Universidade Estadual de Campinas
Instituto de Computação

Daniele Cristina Uchôa Maia Rodrigues

Complex Network Measurements in Graph-based Spatio-Temporal Soccer Match Analysis

Medidas de Redes Complexas na Análise Espaço-Temporal Baseada em Grafos de Jogos de Futebol

CAMPINAS
2017

Daniele Cristina Uchôa Maia Rodrigues

Complex Network Measurements in Graph-based Spatio-Temporal Soccer Match Analysis

Medidas de Redes Complexas na Análise Espaço-Temporal Baseada em Grafos de Jogos de Futebol

Tese apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutora em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Supervisor/Orientador: Prof. Dr. Ricardo da Silva Torres

Este exemplar corresponde à versão final da Tese defendida por Daniele Cristina Uchôa Maia Rodrigues e orientada pelo Prof. Dr. Ricardo da Silva Torres.

CAMPINAS
2017

Agência(s) de fomento e nº(s) de processo(s): Não se aplica.

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

R618c

Rodrigues, Daniele Cristina Uchôa Maia, 1979-
Complex network measurements in graph-based spatio-temporal soccer match analysis / Daniele Cristina Uchôa Maia Rodrigues. – Campinas, SP : [s.n.], 2017.

Orientador: Ricardo da Silva Torres.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Redes complexas. 2. Redes temporais. 3. Futebol. I. Torres, Ricardo da Silva, 1977-. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Medidas de redes complexas na análise espaço-temporal baseada em grafos de jogos de futebol
Palavras-chave em inglês:
Complex networks
Temporal networks
Soccer
Área de concentração: Ciência da Computação
Titulação: Doutora em Ciência da Computação
Banca examinadora:
Ricardo da Silva Torres [Orientador]
Paulo Roberto Pereira Santiago
Silvio Jamil Ferzoli Guimarães
André Santanchè
Milton Shoiti Misuta
Data de defesa: 29-09-2017
Programa de Pós-Graduação: Ciência da Computação

Universidade Estadual de Campinas
Instituto de Computação

Daniele Cristina Uchôa Maia Rodrigues

Complex Network Measurements in Graph-based Spatio-Temporal Soccer Match Analysis

Medidas de Redes Complexas na Análise Espaço-Temporal Baseada em Grafos de Jogos de Futebol

Banca Examinadora:

- Prof. Dr. Ricardo da Silva Torres
  IC - UNICAMP
- Prof. Dr. Paulo Roberto Pereira Santiago
  EEFERP - USP
- Prof. Dr. Silvio Jamil Ferzoli Guimarães
  ICEI- PUC-Minas
- Prof. Dr. André Santanchè
  IC - UNICAMP
- Prof. Dr. Milton Shoiti Misuta
  FCA - UNICAMP

A ata da defesa com as respectivas assinaturas dos membros da banca encontra-se no processo de vida acadêmica do aluno.

Campinas, 29 de setembro de 2017

If I have seen further, it is by standing on the shoulders of giants.

(Sir Isaac Newton)

Acknowledgements

First of all, I would like to thank God for the health and careful family that I have been granted. Without these aces, I would not be able to seize all the great opportunities I have been given in my life.

My husband, Clayton and my daughters, Isadora and Raíssa for your love and patience to my lack of time and attention in the last few years. Your love and comprehension were essential to the completion of this project. To my husband, I am deeply thankful for being so careful to our daughters in so many moments of my absence.

My parents, Arlinda and Elaerto and my brother Carlos, for the incredible and loving family that you are. Thank you also for the support with my children in so many moments. I am very thankful to my parents who, since my early childhood, have taught me to believe that with determination and hard work, dreams can come true. Thank you so much for never letting me give up.

I thank my advisor, Prof. Ricardo da Silva Torres, for all the support received over the last few years. Thank you for all the teachings, patience, encouragement, revisions, and collaborations. Working with you was an incredible opportunity.

I can not forget the support of Prof. Felipe Arruda Moura and Prof. Sergio Cunha with the soccer analysis and insights. Professor Felipe, although far away, has always been extremely helpful and thoughtful in his collaborations.

I thank the professors of PUC-Campinas, who graduated me, and now support me as work colleagues, Prof. Juan, Prof. Pannain, Prof. Freitas, and Prof. Tobar, for the encouragement to complete this work.

I am thankful to the professors of the Institute of Computing at UNICAMP, who contributed to my growth. And also the technical and administrative staff of the IC for their support.

I thank PUC-Campinas for its support during the development of this research.

This work has been partially funded by FAEPEX, CAPES, CNPq, and São Paulo Research Foundation – FAPESP (grant #2016/50250-1).

Resumo

A análise de partidas de futebol é de suma importância na definição de programas de treinamento apropriados e estratégias de jogo. A crescente disponibilidade de dados relacionados ao esporte nos últimos anos, devido ao uso de sistemas de rastreamento modernos, permitiu avanços em análises esportivas, proporcionando aos treinadores informações valiosas para análise de times e partidas. A disponibilidade desses dados, por outro lado, desafia a Ciência a desenvolver ferramentas capazes de armazenar, visualizar e analisar esse grande volume de informações. Análises no futebol são geralmente realizadas usando estatísticas de partidas, eventos do jogo (por exemplo, passes e finalizações) e os dados de localização dos jogadores. Estudos relacionados têm representado os eventos dos jogos como um único grafo, em que os jogadores são vértices e as arestas são ações realizadas entre eles durante a partida. O grafo é então analisado sob a perspectiva de medidas de redes complexas. Embora as abordagens existentes ofereçam informações relevantes sobre as ações táticas ocorridas durante o jogo, revelando alguns padrões táticos, desconsideram os aspectos espaço-temporais inerentes ao esporte, como o posicionamento dos jogadores no campo e o momento no tempo que ações relevantes ocorrem.

Esta tese trata destes problemas ao apresentar um *framework* de análise de jogos de futebol. Para tanto, propõe-se uma nova abordagem para a análise de partidas de futebol, baseada em grafos que considera as características espaço-temporais, intrínsecas a esse esporte dinâmico. As partidas de futebol foram representadas como grafos temporais, codificando a localização dos jogadores em grafos instantâneos. Nestes grafos, os vértices representam os jogadores em sua localização real e as arestas são definidas com base na distância entre eles no campo e na possibilidade de trocas de passes curtos. Demonstramos que essa representação, denominada *Opponent-Aware graph*, que leva em consideração a presença de adversários, e a medida de entropia de diversidade são ferramentas efetivas para determinar o papel dos jogadores atacantes em uma partida e a probabilidade de passes bem-sucedidos. Considerando diferentes medidas de redes complexas em grafos temporais, este estudo também investiga a viabilidade da utilização de medidas de redes complexas e algoritmos de aprendizado de máquina para caracterizar o papel dos jogadores em uma partida. Os resultados permitem caracterizar melhor o processo de tomada de decisão dos jogadores, fornecendo informações relevantes para treinadores e pesquisadores para, possivelmente, melhorar estratégias de treinamento. Este estudo também aborda o problema de visualização de grafos temporais, introduzindo o *Ritmo Visual de Grafos* (do inglês *Graph Visual Rhythm*), uma nova representação baseada em imagem para visualizar padrões de mudança tipicamente encontrados em grafos temporais. Esta representação é baseada no conceito de ritmos visuais, motivada pela sua capacidade de codificar uma grande quantidade de informações contextuais sobre a dinâmica de grafos de forma compacta. A utilização dos ritmos visuais de grafos foi realizada através da criação de uma ferramenta de análise visual para apoiar o processo de tomada de decisão com base em análises de partidas de futebol baseadas em redes complexas.

Abstract

Soccer match analysis is of paramount importance in the definition of appropriate training programs and game strategies. The increasing availability of sport-related data in the recent years, due to the use of modern tracking systems, has allowed advances in sports analytics, providing coaches with valuable information for match and team analysis. The availability of these data, on the other hand, challenges science to develop tools capable of storing, visualizing, and analyzing this large volume of information. Soccer analyses are usually performed using matches' statistics, events (e.g., passes and shots on goal) and players location data. Related studies have been representing the matches' events as a single graph, where players are vertices and edges are actions performed among them during the match. The graph is then analyzed from a complex network measurement perspective. Although this approach provides interesting insights about the tactical actions occurred during the game, revealing some tactical patterns, it disregards the spatio-temporal aspects inherent to the sport, as the positioning of the players on the pitch, and the moment in time when relevant actions occur.

This thesis addresses these shortcomings by presenting a soccer game analysis framework. We propose a new approach for soccer match analysis, based on graphs, that considers the spatio-temporal characteristics, intrinsic to the dynamic of soccer. We propose to represent the match as a temporal graph, by encoding players' location on the pitch into instant graphs. In these graphs, vertices represent players in their real location and edges are defined based on their distance in the field and the possibility of short pass exchanges. We demonstrate that this representation, named *Opponent-Aware graph*, which takes into account the presence of opponents, and the diversity entropy measurement are effective tools for determining the role of attacking players in a match and the probability of successful passes. By taking into account different measurements of complex networks in temporal graphs, this study also investigates the feasibility of using complex network measurements and machine learning algorithms to characterize the role of players in a match. The results allow to further characterize the decision-making process of players, providing interesting insights to coaches and researchers for possibly improving training strategies. This study also addresses the visualization of temporal graphs problem by introducing the *Graph Visual Rhythm*, a novel image-based representation to visualize changing patterns typically found in temporal graphs. This representation is based on the concept of visual rhythms, motivated by its capacity of providing a lot of contextual information about graph dynamics in a compact way. We validate the use of graph visual rhythms through the creation of a visual analytics tool to support the decision-making process based on complex-network-oriented soccer match analysis.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Soccer is one of the most popular and important sport in the world [75]. It is a sport that can be practiced by men, women, and children. Annually professional and amateur soccer practices move large volumes of money, usually associated with tournaments, championships, and negotiations of players between teams. The availability of sport-related data, usually obtained by means of monitoring systems [44, 76], has allowed the growth of initiatives of the scientific community of different areas targeting the creation of effective approaches for sport analysis. Frencken et al. [48], for example, claimed that to better understand the dynamics of a soccer match, it is necessary to identify the variables that capture the flow of the match. The performance of the teams in matches depends on a large extent on the strategic actions adopted by coaches [77] along with the players' organization on the pitch. The quantitative analysis of the games can provide accurate information so that coaches can plan more easily appropriate tactical actions in matches [107], potentially improving the performance of their players and consequently of their teams. Recent work by Fister et al. [45] suggests, for example, that the use of computational intelligence techniques may provide tools to support the decision-making processes of coaches.

The availability of spatio-temporal data from monitoring systems challenges the academic community to develop tools for storing, processing, analyzing, and visualizing this kind of information. The processing and analysis mechanisms should consider the large volume of data generated in order to present summarized and concise results, which improve the quality of historical information for training strategies.

This study presents a review of the literature related to the use of networks for soccer game modeling, and proposes new approaches to soccer matches modeling, considering the spatio-temporal characteristics of the sport, presents the use of complex network measurements as features to characterize matches and players' actions, and also methods for summarizing and visualizing temporal graphs.

1.1 Motivation

Several studies have been conducted in recent years with the goal of analyzing the performance of players and teams in sport matches, along with game style. It is important to highlight that still, there is no formal definition to the concept of style. Hewitt et al. [56] provide a framework with metrics for assessing game style, in an attempt to give it a quantitative meaning. Some studies analyze performance, focusing on statistical analysis, from historical information of team matches during championships, e.g., number of goals and faults; while others perform analysis based on individual players' information during the matches, such as location data, player role, technical actions performed, among others. Sarmento et al. [98] provide an overview of the most used techniques for soccer match analysis.

The analysis of soccer matches requires the exploration of individual and collective information of players on the pitch, since their performances change according to dynamic patterns formed with their teammates and their opponents during the match [107]. The team distribution on the pitch during the match reveals tactical organization and contributes to the productivity of the player [72]. Memmert et al. [72] present an overview of studies on analysis of position data in soccer. In addition, with position data, capturing the interaction between the players is of paramount importance to understand soccer game dynamics. If goals are rare events in most matches, passes, on the other hand are plentiful [55]. Some recent works have been using complex networks (and graphs) techniques to analyze sports. Those techniques, has already been successfully used on analysis of complex systems, such as social networks [63]. In these networks, the relationships between vertices usually reveal interesting patterns [51], which fomented the development of methods associated with the link mining activity. The link mining activities also consider the fact that some complex system structures are highly dynamic, and the understanding of their temporal evolution mechanisms is not yet complete, but can be improved by extracting quantitative measures from those networks [97]. The use of complex network concepts in the analysis of real networks was the basis for the studies of interactions between athletes in collective games, since, in several sports, the relations between the players are fundamental for the good performance of the team as a whole. The use of complex networks concepts in sports was presented, for example, in [5, 86, 93].

Recent researches in soccer analysis have shaped the relationship of players during a soccer match from different perspectives. A study venue explores the extraction of a single graph representing the whole match, or a time interval, whose vertices are players and the edges are relationships between them, such as successful ball passes. From these graphs, it is possible to extract complex networks measurements. In [38], for example, flow networks approaches represent the ball passes between players at specific moments of the game. A second strand of studies analyzes the existence of polygons formed by the positioning of the players on the pitch [24], and how the positioning of these polygons is related to match events [29], [77] and [107].

Araujo and Davis [2] have discussed the lack of studies focusing on the dynamic behavior of sports when modeled as networks, which can only represent an instant or period of time, even though the connections are in constant change. Ribeiro et al. [93] contextualized the use of social networks to model sport teams, showing the potential benefits in the evaluation of interactions and collective performance, as well as some limitations of their use, especially concerning the use of a single graph to model passes accomplished between players. The use of a single graph precludes the characterization of dynamic aspects of soccer matches.

In fact, soccer matches can be represented as temporal networks, since the interaction among players occurs in time, and the mechanism of understanding the evolution of events of interest in a match depends on the temporal ordering of these interactions. Temporal networks are those in which edges between vertices exist only at specific instants in time, and whose analysis depends on the knowledge of these instants [57,67]. Several studies on temporal networks show that there are specific measures that can be extracted from these networks, considering the activation moment of the edges, such as connectivity, paths, distance, diameter, centrality, among others [57, 62]. Examples of applications include social network analysis [16,111], sport analysis [35,38,86,87], and urban planing [80]. Also, it is important to notice that the availability of very large amounts of data, represented as temporal networks, demands the development of appropriate tools for analysis and visualization of temporal pattern changes.

1.2 Hypothesis and Research Questions

By taking into account the importance of the analysis of soccer matches, there is a lack of studies in the area that consider both the spatial aspects of the players on the pitch, and the temporal nature of the game that determines the dynamics of actions during the matches. While a great amount of the related studies use networks and their measures as features to analyze soccer matches, modeling the entire match using a single graph affects the effectiveness of conducted analysis as the temporal dynamics of the sport are ignored. Some concepts, such as the order and the moments of activation of the passes are lost by compacting all the information into a single representation. This model also neglects the spatial aspects of the match, evidenced by the players' positioning on the pitch.

Given the provided context, the main hypothesis addressed in this study is: *Temporal graphs and associated complex-network measurements are effective to model the spatio-temporal dynamics of soccer matches and potentially improve soccer matches analyses.*

Given this hypothesis, the following research questions are defined:

1. Which graph model better captures spatio-temporal aspects of soccer matches based on players' location on the pitch?

2. Which information visualization approach would be suitable for supporting the analysis of temporal changes in dynamic graphs?

3. Which complex networks measurements better characterize events of interest in soccer matches?

4. Which complex networks measurements better characterize the players' role? Is it possible to fully classify players according to their associated complex network measurements?

1.3 Contributions

This study provides contributions in different domains, such as Computer Science and Sport Science. Specific contributions refer to the areas of data science, sports analytics, and graph visualization and are summarized in the following:

1. Sports analytics framework, based on a complex network modeling, which is validated on soccer match analysis tasks (Chapter 3);

2. A novel approach for soccer modeling, named *Opponent-Aware Graphs*, which considers the spatio-temporal aspects of the sport (Chapter 3);

3. A novel approach for visualizing temporal graphs, named *Graph Visual Rhythms* (Chapter 4);

4. A soccer analytics visual tool intended to highlight aspects of the game (Chapter 4);

5. Identification of complex network measurements related to relevant events of soccer matches (Chapter 4);

6. Identification of complex network measurements related to the players' role characterization (Chapter 5).

1.4 Thesis Organization

The remaining of this text is organized in six chapters. Chapter 2 presents key concepts about complex networks, measures typically used in their analysis, studies on modeling and analysis of soccer matches, as well as concepts of visual rhythms. Chapter 3 presents the *Opponent-Aware Graph* representation, a novel method for modeling soccer games as graphs. We also demonstrate the use of Diversity Entropy measurement to assess the roles of players in a match. Chapter 4 introduces the *Graph Visual Rhythm* concept, a novel tool for visualizing temporal graphs. We describe the technique and introduce its use in soccer match analysis tasks. Chapter 5 presents the classification of players, according to their roles in the match. We use Opponent-Aware Graphs to extract complex network measurements, and machine learning algorithms to assign roles to players, according to their measurements. Graph Visual Rhythm images are generated to assess the results. Chapter 6 summarizes the main contributions of the thesis and draws possible directions for future work.

Figure 1.1 presents a schematic flow relating the subjects studied and the main contributions in each chapter.

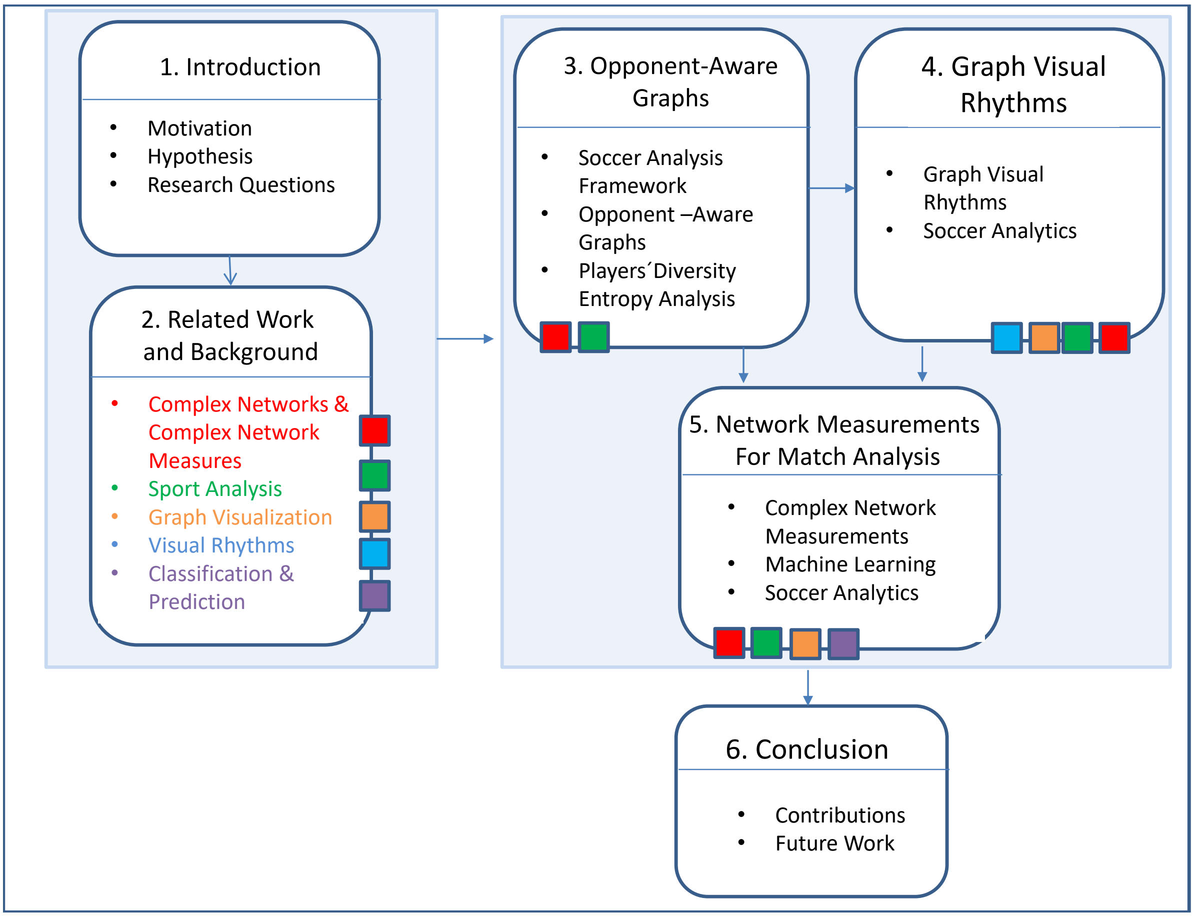


Figure 1.1: Schematic distribution of the thesis content along its chapters. Main concepts colored in Chapter 2 are used as background for Chapters 3, 4, and 5.

