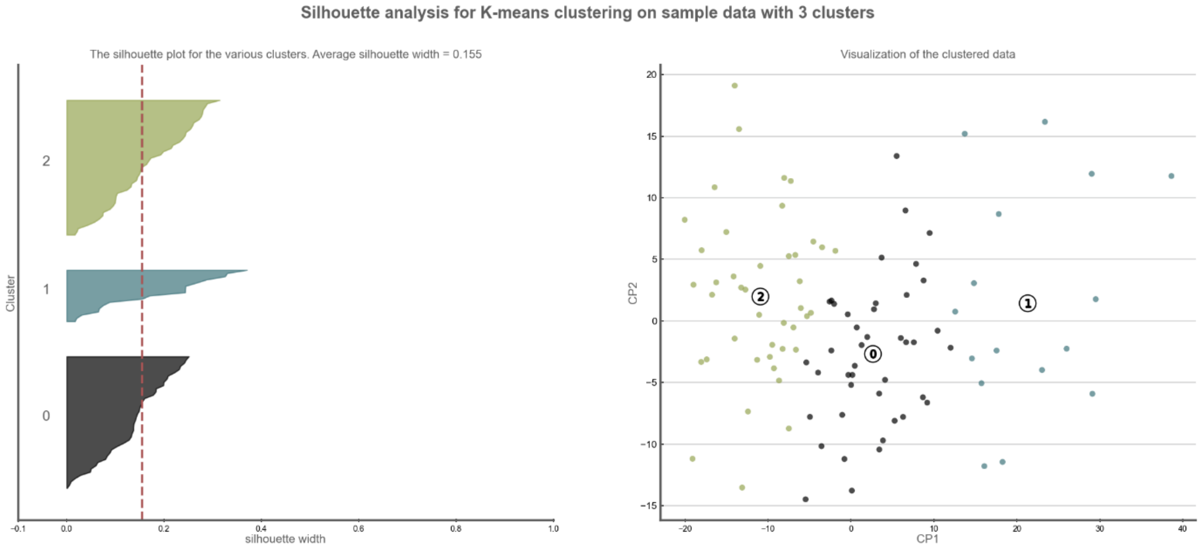


Figure 28: Silhouette analysis and visualisation of the clustered data for the clustering analysis with 3 clusters on the zone passing networks using the degree.

Tables 34.1 and 34.2 show the assignment of each network to each cluster, grouped by team, in the clustering analysis using the clustering coefficient and the degree, respectively. A more detailed analysis is presented in Appendix G. In the clustering analysis using the clustering coefficient, Cluster 0 has 17 objects, whereas Cluster 1 has 50 objects and Cluster 2 has 35 objects. Cluster 0 is composed of networks with lower values of the local clustering coefficient across the 30 zones of the playing field. The zone in which networks have, although low, higher values is the one formed by the second half of the defensive and the first half of midfield sectors $Z_{D-M} = \{2, 3, 8, 9, 14, 15, 20, 21, 26, 27\}$, as can be seen in Figure 29. The local clustering coefficients become lower as it goes further into the field, and so, the offensive sector, formed by $Z_O = \{5, 6, 11, 12, 17, 18, 23, 24, 29,\ 30\}$, is the one that presents the lower values, suggesting not only the teams' difficulty in progressing through the playing field but also the teams' incapacity of playing near the opposing penalty area. Most objects that integrate cluster 0 are networks from national teams that only reached the Group Stage of the competition, which are the case of Finland, Hungary, North Macedonia, Poland, Russia, Scotland, Slovakia and Turkey. In addition, networks from the Czech Republic, Denmark, England, France, and Wales are part of this cluster. Cluster 1 and Cluster 2 are constituted by networks with higher clustering coefficient values across the 30 zones of the playing field compared to the networks in Cluster 0. The central aspect that distinguishes the networks of these two clusters is a tendency to have a higher value of cluster coefficient in the opposing penalty box, $Z_{PB} = \{12, 18, 24\}$, as seen in Figure 29.

In the clustering analysis using the degree, Cluster 0 has 42 objects, while Cluster 1 has 17 objects and Cluster 2 has 43 objects. In all clusters, the second half of the defensive sector and the midfield sector tended to have higher values of degree. In contrast, the outer corridors tended to have lower values of degree. Compared with clusters 0 and 2, cluster 1 had networks with lower values of degree throughout the 30 zones of the playing field. However, these differences were most visible in the second half of the midfield sector and in the offensive sector. Additionally, Cluster 0 differentiates from Cluster

2 by having lower values in the central corridors of the second half of the midfield sector and of the offensive sector, $Z_{CD-CO} = \{9, 10, 11, 15, 16, 17, 21, 22, 23\}$.

Table 34: Assignment of each network to each cluster, grouped by team, in the clustering analysis on the zone passing networks using the (1) clustering coefficient and (2) the degree.

(1)

National Team	Cluster 0	Cluster 1	Cluster 2	Total
Austria		1	3	4
Belgium		4	1	5
Croatia		2	2	4
Czech Republic	1	3	1	5
Denmark	1	3	2	6
England	1	5	1	7
Finland	1	2		3
France	1	2	1	4
Germany		1	3	4
Hungary	1	2		3
Italy		2	5	7
Netherlands		2	2	4
North Macedonia	1		2	3
Poland	2		1	3
Portugal		4		4
Russia	2		1	3
Scotland	1	2		3
Slovakia	1	1	1	3
Spain		2	4	6
Sweden	2		2	4
Switzerland		4	1	5
Turkey	1	2		3
Ukraine		3	2	5
Wales	1	3		4
Total	**17**	**50**	**35**	**102**

(2)

National Team	Cluster 0	Cluster 1	Cluster 2	Total
Austria	1		3	4
Belgium	3		2	5
Croatia	3		1	4
Czech Republic	4	1		5
Denmark	3		3	6
England	4	1	2	7
Finland	2	1		3
France	2		2	4
Germany	1		3	4
Hungary		3		3
Italy		1	6	7
Netherlands	2		2	4
North Macedonia	2	1		3
Poland		1	2	3
Portugal	2		2	4
Russia	1	1	1	3
Scotland	2		1	3
Slovakia	1	1	1	3
Spain			6	6
Sweden		2	2	4
Switzerland	2	1	2	5
Turkey	1	1	1	3
Ukraine	4		1	5
Wales	2	2		4
Total	**42**	**17**	**43**	**102**

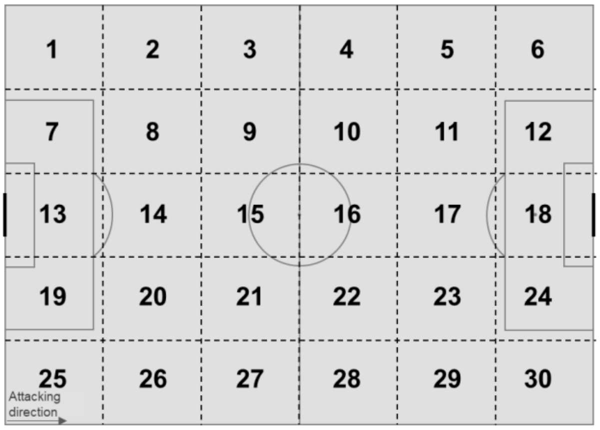


Figure 29: Field of play's division into 30 zones (6 sectors x 5 corridors).

5.2.2.1. Clustering analysis on the playing position-zone passing networks

By applying the methodology described in section 5.1.2., 1463 playing position-zone passing networks were the subject of the clustering analysis. As shown in Figure 30, nine systems of play were identified in this set of networks, namely the 4-3-3; 4-2-3-1; 3-5-2; 3-4-2-1; 3-4-1-2; 4-1-4-1; 3-4-3; 4-4-2 and 4-2-2-2. Firstly, the local clustering coefficient and degree were computed for all networks. Therefore, the combination of the six common positions and 30 zones was used as clustering variables.

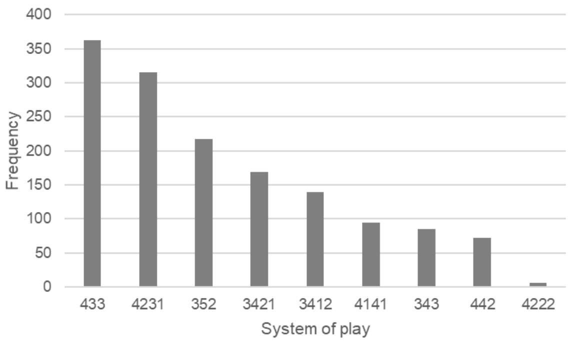
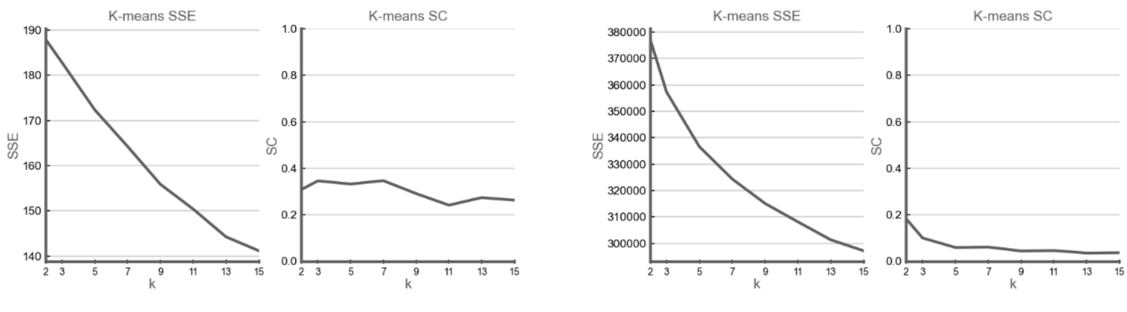


Figure 30: Frequency of each system of play in the 1463 playing position-zone passing networks.

Then, the PCA was executed to reduce the dimensionality of the clustering variables. Again, the criterion that states to include all those PCs up to the predetermined 90% total percentage variance explained was applied. On the one hand, for the clustering analysis using the clustering coefficient, 69 PCs were selected. In this case, the two first PC explains 8.28% of the variance (Holland, 2019). On the other hand, 59 PCs were chosen for the clustering analysis using the degree, and the two first PC explains 26.40% of the variance in this instance.

Similarly, the number of clusters was selected using the elbow method and the silhouette analysis. Figure 31 shows that there is no evident elbow in both k-SSE plots. On the one hand, for the clustering analysis using the clustering coefficient, since $k = 7$ had the highest silhouette coefficient ($SC_{k=7} = 0.347$) $k = 7$ was selected, as can be seen in the k-SC plot in Figure X. On the other hand, the k-SC plot for the clustering analysis using the degree reveals that $k = 2$ ($SC_{k=2} = 0.183$) had the highest silhouette coefficient, followed by $k = 3$ ($SC_{k=3} = 0.100$) and by $k = 7$ ($SC_{k=7} = 0.061$). However, as the main objective was to study the differences and similarities between the nine systems of play, $k = 7$ was chosen as the number of clusters to use in the clustering analysis using the degree. Once more, the clustering analysis results must be interpreted with caution since some objects with a silhouette width have a negative value, which means that they should have been to the wrong cluster.



(1) (2)

Figure 31: (1) k-SSE and k-SC plots using the for the clustering analysis on the playing position-zone passing networks clustering coefficient. (2) k -SSE and k-SC plots for the clustering analysis on the playing position-zone passing networks using the degree.

Figure 32 shows the clustering analysis results using the clustering coefficient. Cluster 0 is constituted of 1354 objects, whereas Cluster 1 and 4 both have 16 objects; Cluster 2 is composed of 33 objects; Cluster 3 is made up of 4 objects; Cluster 5 consists of 8 objects and Cluster 6 has 32 objects. Therefore, since most objects were assigned to the same cluster, it is possible to conclude that the clustering coefficient was a poor feature in partitioning the different objects. In addition, using the clustering analysis, it was impossible to verify the differences and similarities between the different systems of play.

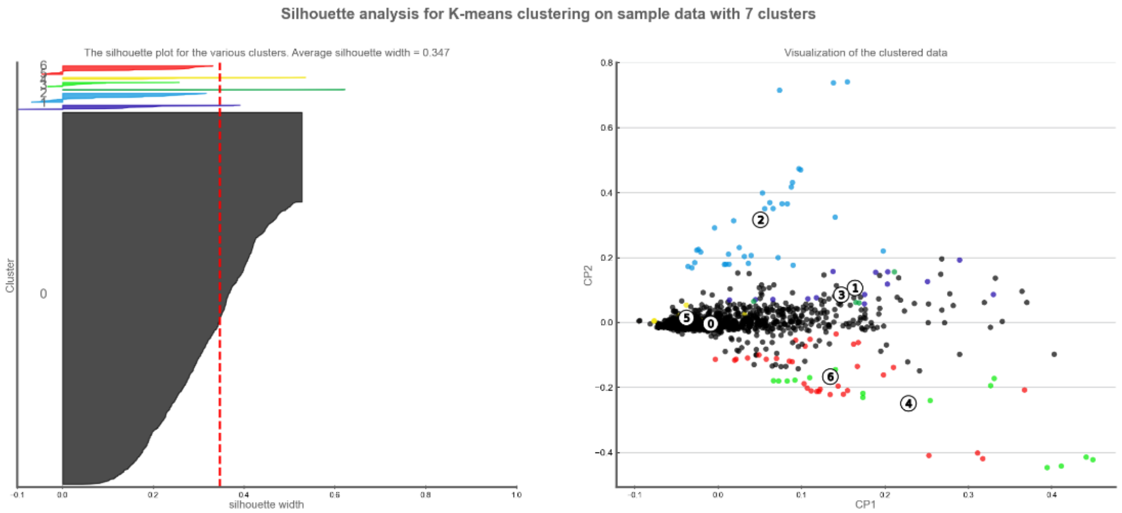


Figure 32: Silhouette analysis and visualisation of the clustered data for the clustering analysis with 7 clusters on the playing position-zone passing networks using the clustering coefficient.

Moreover, there were 617 with a silhouette width lower than the average, of which 28 had negative silhouette width. However, considering this fact, a prudent description of the objects in the different clusters is still possible. First, the Goalkeeper has a positive cluster coefficient within his penalty area, particularly zone 13, and also in the inner corridors of the second half of the defensive sector, $Z_{ID2} = \{8, 14, 20\}$. In region 13, the Goalkeepers in networks belonging to Cluster 2 have a higher cluster coefficient, whereas those in Cluster 4 have a lower cluster coefficient. Alternatively, the Goalkeepers in networks of Cluster 5 have a higher clustering coefficient in zone 7. Second, the Central Defenders have a higher clustering coefficient in the inner corridors of the defensive sector and the first half of the midfield sector, $Z_{ID-IM1} = \{8, 9, 14, 15, 20, 21\}$. However, Central Defenders in the networks of Cluster 3 have a higher clustering coefficient in the inner corridors of the second half of the midfield sector $Z_{IM2} = \{10, 16, 22\}$. Specifically, 3 of these networks are consecutive sliding windows, between minutes 50 and 75, from Portugal in the Group Stage match against Hungary. Thus, this can mean that, during this period of the game, Portugal played mainly in the opponent's half. Third, the External Defenders have a higher clustering coefficient in the outer corridors of the midfield sector and the first half of the offensive sector, $Z_{OM-IO1} = \{3, 4, 5, 27, 28, 29\}$. Fourth, the Central Midfielders present a higher cluster coefficient in the midfield sector, greater in the interior corridors than in the exterior corridors. Fifth, the External Midfielders have a higher clustering coefficient in the outer corridors of the midfield sector and the first half of the offensive sector, $Z_{OM-IO1} = \{3, 4, 5, 27, 28, 29\}$. However, there are networks in Clusters 1 and 4 which also have high values of the clustering coefficient in the inner corridors of the midfield and offensive sectors, $Z_{OM-IO1} = \{10, 11, 22, 23\}$. This indicates a tendency to play inside with the External Midfielders. Finally, the Forwards have a higher clustering coefficient in the inner corridors of the second half of the midfield sector and the offensive sector, $Z_{IM2-IO} = \{10, 11, 15, 16, 22, 23\}$.