Chapter 2

Related Work and Background Concepts

This chapter presents basic concepts that support this thesis. The first section is dedicated to the presentation of basic concepts and measures associated with complex networks. These concepts will be used to model soccer matches, and to characterize events of interest. Next, we survey the most representative related work, aiming to characterize the main methodologies that have been used in sports analysis, focusing on soccer. We then present concepts of graph-based visualization and visual rhythms, which will be explored later in a novel approach to visualize temporal graphs. Next, we describe related studies on classification approaches and prediction usually used in soccer analysis. Finally, the dataset used in this research is detailed.

2.1 Complex Networks

2.1.1 Definition

Complex networks are a representation of systems that have a complex structure of connection among their elements. These systems can be modeled as graphs, where the vertices represent real elements or abstractions of relevant information, and the edges represent connections among the vertices. These links can be physical as a power cable linking two towers, or abstract, as interactions among people in social networks. The analysis of these networks is originated in graph theory studies, with the extraction of measurements in small systems. However, the progress in the use of technologies for the acquisition and storage of large data volumes has allowed the generation of large size graphs, reaching millions of vertices, which facilitated the perception of common characteristics in graphs used in several application areas.

The interest of the scientific community for complex networks began with the extraction of measurements in graphs to characterize the topology of systems. The most relevant models for the generation of complex networks are:

- the Erdős and Rényi model [92], which is based on the generation of random graphs;

- the model of Watts and Strogatz [108], which presents the occurrence of loops of order three (also known as clustering or transitivity) in real networks, demonstrating a significant difference of the Erdős and Rényi random graphs, and the characterization of the "Small World" phenomenon; and
- the model of Barabási and Albert [6], which differentiates real networks from random networks. They show that, in real networks, the degree of vertices follows a scale-free distribution pattern, that is, few network vertices have a high degree of connection, while most of them have a low degree of connection.

Complex networks have been used to model and analyze systems from different disciplines. Among the most common applications are social networks, the internet, metabolic networks, protein networks and genetic networks, telecommunications systems, brain networks, among other areas that have benefited from this approach [13, 33, 81]. Detailed surveys on the concepts of complex networks, and the main types of measures extracted from these networks can be found in [1, 13, 34, 81].

2.1.2 Basic Concepts on Graphs

Graphs can be classified as directed (digraph) and undirected, considering the existence of directions on the edges. Let $G_D = (V, E)$ be a digraph, V be the vertices of G_D, and E be the edges of G_D. For each edge $e \in E$, $e = (v_i, v_t)$, where $v_i \in V$ is the starting vertex, and $v_t \in V$ is the ending vertex.

Let $G = (V, E)$ be a graph, two vertices $v_i \in V$ and $v_j \in V$ are adjacent if there is an edge $e_{ij} \in E$ connecting them. For each edge $e \in E$, we may have a weight $w(e)$ associated. In this case, G is called a weighted graph. If all the edges have equal value $w(e) = 1$, than the graph is unweighted. In many cases, the weight can be associated with the existence of multiple edges between the vertices.

In most networks, any two vertices are usually not adjacent, once only a few number of all possible edges exists [34]. The reachability of nodes in a network is of paramount importance in the network analysis, and this concept is present in many network measures [13]. Considering two non-adjacent vertices of G, $v_i \in V$ and $v_j \in V$, they could be connected by a chain of edges, interleaved with vertices: $W =< v_i, e_i, v_{i+1}, e_{i+1}, \dots v_j >$, called walk. If all the edges and vertex of W are distinct (there is no repetitions), we call this walk a path. For more details on graph concepts, the reader may refer to [14, 109].

2.1.3 Complex Network Measurements

The measurements extracted from complex networks help understanding their behavior, their main topological characteristics, and allow their comparison. The description of different measurements in complex networks can be found in [1, 34]. Although there is a large number of measures, some should be chosen for the analysis of networks according to the nature of the target application [96].

In the following, we present the most relevant measurements for the development of this study. They are related to distance, cycles, entropy, centrality.

Distance Related Measurements

The distance among vertices in a network is an important measure in several types of applications. It usually reveals relevant structural characteristics, since it has great dependence on the internal structure of the network. Distance-related measures are based on paths between the vertices of the network. The length of a path between two vertices is calculated by the number of edges present among vertices.

***Average Distance (l)*:** The average distance measurement l is associated with the average distance between any two network vertices (geodesic distance). More formally:

$$l = \frac{1}{N(N-1))} \sum_{i \neq j} d_{ij} \tag{2.1}$$

where i and j are any two distinct vertices in a network with N vertices. The length (distance) of the minimum path between i and j is called d_{ij}. Low values for l indicate highly connected networks with small average distances between vertices. This measure is not representative for networks that have large number of disconnected vertices, because the value of l will be low, even though the network is disconnected.

***Eccentricity (e)*:** The eccentricity is a vertex measurement and corresponds to the maximum shortest distance from a vertex to all others in the graph. It is usually associated with how easily accessible a vertex is from other vertices. Considering all the vertices of a graph (j), except i, the eccentricity of i is calculated as:

$$e_i = \max(d_{ij}), \tag{2.2}$$

where d_{ij} is the shortest distance between vertex i and j, $j \in V(G)$.

***Global Efficiency (E)*:** This measure is based on the average distance (l) and indicates the efficiency of sending information among vertices. The efficiency is proportional to the inverse of the distance between the vertices, as defined in [66]:

$$E(G) = \frac{1}{N(N-1))} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{2.3}$$

***Local Efficiency (E_{loc})*:** From the Global Efficiency (E) concept (Equation 2.3), it is possible to derive the concept of Local Efficiency (E_{loc}), considering the network's fault tolerance, by verifying the impact on local communication (direct neighbors) when a vertex and all its associated edges are removed. More formally, it is defined as [66]:

$$E_{loc} = \frac{1}{N} \sum_{i \in G} E(G_i), \tag{2.4}$$

where G_i is the subgraph of neighbors of vertex i and N is the vertices count in G_i.

***Vulnerability*:** Vulnerability of a vertex i is calculated as the proportional drop in Global Efficiency E when the vertex is removed. More formally:

$$V_i = 1 - \left(\frac{E_i}{E(G)}\right), \tag{2.5}$$

where E_{total} is the Global Efficiency of the graph, and E corresponds to Global Efficiency after vertex i is removed.

Cycle-Related Measurements

These measures allow the analysis of the structure of cycles and the tendency to form sets among connected vertices.

***Clustering Coefficient (C)*:** Real networks, as opposed to the random ones, usually present in their structure a great amount of loops of order three (triangles) between the vertices. The clustering coefficient characterizes the presence of these loops in graphs. This measure is also known as transitivity. Let i, j, and k be vertices of a graph G. If there are edges e_{ij} and e_{jk}, the transitivity is characterized by the existence of the edge e_{ik}.

The transitivity (clustering coefficient) is calculated as follows:

$$C = \frac{3 \times N_\triangle}{N_3}, \tag{2.6}$$

where $N_\triangle$ is the number of triangles (set of complete subgraphs of 3 vertices), and N_3 is the number of connected triples (set of 3 vertices in which each vertex can be reached from any of the other two). This is a network measure, which encodes the proportion of times in which, in the existence of a connected triple, there is also a triangle.

One can also calculate the transitivity of each vertex individually, verifying the relation between the number of connected triples in which a vertex participates as central and the amount of triangles formed from these connected triples. Let i be a network vertex, the transitivity of i is computed as follows:

$$C_i = \frac{3 \times N_{\triangle i}}{N_{3i}}. \tag{2.7}$$

***Rich Club Coefficient (φ)*:** This measure is also known as preferential attachment, and indicates a trend connection among vertices that have many connections (hubs). The rich club of degree k ($R(k)$) of a network is the set of vertices that have degree greater than k: $R(k) = \{v \in V(G)|k_v > k\}$. The Rich Club coefficient is the proportion of edges (e) between vertices belonging to the Rich Club in relation to the total amount of edges that could exist between them ($|R(k)| \times |R(k)| - 1$), and is calculated as:

$$\varphi(k) = \frac{1}{|R(k)|\,(|R(k)| - 1)} \sum_{i,j \in R(k))} e_{ij} \tag{2.8}$$

Entropy-related Measures

Entropy-related measures for complex networks give clues on the heterogeneity and resilience of networks, helping to understand their properties and operating mechanisms, since they are based on the connections and possibilities of paths among vertices.

***Diversity Entropy (E_h)*:** Diversity Entropy [100, 101] considers the transition probability ($P_h(j, i)$) that a vertex i reaches a vertex j after h steps in a self avoiding random walk. Let Ω be the set of all vertices but i. The normalized diversity entropy of a vertex i is defined as [101]:

$$E_h(\Omega, i) = -\frac{1}{log(N-1)} \sum_{j=1}^{N} \begin{cases} P_h(j,i)log(P_h(j,i)), & \text{if } P_h(j,i) \neq 0, \\ 0, & \text{if } P_h(j,i) = 0. \end{cases} \tag{2.9}$$

Centrality-related Measures

Measures related to centrality usually considers the vertices or the edges of a graph. These measures use the premise that a vertex or edge enrolled in many paths of the network is usually more important.

***Degree (K)*:** The degree of a vertex, say i considering undirected networks, is the number of edges (a) connected to this vertex, and can be defined as:

$$K_i = \sum_j a_{ij}. \tag{2.10}$$

***Betweenness Centrality (B)*:** The betweenness centrality [34] of a vertex u is quantified as the sum over all distinct pairs of vertex i,j of the number of shortest paths from i to j that pass through u ($\theta(i, u, j)$) divided by the total number of shortest paths between i and j ($\theta(i, j)$):

$$B_u = \sum_{ij} \frac{\theta(i,u,j)}{\theta(i,j)}. \tag{2.11}$$

Page Rank (p): Page Rank measures the prestige of a vertice based on the prestige of adjacent vertices that points to it. Let u be a vertex in a graph G, and B_u be the set of vertices connected to u. The page rank value of u is [88]:

$$p(u) = \frac{q}{n} + (1-q) \times \sum_{j \in B_u} \frac{p(j)}{K_{out}(j)}, \tag{2.12}$$

where n is the number of vertices, $K_{out}(j)$ is the outdegree of node j, $j \in B_u$, and q is the damping factor, a probability of performing a random walk or a random jump.

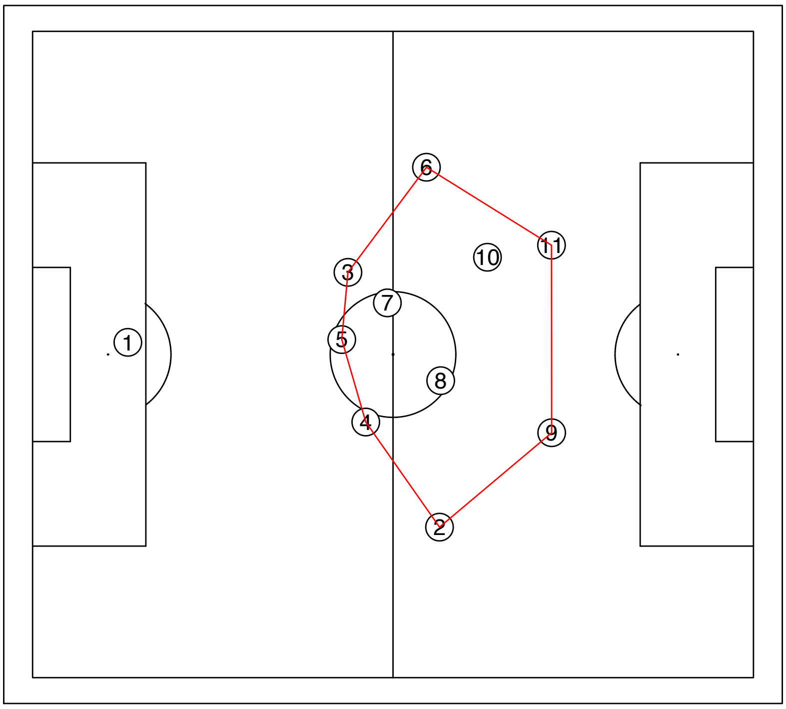


Figure 2.1: Example of a convex hull of a team, based on the location of its players on the pitch.

2.2 Sport Analysis

This section presents studies related to quantitative analysis of player distribution in sports, with emphasis on soccer. Some authors model the information through graph structures to analyze complex network measurements, while others use modeling of geometric structures, taking into account, for example, polygons formed from the physical position of players on the pitch. The following sections describe each of these approaches, and a summary table of related studies is presented at the end (Table 2.1).

2.2.1 Polygon-based Analysis

Several approaches have been proposed aiming to characterize the polygons defined in terms of the location of players on the pitch. Figure 2.1 illustrates a polygon of a team, defined in terms of the convex hull formed by taking into account the location of its players, excluding the goal keeper.

Frencken and Lemmink [47], for example, analyzed the patterns of movement of the centroid and surface area for both teams along the whole match. Their work, however, handled small-sided games. Frencken et al. [48] also present a study on the correlation of centroid measurements and surface area of the team at decisive moments of attacking formation, considering reduced teams in training games. This analysis showed the existence of teams' centroid movement patterns during the match and at decisive moments, as in the goal scoring. In another research venue, Clemente et al. [24] analyzed the correlation between centroid, stretch index, and the surface area of both teams with their tactical behavior. Clemente et al. [26] also surveys computational metrics (centroid, stretch index,

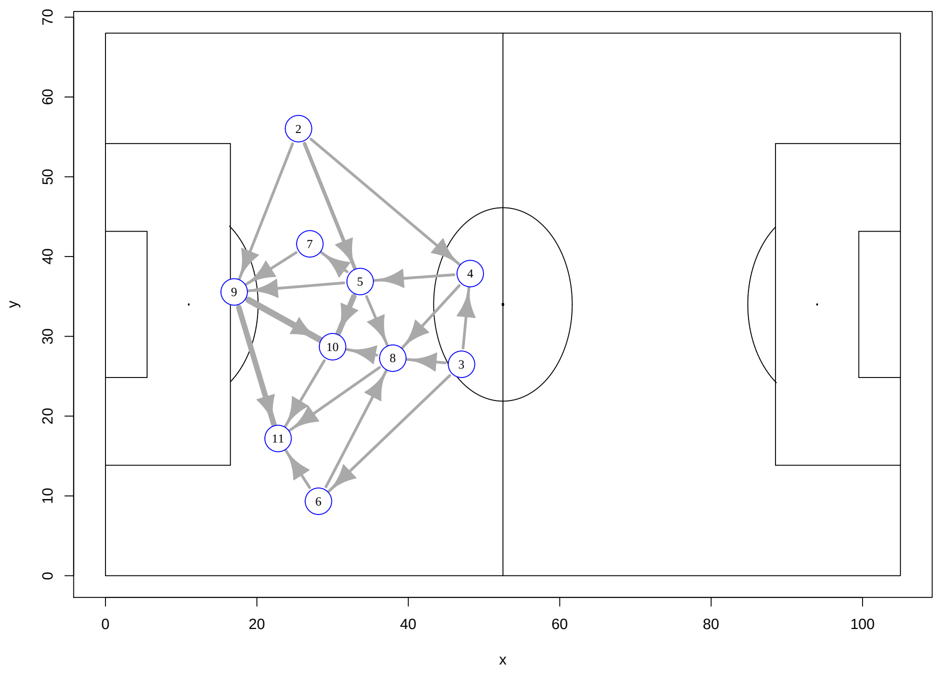


Figure 2.2: Example of flow network of a match. Players are represented as vertices, and edges represent the passes accomplished among them. The number of passes are represented in the edges' arrow width.

effective area of play, among others) and their implementation. More recently, Clemente et al. [31] showed that the goalkeeper position and the use of ball location to determine weights for players given their proximity leads to more useful centroid information. Moura et al. [76, 77] studied the efficiency of the organization of teams in attack and defensive moments, by taking into account the time series related to the surface area and stretch. The area covered by teams on the pitch was also considered in the study reported in [107]. In that case, the authors investigated the differences among different defensive strategies and attacking opportunities.

2.2.2 Network-based Analysis

Networks have been successfully used to model interactions between teammates in sports like basketball [42, 83], aquatic polo [86], cricket [37], among others. Network-based representations have been also explored for soccer match analysis [35, 38, 86, 87]. Duch et al. [38] proposed the construction of a "flow network" of a match. This network is a directed weighted graph where players are vertices, and passes completed among players defined weighted edges. The authors used network measurements (e.g., betweenness centrality) to characterize the influence of each player in the match. Figure 2.2 presents an example of a flow network of ten players in a match. The weight of edges are represented by the width of arrows (edges).

Cotta et al. [35] conducted an analysis of the successful participation of the Spanish national team at the 2010 World Cup, taking into account the complex network structure and space-time nature of its matches. On the spatial aspect, they considered the fact that the same player can perform different roles, according to their location on the pitch. They also considered the fact that a team goes through phases during a match, changing their playing style. The authors characterized the playing style of the Spanish team in relation to their opponent teams in terms of centrality measurement, clustering coefficient, and consecutive passes.

Grund [52] confirmed the hypotheses of previous studies, which state that a high centralization of interactions in a team leads to a decrease in performance, and that the greater the number of interactions in a team, the better is its performance. The experiments were carried out using a data set with 1,520 graphs extracted from 760 matches of the first division of English Premier League, extending previous studies.

Peña and Touchette [87] previously studied the importance of each player in a team and characterized the team playing style using clicks, page rank, and the clustering coefficient graph measures. The small-world network theory was used previously by Passos et al. [86] to characterize the dynamic patterns of water polo players during attacking actions. By using this representation, the authors were able to demonstrate that teams could be characterized as small-world networks, and therefore, features like preferential attachments could be observed in key players. The compatibility of small-world networks concepts and the interactions between soccer players were also investigated in [50].

Networks of passes were explored using several different measurements, which are related to technical and tactical match events. Clemente et al. [29] used network measurements to show which players have a central role in offensive actions. They also analyzed the play style and patterns of interactions of the Switzerland national team in the FIFA World Cup 2014 [30]. In their study, they converted the passes accomplished between players during attacking actions into network graphs. The measurement results revealed interesting analysis on the team's style of play. Clemente and Martins [27] also investigated whether final outcome, tactical positioning, and season affect centrality, density, and clustering coefficients. They found out that tactical positioning seems to be determinant in teammates' interactions. In another research work, Clemente et al. [28] investigated the characterization of midfield player as prominent role in the build-up of attacking plays, when compared to other roles. Malta and Travassos [70] used networks of passes to assess the most important player in defensive/attacking transitions, and the pitch areas where they received most passes. Cintia et al. [23] analyzed soccer team performance using flow networks, and the concept of zone passing network, that models ball displacement among pitch areas. In this case, vertices are zones of the pitch and edges are ball displacement among zones. This concept can be useful in discovering preferred areas, and passes distances.

Gudmundsson and Horton [53] presented a survey of team-based invasion sports, with spatio-temporal data, in which they perform non-trivial computation of these data. Temporal aspects of soccer matches were explored by Gyarmat et al. [55], which propose the characterization of the sequence of passes patterns in passes networks, the flow motifs, to reveal teams' style of playing.

Table 2.1 summarizes research initiatives that exploit complex network measurements in sport analysis, by taking into account the reference, the modeling representation, the

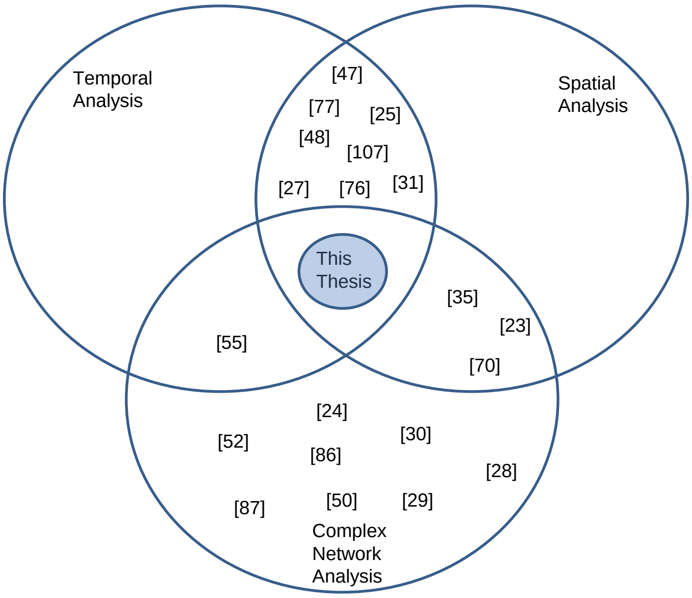


Figure 2.3: Related work, considering the three main analysis aspects of this research: temporal, spatial, and complex networks.

sport, the number of samples considered in the study, and the measurements employed. Figure 2.3 presents the comparison of related work and this thesis, considering the three main analysis aspects of this research: temporal, spatial, and complex networks.