Figure 33 shows the clustering analysis results using the degree. Cluster 0 has 136 objects, whereas Cluster 1 is composed of 316 objects; Cluster 2 is constituted of 107 objects; Cluster 3 consists of 129 objects; Cluster 4 is made up of 126 objects; Cluster 5 is formed of 258 objects, and Cluster 6 has 391 objects. Cluster 6 is the only cluster in which all objects have a positive silhouette width. Indeed, 973 objects have a silhouette width lower than the average, of which 673 have a negative value. Considering this fact and the close-to-zero value of the silhouette coefficient, the results were not interpreted.

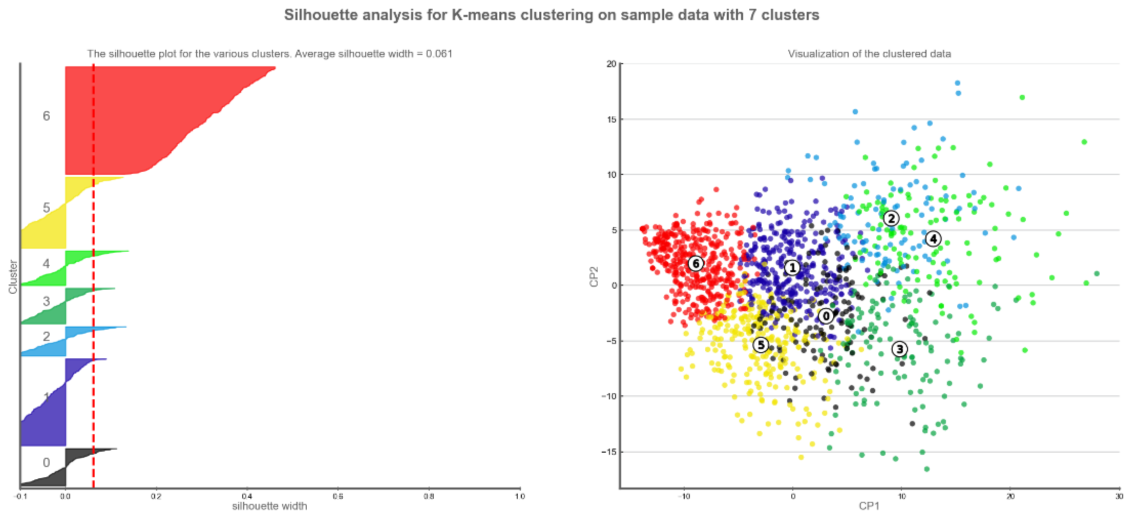


Figure 33: Silhouette analysis and visualisation of the clustered data for the clustering analysis with 7 clusters on the playing position-zone passing networks using degree.

5.3. Discussion

This section discusses the results of zone passing networks analysis and the clustering analyses.

5.3.1. Zone passing networks analysis

Network macro characteristics, such as the density and the average clustering coefficient, have been reported to be valuable descriptors of how football teams play. Also, they can be associated with performance variables, such as the match results achieved and the competition stage reached by teams (Pina *et al.*, 2017). Indeed, few research works have evidenced how passing network characteristics influence the overall performance of a team (Clemente *et al.*, 2015; Passos *et al.*, 2011). Consequently, the work developed in section 5.2.1 extended the work of section 4.2.1 by relating the network's density and average clustering coefficient with, firstly, the general strategy of play, described by α, and the pass statistics and, secondly, with overall team performance variables.

On the one hand, the density showed a very large positive correlation with the number of passes, a nearly perfect correlation with the number of passes completed and a large correlation with the percentage of passes completed. 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 pass is the link between nodes in these networks. Thus, these findings revealed how teams that performed more passes and more successfully originated denser networks. Moreover, the findings revealed that teams that adopt a more possessive type of play, characterised by higher values of α, also generate denser networks. This fact is demonstrated by the very large negative correlation between the density and the parameter α.

On the other hand, the average clustering showed a very large positive correlation with the number of passes and the number of passes completed and a large correlation with the percentage of passes completed. In the same way, teams performing more passes and more successfully have a higher probability of forming triangles around the zones of the playing field (Herrera-Diestra *et al.*, 2020). Furthermore, the findings also revealed that teams characterised by a possessive type of play have higher values of the average clustering coefficient, indicated by the very large negative correlation between the average clustering coefficient and the parameter α. This means that, specifically in zone networks, teams with a possession-based strategy reach the ball to all playing field areas and better connect the field through the network of passes (Herrera-Diestra *et al.*, 2020). Consequently, these results align with the findings of Buldú *et al.* (2019), which revealed that increasing the number of passes improved the passing networks' characteristics.

Then, the work investigated how the networks' macro properties (density and the average clustering coefficient) are related to the overall performance variables (match result and stage reached in the tournament). As a result, no statistical differences were observed between teams that achieved different match results (defeat, draw or victory) regarding the density and the average clustering coefficient. This result did not corroborate the research work of Clemente *et al.* (2015), which found differences in network density between teams that achieved different match results. This disparity, however, could be explained by the differences in the type of networks, competition, and the number of teams studied. Conversely, statistical differences were found between the stage reached in the competition and the networks' macro properties. The results revealed that achieved higher stages of the tournament, namely the Round of 16, Semi-finals and Final, were significantly different from the teams that were eliminated in the first stage of the tournament (Group Stage) concerning the density. In addition, regarding the average clustering coefficient, teams that were eliminated in the Group Stage were significantly different from all the other teams. These findings are consistent with the conclusions of Grund (2012), Clemente *et al.* (2015) and Gonçalves *et al.* (2017), who found that successful teams are associated with high levels and distribution of interactions. Clemente *et al.* (2015) also concluded that high cooperation and interconnectivity could lead to better performance outcomes, as also suggested by the previous results.

5.3.2. Clustering analysis

Section 5.2.2 aimed to study how the systems of play affect networks' characteristics, more specifically, whether the same systems of play generate similar networks. Thus, to accomplish this objective, a clustering analysis was performed using, one the one hand, the local clustering coefficient and, on the other hand, the degree. These two metrics were chosen because they have been reported to characterize well how teams play (Pina *et al.*, 2017). First and foremost, a preliminary clustering analysis was conducted on the 102 zone networks generated in section 5.2.1 to determine if any general differences were observed.

Using the nodes' clustering coefficient as the clustering variables, the networks were divided into 3 clusters. One of the clusters was composed of networks with lower values of the local clustering coefficient across the 30 zones of the field of play. In their respective matches, these teams have a lower probability of forming triangles around the zones of the playing field. However, although low, the higher values were present in the defensive sector's second half and the offensive sector's first half, suggesting the zone in which these teams mainly exchange the ball. Additionally, the local clustering coefficient becomes lower in the offensive sector. Alternatively, the two other clusters present higher values of the local clustering coefficient across the 30 zones of the field of play. The main difference is that one of these clusters has a higher local clustering coefficient within the penalty area, demonstrating how some teams in certain matches can form triangles in the offensive sector's zones and better connect the zones near the opposing goal through the network of passes.

Using the nodes' degree as the clustering variables, the networks were also divided into 3 clusters. The second half of the defensive sector and the midfield sector tended to have higher degree values across all clusters. In opposition, the outer corridors tended to have lower degree values. This result can be explained by the fact that the midfield sector and inner corridors act as bridges to the other sectors and the outer corridors, respectively. Furthermore, one of the clusters was characterised by having networks with lower degree values throughout the 30 zones of the playing field, being these differences most evident in the second half of the midfield sector and the offensive sector, demonstrating the inability to exchange the ball on the opposing half of the playing field.

The cluster analysis on the player position-zone networks did not have the expected results, so it was impossible to capture the differences and similarities between the different systems of play. On the one hand, in the clustering analysis using the clustering coefficient, most objects were assigned to the same cluster. On the other hand, the clustering analysis using the degree had poor performance, as described by the low values of the silhouette coefficient for the different values of the number of clusters. Therefore, it is possible to conclude that the proposed methodology was inappropriate for examining the differences and similarities between the different systems of play was again impossible. Nevertheless, through the clustering analysis using the clustering coefficient, it was possible to observe with caution which zones of the playing field each common position tends to form triangles.

First, the Goalkeeper has a positive clustering coefficient within his penalty area. Second, the Central Defenders have a higher clustering coefficient in the inner corridors of the defensive sector and the first half of the midfield sector. However, in some networks, Central Defenders also have a higher clustering coefficient in the inner corridors of the second half of the midfield sector. Third, the External Defenders have a higher clustering coefficient in the outer corridors of the midfield sector and the first half of the offensive sector. Forth, the Central Midfielders present a higher cluster coefficient in the midfield sector, greater in the interior corridors than in the exterior corridors. Fifth, the External Midfielders had a higher clustering coefficient in the outer corridors of the midfield sector and the first half of the offensive sector. However, in some networks, this common position also has high clustering coefficient values in the inner corridors of the midfield and offensive sectors, indicating a tendency to play inside. Finally, the Forwards have a higher clustering coefficient in the inner corridors of the second half of the midfield sector and the offensive sector.

Chapter 6 – Conclusions, Limitations and Future Work

This chapter summarises this dissertation's conclusions and insights in section 6.1, presenting the main limitations and highlighting opportunities for future work in section 6.2.

6.1. Conclusions

Football has become more professionalised, and match analysis demands have grown. Nowadays, to improve their teams' play and identify weaknesses in the opposition, coaches seek to extract information and produce knowledge about both performances of their team and the opponents. Additionally, the coaches want to know how they can lead their teams to success. As a result, the application of graph theory and network science has emerged in football analysis, providing valuable tools for describing teams' interactive behaviour, organisation and performance that classical analysis based on the performance of individual players does not capture.

This dissertation sought to answer several research questions to validate and extend the literature on passing sequences and network analysis. First, this work studied whether the distribution of successful passes tends to follow a power law distribution and how can this distribution of successful passes explain the general attacking strategy (direct or possessive play). Thus, it was found that approximately 70% of the samples were consistent with the power-law hypothesis. Furthermore, this dissertation proposed to describe the general attacking strategy of football teams through the power law exponent, $-\alpha$. 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.

Second, this work examined how the general attacking strategy and network characteristics relate to each other and how they impact the match result achieved and the stage reached in the tournament. Through statistical studies, the outcomes 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. Moreover, the findings indicated that unsuccessful teams that were eliminated in the first stage of the tournament have higher values of α and lower values of the number of passes, the number of passes completed, percentage of passes completed, density and average clustering coefficient. This suggested that teams embracing direct play are less successful. In addition, the outcomes indicated that, nowadays, teams score more goals from longer passing sequences, demonstrating how football has become more organised, being necessary to exchange the ball more to score goals.

Finally, this work could not unveil how the systems of play affect the characteristics of the networks. Indeed with the clustering analysis approach, it was impossible to reveal the differences and similarities between the different systems of play. Given the importance of the systems of play for player interactions within the team, this issue, which has received little attention in the literature, needs to be continually explored.

Although not all the questions were answered, this dissertation enhances that graph theory and network science are valuable for football analysis by providing relevant insights that can aid coaches.

Therefore, the use of these approaches in the football analysis departments is recommended to extract deeper information about the team at collective and individual levels.

6.2. Limitations and Future Work

This dissertation faces some limitations that should be addressed. First, as this dissertation 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. Second, although 30% of the samples were not consistent with the power-law hypothesis, all the samples were fitted to the power-law distribution. Third, the clustering analysis should have been reproduced for the different combinations of the divisions of the field of play, for different sliding window sizes, and using different clustering methods to make possible a comparative study between them.

As a result, this dissertation contributes to future research with proposals that complement the present work and the literature in general. First, further investigations should replicate this dissertation 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. Another clustering method should be experienced to verify if it 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. This study begins to be possible with the introduction of technology within the ball that allows the collection of the ball's tracking data.

References

Alstott, J., Bullmore, E., & Plenz, D. (2014). Powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0085777

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

Barabási, A. L. (2016). *Network Science*. Cambridge University Press, Cambridge.

Bate, R. (1988). *Football chance: tactics and strategy.* (Routledge, Ed.).

Beauchamp, M. A. (1965). (1965). An improved index of centrality. *Behavioral Science, 10(2), 161-163.*

Bekkers, J., & Dabadghao, S. (2019). Flow motifs in soccer: What can passing behavior tell us? *Journal of Sports Analytics*, *5*(4), 299–311. https://doi.org/10.3233/jsa-190290

Bellman, R. E. (1961). Adaptive Control Processes. *Princeton University Press, Princeton, NJ*.

Bradley, P., Bransen, L., Guerrero, I., Kempe, M., Lago, C., López-Felip, M. A., Pol, R., Shaw, L., & Hans, T. (2021). *Football Analytics 2021* (Barcelona Innovation Hub, Ed.).

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, *30*(2), 136–145. https://doi.org/10.1016/j.socnet.2007.11.001

Buldú, J. M., Busquets, J., Echegoyen, I., & Seirul.lo, F. (2019). Defining a historic football team: Using Network Science to analyze Guardiola's F.C. Barcelona. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-49969-2

Buldú, J. M., Busquets, J., Martínez, J. H., Herrera-Diestra, J. L., Echegoyen, I., Galeano, J., & Luque, J. (2018). Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game. In *Frontiers in Psychology* (Vol. 9, Issue OCT). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2018.01900

Caicedo-Parada, S., Lago-Peñas, C., & Ortega-Toro, E. (2020). Passing networks and tactical action in football: A systematic review. *International Journal of Environmental Research and Public Health*, *17*(18), 1–19. https://doi.org/10.3390/ijerph17186649

Cano, O. (2009). *El modelo de juego del FC Barcelona.* (MCSport, Ed.).

Casal, C. A., Anguera, M. T., Maneiro, R., & Losada, J. L. (2019). Possession in football: More than a quantitative aspect - A mixed method study. *Frontiers in Psychology*, *10*(MAR). https://doi.org/10.3389/fpsyg.2019.00501

Castellano, J. (2000). *Observación y análisis de la acción de juegoen el fútbol*. Universidad del País Vasco.

Chen, L. (2009). Curse of Dimensionality. *In: LIU, L., ÖZSU, M.T. (Eds) Encyclopedia of Database Systems. Springer, Boston, MA. Https://Doi.Org/10.1007/978-0-387-39940-9_133*.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. In *SIAM Review* (Vol. 51, Issue 4, pp. 661–703). https://doi.org/10.1137/070710111

Clemente, F. M., José, F., Oliveira, N., Martins, F. M. L., Mendes, R. S., Figueiredo, A. J., Wong, D. P., & Kalamaras, D. (2016). Network structure and centralization tendencies in professional football teams from Spanish La Liga and english premier leagues. *Journal of Human Sport and Exercise*, *11*(3), 376–389. https://doi.org/10.14198/jhse.2016.113.06

Clemente, F. M., Martins, F. M. L., Kalamaras, D., Wong, D. P., & Mendes, R. S. (2015). General network analysis of national soccer teams in Fifa World Cup 2014. *International Journal of Performance Analysis in Sport*, *15*(1), 80–96. https://doi.org/10.1080/24748668.2015.11868778

Clemente, F. M., Martins, F. M. L., & Mendes, R. S. (2016). Analysis of scored and conceded goals by a football team throughout a season: A network analysis. *Kinesiology*, *48*(1), 103–114. https://doi.org/10.26582/k.48.1.5

Clemente, F. M., Sarmento, H., & Aquino, R. (2020). Player position relationships with centrality in the passing network of world cup soccer teams: Win/loss match comparisons. *Chaos, Solitons and Fractals*, *133*. https://doi.org/10.1016/j.chaos.2020.109625

Clemente, F. M., Silva, F., Martins, F. M. L., Kalamaras, D., & Mendes, R. S. (2016). Performance Analysis Tool for network analysis on team sports: A case study of FIFA Soccer World Cup 2014. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *230*(3), 158–170. https://doi.org/10.1177/1754337115597335

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences Second Edition*.

Cotta, C., Mora, A. M., Merelo, J. J., & Merelo-Molina, C. (2013). A network analysis of the 2010 FIFA world cup champion team play. *Journal of Systems Science and Complexity*, *26*(1), 21–42. https://doi.org/10.1007/s11424-013-2291-2

de Amorim, R. C., & Hennig, C. (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences*, *324*, 126–145. https://doi.org/10.1016/j.ins.2015.06.039

Diquigiovanni, J., & Scarpa, B. (2019). Analysis of association football playing styles: An innovative method to cluster networks. *Statistical Modelling*, *19*(1), 28–54. https://doi.org/10.1177/1471082X18808628

Duarte, R., Araújo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. *Human Movement Science*, *31*(6), 1639–1651. https://doi.org/10.1016/j.humov.2012.03.001

Duch, J., Waitzman, J. S., & Nunes Amaral, L. A. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0010937

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis 5th Edition Cluster Analysis 5th Edition WILEY SERIES IN PROBABILITY AND STATISTICS Cluster Analysis 5th Edition*.

Fagiolo, G. (2007). *Clustering in Complex Directed Networks*.

Fernandez, J., & Bornn, L. (2018). 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*.

FIFA. (2015). *LAWS OF THE GAME 2015/2016*.

Freeman, L. C. (1978). Centrality in Social Networks Conceptual Clarification. In *Social Networks* (Vol. 1).

Gama, J., Couceiro, M., Dias, ; Gonçalo, & Vasco Vaz, ; (2015). SMALL-WORLD NETWORKS IN PROFESSIONAL FOOTBALL: CONCEPTUAL MODEL AND DATA. In *European Journal of Human Movement* (Vol. 35).

Gama, J., Dias, G., Couceiro, M., Belli, R., Vaz, V., Ribeiro, J., & Figueiredo, A. (2016). Networks and centroid metrics for understanding football. *South African Journal for Research in Sport, Physical Education and Recreation*, *38*(2).

Gama, J., Dias, G., Couceiro, M., Sousa, T., & Vaz, V. (2016). Networks metrics and ball possession in professional football. *Complexity*, *21*, 342–354. https://doi.org/10.1002/cplx.21813

Gama, J., Passos, P., Davids, K., Relvas, H., Ribeiro, J., Vaz, V., & Dias, G. (2014). Network analysis and intra-team activity in attacking phases of professional football. *International Journal of Performance Analysis in Sport*, *14*(3), 692–708. https://doi.org/10.1080/24748668.2014.11868752

Garganta, J. (1997). *Modelação táctica do jogo de Futebol Estudo da organização da fase ofensiva em equipas de alto rendimento*.

Golbeck, J. (2015). *Introduction to social media investigation: A hands-on approach.* Syngress.

Goldstein, J. (1999). Emergence as a Construct: History and Issues. *Emergence*, *1*(1), 49–72. https://doi.org/10.1207/s15327000em0101_4

Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004a). *Fitting to the Power-Law Distribution*.

Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004b). Problems with fitting to the power-law distribution. *European Physical Journal B*, *41*(2), 255–258. https://doi.org/10.1140/epjb/e2004-00316-5

Gonçalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jiménez, S., & Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. *PLoS ONE*, *12*(1). https://doi.org/10.1371/journal.pone.0171156

Gould, P., & Gatrell, A. (1979). A Structural Analysis of a Game: The Liverpool v Manchester United Cup Final of 1977. In *Social Networks* (Vol. 2).

Greco, P. J., & Greco, P. (2009). *Tactical Principles of Soccer: concepts and application Tactical Principles of Soccer*. https://doi.org/10.5016/2488

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, *34*(4), 682–690. https://doi.org/10.1016/j.socnet.2012.08.004

Gyarmati, L., Kwak, H., & Rodriguez, P. (2014). *Searching for a Unique Style in Soccer*. http://arxiv.org/abs/1409.0308

Han, J., & Kamber, M. (2006). *Data mining: Concepts and Techniques*. Morgan Kaufmann, 340, 94104-3205.

Hanseth, O., & Lyytinen, K. (2016). Design theory for dynamic complexity in information infrastructures: The case of building internet. In *Enacting Research Methods in Information Systems: Volume 3* (pp. 104–142). Springer International Publishing. https://doi.org/10.1007/978-3-319-29272-4_4

Hartigan, J. A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. In *Source: Journal of the Royal Statistical Society. Series C (Applied Statistics)* (Vol. 28, Issue 1).

Herrera-Diestra, J. L., Echegoyen, I., Martínez, J. H., Garrido, D., Busquets, J., Io, F. S., & Buldú, J. M. (2020). Pitch networks reveal organizational and spatial patterns of Guardiola's F.C. Barcelona. *Chaos, Solitons and Fractals*, *138*. https://doi.org/10.1016/j.chaos.2020.109934

Hewitt, A., Greenham, G., & Norton, K. (2016). Game style in soccer: What is it and can we quantify it? *International Journal of Performance Analysis in Sport*, *16*(1), 355–372. https://doi.org/10.1080/24748668.2016.11868892

Holland, S. M. (2019). *Principal Component Analysis (PCA)*.

Hook, C., & Hughes, M. D. (2001). Patterns of play leading to shots in 'Euro 2000.' *Pass.Com. Cardiff: UWIC, Pp. 295-302*.

Hughes, M. D., Robertson, K., & Nicholson, A. (1988). An analysis of the 1984 World Cup of Association Football. *In Science and Football*.

Hughes, M., & Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, *23*(5), 509–514. https://doi.org/10.1080/02640410410001716779

IFAB. (2021). *Laws of the Game 2021/22*.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jones, P. D., James, N., & Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport*, *4*(1), 98–102. https://doi.org/10.1080/24748668.2004.11868295

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data An Introduction to Cluster Analysis*.

Kempe, M., Vogelbein, M., Memmert, D., & Nopp, S. (2014). Possession vs. Direct Play: Evaluating Tactical Behavior in Elite Soccer. *International Journal of Sports Science*, *2014*(6A), 35–41. https://doi.org/10.5923/s.sports.201401.05

Korte, F., & Lames, M. (2019). Passing Network Analysis of Positional Attack Formations in Handball. *Journal of Human Kinetics*, *70*(1), 209–221. https://doi.org/10.2478/hukin-2019-0044

Korte, F., Lames, M., Link, D., & Groll, J. (2019). Play-by-play network analysis in football. *Frontiers in Psychology*, *10*(JULY). https://doi.org/10.3389/fpsyg.2019.01738

Lago-Peñas, C., & Dellal, A. (2010). Ball possession strategies in elite soccer according to the evolution of the match-score: The influence of situational variables. *Journal of Human Kinetics*, *25*(1), 93–100. https://doi.org/10.2478/v10078-010-0036-z

Li, X., Yu, L., Hang, L., & Tang, X. (2017). The parallel implementation and application of an improved k-means algorithm. *J. Univ. Electron. Sci. Technol. China, 46, 61-68*.

Malta, P., & Travassos, B. (2014). Caraterização da transição defesa-ataque de uma equipa de Futebol. *Motricidade*, *10*(1), 27–37. https://doi.org/10.6063/motricidade.10(1).1544

Martín, A. (2022). *Match Analysis: Final Assessement*. Barça Innovation Hub.

Martín-Barrero, A., & Ignacio Martínez-Cabrera, F. (2019). *El modelo de juego en el fútbol. De la concepción teórica al diseño práctico Game models in soccer. From theoretical conception to practical design*. www.retos.org

Martins, F. M. L., Clemente, F. M., & Couceiro, M. S. (2013). From the individual to the collective analysis at the football game. *Paper Presented at the Proceedings Mathematical Methods in Engineering International Conference, Porto*. https://doi.org/10.5890/JAND.2012.02.001

Mclean, S., Salmon, P. M., Gorman, A. D., Stevens, N. J., & Solomon, C. (2017). A social network analysis of the goal scoring passing networks of the 2016 European Football Championships. *Human Movement Science*, *57*, 400–408. https://doi.org/10.1016/j.humov.2017.10.001

McLean, S., Salmon, P. M., Gorman, A. D., Wickham, J., Berber, E., & Solomon, C. (2018). The effect of playing formation on the passing network characteristics of a professional football team. *Human Movement*, *2018*(5), 14–22. https://doi.org/10.5114/hm.2018.79416

Memmert, D., & Raabe, D. (2018). *Data analytics in football: Positional data collection, modelling and analysis. Routledge.*

Memmert, D., Raabe, D., Schwab, S., & Rein, R. (2019). 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*, *14*(1). https://doi.org/10.1371/journal.pone.0210191

Mendes, B., Clemente, F. M., & Maurício, N. (2018). Variance in Prominence Levels and in Patterns of Passing Sequences in Elite and Youth Soccer Players: A Network Approach. *Journal of Human Kinetics*, *61*(1), 141–153. https://doi.org/10.1515/hukin-2017-0117

Mercé, J. (2004). *Fútbol: el sistema 1.4.4.2.: fundamentos y enseñanza.* (Wanceulen, Ed.).

Milligan, G. W. (1996). *Clustering validation: results and implications for applied analyses*. www.worldscientific.com

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Montgomery, D. C., & Runger, G. C. (2003). *Applied statistics and probability for engineers*. John Wiley & Sons.

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

Narizuka, T., & Yamazaki, Y. (2019). Clustering algorithm for formations in football games. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-48623-1

Newman, M. (2010). *Networks: An Introduction* (Oxford Academic, Ed.).

Olsen, E., & Larsen, O. (1997). Use of match analysis by coaches. *Science and Football III, Pp. 209-220. London: E & FN SPON*.

Pallant, J. (2005). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows (Version 12)* (2nd Edition). Allen & Unwin. www.allenandunwin.com/spss.htm

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*(2), 170–176. https://doi.org/10.1016/j.jsams.2010.10.459

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine, 11, 559-572*.

Peña, J. L., & Navarro, R. S. (2015). *Who can replace Xavi? A passing motif analysis of football players*. http://arxiv.org/abs/1506.07768

Peña, J. L., & Touchette, H. (2012). *A network theory analysis of football strategies*. http://arxiv.org/abs/1206.6904

Pina, T. J., Paulo, A., & Araújo, D. (2017). Network characteristics of successful performance in association football. A study on the UEFA champions league. *Frontiers in Psychology*, *8*(JUL). https://doi.org/10.3389/fpsyg.2017.01173

Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. *Journal of the Royal Statistical Society Series D: The Statistician*, *46*(4), 541–550. https://doi.org/10.1111/1467-9884.00108

Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. In *Source: Journal of the Royal Statistical Society. Series A (General)* (Vol. 131, Issue 4).

Reep, C., Pollard, R., & Benjamin, B. (1971). Skill and chance inball games. *Journal of the Royal Statistical Society*, *134*, 623–329.

Ribeiro, J., Silva, P., Duarte, R., Davids, K., & Garganta, J. (2017). Team Sports Performance Analysed Through the Lens of Social Network Theory: Implications for Research and Practice. *Sports Medicine*, *47*(9), 1689–1696. https://doi.org/10.1007/s40279-017-0695-1

Roessner, U., Nahid, A., Chapman, B., Hunter, A., & Bellgard, M. (2011). Metabolomics - The Combination of Analytical Biochemistry, Biology, and Informatics. In *Comprehensive Biotechnology, Second Edition* (Vol. 1, pp. 447–459). Elsevier Inc. https://doi.org/10.1016/B978-0-08-088504-9.00052-0

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In *Journal of Computational and Applied Mathematics* (Vol. 20).

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika, 31(4), 581-603.*

Sarmento, H., Clemente, F. M., Araújo, D., Davids, K., McRobert, A., & Figueiredo, A. (2018). What Performance Analysts Need to Know About Research Trends in Association Football (2012–2016): A Systematic Review. *Sports Medicine*, *48*(4), 799–836. https://doi.org/10.1007/s40279-017-0836-6

Soriano, E. (2019). El Juego desde las Perspectiva de las Ventajas. In *IDE Universidad*.

Stone, L., Simberloff, D., & Artzy-Randrup, Y. (2019). Network motifs and their origins. *PLoS Computational Biology*, *15*(4), 1–7. https://doi.org/10.1371/journal.pcbi.1006749

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effect of playing tactics on goal scoring in norwegian professional soccer. *Journal of Sports Sciences*, *28*(3), 237–244. https://doi.org/10.1080/02640410903502774

Tenga, A., & Sigmundstad, E. (2011). Characteristics of goal-scoring possessions in open play: Comparing the top, in-between and bottom teams from professional soccer league. *International Journal of Performance Analysis in Sport*, *11*(3), 545–552. https://doi.org/10.1080/24748668.2011.11868572

UEFA. (2018). *Regulations of the UEFA European Football Championship 2018-20*.

Vales-Vásquez. (2012). *Fútbol: Del análisis del juego a la edición de informes técnicos* (MC Sports, Ed.).

Vilar, L., Araújo, D., Davids, K., & Bar-Yam, Y. (2013). Science of winning soccer: Emergent pattern-forming dynamics in association football. *Journal of Systems Science and Complexity*, *26*(1), 73–84. https://doi.org/10.1007/s11424-013-2286-z

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.*

Willy, C., Neugebauer, E. A. M., & Gerngroß, H. (2003). The concept of nonlinearity in complex systems: An additional approach to understand the pathophysiology of severe trauma and shock. In *European Journal of Trauma* (Vol. 29, Issue 1, pp. 11–22). https://doi.org/10.1007/s00068-003-1248-x

Yamamoto, Y., & Yokoyama, K. (2011). Common and unique network dynamics in football games. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0029638

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, *2*(2), 226–235. https://doi.org/10.3390/j2020016

Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.