Table 2.1: Overview of initiatives that exploit complex network measurements in sport analysis.

Authors	Representation	Sport	# of Samples	Measurements
Duch and Amaral [38]	Weighted Graphs (successful passes)	Soccer	30 matches	Betweenness Centrality
Cotta et al. [35]	Weighted Graphs with spatial information	Soccer	3 matches	Number of passes; Clustering Coefficient; Centrality
Passos et al. [86]	Weighted Graphs (passes in attacking actions)	Water Polo	1 match	Small-World; Preferential Attachment
Grund [52]	Weighted Graphs	Soccer	760 matches	Centrality; Density
Peña and Touchette [87]	Weighted Graph (passing network)	Soccer	4 matches	Centrality; Clique; Page Rank; Clustering Coefficient
Clemente et al. [30]	Weighted Graphs (passes in attacking actions)	Soccer	4 matches	Degree; Centrality; Density
Gama et al. [50]	Weighted Graphs (passes in ball possession)	Soccer	30 matches	Connectivity; Clustering coefficient; Small-World
Malta and Travassos [70]	Weighted Graphs (passes in defense/attacking transition)	Soccer	4 matches	Betweenness Centrality; Degree
Gyarmat et al. [55]	Weighted Graphs (passing network)	Soccer	380 matches	Motifs
Clemente et al. [29]	Weighted Graphs (passes in attacking units)	Soccer	1 match	Density; Centrality; Clustering Coefficient; Centroid Player
Clemente et al. [28]	Weightd Graphs (passes in attacking actions)	Soccer	109 matches	Degree; Centrality; Betweenness Centrality
Clemente and Martins [27]	Weighted Graphs (passes in attacking actions)	Soccer	17 matches	Centrality; Density; Clustering Coefficient

Continues on the next page

Authors	Representation	Sport	# of Samples	Measurements
Cintia et al. [23]	Weighted Graphs (passes among players and zone passing network)	Soccer	444 matches	Mean Degree

2.3 Graph Visualization

Many research works have been carried out with the purpose of presenting the importance of visualizing information [40, 104]. Wijk [104] suggests that exploration and presentation of data is the main use of visualization, and the value of visualization in those activities is hard to quantify. In the exploration activity, visualization provides tools so that the human vision can perceive patterns in analyzed data. Visual analytics uses the visualization techniques integrated to analysis algorithms in many disciplines, including data mining, data management and analysis, among others, as defined by Keim et al. [61]. They also claim that one of the visual analytics goals is the possibility of synthesizing massive information allowing insights.

Information modeled as graphs is usually presented as static graphs. However, some data must be represented by temporal/dynamic graphs, or graphs that change over time, specially in areas that observe dynamical or evolving behavior, as social networks, spread of diseases, among others. This challenges researchers for the development of appropriate visual analytics tools.

Beck et al. [7,8] present a survey reporting the most representative methods proposed for visualizing dynamic graphs. According to their survey, the approaches can be subdivided in two macro groups: animation and timeline. Animation approaches consist of a sequence of snapshots of graphs in time, concatenated in an animation. Timeline approaches, in turn, present the sequence of static graphs in an image. Many studies have been carried out to provide comparative evaluations of animation and timeline approaches in different perspectives, as performance, response time, accuracy, among others [3, 39]. Under those macro groups, several approaches have been proposed to the visualization of temporal graphs [8, 15, 16, 60]. Most of the approaches rely on the use of node-link diagrams, where different visual marks (typically circle glyphs) are used for representing vertices and lines to visually represent relations among vertices. Different additional visual properties associated with visual marks (e.g., position, size, length, angle, slope, color, gray scale, texture, shape, animation, blink, motion) are employed to highlight properties associated with both vertices and edges [8]. A typical challenge faced by those initiatives refers to the visualization of huge volumes of data. In these scenarios, complex interaction controls have been proposed to handle occlusion and to support browsing activities over graph data.

2.4 Visual Rhythm

Visual Rhythm is a sampling method widely used to video processing and analysis [21, 54, 82]. Its objective is to transform tridimensional information into bidimensional images by sampling one dimensional information from video frames. Let V be a digital video (in domain $2D+t$) composed of T frames f_t, i.e., $V = (f_t)$, $t \in [1, T]$, where T is the number of frames. Let H and W be, respectively, the height and width from each frame f_t.

The visual rhythm computation consists in using a function to map each f_t into a column of an image in domain $1D+t$. The final image generated is known as visual rhythm image (VR). More formally, the computation of the VR image is defined as follows [54, 82]:

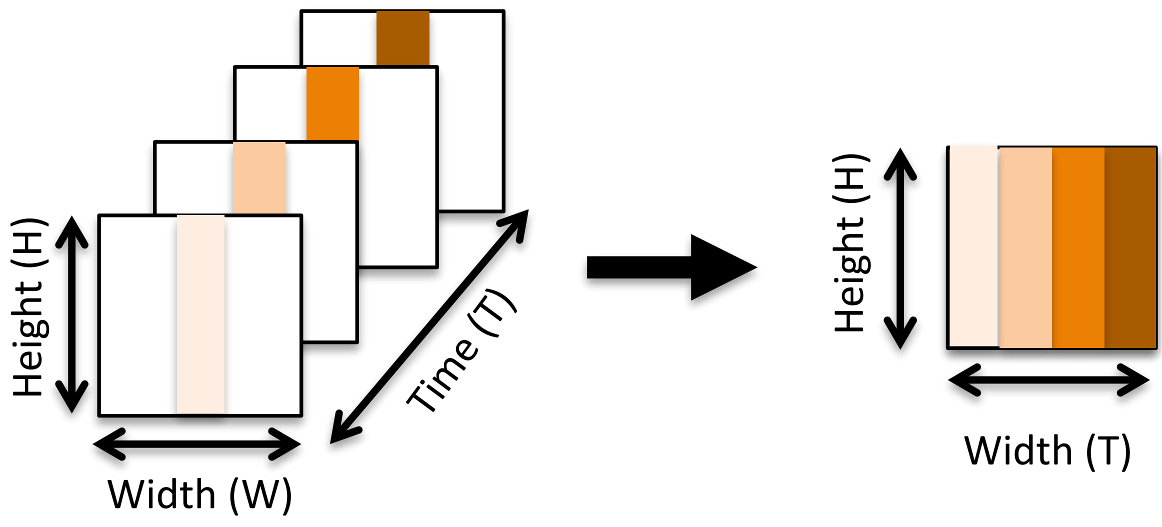


Figure 2.4: Example of visual rhythm computed by extracting the pixel values defined by the central vertical line. In this example, $r_x = 0$, $r_y = 1$, $a = \frac{W}{2}$, and $b = 0$. This leads to a visual rhythm $VR = f_t(\frac{W}{2}, z)$, where $z \in [1, H_{VR}]$ and $t \in [1, T]$, $H_{VR} = H$ is the height of the visual rhythm image, and T is its width.

$$VR(t, z) = f_t(r_x \times z + a, r_y \times z + b), \tag{2.13}$$

where $z \in [1, H_{VR}]$ and $t \in [1, T]$, H_{VR} and T are, respectively, the height (i.e., $H_{VR} = H$) and the width of the visual rhythm image; r_x and r_y are ratios of pixel sampling; a and b are shifts on each frame. Figure 2.4 illustrates the computation of visual rhythm based on the pixel values defined by the vertical line passing in the center of the frame.

A more general definition of visual rhythms assumes that it is possible to use a function $\mathbb{F}$ to represent each frame of a video as point in an n-dimensional space. Let f_t be a frame defined in terms of $\mathbb{D}$, a set of pixels. Function $\mathbb{F}$ is defined as $\mathbb{F} : \mathbb{D} \to \mathbb{R}^n$. For example, a widely used implementation of function $\mathbb{F}$ relies on the computation of the histogram associated with each frame f_t [54]. In this case, the visual rhythm image is a 2D representation encoding all frame histograms as vertical lines, i.e.,

$$VR(t, z) = \mathcal{H}(f_t), \tag{2.14}$$

where $\mathcal{H}(f_t)$ is a function that computes the histogram of frame f_t, $t \in [1, T]$ and $z \in [1, L]$, T is the number of frames and L the number of histogram bins.

2.5 Classification Approaches and Prediction in Soccer Analysis

In the analysis of soccer matches, several studies have been deepened in the characterization of team's performance, and actions that lead to successful performance, like passes and goal shots, to obtain classifiers for prediction. Those studies mainly use spatio-temporal information with machine learning algorithms to search for patterns in this big data.[1] Usually, they consider players' trajectories, match events, and networks of flow of passes.

Bialkowski et al. [12], for example, proposed a method to analyze players' role, according to their location on the pitch along the match. They consider the role to be the area for which each player is responsible on the pitch. They perform individual analysis and formations, considering that the players' role changes on time. Cintia et al. [22] uses flow networks and zone passing networks to propose classifiers to predict the results of matches considering pass-based performance indicators, which summarize the passing behavior of a team, like passes accomplished among players, zone where the passes occurred, among others.

The teams' style identity was investigated by Bialkowski et al. [11]. They used match statistics, location of the ball possession, that they call ball occupancy, and team formation as descriptors to characterize each team's style identity, allowing prediction activities. In this same context, Pappalardo and Cintia [85] used machine learning to investigate team's success according to their performance, defined as a vector of features considering goalkeeping actions, interceptions, passes, shots, among others.

By taking into account successful passes and possibility of goals, Horton et al. [58] proposed a system to perform evaluation of passes based on spatio-temporal match information. They extracted features from players' trajectories to train an SVM classifier for passes that have been previously evaluated by experts. Van Haaren et al. [102] used inductive logic programming with match events data to investigate patterns that could lead to goal attempts. In another research, Van Haaren et al. [103] also investigated event sequences that could lead to goal attempts, in order to discover attacking strategies. Fernando et al. [41] clustered players' trajectories to compare scoring methods of different teams. Lucey et al. [69], in turn, estimated the chances of scoring a goal, using logistic regression and considering the spatio-temporal information before a shot.

2.6 Data Collection and Dataset

We used a dataset related to ten official soccer matches of the Brazilian Professional First League Championship. This dataset is comprised of the location (defined in terms of the coordinate) of players of both teams. The dataset also contains a list of technical actions performed during a match (e.g., shots on goal and passes) along with a timestamp, which encodes when that particular event occurred. The players' location data were collected in a rate of 30 frames per second, amounting to at least 162,000 frames per match. Notice that technical actions data are qualitatively and quantitatively different from players location. The amount of data is smaller, once a technical action is captured when an event occurs. On the other hand, they lead to rich analysis because they carry details of the type of event and the location of the player who performed the action [53].

The players' location was tracked using the Dvideo Software [44]. DVideo software has an average error of 0.3m for the player position determination, and an average error

[1] Rein and Memmert [91] provide a relevant discussion on big data and soccer analysis.

of 1.4% for the distance covered by players [44]. Around 94% of the locations were determined automatically. Trained operators handled the remaining complex tracking situations (e.g., occlusions).

Technical actions, in turn, were defined manually by expert operators. The intra-rater data reproducibility analysis was performed with a 15-day interval between the test and retest. For inter-rater reproducibility, two independent analyses were performed. The two raters have at least 70 hours of experience with the registration process of the technical actions. The operators registered approximately 1,450 events, considering 15 different technical actions. The reliability of the data was evaluated using the kappa coefficient [32]. The values of agreement were $k = 0.9777$ for intra- and $k = 0.9390$ for inter-rater. The intraclass correlation coefficient (ICC) and the 95% confidence intervals (CI) were also calculated to verify the agreement of the measurements [71]. The values obtained were 0.9998 (CI: 0.9995-0.9999) and 0.9995 (CI: 0.9987-0.9998) for intra and inter-rater, respectively. The kappa coefficient and ICC for both situations are considered almost perfect agreement according to the interpretation suggested by Landis and Koch [64]. Furthermore, to supplement this assessment, we quantified the errors frame-by-frame and calculated the percentage error referring to the total frames, for intra (3.36%) and inter-rater (4.79%). Hughes et al. [59] suggest that an acceptable percentage of error is $< 5\%$. Further information about this process is available in [76]. In this work, we only used the first half time of each game in the dataset, as we wanted to avoid dealing with any noisy data caused by substitutions.

Chapter 3

Opponent-Aware Graphs in Soccer Analysis

3.1 Introduction

Graph-based approaches have been successfully used for sports analyses. In fact, some works have been dedicated to the characterization of match dynamics and its complexity using the complex network theory [5, 86]. One drawback of those initiatives relies on the fact that the graph representation explored is usually based on passes. Players are modeled as vertices, while edges linking players are defined if a given player has passed the ball to a teammate. Furthermore, only one graph is constructed to represent the whole match, which limits the use of this representation in the understanding of the match dynamics over time. In fact, a match could be represented as a temporal network, in which edges only exist in specific instants of time. Using this approach allows the investigation of the correlation of graph-based specific measures (such as connectivity, paths, distance, diameter, centrality, among others) with the activation window of the edges [57,62]. These analysis are important once they improve soccer dynamics comprehension, considering the temporal aspects of complex players' interactions. Its use may also be important in the identification, characterization, understanding, and possibly decision-making process related to temporal soccer patterns (e.g., passes and tactical strategy).

One of the research questions of this thesis concerns to defining which graph model better captures spatio-temporal aspects of soccer matches based on players' location on the pitch. This chapter addresses this issue by presenting our proposal to the characterization, over time, of multiple graphs defined in terms of the position of players on the pitch, which allows the analysis of both spatial and temporal characteristics of soccer matches. We also propose a novel graph-based representation, named *Opponent-aware Graph*, which takes into account the position of opponents in the match dynamic analysis. Finally, we investigate the use of the complex network diversity entropy measures in the analysis of those graphs. To the best of our knowledge, this is the first work dedicated to the use of this measure in the context of soccer match analysis.

Performed analyses considering nine professional soccer matches demonstrated that the proposed opponent-aware graph representation is appropriate for soccer match dynamic analysis. We also demonstrated that there is a correlation between the roles of players in a match with the diversity entropy scores. Finally, we verified that there is a correlation between the diversity entropy scores and the frequency of occurrence of successful passes between players, which is an important issue for the soccer analysis considering the high demand on the decision-making process of attacking players. Also, it provides interesting insights to coaches and researchers on training strategies.

The remainder of this chapter is organized as follows. Section 3.2 introduces the proposed opponent-aware graph representation, as well as the diversity entropy complex network measure. Next, Section 3.2 describes the analysis framework proposed for soccer match analysis context. Finally, Section 3.3 discusses results obtained.

3.2 Materials and Methods

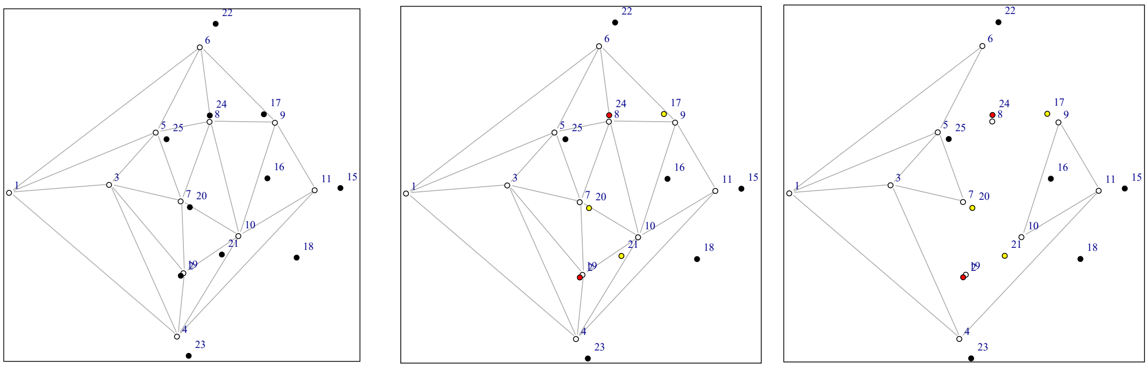
This section describes our graph model proposal, the opponent-aware graph representation (Section 3.2.1), the diversity entropy measure (Section 3.3.2), the analysis framework used in this work (Section 3.2.3), as well as the statistical analysis performed (Section 3.2.4).

3.2.1 Opponent-Aware Graph-based Analysis

Our analyses are based on the characterization of players according to their location on the pitch over time. We use a graph-based representation to encode the location change patterns. Let $G_i = (V_i, E_i)$ be a weighted graph at timestamp $t_i \in T$ composed of a set of vertices, V_i, and a set of edges, E_i. According to this representation, a vertex $v \in V_i$ represents a player, whereas an edge $e_{jk} \in E_i$ connecting two vertices $v_j \in E_i$ and $v_k \in E_i$ is defined based on the location of players (v_j and v_k) of the same team at timestamp i. We refer to the graph defined at a particular timestamp i (say G_i) as *instant graph.* The weight $w(e_{jk})$ is defined by the Euclidean distance of players j and k in the field.

Our goal is to represent the possibility of passes among players at each instant of a match. By building one graph per team for each instant of time considering the players' location, it is possible to capture the space and temporal nature of the match dynamics. In this study, we propose an opponent-aware graph representation, which encodes the possibility of passes among players, considering the position of opponents. The presence of opponent players nearby reduces the chances of successful passes among teammates.

Initially, edges are defined by computing the Delaunay triangulation [46] considering the location of players of the same team. Note that the Delaunay triangulation, for the situation in which all the vertices are contained in the same plan, defines a planar graph, with no edges crossing each other. In this way, we obtain a graph that determines the neighborhood of each vertex, representing the shortest paths for ball passing among teammates. In the following, we consider the proximity of opponent players to remove edges from the triangulation graph, once opponent players nearby disrupt the ball flow among teammates. Let $\mathcal{L}_{j,k}$ be the segment line bounded by players j and k and $p = (x, y)$ be the closest opponent player to $\mathcal{L}_{j,k}$. An edge e_{jk} is removed from E_i, if $d(\mathcal{L}_{j,k}, p) \leq \mathcal{T}$, where d is the Euclidean distance between the opponent player and E_i and $\mathcal{T}$ is a predefined threshold (in this case, we used 1.0 m as threshold). We also remove edges



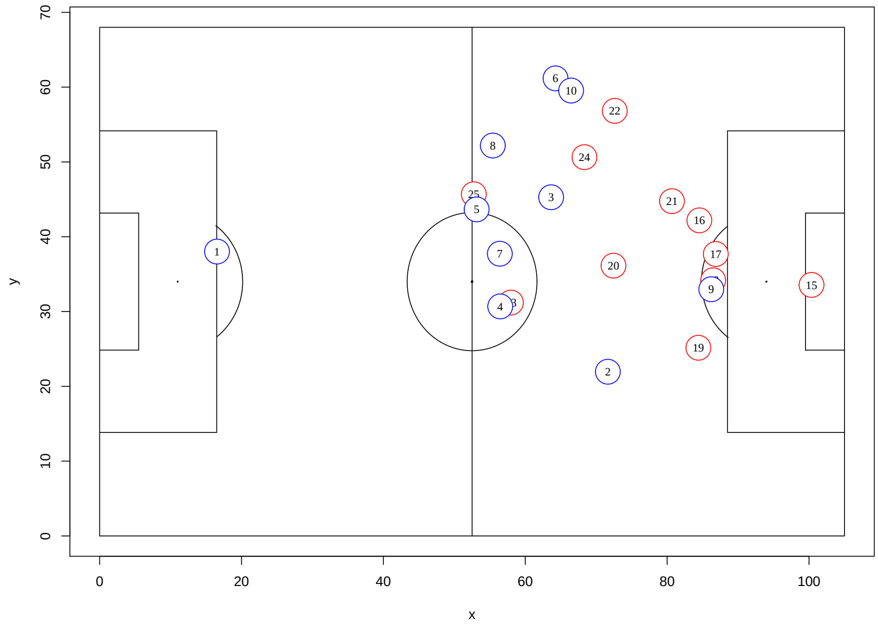
(a) (b) (c)

Figure 3.1: Opponent-Aware Graph Computation. (a) Delaunay triangulation graph of Team A (players 1 to 11), with vertex representing players and edges representing the possible flow of passes. (b) Observation of the location of opponent players. Yellow and red vertices (opponents) might block passes, i.e., passes among players of Team A, which are close to opponents, are less probable to happen. (c) Resulting graph after edge removals.