Appendix

Appendix A – UEFA EURO 2020's matches

#	match_id	match_date	group*	competition_stage_name	home_team_home_team_name	away_team_away_team_name	home_score	away_score	stadium_name	stadium_country_name
1	3788741	11/06/2021	Group A	Group Stage	Turkey	Italy	0	3	Stadio Olimpico	Italy
2	3788742	12/06/2021	Group B	Group Stage	Denmark	Finland	0	1	Parken	Denmark
3	3788743	12/06/2021	Group B	Group Stage	Belgium	Russia	3	0	Saint-Petersburg Stadium	Russia
4	3788744	12/06/2021	Group A	Group Stage	Wales	Switzerland	1	1	Bakı Olimpiya Stadionu	Azerbaijan
5	3788745	13/06/2021	Group D	Group Stage	England	Croatia	1	0	Wembley Stadium	England
6	3788746	13/06/2021	Group C	Group Stage	Netherlands	Ukraine	3	2	Johan Cruijff Arena (Amsterdam)	Netherlands
7	3788747	13/06/2021	Group C	Group Stage	Austria	North Macedonia	3	1	Arena Naţională	Romania
8	3788749	14/06/2021	Group E	Group Stage	Poland	Slovakia	1	2	Saint-Petersburg Stadium	Russia
9	3788750	14/06/2021	Group E	Group Stage	Spain	Sweden	0	0	Estadio de La Cartuja	Spain
10	3788748	14/06/2021	Group D	Group Stage	Scotland	Czech Republic	0	2	Hampden Park	Scotland
11	3788751	15/06/2021	Group F	Group Stage	France	Germany	1	0	Allianz Arena	Germany
12	3788752	15/06/2021	Group F	Group Stage	Hungary	Portugal	0	3	Puskás Aréna	Hungary
13	3788753	16/06/2021	Group B	Group Stage	Finland	Russia	0	1	Saint-Petersburg Stadium	Russia
14	3788754	16/06/2021	Group A	Group Stage	Italy	Switzerland	3	0	Stadio Olimpico	Italy
15	3788755	16/06/2021	Group A	Group Stage	Turkey	Wales	0	2	Bakı Olimpiya Stadionu	Azerbaijan
16	3788758	17/06/2021	Group C	Group Stage	Ukraine	North Macedonia	2	1	Arena Naţională	Romania
17	3788757	17/06/2021	Group B	Group Stage	Denmark	Belgium	1	2	Parken	Denmark
18	3788756	17/06/2021	Group C	Group Stage	Netherlands	Austria	2	0	Johan Cruijff Arena (Amsterdam)	Netherlands
19	3788759	18/06/2021	Group D	Group Stage	England	Scotland	0	0	Wembley Stadium	England
20	3788761	18/06/2021	Group E	Group Stage	Sweden	Slovakia	1	0	Saint-Petersburg Stadium	Russia
21	3788760	18/06/2021	Group D	Group Stage	Croatia	Czech Republic	1	1	Hampden Park	Scotland
22	3788763	19/06/2021	Group F	Group Stage	Hungary	France	1	1	Puskás Aréna	Hungary
23	3788764	19/06/2021	Group F	Group Stage	Portugal	Germany	2	4	Allianz Arena	Germany
24	3788762	19/06/2021	Group E	Group Stage	Spain	Poland	1	1	Estadio de La Cartuja	Spain
25	3788765	20/06/2021	Group A	Group Stage	Switzerland	Turkey	3	1	Bakı Olimpiya Stadionu	Azerbaijan
26	3788766	20/06/2021	Group A	Group Stage	Italy	Wales	1	0	Stadio Olimpico	Italy
27	3788768	21/06/2021	Group B	Group Stage	Finland	Belgium	0	2	Saint-Petersburg Stadium	Russia
28	3788767	21/06/2021	Group C	Group Stage	Ukraine	Austria	0	1	Arena Naţională	Romania
29	3788769	21/06/2021	Group B	Group Stage	Russia	Denmark	1	4	Parken	Denmark
30	3788770	21/06/2021	Group C	Group Stage	North Macedonia	Netherlands	0	3	Johan Cruijff Arena (Amsterdam)	Netherlands
31	3788771	22/06/2021	Group D	Group Stage	Croatia	Scotland	3	1	Hampden Park	Scotland
32	3788772	22/06/2021	Group D	Group Stage	Czech Republic	England	0	1	Wembley Stadium	England
33	3788774	23/06/2021	Group F	Group Stage	Germany	Hungary	2	2	Allianz Arena	Germany
34	3788773	23/06/2021	Group F	Group Stage	Portugal	France	2	2	Puskás Aréna	Hungary
35	3788775	23/06/2021	Group E	Group Stage	Slovakia	Spain	0	5	Estadio de La Cartuja	Spain
36	3788776	23/06/2021	Group E	Group Stage	Sweden	Poland	3	2	Saint-Petersburg Stadium	Russia
37	3794685	26/06/2021		Round of 16	Italy	Austria	2	1	Wembley Stadium	England
38	3794689	26/06/2021		Round of 16	Wales	Denmark	0	4	Johan Cruijff Arena (Amsterdam)	Netherlands
39	3794687	27/06/2021		Round of 16	Belgium	Portugal	1	0	Estadio de La Cartuja	Spain
40	3794690	27/06/2021		Round of 16	Netherlands	Czech Republic	0	2	Puskás Aréna	Hungary
41	3794686	28/06/2021		Round of 16	Croatia	Spain	3	5	Parken	Denmark
42	3794691	28/06/2021		Round of 16	France	Switzerland	3	3	Arena Naţională	Romania
43	3794688	29/06/2021		Round of 16	England	Germany	2	0	Wembley Stadium	England
44	3794692	29/06/2021		Round of 16	Sweden	Ukraine	1	2	Hampden Park	Scotland
45	3795107	02/07/2021		Quarter-finals	Belgium	Italy	1	2	Allianz Arena	Germany
46	3795108	02/07/2021		Quarter-finals	Switzerland	Spain	1	1	Saint-Petersburg Stadium	Russia
47	3795187	03/07/2021		Quarter-finals	Ukraine	England	0	4	Stadio Olimpico	Italy
48	3795109	03/07/2021		Quarter-finals	Czech Republic	Denmark	1	2	Bakı Olimpiya Stadionu	Azerbaijan
49	3795220	06/07/2021		Semi-finals	Italy	Spain	1	1	Wembley Stadium	England
50	3795221	07/07/2021		Semi-finals	England	Denmark	2	1	Wembley Stadium	England
51	3795506	11/07/2021		Final	Italy	England	1	1	Wembley Stadium	England

Data provided by StatsBomb

Appendix B – Passing sequences' distribution

#	id	match_id	team_name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
1	3788741_Italy	3788741	Italy	14	11	6	9	7	6	5	4	1	1	3	4			2	1	1	1	2		1		1		1		1							1																						
2	3788741_Turkey	3788741	Turkey	29	21	8	8	3	3	1	1	3	1	4			1	1							1																																		
3	3788742_Denmark	3788742	Denmark	16	15	8	4	3	8	5	6	1	2	3	2	3	1	1	2	2		1	1		1			1				1																									
4	3788742_Finland	3788742	Finland	17	14	9	2	5	2	2	2			2							2																																						
5	3788743_Belgium	3788743	Belgium	28	8	6	7	4	5	4	1	4	4	2	4	5		1	1	1	2		1		1				1		1	1									1			1															
6	3788743_Russia	3788743	Russia	20	14	7	5	7	4	2	1		2	1	2			1	1		2																																				
7	3788744_Switzerland	3788744	Switzerland	15	5	11	6	7	2	6	4	3	5	1	2	2	2		1				1			2			1																												
8	3788744_Wales	3788744	Wales	18	8	13	8	5	2	3	2			3		1			2																																						
9	3788745_Croatia	3788745	Croatia	9	16	5	6	7	2	1	2	3	1	2	2	1	1		3		3			1	1																																
10	3788745_England	3788745	England	22	11	6	2	2		1	3	2		1		2	1		1	2		2	1						1								1									1											
11	3788746_Netherlands	3788746	Netherlands	20	6	4	5	4	6	5	6	2	3	3	3	2	1		2	3	1	3			1		1					1				1																					
12	3788746_Ukraine	3788746	Ukraine	24	13	7	5	3	4	1	4	1	3	2	2		1	1	2				1												1																						
13	3788747_Austria	3788747	Austria	21	15	7	6	7	6	9	7	1	1	7	1				3		1		2																																		
14	3788747_North Macedonia	3788747	North Macedonia	25	23	13	3	8	4			3	1	3	1																																										
15	3788748_Czech Republic	3788748	Czech Republic	40	24	10	3	7	1	1	3	1	2		1	1	1																																								
16	3788748_Scotland	3788748	Scotland	27	14	18	5	2	3	6	7	4	2		1		2			1	1																																				
17	3788749_Poland	3788749	Poland	21	10	16	6	3	3	4	3	3	3	3		2		4	2	1		1				1	1																														
18	3788749_Slovakia	3788749	Slovakia	26	10	5	6	1	4	4	4	2	1		1	1	2		1	1			1		1							1																									
19	3788750_Spain	3788750	Spain	5	4	6	2	5	4	4	4	6	3	1	3	2	3	2	2	1	1	2	1	1			2	1		1				1	1	1			1							2											
20	3788750_Sweden	3788750	Sweden	21	12	6	5	1	1																																																	
21	3788751_France	3788751	France	24	12	8	5	3	2	4	2	4						1			2	1	1	1									1					1																			
22	3788751_Germany	3788751	Germany	6	8	9	2	6	1	5	5	7	6	3	2		2	1	1	2	2	3	2			2	1																														
23	3788752_Hungary	3788752	Hungary	25	17	12	11	7	7	1	1			1																																										
24	3788752_Portugal	3788752	Portugal	38	10	11	4	8	6	3	6	2	2	4	2	3		1	2		1					1	1	1		1						2																					
25	3788753_Finland	3788753	Finland	28	15	6	10	7	6	5	3		3				2	1									1																													
26	3788753_Russia	3788753	Russia	35	14	9	6	7	8	4	4	5	3	1	1			2			1	1				1							1						1																		
27	3788754_Italy	3788754	Italy	16	17	5	3	7	3	6	2	3	1	2	2	2	1	2		1		1	1	1						1			1																								
28	3788754_Switzerland	3788754	Switzerland	15	8	10	9	4	3	1	5	6	6	3	1	3		1	2	1	1			1																																	
29	3788755_Turkey	3788755	Turkey	23	11	7	4	9	1	2	5	4	2	4		2		3	1	1							1																														1		
30	3788755_Wales	3788755	Wales	31	10	12	4	7	3	5		2	1	2				1																																							
31	3788756_Austria	3788756	Austria	19	13	8	5	6	11	7	5	3	2	2	2					1	1					2																															
32	3788756_Netherlands	3788756	Netherlands	14	14	13	3	6	4	3	4	1	3	2	1	4	1	3						1																																	
33	3788757_Belgium	3788757	Belgium	20	10	7	10	7	3	3	6	2	2	3	2		2	2		2	1	1			1		2																														
34	3788757_Denmark	3788757	Denmark	29	15	11	5	6	3	4	2	5	3	3	1	1		3		1	1		1																																		
35	3788758_North Macedonia	3788758	North Macedonia	35	14	8	3	5	7	3	2	4	2		1		1	2	2		1																																				
36	3788758_Ukraine	3788758	Ukraine	17	10	7	5	6	3	5	3	5	1			1	1	1	1	1	1	1			1													1																				
37	3788759_England	3788759	England	18	6	8	2	5	2	1	3	2	1	4	3	1	1	3	2		2						1			1		1																										1
38	3788759_Scotland	3788759	Scotland	16	13	7	4	2	3	1	4	1		4						1		1					2		1																												
39	3788760_Croatia	3788760	Croatia	19	12	9	8	7	4	8	4	1	1	1	1	3		2			1																																				
40	3788760_Czech Republic	3788760	Czech Republic	21	14	12	6	3	4	3	1	1	2	2	1		1	1		2			1	1			1																														
41	3788761_Slovakia	3788761	Slovakia	37	9	6	6	3	3	3	1	3	4	1	2	1	1	1	1	3			1					1	1	2	1					1																					
42	3788761_Sweden	3788761	Sweden	9	9	5	5	1	5	3	3	2	2			2	1			2	1	1								1	1						1																				
43	3788762_Poland	3788762	Poland	28	14	13	2	3	2	1		1																																													
44	3788762_Spain	3788762	Spain	18	10	12	5	3	8	5	4	2	2	6	1	3	3	2	2	2						3		2			1																												
45	3788763_France	3788763	France	31	6	4	4	6	7	4	2	3	4	6	3	1	2	5	3		1	1				1		1																														
46	3788763_Hungary	3788763	Hungary	13	15	6	6	2	5	4	4	2			1	1	1	1	1				1																																		
47	3788764_Germany	3788764	Germany	12	5	4	8	4	4	4	3	1	3	3	1		2	3				2	1	2	1			1		1															1												
48	3788764_Portugal	3788764	Portugal	24	12	10	7	7	6	6	4	2		2				3	1					1																																	
49	3788765_Switzerland	3788765	Switzerland	25	17	10	8	3	3	9	4	2	2	2	1	1		1	1								1		2																												
50	3788765_Turkey	3788765	Turkey	36	15	7	6	3	7	1	3	2	4	2	3		1	2	1		1				1		1																														
51	3788766_Italy	3788766	Italy	14	3	9		4	6	3	6	2	5	2	1		3	1	1	1	2	1	1	1	1						1		1				1																				
52	3788766_Wales	3788766	Wales	33	14	7	1	3		1	2	3		1	1				1					1																																	

(continues on the next page)

#	id	match_id	team_name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
53	3788767_Austria	3788767	Austria	23	15	12	11	7	3	3	8	1				2	2	3	2																																								
54	3788767_Ukraine	3788767	Ukraine	35	17	12	4	6	5	5	5	4	1		1	3	2							1																																			
55	3788768_Belgium	3788768	Belgium	15	7	6	8	3	6	3	1	3		3	1	3	2	3	4	3		3	3			1																		1															
56	3788768_Finland	3788768	Finland	21	7	6	8	2	2	1	1	4	2	1	1	2		1				1			1	2		2																															
57	3788769_Denmark	3788769	Denmark	18	24	6	7	4	5	2	3	2	1		2	2	2	3		3			1		1																																		
58	3788769_Russia	3788769	Russia	26	21	9	5	6		5	1	1																																																
59	3788770_Netherlands	3788770	Netherlands	18	17	5	8	6	4	1	4	2	1	2		1	2	4	1			2			1	2			1										1																		
60	3788770_North Macedonia	3788770	North Macedonia	29	11	12	8	7	1	3	3	2	2		2	2	1										1					1																									
61	3788771_Croatia	3788771	Croatia	24	10	9	7	5	3	1	2	4	1	5	6	2	3	1	4				1	1	2		1																														
62	3788771_Scotland	3788771	Scotland	21	17	3	2	5	2	2	3	6	1	1		2			1																																						
63	3788772_Czech Republic	3788772	Czech Republic	13	10	11	3	7	2		3	5	2	3	1	1	1		1																	1																					
64	3788772_England	3788772	England	16	3	3	6	7	3	1	2	2	4	3	3	2		1	2	3				1								1			1																							
65	3788773_France	3788773	France	17	5	1	3	1	6	2	4	1	3	4		4	3		2		1	2					1		1	2								1		1																	
66	3788773_Portugal	3788773	Portugal	11	6	4	3	1	6	2	1	7	5	1			1	1	3		1	2	2	1		1				2																											
67	3788774_Germany	3788774	Germany	18	5	9	11	5	3	4		4	4	1	2	4	1	2	3	1	1	1	1	1		1		1											1				1														1	
68	3788774_Hungary	3788774	Hungary	39	14	5	8	2	3			1	1	1		1																																										
69	3788775_Slovakia	3788775	Slovakia	26	10	5	8	7	6		2		1	1	1	2			1	1				1																																	
70	3788775_Spain	3788775	Spain	14	4	9	10	6	5	3	5	4	3	3	3	1		1		3				1		2	1	1	1					1			1																				
71	3788776_Poland	3788776	Poland	18	12	9	13	8	7	9		4	2	1	1			2	1	2																																					
72	3788776_Sweden	3788776	Sweden	28	18	8	4	2	1		2	3																																														
73	3794685_Austria	3794685	Austria	26	12	12	12	4	3	4	2		3	1	1	1	1		1	1		2		2																																	
74	3794685_Italy	3794685	Italy	16	18	9	7	4	6	5	3	3	5			1	3	2	1	2														1	1	1																					
75	3794686_Croatia	3794686	Croatia	28	10	8	9	8	6	3	1	1	3						2																																						
76	3794686_Spain	3794686	Spain	26	8	7	6	4	4	5	4	6	4	6	1	2	1	4	1			1		1					1			1																									
77	3794687_Belgium	3794687	Belgium	22	12	4	4	3	3	2	3	3	3	2	2	1		1		1												1								1	1																	
78	3794687_Portugal	3794687	Portugal	17	8	7	10	2	4	1	4	2		2	1	4	1	3	3	1	1	3		1	1							1																									
79	3794688_England	3794688	England	44	12	5	2	7	2	2	3	5	3	2	1	4					1	1		1																																	
80	3794688_Germany	3794688	Germany	7	8	16	8	4	5	4	2	2	2	1		2	2	1	3	1		2					1				1																										
81	3794689_Denmark	3794689	Denmark	18	11	14	5	1	4	1	2	2	2	3	1	1				1	1		2		1									1																							
82	3794689_Wales	3794689	Wales	18	10	4	5	2	3	7	1	2		2	3	2		1		2	1	1																																			
83	3794690_Czech Republic	3794690	Czech Republic	22	14	3	4	8	3	4	3	3		1	2		1			2	1																																				
84	3794690_Netherlands	3794690	Netherlands	19	19	10	4	8	4	2	3	2	3						2			2																																			
85	3794691_France	3794691	France	8	11	7	2	3	4	4	2	2	1	5	2			2	1	2	2	1	3	1																																	
86	3794691_Switzerland	3794691	Switzerland	30	10	3	5	1	4	6	2	1	3	2	2	2			1	1	2					1																															
87	3794692_Sweden	3794692	Sweden	17	8	5		4	3	5	6	4	1	2	1	2	2	2	2	1	2	1	1											1			1																				
88	3794692_Ukraine	3794692	Ukraine	14	11	5	2	2	1	1	3	3		2	1	1	3	1	1			1	2	1		1	1				1		1																			1					
89	3795107_Belgium	3795107	Belgium	23	12	8	8	4	2	3	3	3	3				4	1		2			1													1																					
90	3795107_Italy	3795107	Italy	14	8	6	6	4	4	4	3	2	2	2		2		3	1			1	1		1	1			1	1							1																				
91	3795108_Spain	3795108	Spain	13	5	4	9	7	6	2	8	4	4	1	1	2	1	6			1	1							1			1			1	1																1					
92	3795108_Switzerland	3795108	Switzerland	28	18	9	8	6	2	4	1	2				1		1																																								
93	3795109_Czech Republic	3795109	Czech Republic	26	8	8	5	7	6	1	4	3	4	3	2	1					1		1		1																																
94	3795109_Denmark	3795109	Denmark	25	13	11	5	4	3	4	1	2			1	1	1		1		1	2																																			
95	3795187_England	3795187	England	13	6	5	3	5		1	3		3	1	1	2		5	1		1			2	1			1	1			1				1	1		1														1					
96	3795187_Ukraine	3795187	Ukraine	6	2	2	7	5	2	3	1	1		1	6	1	2	4			1					1	1	1	1			2	1			1																					
97	3795220_Italy	3795220	Italy	32	19	5	9	6	4	2	4		3	2																																												
98	3795220_Spain	3795220	Spain	56	15	6	7	6	6	2	4	1	1	3	3	1	1			2	4			1		3		1				1												1															
99	3795221_Denmark	3795221	Denmark	25	14	7	2	4		3		2	1	1	1	3	1		1										2								1																				
100	3795221_England	3795221	England	21	6	9	3	10	6	1	1	3	1		2	2			2		2							1	1	1										1								1											
101	3795506_England	3795506	England	27	10	12	6	6	1	3	2	2	1			2													1																											
102	3795506_Italy	3795506	Italy	11	9	5	7	4	4	1	3	1	3	4	1	2	3	1	1	2	2				1	1	1			1				1																1			1					

Appendix C – Fitting the passing sequences to the power-law distribution (90 min[11])

#	id	match_id	match_date	group	competition_stage_name	team_name	opponent	result	goals scored	nu_passes	nu_passes_completed	per_passes_completed	α	x_{min}	$\alpha_{(opt)}$	$x_{min(opt)}$	p-value
1	3788741_Italy	3788741	2021-06-11	Group A	Group Stage	Italy	Turkey	victory	3	669	577	86%	**1.502**	1	2.169	4	0.062
2	3788741_Turkey	3788741	2021-06-11	Group A	Group Stage	Turkey	Italy	defeat	0	390	307	79%	**1.740**	1	2.120	2	0.116
3	3788742_Denmark	3788742	2021-06-12	Group B	Group Stage	Denmark	Finland	defeat	0	678	567	84%	**1.519**	1	5.192	16	0.000
4	3788742_Finland	3788742	2021-06-12	Group B	Group Stage	Finland	Denmark	victory	1	307	205	67%	**1.721**	1	2.987	5	0.661
5	3788743_Belgium	3788743	2021-06-12	Group B	Group Stage	Belgium	Russia	victory	3	763	675	88%	**1.533**	1	3.352	12	0.250
6	3788743_Russia	3788743	2021-06-12	Group B	Group Stage	Russia	Belgium	defeat	0	401	292	73%	**1.660**	1	1.985	2	0.008
7	3788744_Switzerland	3788744	2021-06-12	Group A	Group Stage	Switzerland	Wales	draw	1	563	477	85%	**1.516**	1	3.407	9	0.624
8	3788744_Wales	3788744	2021-06-12	Group A	Group Stage	Wales	Switzerland	draw	1	329	257	78%	**1.660**	1	2.524	3	0.625
9	3788745_Croatia	3788745	2021-06-13	Group D	Group Stage	Croatia	England	defeat	0	483	405	84%	**1.527**	1	1.848	2	0.000
10	3788745_England	3788745	2021-06-13	Group D	Group Stage	England	Croatia	victory	1	470	391	83%	**1.617**	1	1.617	1	0.044
11	3788746_Netherlands	3788746	2021-06-13	Group C	Group Stage	Netherlands	Ukraine	victory	3	705	626	89%	**1.492**	1	5.000	16	0.000
12	3788746_Ukraine	3788746	2021-06-13	Group C	Group Stage	Ukraine	Netherlands	defeat	2	449	367	82%	**1.637**	1	3.753	10	0.590
13	3788747_Austria	3788747	2021-06-13	Group C	Group Stage	Austria	North Macedonia	victory	3	605	500	83%	**1.562**	1	3.027	6	0.029
14	3788747_North Macedonia	3788747	2021-06-13	Group C	Group Stage	North Macedonia	Austria	defeat	1	351	268	76%	**1.752**	1	2.290	2	0.052
15	3788748_Czech Republic	3788748	2021-06-14	Group D	Group Stage	Czech Republic	Scotland	victory	2	386	270	70%	**1.864**	1	2.307	2	0.091
16	3788748_Scotland	3788748	2021-06-14	Group D	Group Stage	Scotland	Czech Republic	defeat	0	503	386	77%	**1.654**	1	3.990	7	0.538
17	3788749_Poland	3788749	2021-06-14	Group E	Group Stage	Poland	Slovakia	defeat	1	590	491	83%	**1.566**	1	2.765	7	0.053
18	3788749_Slovakia	3788749	2021-06-14	Group E	Group Stage	Slovakia	Poland	victory	2	432	359	83%	**1.644**	1	2.679	6	0.599
19	3788750_Spain	3788750	2021-06-14	Group E	Group Stage	Spain	Sweden	draw	0	946	841	89%	**1.387**	1	2.859	12	0.393
20	3788750_Sweden	3788750	2021-06-14	Group E	Group Stage	Sweden	Spain	draw	0	173	94	54%	**2.025**	1	6.315	4	0.000
21	3788751_France	3788751	2021-06-15	Group F	Group Stage	France	Germany	victory	1	463	375	81%	**1.644**	1	1.857	2	0.079
22	3788751_Germany	3788751	2021-06-15	Group F	Group Stage	Germany	France	defeat	0	733	644	88%	**1.432**	1	2.999	8	0.149
23	3788752_Hungary	3788752	2021-06-15	Group F	Group Stage	Hungary	Portugal	defeat	0	324	242	75%	**1.758**	1	6.692	6	0.000
24	3788752_Portugal	3788752	2021-06-15	Group F	Group Stage	Portugal	Hungary	victory	3	713	627	88%	**1.602**	1	3.000	10	0.198
25	3788753_Finland	3788753	2021-06-16	Group B	Group Stage	Finland	Russia	defeat	0	423	344	81%	**1.681**	1	3.368	6	0.687
26	3788753_Russia	3788753	2021-06-16	Group B	Group Stage	Russia	Finland	victory	1	615	511	83%	**1.637**	1	2.821	6	0.680
27	3788754_Italy	3788754	2021-06-16	Group A	Group Stage	Italy	Switzerland	victory	3	554	475	86%	**1.548**	1	3.434	11	0.816
28	3788754_Switzerland	3788754	2021-06-16	Group A	Group Stage	Switzerland	Italy	defeat	0	563	483	86%	**1.520**	1	4.211	9	0.406
29	3788755_Turkey	3788755	2021-06-16	Group A	Group Stage	Turkey	Wales	defeat	0	559	467	84%	**1.581**	1	3.203	8	0.860
30	3788755_Wales	3788755	2021-06-16	Group A	Group Stage	Wales	Turkey	victory	2	344	256	74%	**1.765**	1	3.402	5	0.327
31	3788756_Austria	3788756	2021-06-17	Group C	Group Stage	Austria	Netherlands	defeat	0	542	448	83%	**1.568**	1	3.395	6	0.983
32	3788756_Netherlands	3788756	2021-06-17	Group C	Group Stage	Netherlands	Austria	victory	2	500	405	81%	**1.562**	1	7.918	13	0.000
33	3788757_Belgium	3788757	2021-06-17	Group B	Group Stage	Belgium	Denmark	victory	2	602	517	86%	**1.543**	1	2.735	7	0.081
34	3788757_Denmark	3788757	2021-06-17	Group B	Group Stage	Denmark	Belgium	defeat	1	519	437	84%	**1.635**	1	4.043	9	0.357
35	3788758_North Macedonia	3788758	2021-06-17	Group C	Group Stage	North Macedonia	Ukraine	defeat	1	440	365	83%	**1.705**	1	1.705	1	0.001
36	3788758_Ukraine	3788758	2021-06-17	Group C	Group Stage	Ukraine	North Macedonia	victory	2	485	409	84%	**1.562**	1	4.389	15	0.001
37	3788759_England	3788759	2021-06-18	Group D	Group Stage	England	Scotland	draw	0	580	513	88%	**1.518**	1	3.359	11	0.854
38	3788759_Scotland	3788759	2021-06-18	Group D	Group Stage	Scotland	England	draw	0	383	309	81%	**1.619**	1	1.908	2	0.118
39	3788760_Croatia	3788760	2021-06-18	Group D	Group Stage	Croatia	Czech Republic	draw	1	465	378	81%	**1.596**	1	3.767	7	0.767
40	3788760_Czech Republic	3788760	2021-06-18	Group D	Group Stage	Czech Republic	Croatia	draw	1	469	368	78%	**1.631**	1	2.155	3	0.139
41	3788761_Slovakia	3788761	2021-06-18	Group E	Group Stage	Slovakia	Sweden	defeat	0	631	556	88%	**1.620**	1	1.620	1	0.002
42	3788761_Sweden	3788761	2021-06-18	Group E	Group Stage	Sweden	Slovakia	victory	1	464	383	83%	**1.507**	1	2.488	6	0.348
43	3788762_Poland	3788762	2021-06-19	Group E	Group Stage	Poland	Spain	draw	1	242	146	60%	**1.953**	1	3.543	3	0.539
44	3788762_Spain	3788762	2021-06-19	Group E	Group Stage	Spain	Poland	draw	1	754	658	87%	**1.501**	1	3.720	11	0.655
45	3788763_France	3788763	2021-06-19	Group F	Group Stage	France	Hungary	draw	1	681	605	89%	**1.545**	1	7.231	15	0.000
46	3788763_Hungary	3788763	2021-06-19	Group F	Group Stage	Hungary	France	draw	1	369	293	79%	**1.603**	1	3.172	6	0.718
47	3788764_Germany	3788764	2021-06-19	Group F	Group Stage	Germany	Portugal	victory	4	596	535	90%	**1.471**	1	5.398	19	0.000
48	3788764_Portugal	3788764	2021-06-19	Group F	Group Stage	Portugal	Germany	defeat	2	442	373	84%	**1.631**	1	3.258	6	0.467
49	3788765_Switzerland	3788765	2021-06-20	Group A	Group Stage	Switzerland	Turkey	victory	3	516	441	85%	**1.622**	1	3.355	7	0.864
50	3788765_Turkey	3788765	2021-06-20	Group A	Group Stage	Turkey	Switzerland	defeat	1	510	439	86%	**1.671**	1	3.976	10	0.514
51	3788766_Italy	3788766	2021-06-20	Group A	Group Stage	Italy	Wales	victory	1	629	574	91%	**1.467**	1	2.540	7	0.050
52	3788766_Wales	3788766	2021-06-20	Group A	Group Stage	Wales	Italy	defeat	0	274	211	77%	**1.881**	1	1.881	1	0.046

Data provided by StatsBomb

(continues on the next page)

[11] Note that the variable *result* is the match result after the regular time (90 min)