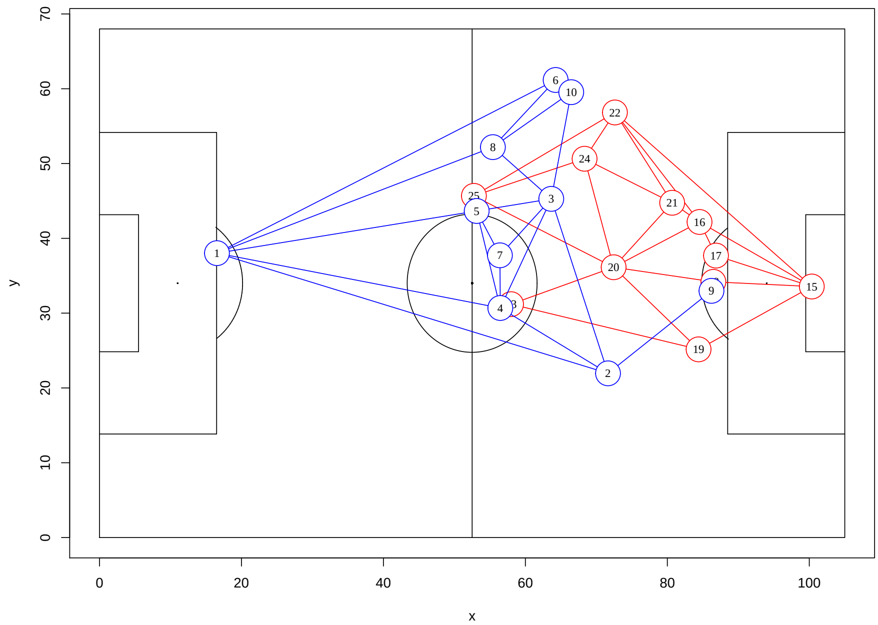
based on the proximity of a player and his opponents. Let $p = (x, y)$ be the closest opponent player to a vertex v_j. All the incident edges e to the vertex v_j are removed if $d(v_j, p) \leq \mathcal{T}$ (in this case, we used 0.5 m as threshold).

Figure 3.1(a) shows a graph for a team (say A) after computing the Delaunay triangulation considering the location of its players. This figure also includes vertices of a second team (say B). Players of teams A and B are represented by white and black circles, respectively, i.e., players labeled from 1 to 11 (1 is the goalkeeper) belong to Team A, while players whose labels range from 15 to 25 (15 is the goalkeeper) belong to Team B.

Successful performance in soccer matches may depend on a continuous flow of passes between players. Observing the positions of opponent players in the Delaunay graph, it is possible to notice that some passes (represented by edges) are less probable to happen. For example, in Figure 3.1(a), a pass between players 7 and 10 (both from Team A) might be blocked by player 20 (Team B). The same is true for passes between players 4 and 10, which might be blocked by player 21. Note also that all passes from player 2 would probably be blocked by player 19, while all passes from player 8 would be blocked by player 24. In Figure 3.1(b), opponent players near segment lines of passes between teammates are represented as yellow vertices, while opponents with short distances from a player (direct marking) are represented in red. Figure 3.1(c) shows the resulting opponent-aware graph after the process of removing edges. In Figure 3.2(a), we present players' position in field for a specific instant of time, while in Figure 3.2(b), the respective opponent-aware graphs for both teams. We believe that this approach can represent in a more suitable way the possibilities of short passes between teammates according to opponent defensive strategies. Note also that this approach is not intended for analysis considering over-the-top passes among players.



(a)



(b)

Figure 3.2: Opponent-Aware graph at an instant of time. (a) Players' position in field, considering Team A (players 1 to 11, in blue), and B (players 15 to 25, in red). (b) Resulting opponent-aware graphs for each team.

3.2.2 Players' Diversity Entropy Computation

Dynamics of passes among players in a match is an important characteristic to achieve successful results. In this study, we use the diversity entropy [100, 101] to characterize the possibility of passes among players, as well as their roles in a match as a defender, midfield, or forward. In this context, we used the diversity entropy of each player as a variable to characterize the dynamic nature of the match.

Diversity Entropy (as defined in Section 2.1.3 – Eq. 2.9) considers the transition probability $P_h(j, i)$ that a vertex i reaches a vertex j after h steps in a self-avoiding random walk. Let Ω be the set of all vertices except i. This concept is applied for each player in an instant graph of the match. Figure 3.3 illustrates two examples of diversity entropy computation for a midfield and a forward player. In both cases, we are interested in computing the diversity entropy of the vertex highlighted in yellow. For each scenario, we represent in red the accessible vertices, i.e., those vertices that might be accessible with a self-avoiding random walk of size 2 (i.e., $h = 2$). Each vertex accessed with random walk has their transition probabilities shown in the figure. For example, from vertex A, there is a probability of $\frac{1}{3}$ to reach each of it's neighbors (B, C, and D). From vertex B, there is a probability of $\frac{1}{2}$ to reach vertices E and F. So, we compute the probability of $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$ to reach vertex E from vertex A. On the other hand, from vertex C the probability is equal to 1 to reach vertex F. From vertex A, it is also possible to reach vertex F from two paths: B and C. Therefore, the probability to reach vertex F is $\frac{1}{6} + \frac{1}{3} = \frac{1}{2}$. Considering the scenario depicted in Figure 3.3(a), the diversity entropy calculated, according to Eq. 2.9 for player A is: $E_h(\Omega, A) = -\frac{\frac{1}{6} \times log(\frac{1}{6}) + \frac{1}{2} \times log(\frac{1}{2}) + \frac{1}{3} \times log(\frac{1}{3})}{log(11-1)} = 0.44$.

Vertex A (in yellow) has a higher diversity entropy ($E_h = 0.44$) because many more vertices are accessible (higher diversity). A very different diversity entropy score is observed for the yellow vertex of Figure 3.3(b). In this case, $E_h = 0$, i.e., no diversity is observed as only a single vertex is accessible by the random walk. If a player has no edges, i.e., he is completely marked by opponents, his entropy is also $E_h = 0$.

3.2.3 Analysis Framework

In order to perform soccer match analysis, we propose the analysis framework shown in Figure 3.4. This framework is comprised of four steps:

(a) Extraction of players' location over time: This step is accomplished by using the DVideo software [44] applied to nine official soccer matches. The role of each player is then characterized according to their patterns of movements in pitch along the match. Defensive players are likely to be found in different positions to those observed for offensive players. This phenomenon is illustrated in Figure 3.5, which shows typical position maps of defensive, midfield, and offensive players based on their locations. These position maps were used to determine the players' role in the match.

(b) Graph construction based on the obtained locations: This step is implemented using the opponent-aware graph representation approach, described in Section 3.2.1. We

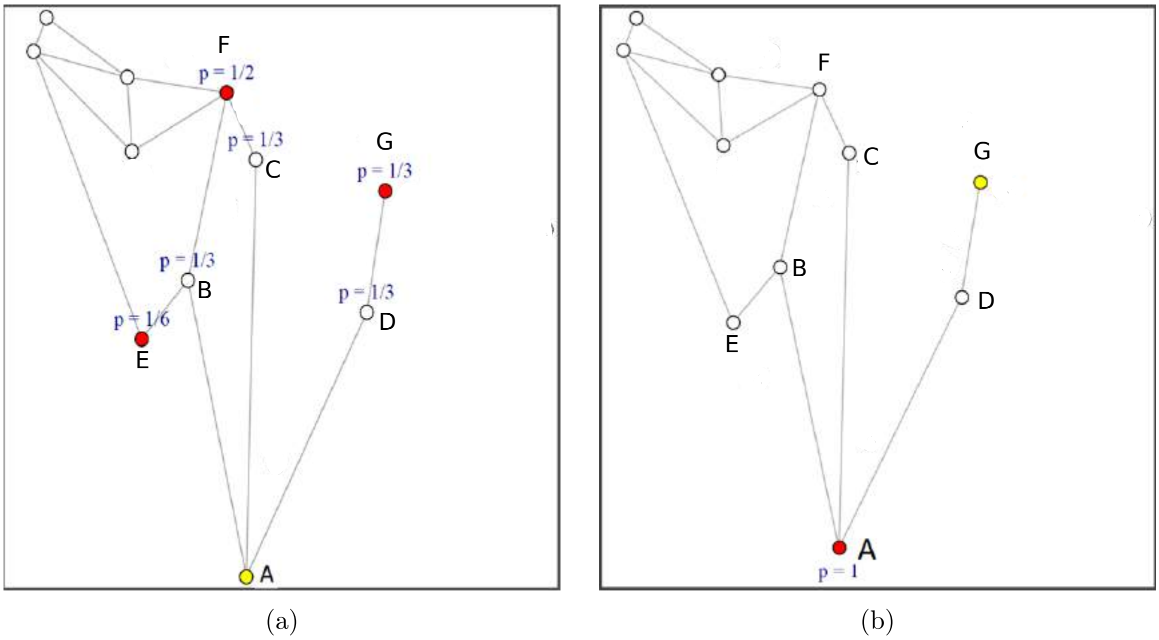


Figure 3.3: Diversity entropy of players in an instant graph, considering a self-avoiding random walk ($h = 2$). (a) Diversity entropy of a midfield player (in yellow). In this case, three players are accessible (in red), leading to a diversity entropy equal to 0.44 (transitions probabilities of each vertex accessed throw random walk are shown). (b) Diversity entropy of a forward player (in yellow). In this case, only one single player is accessible, leading to a diversity entropy equal to 0.

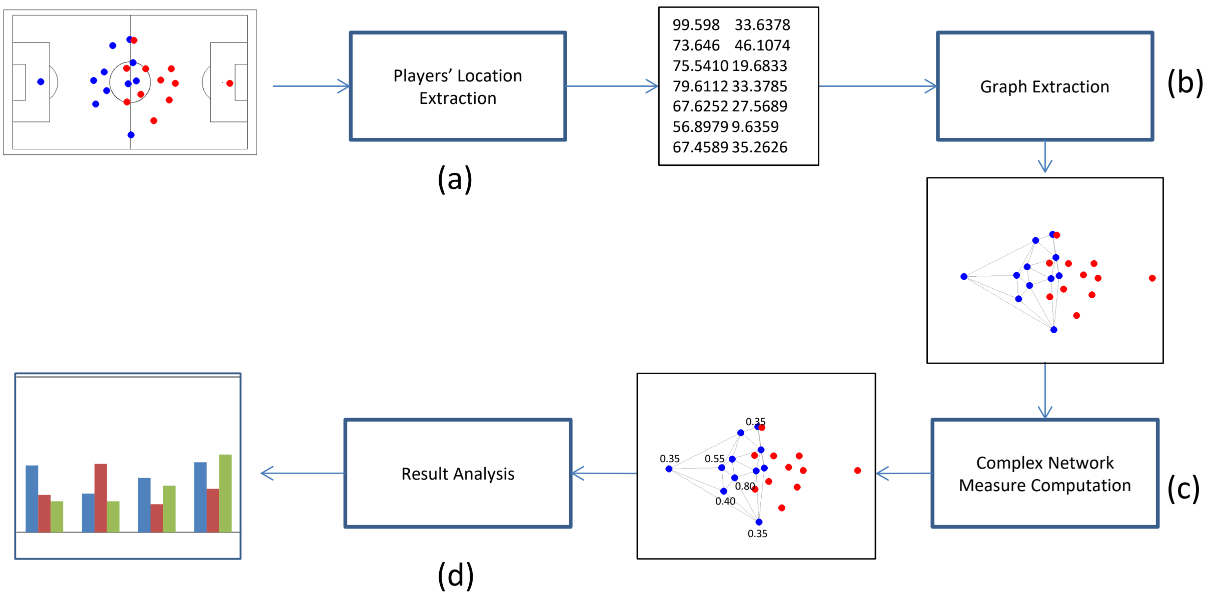
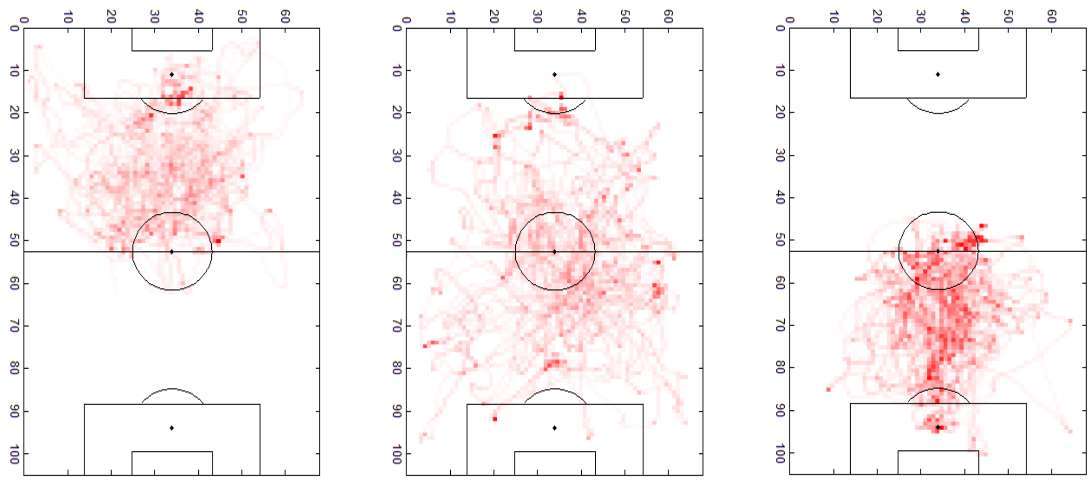


Figure 3.4: Analysis framework.

have computed two graphs (one for each team) in each frame of all matches. For a half time of a match, we have to handle approximately 83 thousand graphs.

(c) Computation of the diversity entropy for each player: This step considers the ap-



(a) (b) (c)

Figure 3.5: Typical position map of players on the pitch. The opponent's goal area is at the bottom region of each figure. Different players present different location patterns of movements along the match. (a) Defensive player; (b) Midfield player; (c) Forward player.

proach described previously. We computed diversity entropy of each player for each instant of time. Figure 3.6 contains a small excerpt of the variation of the diversity entropy of a player over time. The figure also presents some of the match events about which we have information. In this chapter, in special, we analyze the use of our graph representation to characterize moments when passes among players are performed. Figure 3.7 depicts an example of a defensive player's trajectory, during 330 frames (11 seconds). Red pixels represent high diversity entropy values, while yellow pixels are associated with low diversity entropy values. In the starting location, the player has a low diversity entropy (0). Figure 3.7(c) presents the corresponding graph, where player 4 (in blue) is completely blocked by player 20 (from the opponent team). Figure 3.7(b) presents the graph corresponding to the last player position in the trajectory, where his diversity entropy score is high, and it is possible to see that the player has many neighbors nearby.

(d) Analyses of diversity entropy scores and their association with match events: this step concerns the evaluation of players' performance based on their diversity entropy scores and match events. We performed two analysis: we verified if the diversity entropy measures observed for players are somehow related to their roles as a defender, midfield, or forward; and we verify if the players' mean diversity entropy measures are correlated with the frequency of their participation in successful passes.

In our analysis, we compare the OA approach with a baseline representation defined in terms of the Delaunay triangulation graphs (DT) in all analysis performed. The objective is to demonstrate that considering the position of opponents to construct

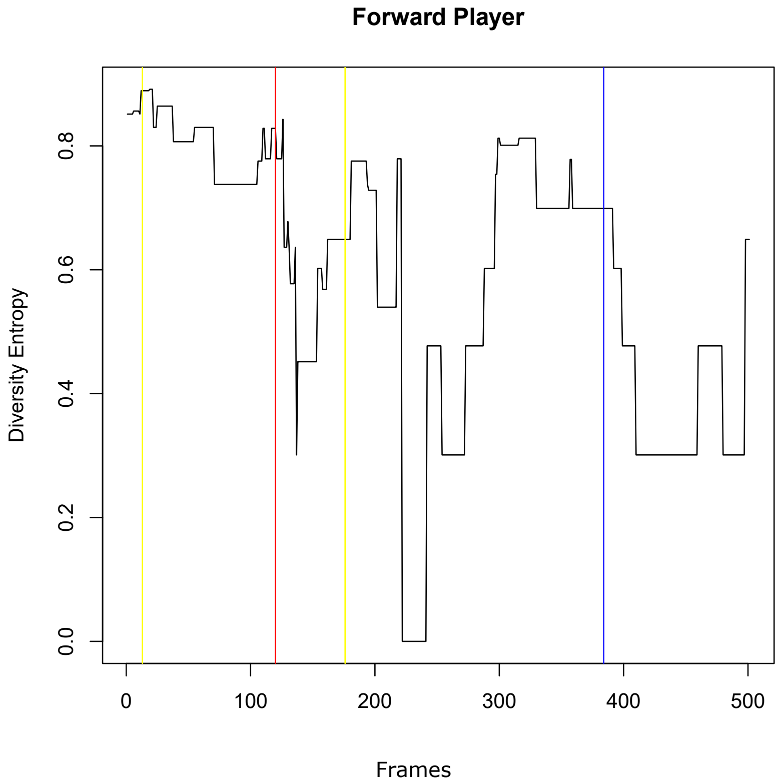
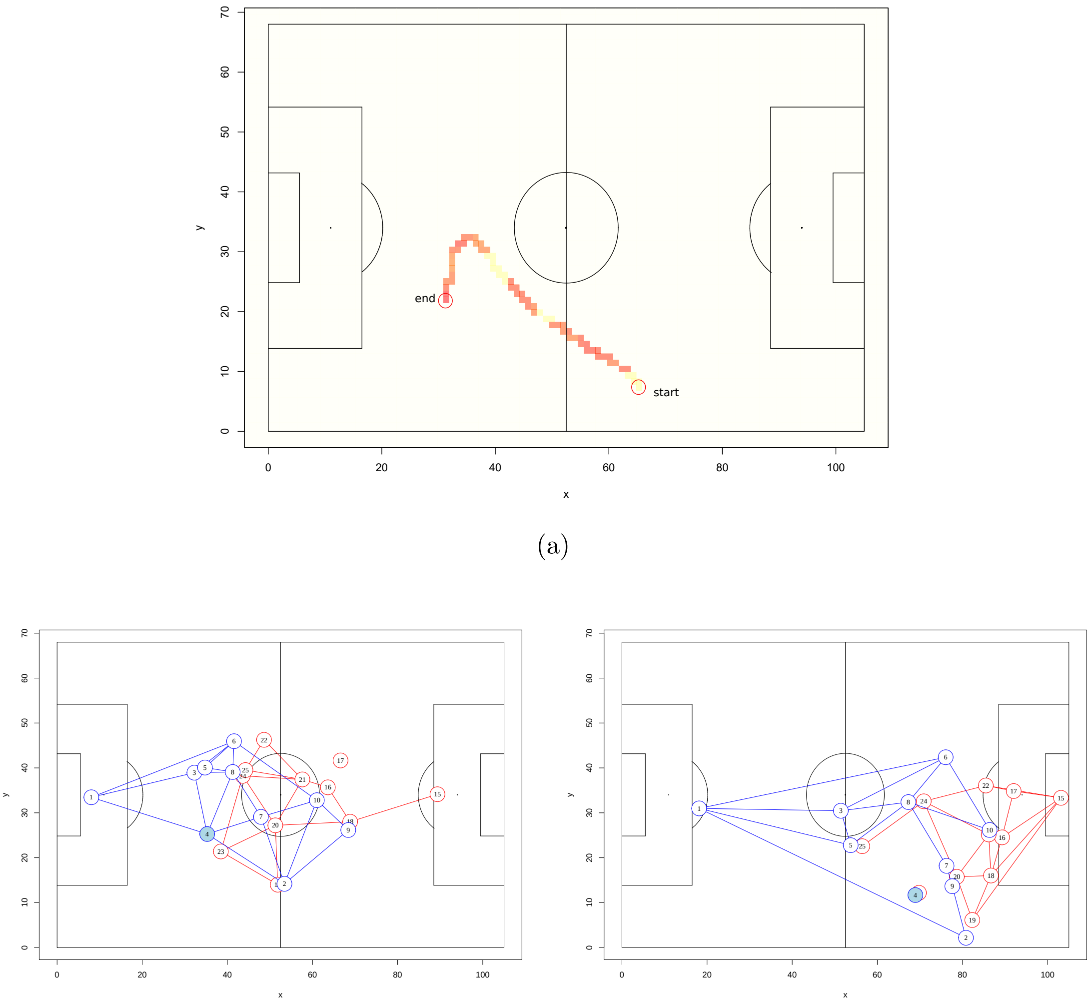


Figure 3.6: The diversity entropy time series for a forward player during 500 frames. We highlight in yellow two instants when the forward player recovered the ball from an opponent. The red line, in turn, indicates the moment when the player lost the ball. Finally, the blue line indicates that the player receives a ball pass from a teammate.

location-based graphs allows a more comprehensive analysis of complex dynamic match patterns resulting of the interaction of players and their opponents.