#	id	match_id	match_date	group	competition_stage_name	team_name	opponent	result	goals scored	nu_passes	nu_passes_completed	per_passes_completed	α	x_{min}	$\alpha_{(opt)}$	$x_{min(opt)}$	p-value
53	3788767_Austria	3788767	2021-06-21	Group C	Group Stage	Austria	Ukraine	victory	1	507	402	79%	**1.624**	1	2.270	3	0.007
54	3788767_Ukraine	3788767	2021-06-21	Group C	Group Stage	Ukraine	Austria	defeat	0	499	402	81%	**1.691**	1	3.616	7	0.546
55	3788768_Belgium	3788768	2021-06-21	Group B	Group Stage	Belgium	Finland	victory	2	714	639	89%	**1.473**	1	5.270	15	0.000
56	3788768_Finland	3788768	2021-06-21	Group B	Group Stage	Finland	Belgium	defeat	0	460	379	82%	**1.594**	1	2.696	8	0.080
57	3788769_Denmark	3788769	2021-06-21	Group B	Group Stage	Denmark	Russia	victory	4	542	444	82%	**1.595**	1	1.938	2	0.016
58	3788769_Russia	3788769	2021-06-21	Group B	Group Stage	Russia	Denmark	defeat	1	289	197	68%	**1.838**	1	2.471	2	0.177
59	3788770_Netherlands	3788770	2021-06-21	Group C	Group Stage	Netherlands	North Macedonia	victory	3	619	527	85%	**1.550**	1	4.101	14	0.092
60	3788770_North Macedonia	3788770	2021-06-21	Group C	Group Stage	North Macedonia	Netherlands	defeat	0	439	360	82%	**1.679**	1	2.269	3	0.332
61	3788771_Croatia	3788771	2021-06-22	Group D	Group Stage	Croatia	Scotland	victory	3	671	594	89%	**1.537**	1	4.276	11	0.174
62	3788771_Scotland	3788771	2021-06-22	Group D	Group Stage	Scotland	Croatia	defeat	1	353	264	75%	**1.685**	1	6.054	9	0.000
63	3788772_Czech Republic	3788772	2021-06-22	Group D	Group Stage	Czech Republic	England	defeat	0	431	335	78%	**1.571**	1	4.277	9	0.197
64	3788772_England	3788772	2021-06-22	Group D	Group Stage	England	Czech Republic	victory	1	554	462	83%	**1.500**	1	3.556	10	0.780
65	3788773_France	3788773	2021-06-23	Group F	Group Stage	France	Portugal	draw	2	617	570	92%	**1.474**	1	3.092	11	0.190
66	3788773_Portugal	3788773	2021-06-23	Group F	Group Stage	Portugal	France	draw	2	556	505	91%	**1.462**	1	6.088	18	0.000
67	3788774_Germany	3788774	2021-06-23	Group F	Group Stage	Germany	Hungary	draw	2	783	698	89%	**1.486**	1	3.378	13	0.939
68	3788774_Hungary	3788774	2021-06-23	Group F	Group Stage	Hungary	Germany	draw	2	255	185	73%	**1.977**	1	1.977	1	0.018
69	3788775_Slovakia	3788775	2021-06-23	Group E	Group Stage	Slovakia	Spain	defeat	0	368	293	80%	**1.690**	1	2.635	4	0.585
70	3788775_Spain	3788775	2021-06-23	Group E	Group Stage	Spain	Slovakia	victory	5	702	631	90%	**1.476**	1	2.734	8	0.260
71	3788776_Poland	3788776	2021-06-23	Group E	Group Stage	Poland	Sweden	defeat	2	557	425	76%	**1.577**	1	3.314	6	0.453
72	3788776_Sweden	3788776	2021-06-23	Group E	Group Stage	Sweden	Poland	victory	3	278	163	59%	**1.922**	1	2.534	2	0.430
73	3794685_Austria	3794685	2021-06-26		Round of 16	Austria	Italy	draw	1	497	409	82%	**1.636**	1	2.212	3	0.124
74	3794685_Italy	3794685	2021-06-26		Round of 16	Italy	Austria	draw	2	604	530	88%	**1.547**	1	2.589	6	0.359
75	3794686_Croatia	3794686	2021-06-28		Round of 16	Croatia	Spain	draw	3	356	284	80%	**1.711**	1	3.445	5	0.924
76	3794686_Spain	3794686	2021-06-28		Round of 16	Spain	Croatia	draw	5	661	581	88%	**1.541**	1	3.615	9	0.872
77	3794687_Belgium	3794687	2021-06-27		Round of 16	Belgium	Portugal	victory	1	471	394	84%	**1.608**	1	2.942	8	0.817
78	3794687_Portugal	3794687	2021-06-27		Round of 16	Portugal	Belgium	defeat	0	630	542	86%	**1.511**	1	5.612	15	0.000
79	3794688_England	3794688	2021-06-29		Round of 16	England	Germany	victory	2	476	395	83%	**1.732**	1	4.148	9	0.006
80	3794688_Germany	3794688	2021-06-29		Round of 16	Germany	England	defeat	0	558	470	84%	**1.496**	1	2.125	3	0.078
81	3794689_Denmark	3794689	2021-06-26		Round of 16	Denmark	Wales	victory	4	465	378	81%	**1.600**	1	2.125	3	0.050
82	3794689_Wales	3794689	2021-06-26		Round of 16	Wales	Denmark	defeat	0	420	343	82%	**1.585**	1	4.597	11	0.000
83	3794690_Czech Republic	3794690	2021-06-27		Round of 16	Czech Republic	Netherlands	victory	2	415	313	75%	**1.648**	1	3.341	7	0.608
84	3794690_Netherlands	3794690	2021-06-27		Round of 16	Netherlands	Czech Republic	defeat	0	440	323	73%	**1.652**	1	2.919	5	0.454
85	3794691_France	3794691	2021-06-28		Round of 16	France	Switzerland	draw	3	535	465	87%	**1.479**	1	8.689	17	0.000
86	3794691_Switzerland	3794691	2021-06-28		Round of 16	Switzerland	France	draw	3	435	369	85%	**1.654**	1	3.939	10	0.267
87	3794692_Sweden	3794692	2021-06-29		Round of 16	Sweden	Ukraine	draw	1	613	522	85%	**1.502**	1	4.337	14	0.002
88	3794692_Ukraine	3794692	2021-06-29		Round of 16	Ukraine	Sweden	draw	2	563	487	87%	**1.511**	1	4.395	19	0.020
89	3795107_Belgium	3795107	2021-07-02		Quarter-finals	Belgium	Italy	defeat	1	482	395	82%	**1.616**	1	2.058	3	0.013
90	3795107_Italy	3795107	2021-07-02		Quarter-finals	Italy	Belgium	victory	2	546	486	89%	**1.509**	1	2.366	6	0.151
91	3795108_Spain	3795108	2021-07-02		Quarter-finals	Spain	Switzerland	draw	1	736	655	89%	**1.462**	1	2.866	8	0.354
92	3795108_Switzerland	3795108	2021-07-02		Quarter-finals	Switzerland	Spain	draw	1	333	247	74%	**1.778**	1	3.061	4	0.464
93	3795109_Czech Republic	3795109	2021-07-03		Quarter-finals	Czech Republic	Denmark	defeat	1	508	393	77%	**1.614**	1	4.076	9	0.157
94	3795109_Denmark	3795109	2021-07-03		Quarter-finals	Denmark	Czech Republic	victory	2	398	307	77%	**1.692**	1	2.249	3	0.269
95	3795187_England	3795187	2021-07-03		Quarter-finals	England	Ukraine	victory	4	648	589	91%	**1.463**	1	4.103	21	0.270
96	3795187_Ukraine	3795187	2021-07-03		Quarter-finals	Ukraine	England	defeat	0	600	540	90%	**1.410**	1	3.059	11	0.163
97	3795220_Italy	3795220	2021-07-06		Semi-finals	Italy	Spain	draw	1	358	273	76%	**1.773**	1	2.980	4	0.231
98	3795220_Spain	3795220	2021-07-06		Semi-finals	Spain	Italy	draw	1	732	648	89%	**1.686**	1	1.686	1	0.002
99	3795221_Denmark	3795221	2021-07-07		Semi-finals	Denmark	England	draw	1	410	329	80%	**1.687**	1	1.900	2	0.061
100	3795221_England	3795221	2021-07-07		Semi-finals	England	Denmark	draw	2	563	482	86%	**1.559**	1	2.384	5	0.250
101	3795506_England	3795506	2021-07-11		Final	England	Italy	draw	1	344	261	76%	**1.739**	1	3.463	7	0.707
102	3795506_Italy	3795506	2021-07-11		Final	Italy	England	draw	1	703	633	90%	**1.462**	1	3.158	13	0.832

Data provided by StatsBomb

Appendix D– Homoscedasticity plots

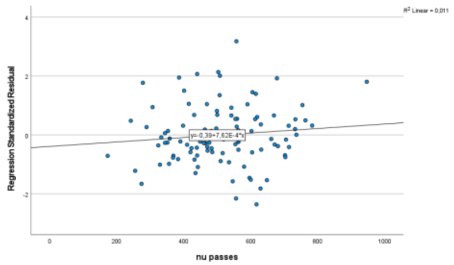


Figure E1: Homoscedasticity plot of nu passes vs nu passes completed

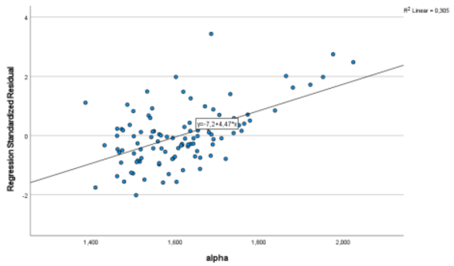


Figure E2: Homoscedasticity plot of parameter α vs nu passes completed

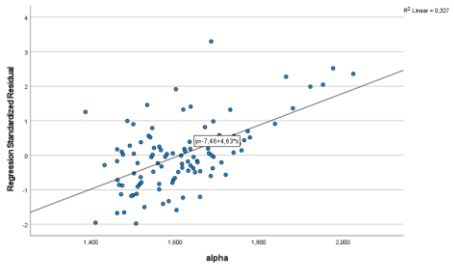


Figure E3: Homoscedasticity plot of parameter α vs nu passes

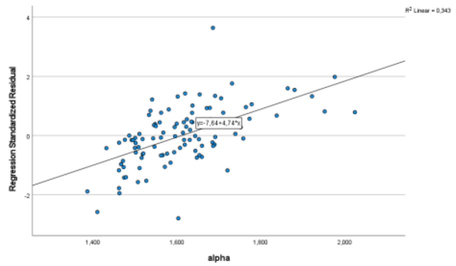


Figure E4: Homoscedasticity plot of parameter α vs density

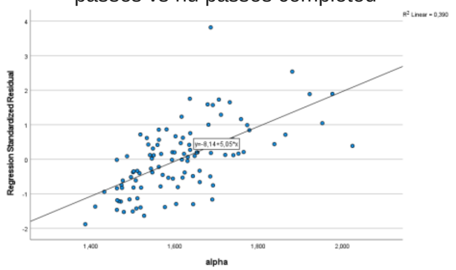


Figure E5: Homoscedasticity plot of parameter α vs average clustering coefficient

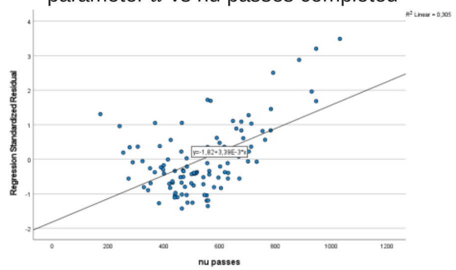


Figure E6: Homoscedasticity plot of nu passes vs density

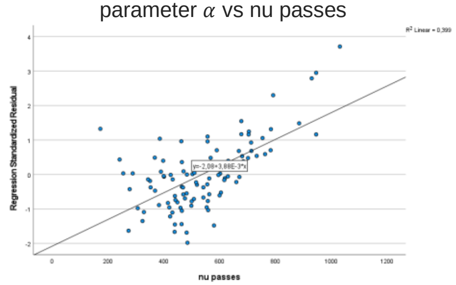


Figure E7: Homoscedasticity plot of nu passes vs average clustering coefficient

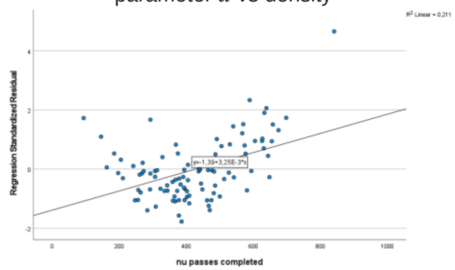


Figure E8: Homoscedasticity plot of nu passes completed vs density

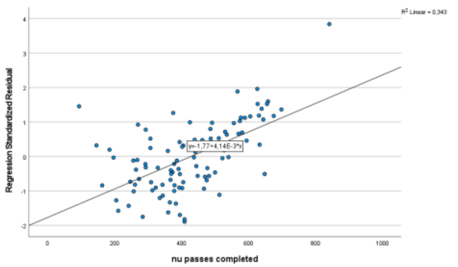


Figure E9: Homoscedasticity plot of nu passes completed vs average clustering coefficient

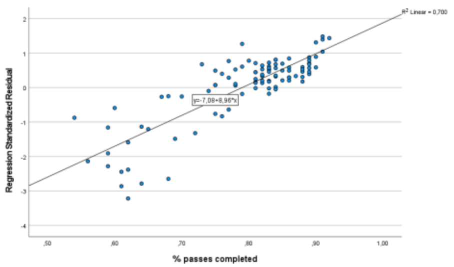


Figure E10: Homoscedasticity plot of % passes completed vs density

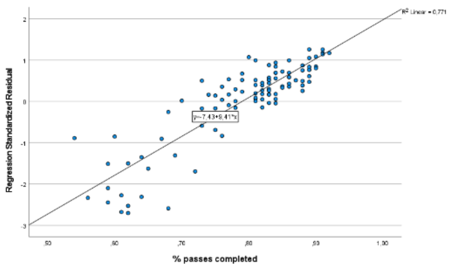


Figure E11: Homoscedasticity plot of % passes completed vs average clustering coefficient