To identify patterns of different players' roles based on their diversity entropy, we computed the diversity entropy for all players of the different matches considered in our study, and also computed the two most important components found by applying the Principal Component Analysis (PCA) technique. The diversity entropy time series of each player along the match was used as input of the PCA function. This approach was applied to the first half time of the nine matches as we did not want consider the effect of substitutions in our analysis.

Diversity Entropy quantifies the number of effectively accessible vertices at a given distance in steps. Therefore, it is possible to notice that if a vertex has higher values of diversity entropy, it can access more vertices, and consequently, it can spread information with more efficiency. In this sense, in the soccer context, it is possible to question if players with higher diversity entropy scores are more likely to be involved with successful passes. We, therefore, also investigate the correlation between diversity entropy scores and the frequency of occurrence of passes among players. In order to investigate the relation between the Diversity Entropy measure and possibility of passes, the Pearson Correlation scores were calculated. We first computed diversity entropy mean scores for each player, considering both teams from



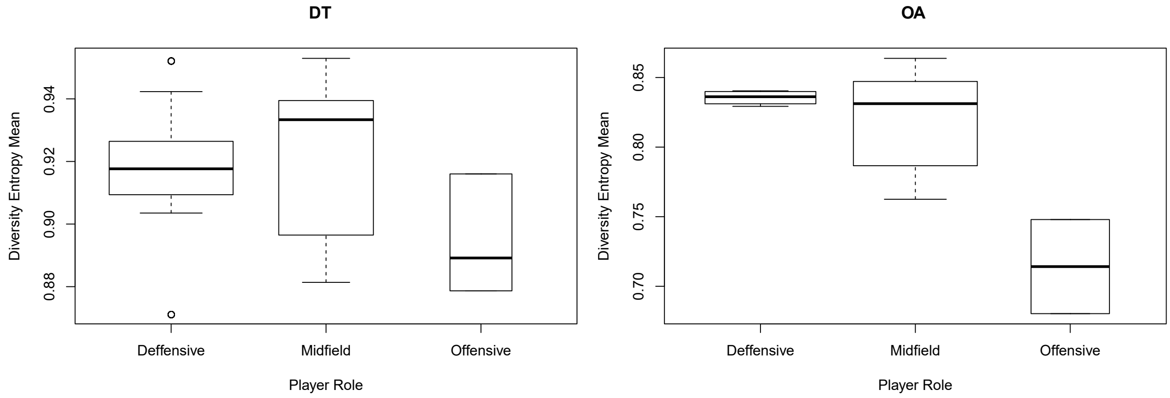
(a) (b) (c)

Figure 3.7: A small example representing a defensive player's trajectory and his corresponding diversity entropy scores represented as a heat map. (a) The diversity entropy for a defensive player according to his trajectory in the field. Dark red pixels represent high diversity, while light yellow ones represent low diversity. (b) Graph corresponding to an instant of time in which the defensive player (in blue) has a low diversity entropy score. (c) Graph corresponding to an instant of time in which the defensive player (in blue) has a high diversity entropy score.

nine different matches. Also, each player in each match was classified according to its position on the pitch. Again, our analysis considers only the first half of all matches. We also do not consider goalkeepers in our evaluation.

3.2.4 Statistical Analysis

We conduced an one-way ANOVA test to assess the complexity of decision making, by grouping players from the nine matches in three classes: defensive, midfield, and attacking. We also collected the entropy measures in the moments those players performed a pass. In this test, we used the entropy measure as the independent variable. We obtained 697,



(a) (b)

Figure 3.8: Boxplots considering diversity entropy mean values according to players' role. (a) Box-plot for the Delaunay triangulation (DT) and (b) Box-plot for the proposed Opponent-Aware graph representations.

1866, and 622 passes for defensive, midfield, and attacking players, respectively. When differences were found in the F-test, Tukey's honestly significant difference criterion was performed as a post hoc test.

In order to verify the correlation between diversity entropy and passes accomplished by players, we computed the Spearman's rank correlation between diversity entropy mean scores and the frequency of passes by each player. The statistical significance was set at 5% for all analyses.

3.3 Results and Discussion

3.3.1 Diversity Entropy Measures and Players' Roles

Figure 3.8 shows the boxplots considering both methods, OA and DT, with the distribution of diversity entropy for players, according to their role. It is possible to observe that the proposed Opponent-Aware graph technique (OA) leads to differences in results between groups, specially highlighting the forward players. Figure 3.9 illustrates the same phenomena with a different perspective. In this figure, we plot the time series related to the mean diversity entropy scores observed for defensive (red), midfield (green), and forward (blue) players for Team B in the half time of Match 3. This graph refers to the first 2000 frames of the match. As we can observe, the average scores observed for forward players are lower that those observed for defensive and midfield players. Similar results were observed for the other matches. Lower entropy values mean that there are few options to pass the ball to.

These circumstances may represent a more complex task to the decision-making process, once players need to collect information about the entire environment and then to select these few teammate options to perform the pass. Previous studies reported that experienced players are better than novices in terms of decision making, pattern recognition, anticipation during the game, visual search, and selection [36, 84]. Therefore, attacking players demands for passing decision-making may be higher than those of other positional

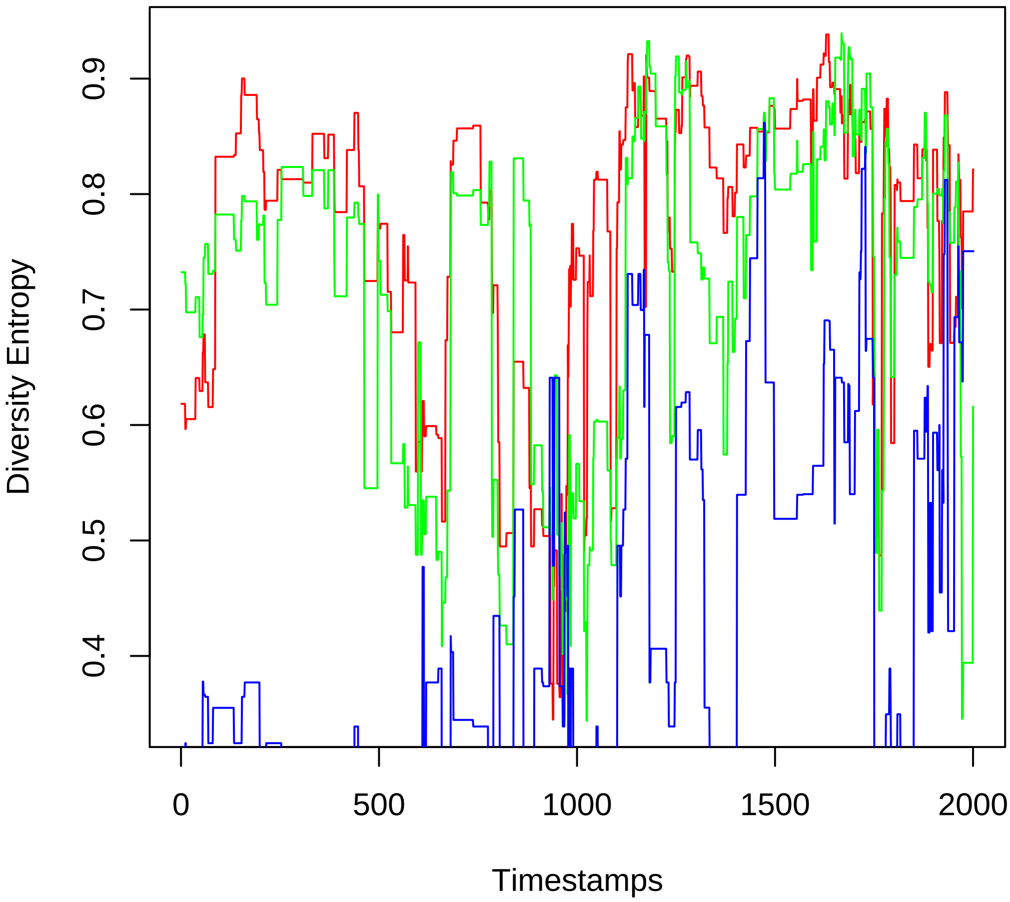


Figure 3.9: Average diversity entropy time series. Average diversity entropy time series for defensive (red), midfield (green), and forward (blue) players for Team B in the half time of Match 3.

groups, as entropy values demonstrated. This result is in accordance with the common perception that forward players have in general less options to make passes, which lead to the need of making decisions much faster than players with other roles. Once we found attacking players has lower entropy measures during the game, it is possible to analyze the complexity of decision making during ball passing for each class of players.

Table 3.1 presents the mean diversity entropy values and standard deviation for each group of players (defensive, midfield, and attacking) in the moments those players performed a pass. After performing the ANOVA test, we highlight that there is no statistical significance difference among defensive and midfield players ($p = 0.99$). The same is not true for attacking players, whose entropy scores are significantly different from the ones observed for defensive and midfield players ($p < 0.001$ for both). This outcome has basically two important practical implications. The first one is related to the demands to decision-making (and consequently to perform the technical actions) according to the position. Coaches may plan specific drills to each position group, considering that attack-

Table 3.1: Entropy mean values and standard deviation for defensive, midfield, and attacking groups.

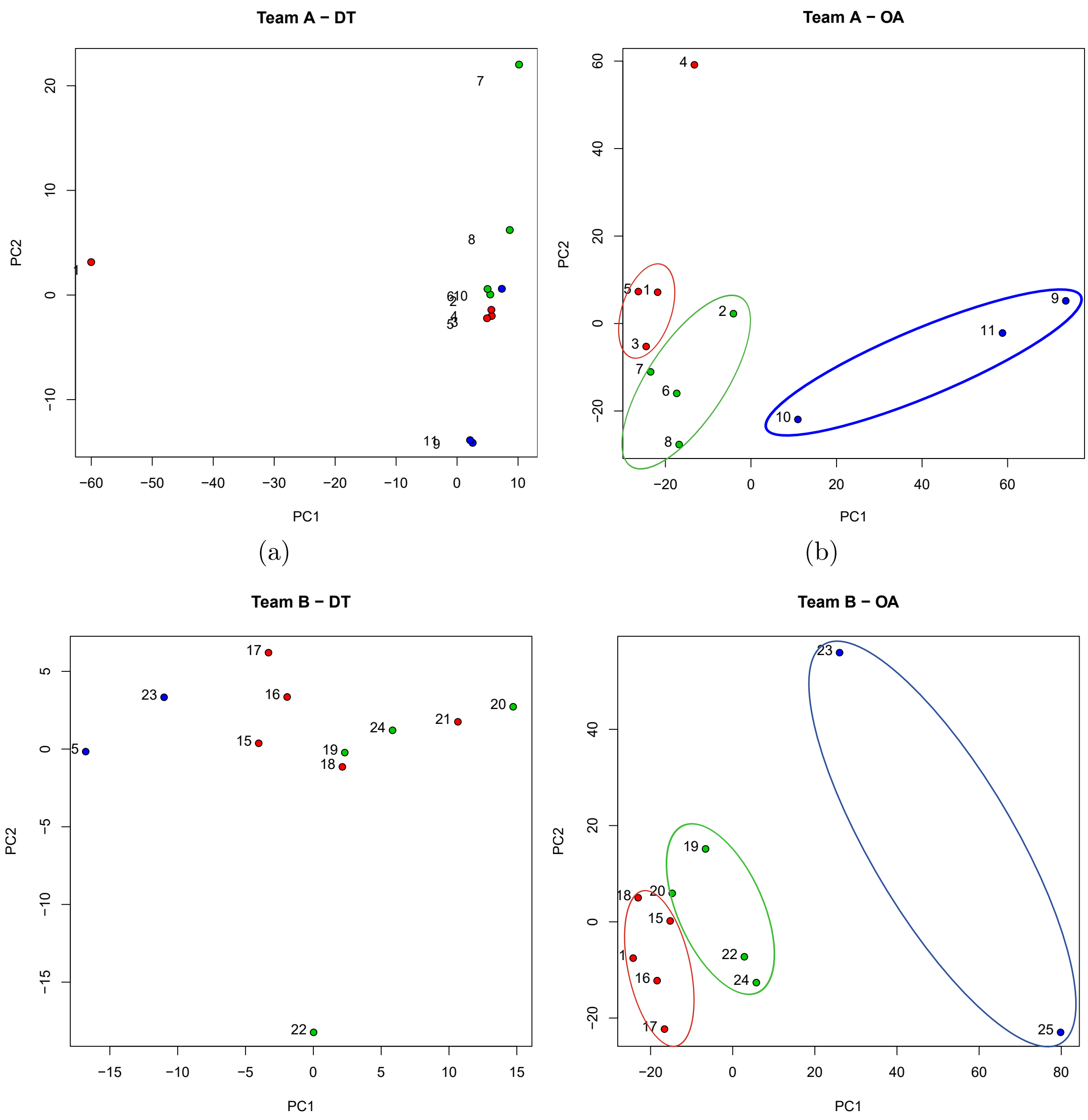
	Mean Value	Std. Deviation
Defensive players	0.75	0.25
Midfield players	0.75	0.26
Attacking players	0.67	0.30

ing players may lead with more complex conditions, with lower possibilities of teammates to receive their passes. Secondly, our results provide important insights for researches interested in evaluating and training players decision-making strategies using video-based methods, as previously reported in literature [65, 68, 110]. Therefore, videos and animations may simulate match conditions considering the complexity for decision-making to each position group.

We performed the Principal Component Analysis (PCA) for 3 matches, as example, considering the Opponent-Aware method proposed, and the Delaunay Triangulation baseline. This analysis helped to highlight the existing diversity entropy measure patterns in time. We used the diversity entropy time series as attributes for each player. In this sense, each player had his diversity entropy scored around 80 thousand values. Figures 3.10, 3.11, and 3.12 present the two main components of PCA axes for each team considering each of the three different matches. In those plots, red, green, and blue vertices stand for defensive, midfield, and forward players, respectively. As it can be observed, patterns (highlighted by ellipses) regarding the variation of diversity entropy over time only can be observed when the proposed OA representation is used (Figures 3.10(b), 3.10(d), 3.11(b), 3.11(d), 3.12(b), and 3.12(d)). For OA plots, midfield players (highlighted in green) are clustered together, which demonstrates that in general they have similar diversity entropy variation over time. The same was observed for defensive players (in red) for all matches considered. According to those plots, it is also possible to notice that forward players (in blue) are not associated with a common pattern. This result might be related to the fact that offensive players usually have opponents nearby, which affects their diversity entropy. In fact, the observed diversity entropy measures for forward players are usually lower than those observed for midfield and defensive players.

3.3.2 Correlation of Diversity Entropy Scores with the Occurrence of passes

Figure 3.13 shows the distribution diagram considering the diversity entropy mean and frequency of passes performed for each player. Pearson Correlation score is 0.45 ($p< 0.001$). It is possible to observe that diversity entropy is moderately correlated with the occurrence of passes for both approaches, and the results are statistically significant. The higher the diversity entropy score observed for a given player, the greater are the successful passes he performs. In other words, when a given player has many teammates as options to perform a pass (high entropy), greater are the chances to perform a correct pass. Thus, players who move constantly may help the teammate with the ball possession in creating opportunities to receive the pass. This behavior explains why, when attacking,



(a) (b) (c) (d)

Figure 3.10: Match 1: Comparative PCA axes plot. Comparative PCA axes plot of Delaunay triangulation graphs (DT) and opponent-aware graphs (OA) for each team on a halftime match. Red points are defensive players, green points are midfield players, and blue points are attacking players.

teams usually distribute the players across the pitch increasing the surface area and stretch indexes in order to create passing opportunities [74,76]. On the other hand, the defending teams try to follow the attacking team, with an in-phase synchronous configuration [79,99], in order to decrease the passing options of the opponent.

Another possible application of these results refer to the definition of appropriate marking strategies in defensive actions. A team may opt for a man-to-man marking with the objective of decreasing the diversity entropy of key players, impacting their performance in making successful passes. In some situations, on the other side, zonal marking strategies may be used for specific areas of the pitch in which non-skilled players (those unable to make fast decisions) are more likely to be present.

We can also observe peculiar behavior of player 4 in Figure 3.10(b) when compared