Appendix E – Descriptive analysis of the different-sized zone networks

id	results_zone_all_3_3	results_zone_all_3_5	results_zone_all_4_3	results_zone_all_4_5	results_zone_all_6_3	results_zone_all_6_5
type_analysis	zone	zone	zone	zone	zone	zone
play_pattern	all	all	all	all	all	all
nu_sectors	3	3	4	4	6	6
nu_corridors	3	5	3	5	3	5
nu_zones	9	15	12	20	18	30
nu_nodes_mean	9	15	12	20	18	30
nu_nodes_median	9	15	12	20	18	30
nu_nodes_max	9	15	12	20	18	30
nu_nodes_min	9	15	12	20	18	30
nu_nodes_q1	9	15	12	20	18	30
nu_nodes_q3	9	15	12	20	18	30
nu_nodes_iqr	0	0	0	0	0	0
nu_nodes_ub	9	15	12	20	18	30
nu_nodes_lb	9	15	12	20	18	30
nu_edges_mean	50.35294118	100.7352941	70.04901961	131.3921569	108.6568627	181.6960784
nu_edges_median	50	102	71	134.5	110	185.5
nu_edges_max	60	121	82	162	133	238
nu_edges_min	34	55	43	60	60	76
nu_edges_q1	49	95.25	67	124.25	104	167.5
nu_edges_q3	54	108.75	74	143	117	204
nu_edges_iqr	5	13.5	7	18.75	13	36.5
nu_edges_ub	61.5	129	84.5	171.125	136.5	258.75
nu_edges_lb	41.5	75	56.5	96.125	84.5	112.75
nu_isolates_mean	0	0	0	0.019607843	0.039215686	0.235294118
nu_isolates_median	0	0	0	0	0	0
nu_isolates_max	0	0	0	1	2	3
nu_isolates_min	0	0	0	0	0	0
nu_isolates_q1	0	0	0	0	0	0
nu_isolates_q3	0	0	0	0	0	0
nu_isolates_iqr	0	0	0	0	0	0
nu_isolates_ub	0	0	0	0	0	0
nu_isolates_lb	0	0	0	0	0	0
density_mean	0.699346405	0.479691877	0.530674391	0.345768834	0.355087787	0.208846067
density_median	0.694444444	0.485714286	0.537878788	0.353947368	0.359477124	0.213218391
density_max	0.833333333	0.576190476	0.621212121	0.426315789	0.434640523	0.273563218
density_min	0.472222222	0.261904762	0.325757576	0.157894737	0.196078431	0.087356322
density_q1	0.680555556	0.453571429	0.507575758	0.326973684	0.339869281	0.192528736
density_q3	0.75	0.517857143	0.560606061	0.376315789	0.382352941	0.234482759
density_iqr	0.069444444	0.064285714	0.053030303	0.049342105	0.04248366	0.041954023
density_ub	0.854166667	0.614285714	0.640151515	0.450328947	0.446078431	0.297413793
density_lb	0.576388889	0.357142857	0.428030303	0.252960526	0.276143791	0.129597701
nu_triangles_mean	28.79411765	84.25490196	44.79411765	112.9019608	79.32352941	152.9019608
nu_triangles_median	28.5	83.5	45	114	80	151.5
nu_triangles_max	45	128	73	177	131	250
nu_triangles_min	10	20	16	15	30	19
nu_triangles_q1	24	71.25	38	98.25	68.25	122.75
nu_triangles_q3	34	95.75	51	132.75	91	182
nu_triangles_iqr	10	24.5	13	34.5	22.75	59.25
nu_triangles_ub	49	132.5	70.5	184.5	125.125	270.875
nu_triangles_lb	9	34.5	18.5	46.5	34.125	33.875
nu_cc_mean	1	1	1	1.039215686	1.049019608	1.284313725
nu_cc_median	1	1	1	1	1	1
nu_cc_max	1	1	1	2	3	4
nu_cc_min	1	1	1	1	1	1
nu_cc_q1	1	1	1	1	1	1
nu_cc_q3	1	1	1	1	1	1
nu_cc_iqr	0	0	0	0	0	0
nu_cc_ub	1	1	1	1	1	1
nu_cc_lb	1	1	1	1	1	1
avg_clust_coef_mean	0.682988508	0.563488326	0.60546674	0.490997308	0.500021057	0.382706615
avg_clust_coef_median	0.698125568	0.577149924	0.612357925	0.502061945	0.508697536	0.393920414
avg_clust_coef_max	0.803127889	0.706603732	0.745203913	0.629693919	0.637381206	0.553751526
avg_clust_coef_min	0.412798856	0.28245381	0.386135037	0.15836335	0.285576268	0.086072
avg_clust_coef_q1	0.650328638	0.533164613	0.55795364	0.457131935	0.453184993	0.34043304
avg_clust_coef_q3	0.734614769	0.611148055	0.661108835	0.54070218	0.54833246	0.437086323
avg_clust_coef_iqr	0.084286132	0.077983442	0.103155195	0.083570245	0.095147467	0.096653283
avg_clust_coef_ub	0.861043967	0.728123218	0.815841629	0.666057548	0.691053661	0.582066247
avg_clust_coef_lb	0.52389944	0.41618945	0.403220847	0.331776567	0.310463792	0.195453116

Appendix F – Zone 6x5 (90 min[12])

#	id	match_id	match_date	group	competition_stage_name	team_name	opponent	result	goals_scored	nu_nodes	nu_edges	nu_isolates	density	nu_triangles	nu_connected_components	average_clustering_coefficient
1	3788741_Italy	3788741	2021-06-11	Group A	Group Stage	Italy	Turkey	victory	3	30	209	0	0.240229885	179	1	0.419054769
2	3788741_Turkey	3788741	2021-06-11	Group A	Group Stage	Turkey	Italy	defeat	0	30	154	0	0.177011494	122	1	0.290475416
3	3788742_Denmark	3788742	2021-06-12	Group B	Group Stage	Denmark	Finland	defeat	0	30	196	0	0.225287356	142	1	0.371506042
4	3788742_Finland	3788742	2021-06-12	Group B	Group Stage	Finland	Denmark	victory	1	30	124	3	0.142528736	69	4	0.306004408
5	3788743_Belgium	3788743	2021-06-12	Group B	Group Stage	Belgium	Russia	victory	3	30	220	0	0.252873563	235	1	0.48030078
6	3788743_Russia	3788743	2021-06-12	Group B	Group Stage	Russia	Belgium	defeat	0	30	160	0	0.183908046	105	1	0.305886213
7	3788744_Switzerland	3788744	2021-06-12	Group A	Group Stage	Switzerland	Wales	draw	1	30	194	0	0.222988506	159	1	0.432899359
8	3788744_Wales	3788744	2021-06-12	Group A	Group Stage	Wales	Switzerland	draw	1	30	152	0	0.174712644	103	1	0.325882005
9	3788745_Croatia	3788745	2021-06-13	Group D	Group Stage	Croatia	England	defeat	0	30	174	1	0.2	128	2	0.350628669
10	3788745_England	3788745	2021-06-13	Group D	Group Stage	England	Croatia	victory	1	30	177	0	0.203448276	141	1	0.378100208
11	3788746_Netherlands	3788746	2021-06-13	Group C	Group Stage	Netherlands	Ukraine	victory	3	30	214	0	0.245977011	213	1	0.405972333
12	3788746_Ukraine	3788746	2021-06-13	Group C	Group Stage	Ukraine	Netherlands	defeat	2	30	184	0	0.211494253	157	1	0.379598073
13	3788747_Austria	3788747	2021-06-13	Group C	Group Stage	Austria	North Macedonia	victory	3	30	213	0	0.244827586	197	1	0.4547707
14	3788747_North Macedonia	3788747	2021-06-13	Group C	Group Stage	North Macedonia	Austria	defeat	1	30	146	1	0.167816092	117	2	0.283769669
15	3788748_Czech Republic	3788748	2021-06-14	Group D	Group Stage	Czech Republic	Scotland	victory	2	30	141	1	0.162068966	76	2	0.22968586
16	3788748_Scotland	3788748	2021-06-14	Group D	Group Stage	Scotland	Czech Republic	defeat	0	30	199	0	0.228735632	194	1	0.361706996
17	3788749_Poland	3788749	2021-06-14	Group E	Group Stage	Poland	Slovakia	defeat	1	30	181	0	0.208045977	139	3	0.370947856
18	3788749_Slovakia	3788749	2021-06-14	Group E	Group Stage	Slovakia	Poland	victory	2	30	172	0	0.197701149	114	1	0.32903389
19	3788750_Spain	3788750	2021-06-14	Group E	Group Stage	Spain	Sweden	draw	0	30	208	0	0.23908046	197	1	0.444687147
20	3788750_Sweden	3788750	2021-06-14	Group E	Group Stage	Sweden	Spain	draw	0	30	76	1	0.087356322	19	3	0.086072
21	3788751_France	3788751	2021-06-15	Group F	Group Stage	France	Germany	victory	1	30	161	0	0.185057471	119	1	0.279882535
22	3788751_Germany	3788751	2021-06-15	Group F	Group Stage	Germany	France	defeat	0	30	226	0	0.259770115	247	1	0.465678121
23	3788752_Hungary	3788752	2021-06-15	Group F	Group Stage	Hungary	Portugal	defeat	0	30	136	1	0.156321839	85	2	0.338753405
24	3788752_Portugal	3788752	2021-06-15	Group F	Group Stage	Portugal	Hungary	victory	3	30	213	0	0.244827586	207	1	0.430184442
25	3788753_Finland	3788753	2021-06-16	Group B	Group Stage	Finland	Russia	defeat	0	30	173	0	0.198850575	158	1	0.389298345
26	3788753_Russia	3788753	2021-06-16	Group B	Group Stage	Russia	Finland	victory	1	30	196	0	0.225287356	182	1	0.408270069
27	3788754_Italy	3788754	2021-06-16	Group A	Group Stage	Italy	Switzerland	victory	3	30	209	0	0.240229885	191	1	0.450742991
28	3788754_Switzerland	3788754	2021-06-16	Group A	Group Stage	Switzerland	Italy	defeat	0	30	196	0	0.225287356	157	1	0.404910918
29	3788755_Turkey	3788755	2021-06-16	Group A	Group Stage	Turkey	Wales	defeat	0	30	210	0	0.24137931	221	1	0.458701463
30	3788755_Wales	3788755	2021-06-16	Group A	Group Stage	Wales	Turkey	victory	2	30	157	0	0.18045977	112	1	0.276908324
31	3788756_Austria	3788756	2021-06-17	Group C	Group Stage	Austria	Netherlands	defeat	0	30	197	0	0.226436782	184	1	0.426231392
32	3788756_Netherlands	3788756	2021-06-17	Group C	Group Stage	Netherlands	Austria	victory	2	30	182	0	0.209195402	177	1	0.408074494
33	3788757_Belgium	3788757	2021-06-17	Group B	Group Stage	Belgium	Denmark	victory	2	30	200	1	0.229885057	196	2	0.419498631
34	3788757_Denmark	3788757	2021-06-17	Group B	Group Stage	Denmark	Belgium	defeat	1	30	185	0	0.212643678	167	1	0.390016031
35	3788758_North Macedonia	3788758	2021-06-17	Group C	Group Stage	North Macedonia	Ukraine	defeat	1	30	181	0	0.208045977	149	1	0.41362375
36	3788758_Ukraine	3788758	2021-06-17	Group C	Group Stage	Ukraine	North Macedonia	victory	2	30	195	0	0.224137931	181	1	0.472593542
37	3788759_England	3788759	2021-06-18	Group D	Group Stage	England	Scotland	draw	0	30	207	0	0.237931034	238	1	0.498309617
38	3788759_Scotland	3788759	2021-06-18	Group D	Group Stage	Scotland	England	draw	0	30	173	0	0.198850575	151	1	0.345471945
39	3788760_Croatia	3788760	2021-06-18	Group D	Group Stage	Croatia	Czech Republic	draw	1	30	186	1	0.213793103	182	2	0.361671641
40	3788760_Czech Republic	3788760	2021-06-18	Group D	Group Stage	Czech Republic	Croatia	draw	1	30	173	0	0.198850575	111	1	0.365243689
41	3788761_Slovakia	3788761	2021-06-18	Group E	Group Stage	Slovakia	Sweden	defeat	0	30	210	0	0.24137931	199	1	0.413956589
42	3788761_Sweden	3788761	2021-06-18	Group E	Group Stage	Sweden	Slovakia	victory	1	30	179	1	0.205747126	155	2	0.427235175
43	3788762_Poland	3788762	2021-06-19	Group E	Group Stage	Poland	Spain	draw	1	30	98	2	0.112643678	53	3	0.181317899
44	3788762_Spain	3788762	2021-06-19	Group E	Group Stage	Spain	Poland	draw	1	30	213	0	0.244827586	211	1	0.44647925
45	3788763_France	3788763	2021-06-19	Group F	Group Stage	France	Hungary	draw	1	30	209	0	0.240229885	176	1	0.435375686
46	3788763_Hungary	3788763	2021-06-19	Group F	Group Stage	Hungary	France	draw	1	30	124	1	0.142528736	66	2	0.311712886
47	3788764_Germany	3788764	2021-06-19	Group F	Group Stage	Germany	Portugal	victory	4	30	205	0	0.235632184	177	1	0.418309989
48	3788764_Portugal	3788764	2021-06-19	Group F	Group Stage	Portugal	Germany	defeat	2	30	180	0	0.206896552	138	1	0.371493449
49	3788765_Switzerland	3788765	2021-06-20	Group A	Group Stage	Switzerland	Turkey	victory	3	30	185	0	0.212643678	144	1	0.384030727
50	3788765_Turkey	3788765	2021-06-20	Group A	Group Stage	Turkey	Switzerland	defeat	1	30	184	0	0.211494253	160	1	0.363661276
51	3788766_Italy	3788766	2021-06-20	Group A	Group Stage	Italy	Wales	victory	1	30	204	0	0.234482759	147	1	0.422344165
52	3788766_Wales	3788766	2021-06-20	Group A	Group Stage	Wales	Italy	defeat	0	30	135	1	0.155172414	71	2	0.326320035

Data provided by StatsBomb

(continues on the next page)

[12] Note that the variable *result* is the match result after the regular time (90 min)

#	id	match_id	match_date	group	competition_stage_name	team_name	opponent	result	goals_scored	nu_nodes	nu_edges	nu_isolates	density	nu_triangles	nu_connected_components	average_clustering_coefficient
53	3788767_Austria	3788767	2021-06-21	Group C	Group Stage	Austria	Ukraine	victory	1	30	190	0	0.218390805	151	1	0.350531303
54	3788767_Ukraine	3788767	2021-06-21	Group C	Group Stage	Ukraine	Austria	defeat	0	30	187	1	0.214942529	173	2	0.415323548
55	3788768_Belgium	3788768	2021-06-21	Group B	Group Stage	Belgium	Finland	victory	2	30	200	0	0.229885057	181	1	0.445526744
56	3788768_Finland	3788768	2021-06-21	Group B	Group Stage	Finland	Belgium	defeat	0	30	175	0	0.201149425	152	1	0.395936581
57	3788769_Denmark	3788769	2021-06-21	Group B	Group Stage	Denmark	Russia	victory	4	30	193	0	0.22183908	165	1	0.408287644
58	3788769_Russia	3788769	2021-06-21	Group B	Group Stage	Russia	Denmark	defeat	1	30	129	0	0.148275862	81	1	0.233706353
59	3788770_Netherlands	3788770	2021-06-21	Group C	Group Stage	Netherlands	North Macedonia	victory	3	30	207	0	0.237931034	163	1	0.437656535
60	3788770_North Macedonia	3788770	2021-06-21	Group C	Group Stage	North Macedonia	Netherlands	defeat	0	30	182	0	0.209195402	151	1	0.426070611
61	3788771_Croatia	3788771	2021-06-22	Group D	Group Stage	Croatia	Scotland	victory	3	30	222	0	0.255172414	199	1	0.466243912
62	3788771_Scotland	3788771	2021-06-22	Group D	Group Stage	Scotland	Croatia	defeat	1	30	155	0	0.17816092	136	2	0.296377783
63	3788772_Czech Republic	3788772	2021-06-22	Group D	Group Stage	Czech Republic	England	defeat	0	30	171	0	0.196551724	143	1	0.387066986
64	3788772_England	3788772	2021-06-22	Group D	Group Stage	England	Czech Republic	victory	1	30	206	0	0.236781609	176	1	0.379140383
65	3788773_France	3788773	2021-06-23	Group F	Group Stage	France	Portugal	draw	2	30	192	1	0.220689655	163	2	0.439757672
66	3788773_Portugal	3788773	2021-06-23	Group F	Group Stage	Portugal	France	draw	2	30	188	1	0.216091954	156	2	0.411193511
67	3788774_Germany	3788774	2021-06-23	Group F	Group Stage	Germany	Hungary	draw	2	30	219	0	0.251724138	250	1	0.485216143
68	3788774_Hungary	3788774	2021-06-23	Group F	Group Stage	Hungary	Germany	draw	2	30	116	1	0.133333333	53	2	0.213519852
69	3788775_Slovakia	3788775	2021-06-23	Group E	Group Stage	Slovakia	Spain	defeat	0	30	152	3	0.174712644	113	4	0.252641903
70	3788775_Spain	3788775	2021-06-23	Group E	Group Stage	Spain	Slovakia	victory	5	30	219	0	0.251724138	226	1	0.451093554
71	3788776_Poland	3788776	2021-06-23	Group E	Group Stage	Poland	Sweden	defeat	2	30	179	0	0.205747126	147	1	0.327173273
72	3788776_Sweden	3788776	2021-06-23	Group E	Group Stage	Sweden	Poland	victory	3	30	118	0	0.135632184	57	1	0.255140905
73	3794685_Austria	3794685	2021-06-26		Round of 16	Austria	Italy	draw	1	30	191	0	0.21954023	181	1	0.468761372
74	3794685_Italy	3794685	2021-06-26		Round of 16	Italy	Austria	draw	2	30	193	0	0.22183908	160	1	0.410525766
75	3794686_Croatia	3794686	2021-06-28		Round of 16	Croatia	Spain	draw	3	30	169	0	0.194252874	151	1	0.383403266
76	3794686_Spain	3794686	2021-06-28		Round of 16	Spain	Croatia	draw	5	30	229	0	0.263218391	245	1	0.444968046
77	3794687_Belgium	3794687	2021-06-27		Round of 16	Belgium	Portugal	victory	1	30	174	0	0.2	140	1	0.337844153
78	3794687_Portugal	3794687	2021-06-27		Round of 16	Portugal	Belgium	defeat	0	30	213	0	0.244827586	227	1	0.440943821
79	3794688_England	3794688	2021-06-29		Round of 16	England	Germany	victory	2	30	184	0	0.211494253	149	1	0.388745487
80	3794688_Germany	3794688	2021-06-29		Round of 16	Germany	England	defeat	0	30	213	0	0.244827586	197	1	0.441925385
81	3794689_Denmark	3794689	2021-06-26		Round of 16	Denmark	Wales	victory	4	30	194	0	0.222988506	134	1	0.404179798
82	3794689_Wales	3794689	2021-06-26		Round of 16	Wales	Denmark	defeat	0	30	170	1	0.195402299	139	2	0.371618473
83	3794690_Czech Republic	3794690	2021-06-27		Round of 16	Czech Republic	Netherlands	victory	2	30	155	0	0.17816092	106	1	0.360740263
84	3794690_Netherlands	3794690	2021-06-27		Round of 16	Netherlands	Czech Republic	defeat	0	30	167	0	0.191954023	125	2	0.363066203
85	3794691_France	3794691	2021-06-28		Round of 16	France	Switzerland	draw	3	30	191	0	0.21954023	179	1	0.394673873
86	3794691_Switzerland	3794691	2021-06-28		Round of 16	Switzerland	France	draw	3	30	155	0	0.17816092	101	1	0.371423314
87	3794692_Sweden	3794692	2021-06-29		Round of 16	Sweden	Ukraine	defeat	1	30	202	0	0.232183908	198	1	0.446834851
88	3794692_Ukraine	3794692	2021-06-29		Round of 16	Ukraine	Sweden	draw	2	30	188	0	0.216091954	161	1	0.393166956
89	3795107_Belgium	3795107	2021-07-02		Quarter-finals	Belgium	Italy	defeat	1	30	190	0	0.218390805	148	1	0.452895448
90	3795107_Italy	3795107	2021-07-02		Quarter-finals	Italy	Belgium	victory	2	30	204	0	0.234482759	207	1	0.37749886
91	3795108_Spain	3795108	2021-07-02		Quarter-finals	Spain	Switzerland	draw	1	30	221	0	0.254022989	214	1	0.448475468
92	3795108_Switzerland	3795108	2021-07-02		Quarter-finals	Switzerland	Spain	draw	1	30	155	0	0.17816092	108	1	0.30408934
93	3795109_Czech Republic	3795109	2021-07-03		Quarter-finals	Czech Republic	Denmark	defeat	1	30	183	0	0.210344828	139	1	0.408745907
94	3795109_Denmark	3795109	2021-07-03		Quarter-finals	Denmark	Czech Republic	victory	2	30	157	0	0.18045977	91	1	0.275807241
95	3795187_England	3795187	2021-07-03		Quarter-finals	England	Ukraine	victory	4	30	184	0	0.211494253	151	1	0.427126759
96	3795187_Ukraine	3795187	2021-07-03		Quarter-finals	Ukraine	England	defeat	0	30	186	1	0.213793103	150	2	0.4568334
97	3795220_Italy	3795220	2021-07-06		Semi-finals	Italy	Spain	draw	1	30	146	0	0.167816092	90	1	0.316904552
98	3795220_Spain	3795220	2021-07-06		Semi-finals	Spain	Italy	draw	1	30	238	0	0.273563218	240	1	0.553751526
99	3795221_Denmark	3795221	2021-07-07		Semi-finals	Denmark	England	draw	1	30	152	0	0.174712644	95	1	0.335824429
100	3795221_England	3795221	2021-07-07		Semi-finals	England	Denmark	draw	2	30	207	0	0.237931034	183	1	0.448552261
101	3795506_England	3795506	2021-07-11		Final	England	Italy	draw	1	30	145	0	0.166666667	83	1	0.31789832
102	3795506_Italy	3795506	2021-07-11		Final	Italy	England	draw	1	30	201	0	0.231034483	150	1	0.497960127

Appendix G – Clustering analyses on the zone passing networks ($k = 3$)

Table G1: Clustering analysis using the clustering coefficient on the zone passing networks ($k = 3$)

National Team/match_id	Cluster 0	Cluster 1	Cluster 2	Total
Austria		**1**	**3**	**4**
3788747_Austria			1	1
3788756_Austria			1	1
3788767_Austria			1	1
3794685_Austria		1		1
Belgium		**4**	**1**	**5**
3788743_Belgium		1		1
3788757_Belgium		1		1
3788768_Belgium			1	1
3794687_Belgium		1		1
3795107_Belgium		1		1
Croatia		**2**	**2**	**4**
3788745_Croatia			1	1
3788760_Croatia		1		1
3788771_Croatia			1	1
3794686_Croatia		1		1
Czech Republic	**1**	**3**	**1**	**5**
3788748_Czech Republic	1			1
3788760_Czech Republic		1		1
3788772_Czech Republic		1		1
3794690_Czech Republic		1		1
3795109_Czech Republic			1	1
Denmark	**1**	**3**	**2**	**6**
3788742_Denmark			1	1
3788757_Denmark			1	1
3788769_Denmark		1		1
3794689_Denmark		1		1
3795109_Denmark	1			1
3795221_Denmark		1		1
England	**1**	**5**	**1**	**7**
3788745_England		1		1
3788759_England			1	1
3788772_England		1		1
3794688_England		1		1
3795187_England		1		1
3795221_England		1		1
3795506_England	1			1
Finland	**1**	**2**		**3**
3788742_Finland	1			1
3788753_Finland		1		1
3788768_Finland		1		1
France	**1**	**2**	**1**	**4**
3788751_France	1			1
3788763_France			1	1
3788773_France		1		1
3794691_France		1		1
Germany		**1**	**3**	**4**
3788751_Germany		1		1
3788764_Germany			1	1
3788774_Germany			1	1
3794688_Germany			1	1
Hungary	**1**	**2**		**3**
3788752_Hungary		1		1
3788763_Hungary		1		1
3788774_Hungary	1			1
Italy		**2**	**5**	**7**
3788741_Italy		1		1
3788754_Italy			1	1
3788766_Italy			1	1
3794685_Italy			1	1
3795107_Italy			1	1
3795220_Italy		1		1
3795506_Italy			1	1
Netherlands		**2**	**2**	**4**
3788746_Netherlands			1	1
3788756_Netherlands		1		1
3788770_Netherlands		1		1
3794690_Netherlands			1	1
North Macedonia	**1**		**2**	**3**
3788747_North Macedonia	1			1
3788758_North Macedonia			1	1
3788770_North Macedonia			1	1
Poland	**2**		**1**	**3**
3788749_Poland			1	1
3788762_Poland	1			1
3788776_Poland	1			1
Portugal		**4**		**4**
3788752_Portugal		1		1
3788764_Portugal		1		1
3788773_Portugal		1		1
3794687_Portugal		1		1
Russia	**2**		**1**	**3**
3788743_Russia	1			1
3788753_Russia			1	1
3788769_Russia	1			1
Scotland	**1**	**2**		**3**
3788748_Scotland		1		1
3788759_Scotland		1		1
3788771_Scotland	1			1
Slovakia	**1**	**1**	**1**	**3**
3788749_Slovakia		1		1
3788761_Slovakia			1	1
3788775_Slovakia	1			1
Spain		**2**	**4**	**6**
3788750_Spain			1	1
3788762_Spain			1	1
3788775_Spain		1		1
3794686_Spain		1		1
3795108_Spain			1	1
3795220_Spain			1	1
Sweden	**2**		**2**	**4**
3788750_Sweden	1			1
3788761_Sweden			1	1
3788776_Sweden	1			1
3794692_Sweden			1	1
Switzerland		**4**	**1**	**5**
3788744_Switzerland			1	1
3788754_Switzerland		1		1
3788765_Switzerland		1		1
3794691_Switzerland		1		1
3795108_Switzerland		1		1
Turkey	**1**	**2**		**3**
3788741_Turkey	1			1
3788755_Turkey		1		1
3788765_Turkey		1		1
Ukraine		**3**	**2**	**5**
3788746_Ukraine		1		1
3788758_Ukraine			1	1
3788767_Ukraine		1		1
3794692_Ukraine		1		1
3795187_Ukraine			1	1
Wales	**1**	**3**		**4**
3788744_Wales		1		1
3788755_Wales	1			1
3788766_Wales		1		1
3794689_Wales		1		1
Total	**17**	**50**	**35**	**102**

Table G2: Clustering analysis using the degree on the zone passing networks ($k = 3$)

National Team/match_id	Cluster 0	Cluster 1	Cluster 2	Total
Austria	**1**		**3**	**4**
3788747_Austria			1	1
3788756_Austria			1	1
3788767_Austria			1	1
3794685_Austria	1			1
Belgium	**3**		**2**	**5**
3788743_Belgium			1	1
3788757_Belgium	1			1
3788768_Belgium			1	1
3794687_Belgium	1			1
3795107_Belgium	1			1
Croatia	**3**		**1**	**4**
3788745_Croatia	1			1
3788760_Croatia	1			1
3788771_Croatia			1	1
3794686_Croatia	1			1
Czech Republic	**4**	**1**		**5**
3788748_Czech Republic		1		1
3788760_Czech Republic	1			1
3788772_Czech Republic	1			1
3794690_Czech Republic	1			1
3795109_Czech Republic	1			1
Denmark	**3**		**3**	**6**
3788742_Denmark			1	1
3788757_Denmark			1	1
3788769_Denmark			1	1
3794689_Denmark	1			1
3795109_Denmark	1			1
3795221_Denmark	1			1
England	**4**	**1**	**2**	**7**
3788745_England	1			1
3788759_England			1	1
3788772_England	1			1
3794688_England	1			1
3795187_England	1			1
3795221_England			1	1
3795506_England		1		1
Finland	**2**	**1**		**3**
3788742_Finland		1		1
3788753_Finland	1			1
3788768_Finland	1			1
France	**2**		**2**	**4**
3788751_France	1			1
3788763_France			1	1
3788773_France	1			1
3794691_France			1	1
Germany	**1**		**3**	**4**
3788751_Germany			1	1
3788764_Germany			1	1
3788774_Germany			1	1
3794688_Germany	1			1
Hungary		**3**		**3**
3788752_Hungary		1		1
3788763_Hungary		1		1
3788774_Hungary		1		1
Italy		**1**	**6**	**7**
3788741_Italy			1	1
3788754_Italy			1	1
3788766_Italy			1	1
3794685_Italy			1	1
3795107_Italy			1	1
3795220_Italy		1		1
3795506_Italy			1	1
Netherlands	**2**		**2**	**4**
3788746_Netherlands			1	1
3788756_Netherlands	1			1
3788770_Netherlands			1	1
3794690_Netherlands	1			1
North Macedonia	**2**	**1**		**3**
3788747_North Macedonia		1		1
3788758_North Macedonia	1			1
3788770_North Macedonia	1			1
Poland		**1**	**2**	**3**
3788749_Poland			1	1
3788762_Poland		1		1
3788776_Poland			1	1
Portugal	**2**		**2**	**4**
3788752_Portugal			1	1
3788764_Portugal	1			1
3788773_Portugal	1			1
3794687_Portugal			1	1
Russia	**1**	**1**	**1**	**3**
3788743_Russia	1			1
3788753_Russia			1	1
3788769_Russia		1		1
Scotland	**2**		**1**	**3**
3788748_Scotland			1	1
3788759_Scotland	1			1
3788771_Scotland	1			1
Slovakia	**1**	**1**	**1**	**3**
3788749_Slovakia	1			1
3788761_Slovakia			1	1
3788775_Slovakia		1		1
Spain			**6**	**6**
3788750_Spain			1	1
3788762_Spain			1	1
3788775_Spain			1	1
3794686_Spain			1	1
3795108_Spain			1	1
3795220_Spain			1	1
Sweden		**2**	**2**	**4**
3788750_Sweden		1		1
3788761_Sweden			1	1
3788776_Sweden		1		1
3794692_Sweden			1	1
Switzerland	**2**	**1**	**2**	**5**
3788744_Switzerland			1	1
3788754_Switzerland	1			1
3788765_Switzerland			1	1
3794691_Switzerland	1			1
3795108_Switzerland		1		1
Turkey	**1**	**1**	**1**	**3**
3788741_Turkey		1		1
3788755_Turkey			1	1
3788765_Turkey	1			1
Ukraine	**4**		**1**	**5**
3788746_Ukraine	1			1
3788758_Ukraine			1	1
3788767_Ukraine	1			1
3794692_Ukraine	1			1
3795187_Ukraine	1			1
Wales	**2**	**2**		**4**
3788744_Wales		1		1
3788755_Wales	1			1
3788766_Wales		1		1
3794689_Wales	1			1
Grand Total	**42**	**17**	**43**	**102**