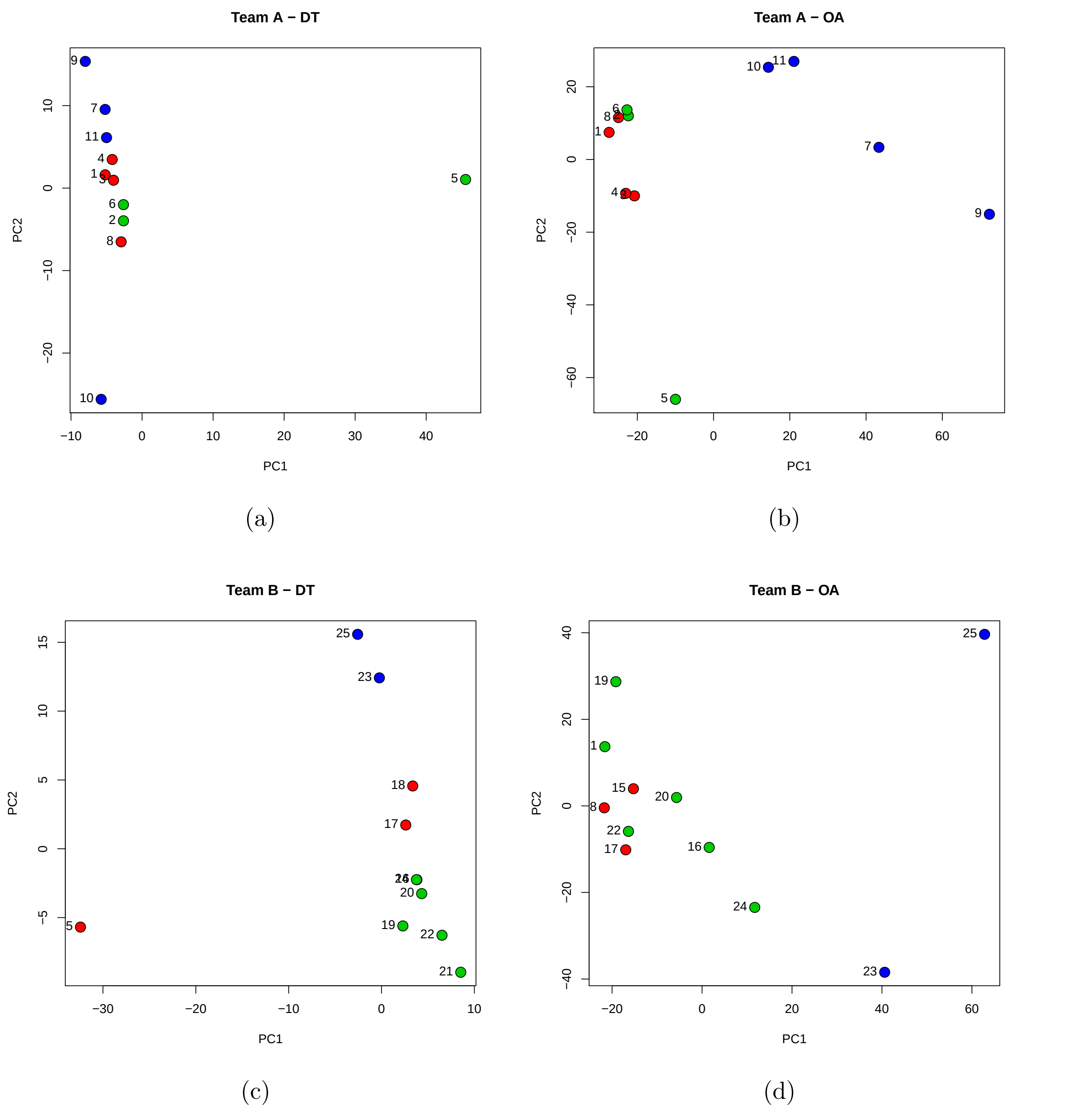
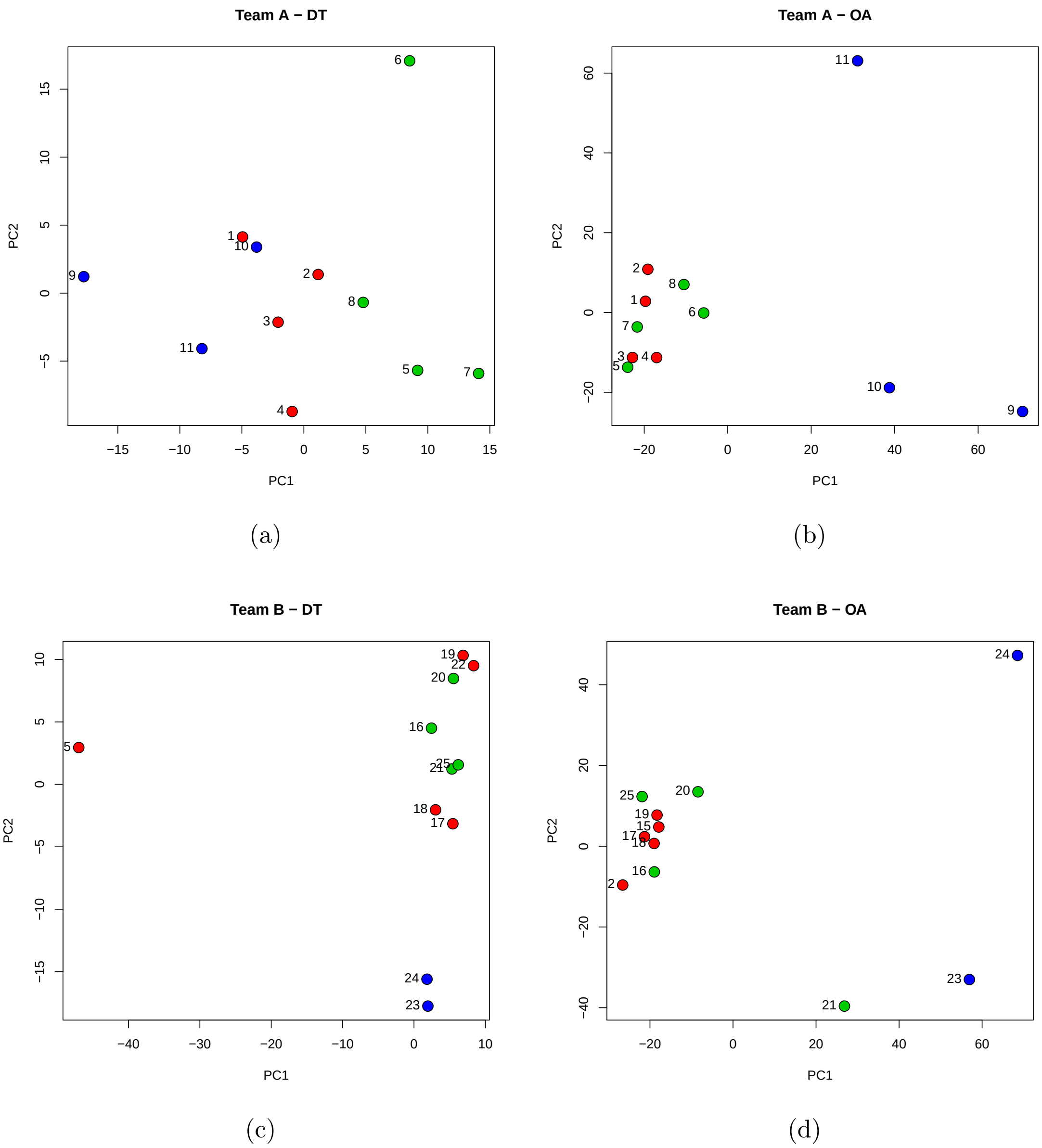


Figure 3.11: Match 2: Comparative PCA axes plot. Comparative PCA axes plot of Delaunay triangulation graphs (DT) and opponent-aware graphs (OA) for each team on a halftime match. Red points are defensive players, green points are midfield players, and blue points are attacking players.

to his teammates. Similarly to attacking players, player 4 (a defending player) has a quite different diversity entropy when compared to defensive players. Table 3.2 shows pass scores for players of Team A in Match 1. Highlighted in bold, it is possible to see that, as forward players, player 4 performs less passes than other players. In Table 3.3, we correlated the match statistics with entropy scores, for this same match. For each player, we calculated the percentage of right and wrong passes he performed in high and low entropy. Highlighted in bold, we found that, as attacking players, player 4 performs many passes in low entropy, which is not the common behavior for a defensive player. According to these findings, we believe that player 4 performs as a sweeper (or libero).



(a) (b) (c) (d)

Figure 3.12: Match 3: Comparative PCA axes plot. Comparative PCA axes plot of Delaunay triangulation graphs (DT) and opponent-aware graphs (OA) for each team on a halftime match. Red points are defensive players, green points are midfield players, and blue points are attacking players.

Figure 3.14 presents a time series of player 4's diversity entropy at the moment he performs his passes. The blue lines refer to successful passes, while the red lines are associated with unsuccessful passes. In the figure, passes a and c were successful passes with low entropy, and pass b was an unsuccessful one with low entropy. Thus, individual entropy analysis are extremely valuable for coaches in order to understand the circumstances that each player face during the match and how the teammates are organized on the pitch in order to increase the chances for successful passing sequences. Table 3.2 also shows that attacking players often make fewer wrong passes proportionally when compared to other players, even presenting, generally, lower entropy values. We must emphasize, therefore, that those

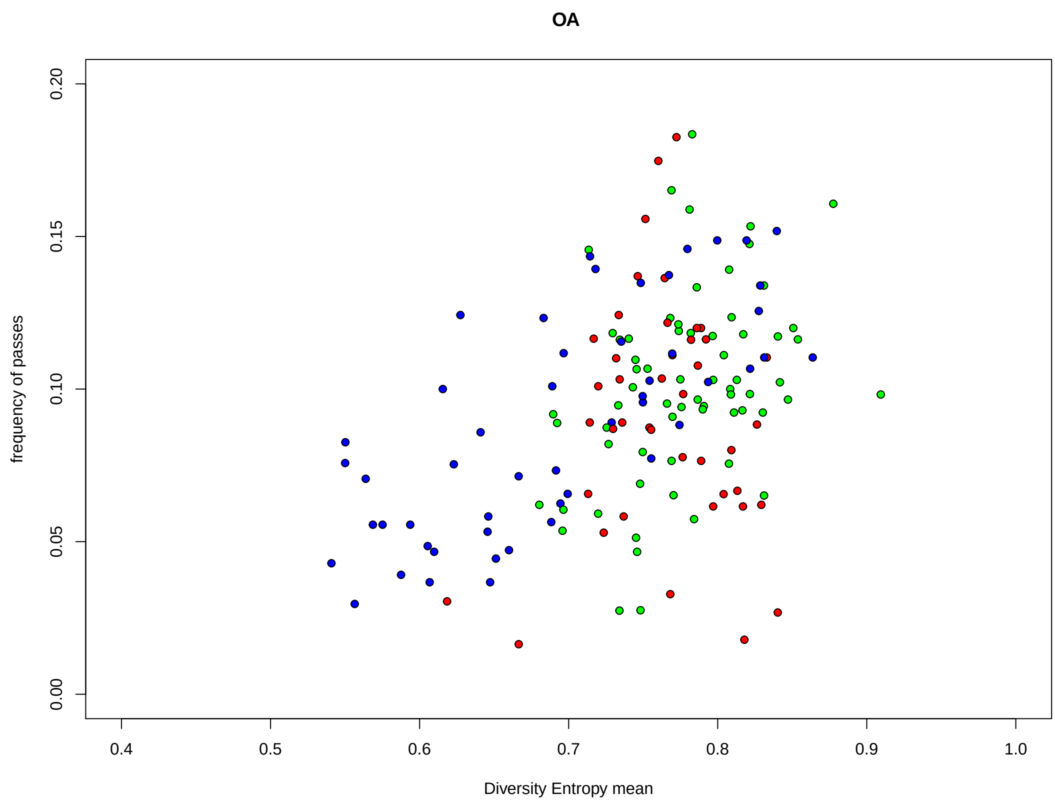


Figure 3.13: Distribution diagram for Diversity Entropy Mean and the frequency of passes.

players perform well in situations when they must take highly complex decisions.

Table 3.2: Statistics of passes from different players in Match 1.

	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Unsuccessful Passes	3	14	6	6	7	10	6	2	9	3
Successful Passes	22	21	10	25	23	21	15	14	10	9
Total Passes	25	35	**16**	31	30	31	21	**16**	**19**	**12**

Table 3.3: Statistics of passes and players' entropy in Match 1

	P2 (M)	P3 (D)	P4 (D)	P5 (D)	P6 (M)	P7 (M)	P8 (M)	P9 (F)	P10 (F)	P11 (F)
Unsuccessful Passes Low Entropy	0.04	0.06	0.06	0.00	0.03	0.00	0.00	0.00	0.00	0.08
Successful Passes Low Entropy	0.04	0.03	0.13	0.07	0.00	0.07	0.05	0.15	0.21	0.25
Total Passes Low Entropy	0.08	0.09	**0.19**	0.07	0.03	0.07	0.05	**0.15**	**0.21**	**0.33**
Unsuccessful Passes High Entropy	0.08	0.34	0.31	0.19	0.20	0.32	0.28	0.15	0.47	0.17
Successful Passes High Entropy	0.84	0.57	0.50	0.74	0.77	0.61	0.67	0.70	0.32	0.50
Total Passes High Entropy	0.92	0.91	0.81	0.93	0.97	0.93	0.95	0.85	0.79	0.67

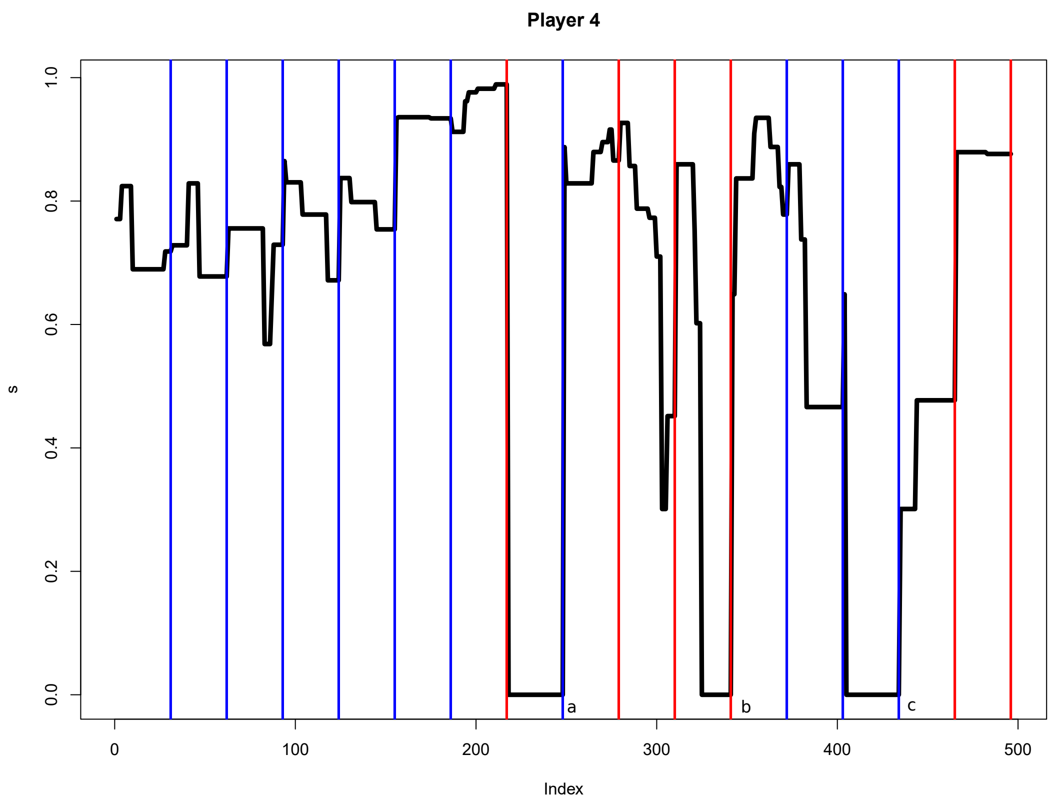


Figure 3.14: Time series of the diversity entropy of Player 4 of Match 1, considering the occurrence of passes. Blue lines indicate successful passes, while red lines are associated with unsuccessful passes accomplished by this player. Passes a, b, and c were accomplished in low entropy.

Chapter 4

Graph Visual Rhythms

4.1 Introduction

This chapter addresses the research question related to the investigation of information visualization approaches which would be suitable for supporting the analysis of temporal changes in dynamic graphs. In this chapter, we are interested in visually representing temporal graphs, with the objective of supporting the identification and analysis of temporal pattern changes. In this context, several approaches have been proposed to the visualization of temporal graphs [8, 15, 16, 60]. Most of the approaches rely on the use of node-link diagrams, where different visual marks (typically circle glyphs) are used to represent visually vertices and lines to encode existing relations among vertices. Different additional visual properties associated with visual marks (e.g., position, size, length, angle, and slope) are employed to highlight or emphasize properties of both vertices and edges [8]. Those initiatives, however, still do not address properly the visualization of huge volumes of data. In these situations, interaction controls are proposed so that users may handle occlusion or even perform browse activities over graph data.

Here, this problem is addressed from a different perspective. We propose a graph-to-image transformation, so that large volumes of sequence graphs can be visually represented in a compact way, enabling fast and easy visual identification of pattern changes. Our solution relies on the use of the *visual rhythm* representation [82], introduced in Section 2.4. This approach has been typically used for efficient video data processing and analysis [10, 89], as it allows the representation of the whole video content by means of an image, whose columns are defined by the extraction of features from pixels of frames. In this chapter, we extend this idea by encoding properties of graph sequences, leading to a representation we name *graph visual rhythm*. For each instant of time, graph properties are represented as a column of an image, allowing the compact representation of important graph features associated with changes of vertices and edges over time. Our solution is somehow similar to previous initiatives focusing on encoding graph dynamics using matrix representations [4, 18, 105]. Different from those initiatives, however, our approach does not rely on radial layouts, nor on complex representations such as small multiples and stacked matrices. To the best of our knowledge, this is the first attempt to encode complex temporal graph changes in an easy-to-interpret single image representation. We validate the method in the context of soccer match analysis. In this chapter, we describe the use of graph visual rhythms defined in terms of complex network measures for understanding complex temporal patterns associated with the match dynamics.

In summary, the contributions of this chapter are twofold: (i) the introduction of a novel compact visual representation for temporal graphs, named graph visual rhythm; and (ii) the presentation of different scenarios of its use in the context of the analysis of real soccer matches using complex network measures.

4.1.1 Complex Network Measurements

Soccer is one of the most difficult sports to analyze quantitatively due to the complexity of the play and to the nearly uninterrupted flow of the ball during the match. Indeed, unlike other sports, in which individual game-related statistics may properly represent player performance, in soccer it is not trivial to define quantitative measures of an individual contribution [38]. Moreover, simple statistics such as number of assists or number of shots may not provide a reliable measure of a player's true impact on team performance and, consequently, the outcomes of a match [38, 78]. Instead, the real contribution of a given player sometimes is hidden in the plays of the team, such as participating from a passing sequence to a shot on goal [38]. This type of information is important to detail the role of a team member on team performance. Thus, this study uses complex network measurements for extracting features from graphs to represent individual behavior and thus to represent team performance using a visual analytical tool. Two measurements were considered in this work: Diversity Entropy and Betweenness Centrality, which are presented in Section 2.1.3.

The dynamic aspects associated with passes among players in a match (such as the 'ball flow' among the players of a team) are important cues for game tactical analysis [38]. We used the diversity entropy as a variable to characterize the dynamic nature of the match, characterizing the possibility of passes among players. In this study, the centrality of players in a match is related to his role in the passing flow along the time. We used betweenness centrality to characterize the role of players in terms of the graph shortest paths within which the players are involved.

4.2 Graph Visual Rhythms

We define a temporal graph $\mathcal{G}$ as a sequence $\mathcal{G} = \langle G_1, G_2, \ldots, G_T \rangle$, where $G_t = (V_t, E_t)$ is a weighted graph at timestamp $t \in [1, T]$ composed of a set of vertices, V_t, and a set of edges, E_t. We refer to the graph defined at a particular timestamp t (say G_t) as an *instant graph*. The edges E_t of an instant graph link vertices according to the Opponent-Aware Graphs technique defined in Section 3.2.1 By building one graph for each instant of time considering the vertices' interaction, it is possible to capture the temporal nature of the graph dynamics. Our goal is to represent the interaction among vertices at each instant using a visual rhythm representation GVR. We follow a similar formulation employed in Eq. 2.14 to define GVR:

$$GVR(t, z) = \mathcal{F}(G_t), \tag{4.1}$$

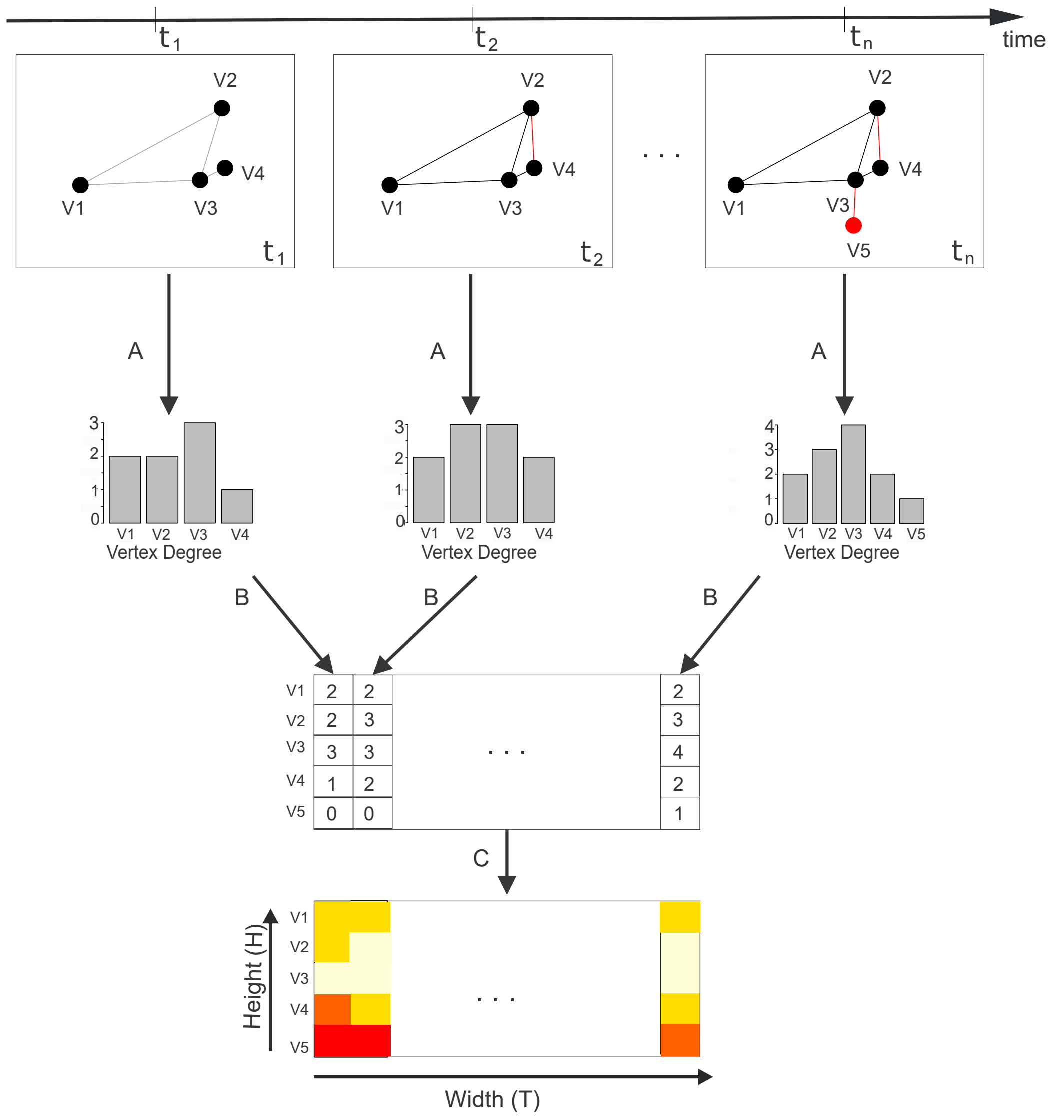


Figure 4.1: Flowchart illustrating how a graph visual rhythm is extracted.

where $\mathcal{F}_{G_t} : \mathcal{G} \rightarrow \mathbb{R}^n$ is a function that represents a graph $G_t \in \mathcal{G}$ as a point in an n-dimensional space, $t \in [1, T]$ and $z \in [1, n]$.

Figure 4.1 illustrates the computation of a graph visual rhythm for a temporal graph. Changes in the graph sequence are highlighted in red. For example, at timestamp t_2, an edge linking vertices v_2 and v_4 is created. At timestamp t_n, vertex v_5 is created along with an edge from v_5 to v_3. In this example, function $\mathcal{F}_{G_t}$ computes the degree of vertices for each instant of time (arrows labeled with A). The degree information is later used to create the graph visual rhythm image (arrows B). Again, different visual properties (e.g., color, opacity) may be used to highlight graph changes. In the case of the example, a heatmap-based color layout is employed (arrow C).

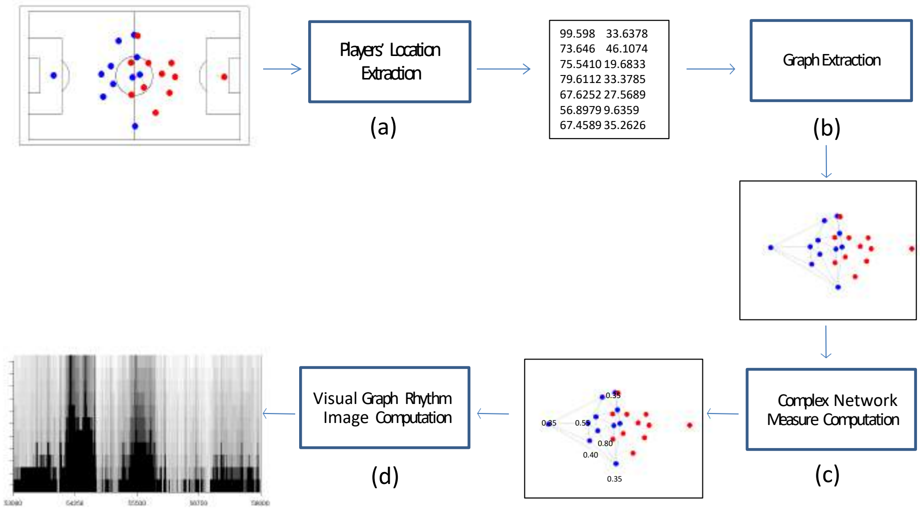


Figure 4.2: Analysis framework.

4.3 Case Study: Soccer Match Analysis

This study is based on the use of graph visual rhythms for identifying events on temporal graphs associated with soccer matches.

4.3.1 Soccer Match Analysis Framework

The graph-based soccer match analysis framework defined in Section 3.2.3 is slightly modified in this chapter. Its four steps, showed in Figure 4.2, are detailed in the following:

(a) Extraction of players' location in field over time: This step is accomplished using the DVideo software [43, 44] applied to official soccer matches. This process starts with soccer match videos and results in files containing players xy location on the pitch and annotation related to match events, such as passes accomplished, fouls, shots on goal, among others. The extraction frame rate is 30 frames per second, so for a typical 45-minute half time of a match, we have 81,000 frames. We used a dataset related to two official soccer matches (referred to as Match 1 and Match 2 along the chapter) of the Brazilian Professional First League Championship.

(b) Graph Extraction: This step builds graphs from soccer match frames. In our experiments, two different kinds of graphs are built: Opponent-Aware Instant Graphs and Flow Networks, which are detailed in Section 4.3.2.

(c) Complex Network Measure Computation: This step comprises the approach described in Section 4.1.1. Basically, complex network measures are computed from graphs obtained in Step (b). In this experiment, two measures are considered: Diversity Entropy and Betweeness Centrality. In this context, these measures are extracted by $\mathcal{F}_{G_t}$, the function that encodes one graph into a column of a graph visual rhythm.

(d) Visual Graph Rhythm Image Computation: Here, the result analysis step defined in Figure 4.2 is concerned with the creation of graph visual rhythm images. From those images, it is possible to analyze patterns that represent match events such as attacking and defensive strategies from each team.

4.3.2 Soccer Temporal Graphs

Our analyses are based on the characterization of interaction among players along the match. Let $\mathcal{G}'$ be a sequence $\mathcal{G}' = \langle G_1, G_2, \ldots, G_T \rangle$. A vertex $v \in V_t$ is associated with a player, whereas an edge $e_{jk} \in E_t$ connecting two vertices $v_j \in E_t$ and $v_k \in E_t$ is defined based on the location (or any other relation) of players (v_j and v_k) of the same team at timestamp t. The weight $w(e_{jk})$ may encode different properties of the interaction of players, such as their distance – possibly measured by the Euclidean distance of players j and k in the field – or the number of passes between them.

Given the importance of interaction between players on soccer matches, we consider two different approaches for constructing temporal graphs: Opponent-Aware Instant Graphs and Flow Networks. Both of them take into consideration passes between players from the same team. Instant graphs represent possibilities of passes according to the players' position on the pitch at each timestamp, while Flow Networks represent all accomplished passes between players in a time interval.

Opponent-Aware Instant Graphs

In this representation, for each time stamp, it is computed the Opponent-Aware Graph as defined in Section 3.2.1, considering the players' position and their opponents on the pitch. Two graphs are computed, one for each team. Figure 4.3 shows examples of instant graphs. Blue vertices (players labeled from 1 to 11) and edges represent Team A, while red ones (players labeled from 15 to 25) represent Team B.

Flow Networks

One important research venue refers to the identification of interaction patterns among players for a given time interval. One common approach relies on the use of Flow Network [38]. Flow network graphs can be defined as $G_{t_i,t_j}(V, E)$, in which vertices are players from a team, and weighted edges represent passes accomplished between them during a time interval $[t_i, t_j]$.

We extend this approach by proposing ball possession flow networks. Those networks show paths that only happen in time [19, 97], which means that no instant graph has all the edges shown in a flow network. Basically, we extract different flow networks, which represent ball passing among teammate players within the time interval in which they have ball possession. Figure 4.4 shows two possession flow networks. In Graph (a), the team has the ball, and accomplishes eight passes among teammates. Graph (b) illustrates the match situation in which a team has the ball possession, but no passes are accomplished until losing the ball possession again.

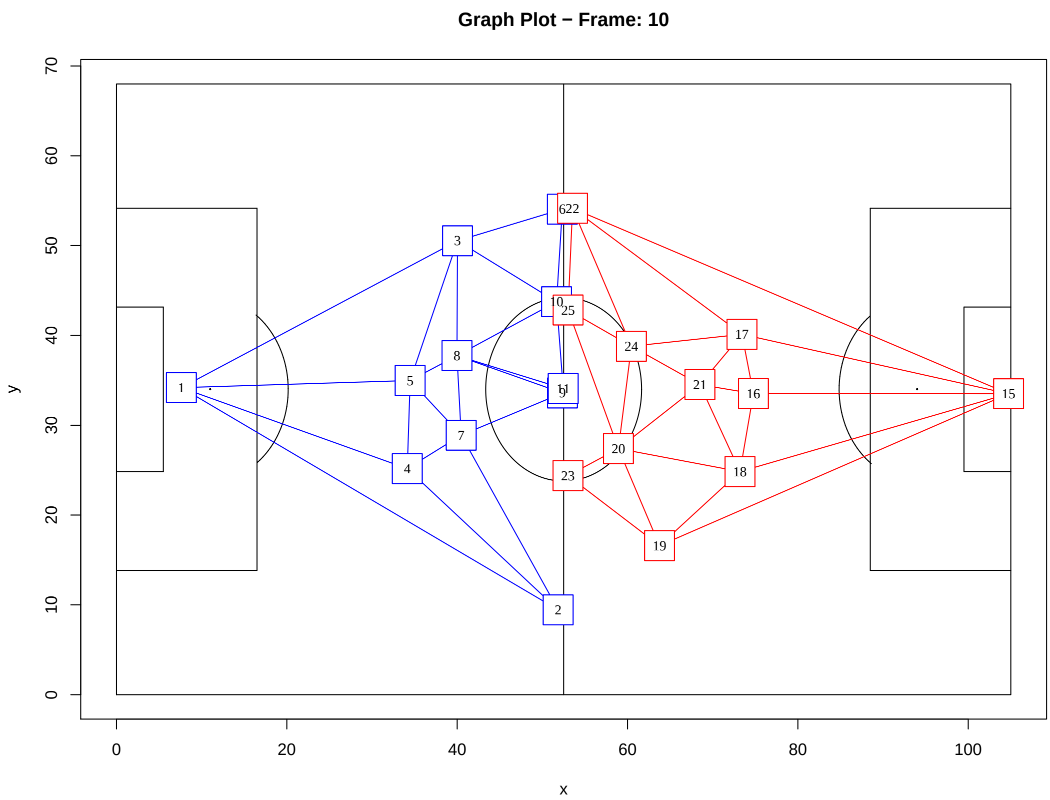
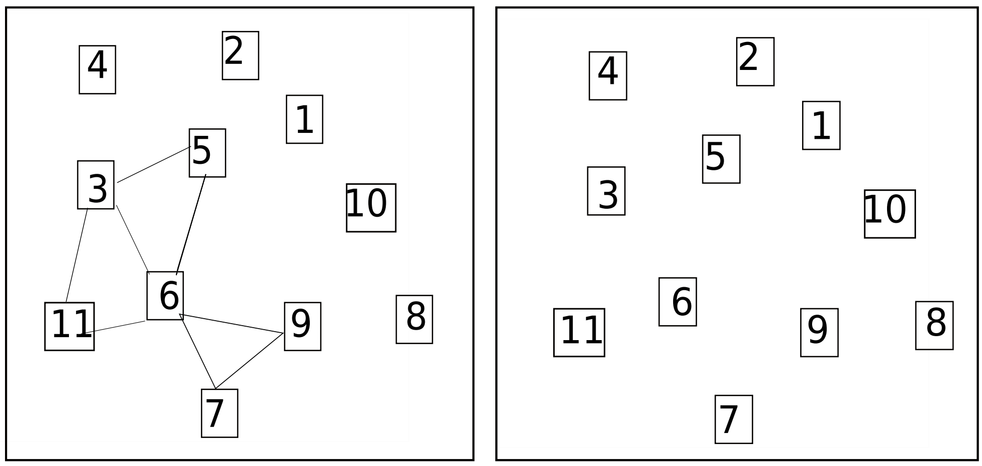


Figure 4.3: Examples of Opponent-Aware instant graphs of two teams (represented in blue and red).

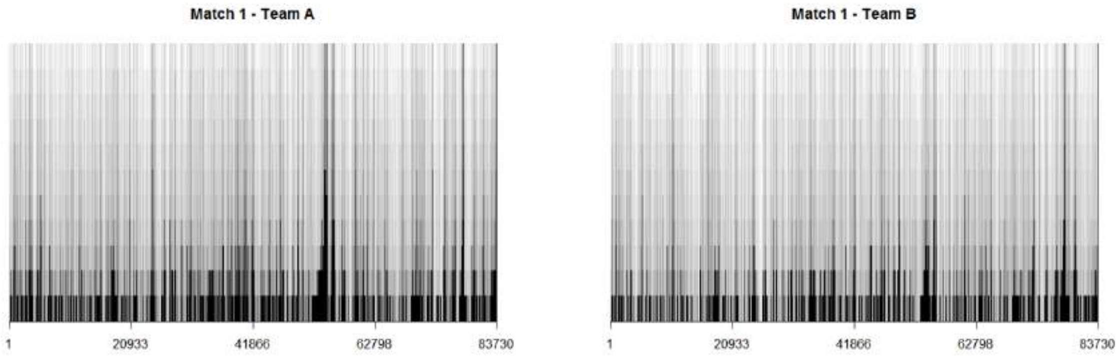


(a) (b)

Figure 4.4: Examples of Ball Possession Flow Networks. (a) Graph from a team that performed eight passes among teammates during a ball possession interval. (b)In another ball possession interval, no passes were performed.

4.4 Results and Discussion

This section discusses several usage scenarios in which graph visual rhythms are used to identify visual temporal patterns related to teams' strategies when defending or attacking.



(a) (b)

Figure 4.5: Graph visual rhythm images for teams of **Match 1**.

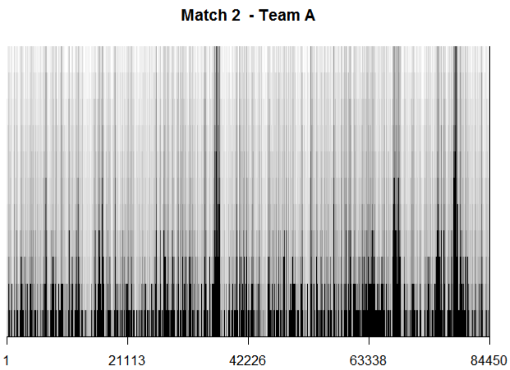
4.4.1 Defensive Patterns

The first usage scenario considers the use of graph visual rhythms in the identification of defensive patterns.

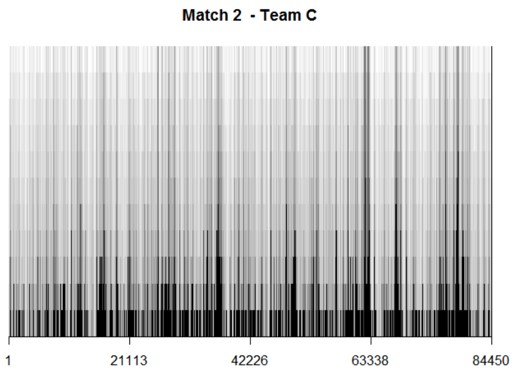
Considering the first half time of a match, we generated the Opponent-Aware Instant Graphs and computed the Diversity Entropy from each vertex in the instant graph ($\mathcal{F}_{G_t}$). For each instant graph, diversity entropy values were ordered and linked together vertically resulting in an image GVR_{xy}, where x is equal to the amount of frames from the match and y is amount of players from graph (11 players in each team). Diversity Entropy values between 0 and 1 in GVR_{xy} were normalized to 0 to 255, generating a grey-scale image. In this case, lower entropy values are darker, and higher values are lighter. For the player who has ball possession, entropy values may be associated with the 'complexity' of the decision-making scenario. If entropy is high, it means that the player has many options (i.e., teammates) to interact with and this is a less complex situation in case that the player has to perform a pass as fast as possible. On the other hand, lower entropy values may represent few teammates to interact with. This complex situation requires the player to evaluate this scenario more carefully, identify who are these few options of interaction, and thus make the decision to perform a pass.

Figures 4.5 and 4.6 present the resulting graph visual rhythms images obtained for teams of two matches (Match 1 and Match 2). In both matches, Team A is the same. It is possible to notice a clear pattern, defined in terms of vertical darker blocks, that distinguish all images. Given the performance of Team A in both matches, we can observe that there are darker regions for Match 2, which means that players of Team A in this match were usually not free, i.e., there were opponents close to them more frequently.

By zooming in the graph visual rhythms of Figure 4.5 for the frames in the range defined between 55000 and 58000, we obtain the images shown in Figure 4.7. We plotted the corresponding graph from the instant highlighted in red, to analyze the game strategy employed in the time period related to a darker block. It is possible to observe that for a darker block, Team A (in blue) is compressed in a defensive strategy while Team B (in red) is attacking. Team B is well positioned in the field with many possibilities of passes



(a) Match 2 – Team A



(b) Match 2 – Team C

Figure 4.6: Graph visual rhythm images for teams of **Match 2**.

among players, which is represented in its own graph visual rhythm image. Thus, entropy values may represent team strategy both in attacking and defending perspectives. The distances among teammates define the team compactness on the pitch during attacking and defending actions [76, 77]. In this case, Team-A players occupy the field in a very compact way, without clear purpose of performing a man-to-man marking. This strategy may favor the attacking team in order to allow a greater number of options for passes between players. If this condition is maintained over time (which is easily detected in the graph visual rhythm image), it may indicate a technical and tactical superiority for the team with lower entropy values.

One goal was scored by Team A of Match 1 at frame 24825. Figure 4.8 shows the graph visual rhythm images associated with this moment. It is interesting to notice that Team A was attacking, but, differently from the situation depicted in Figure 4.7, both teams have higher diversity entropy scores. In this case, this phenomenon is observed due to the fact that the goal was originated from a corner kick.

4.4.2 Most Valued Player: A Centrality-Oriented Perspective

We conducted a preliminary study considering the computation of the betweeness centrality applied to instant graphs. The intention here is to support the identification of players whose centrality scores are higher during the match, and so they could be considered more valued than others in the game strategy. Figure 4.9 shows the centrality-based graph visual rhythm image for Match 2, using darker colors to highlight players with higher centrality scores. It can be noticed that during almost all the match, centrality scores are low for most of the players. The low and homogeneous centrality scores show that there was no 'star topology.' Each had nearly the same connectivity, indicating that the teams did not depend on one single player [25].

We can also notice that for Player 7 of Team A, there were darker pixels along the match. By analyzing his performance during the game, we could realize that this player was involved in more passes than all the others (47 passes in the first half time, while the team's passes average was 33.1), which could mean that he was well positioned on the

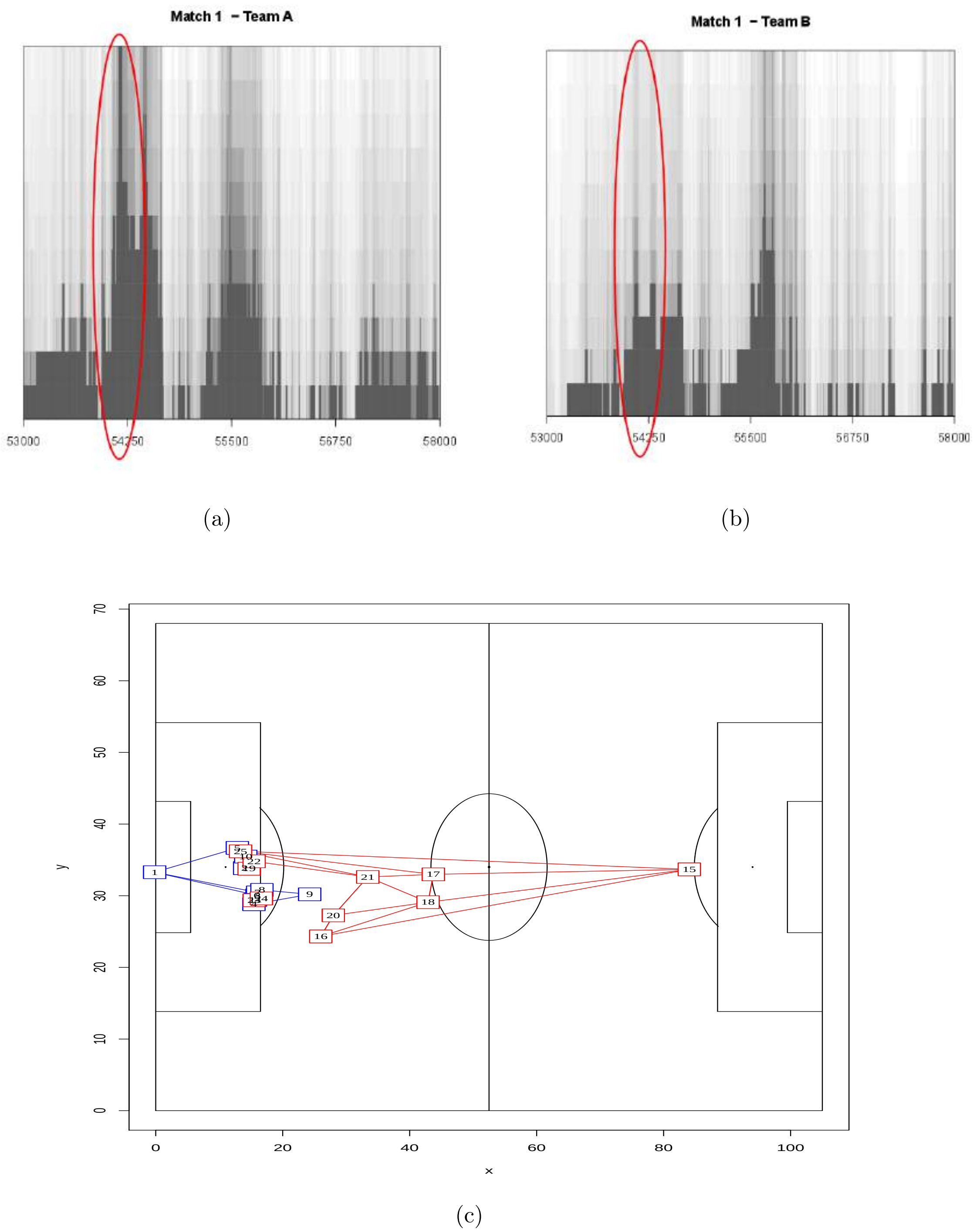
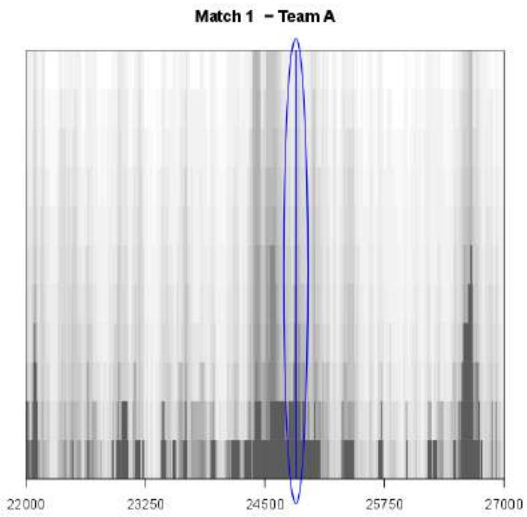


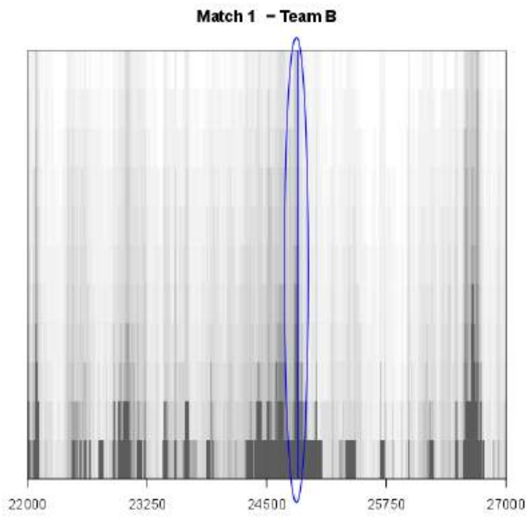
Figure 4.7: Graph Visual Rhythm in details: Highlighted dark block and the corresponding match situation. Team A (in blue) is compressed in a defensive strategy while Team B (in red) is attacking.

pitch during ball possession.

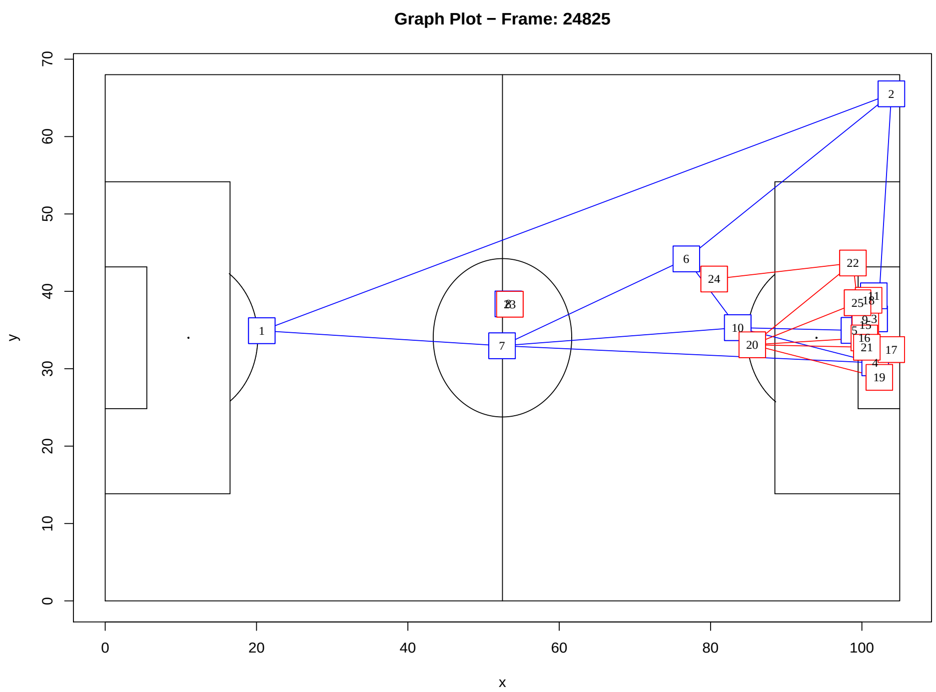
This same pattern is observed for Team B. It is possible to observe a darker pixel line for Player 6, who was also the one involved in more passes (42 passes in the first half time against the team's passes average of 26.4). We observed similar patterns for other matches: players with higher centrality scores, when compared to teammates, are



(a)



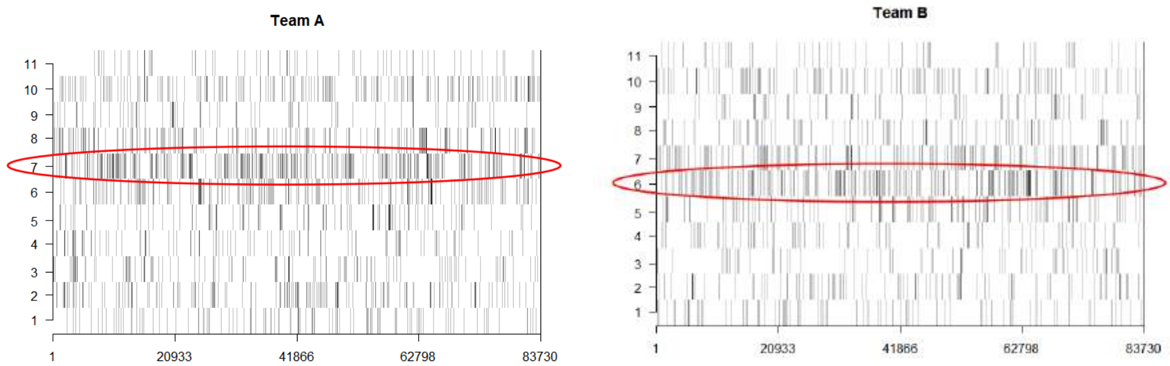
(b)



(c)

Figure 4.8: Graph visual rhythms of teams at a goal event timestamp.

involved more frequently with successful passes. Thus, graph visual rhythm images allow to identify players who had influential contributions in a specific match, and, applied during the entire tournament, they may help coaches to identify the most important players. In other words, it helps to answer in an objective manner whether, for example, the most famous players fulfilled the expectations placed on them [38]. However, in a collective evaluation, both entropy and centrality may be interpreted with caution. The work of [52] showed in 760 matches in the English Premier League that high levels of



(a) Match 2 – Team A

(b) Match 2 – Team C

Figure 4.9: Graph visual rhythm images based on the players' centrality for **Match 2**. We highlighted in red the players with higher centrality scores.

interaction (i.e., passing rate) lead to increased team performance. However, centralized interaction patterns lead to decreased team performance. In fact, even in a social context, teams with denser networks had a tendency to perform better and remain more viable.

Furthermore, these variables, analyzed over the match, may help also to identify who are the players more affected by fatigue. By decreasing the number of players involved, it is possible to allow some players to rest actively. Moreover, it can characterize teams' attacking strategies. The direct play may increase centrality among some players and involves a lot of participation from forwards and strikers, for example as discussed in [25].

4.4.3 Patterns of Passes

We also investigated the possibility of using graph visual rhythms for analyzing patterns of passes. In this case, we have employed graphs defined by Flow Networks. From soccer matches, we computed the Ball Possession Flow Networks, in which vertices are players and edges are passes accomplished among teammates while the team has the ball possession. Considering the first half time from each analyzed match, it is possible to construct N different flow network graphs, considering all N time intervals in which each team has the ball possession. We compute the graph visual rhythm image for each team in a match. This image contains all passes accomplished among teammates in each ball possession flow network, i.e., in this case, $\mathcal{F}_{G_t}$ computes the occurrences of passes among players. Pixels representing an specific pass performed in a network were colored according to the location on the pitch where the ball passing occurred.

Figure 4.10 shows color patterns used. We divided the pitch in four sections, where the defensive area of a team was colored in cold colors (light blue and dark blue), while attacking area of a team (the opponent's pitch) was colored in hot colors (light red and dark red). Also, when a network does not have any edges (no passes performed), all its pixels are grey. Using this color pattern, it is possible to visually understand patterns of passes for a team.

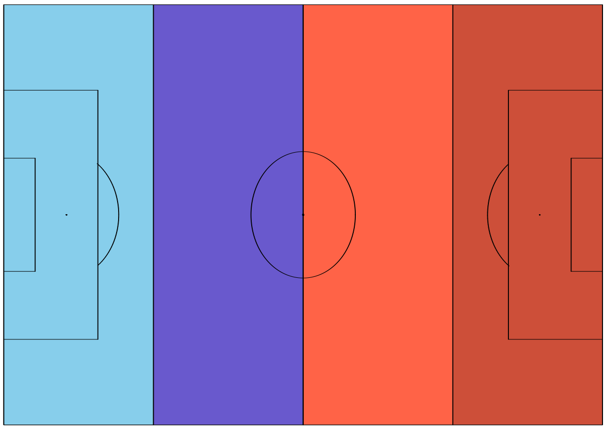


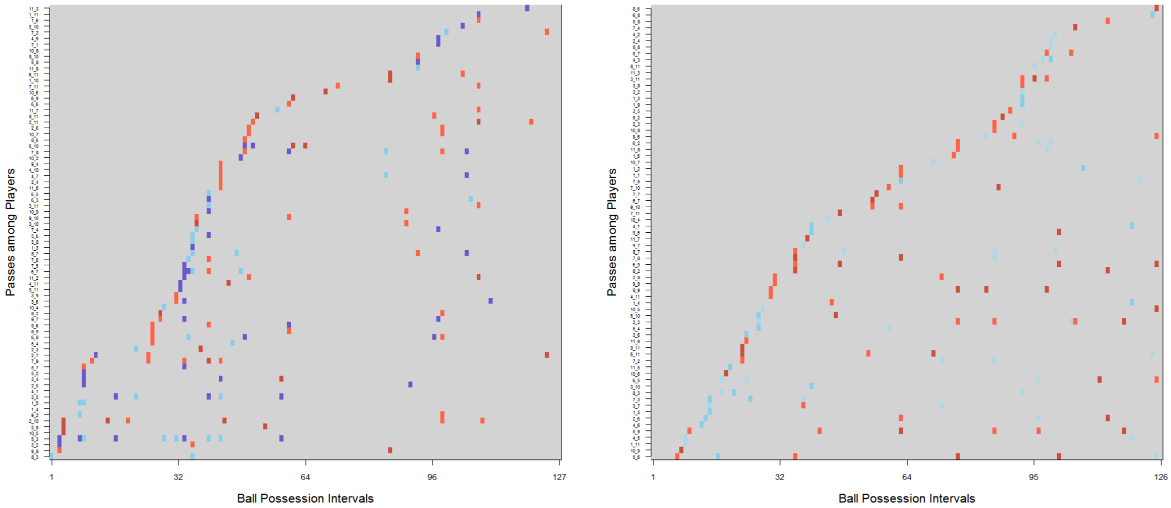
Figure 4.10: Pitch color patterns: defensive area in cold colors, while attacking area in hot colors.

We analyzed the first half time of the same two matches. The resulting graph visual rhythm images are shown in Figures 4.11 and 4.13. The Y axis has labels of players involved in successful passes (e.g., passes from player 3 to player 10), and the X axis refers to the flow networks considering ball possession. It is possible to notice some patterns in each match. During Match 1 (Figure 4.11), Team A performed more passes among teammates than Team B (more colored pixels in image of Team A). Also, Team A has longer vertical lines of pixels colored, which means that for each ball possession, many passes were activated involving many players. Figure 4.12 presents a graph visual rhythm for this same match, but now with y-axis representing passes ordered by players (from player 1 to 11). It is also possible to notice that many passes (vertical pixel lines) involve both defensive, middle, and forward players, in different field regions. Furthermore, some passes occurred many times along networks, and some of them only in its defensive area (cold-colored pixels). Team B has performed less ball passes, and passes in a single ball possession period involve only two players. Most of those passes occur in the attacking area (predominance of hot-colored pixels).

During Match 2 (Figure 4.13), Team B has performed more passes than Team A. It is interesting to notice that many of them occurred in its defensive area (predominance of cold-colored pixels), while team A performs more passes in the attacking area. Figure 4.14 presents a graph visual rhythm for this same match, but now with y-axis representing passes ordered by players (from player 1 to 11). It can be noticed that ball possession involves few players. In this match, Team C performed passes involving more players, from defensive to forward players. Note also that for two different matches, Team A has very different performance in terms of patterns of passes (see Figures 4.11(a) and 4.13(a)).

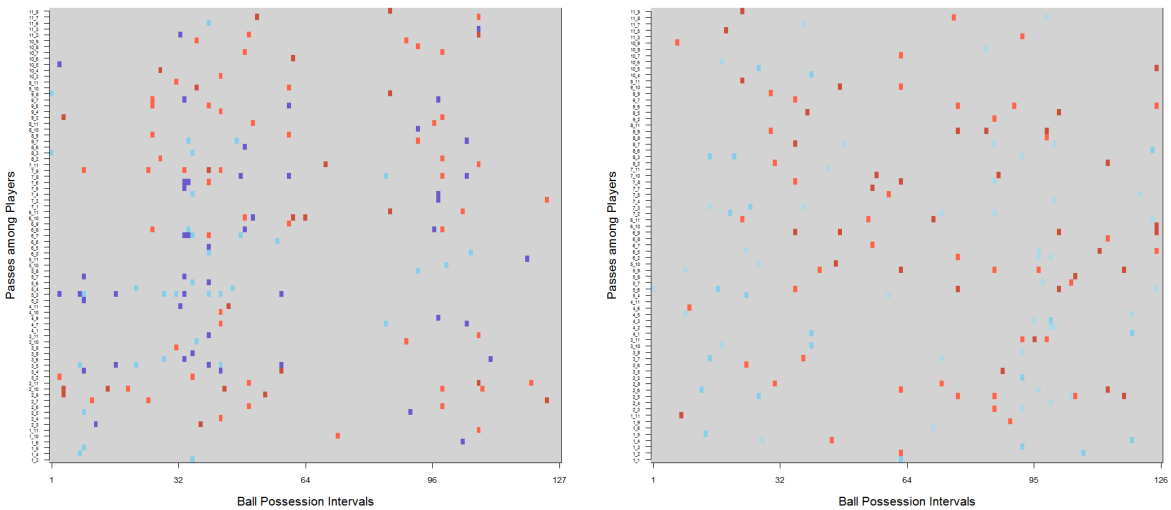
4.4.4 Pass Patterns in Attack Actions

It is also possible to create graph visual rhythm images considering a subset of players. For example, it might be interesting to show pass patterns involving only forward players. With this purpose, we created graph visual rhythms from passes involving only players with role, which is depicted in Figure 4.15. In this case, we refer to Match 1. We can



(a) Match 1 – Team A (b) Match1 – Team B

Figure 4.11: Graph visual rhythms encoding the patterns of passes of teams in **Match 1**.



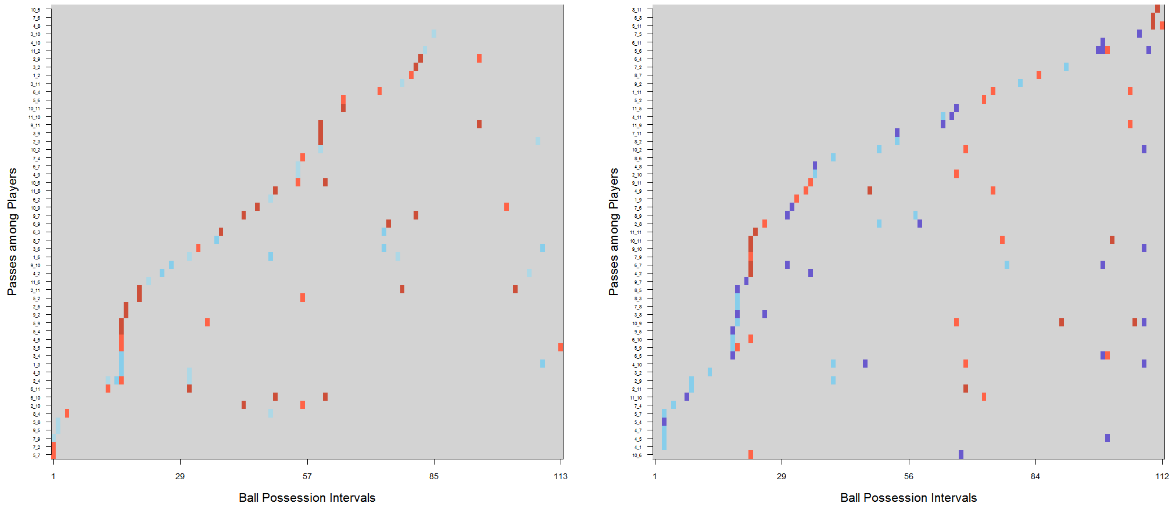
(a) Match 1 – Team A (b) Match1 – Team B

Figure 4.12: Graph visual rhythms (ordered by players) encoding the patterns of passes of teams in **Match 1**.

observe that not only did forward players of Team A accomplish more passes than players of Team B, but also they accomplished those passes in the opponent area (predominance of hot-colored pixels). We can conclude that Team A exploited more frequently the strategy of using multiple passes in attacking actions.

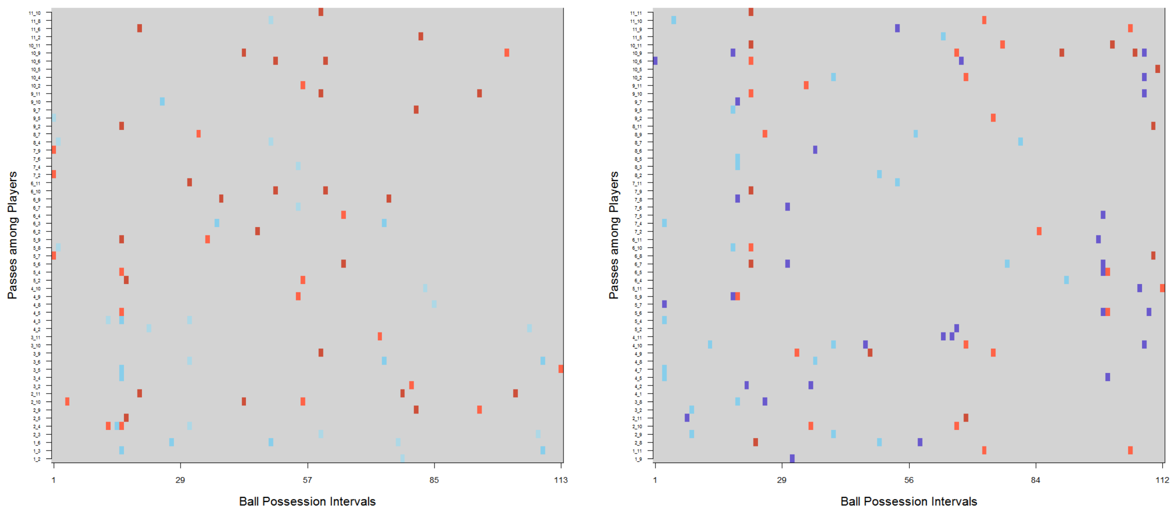
4.4.5 Soccer Visual Analytics Tool

We have created a soccer visual analytics tool that integrates the different graph extraction approaches, and visual rhythm image computation algorithms described in this chapter.



(a) Match 2 – Team A (b) Match 2 – Team C

Figure 4.13: Graph visual rhythms encoding the patterns of passes of teams in **Match 2**.



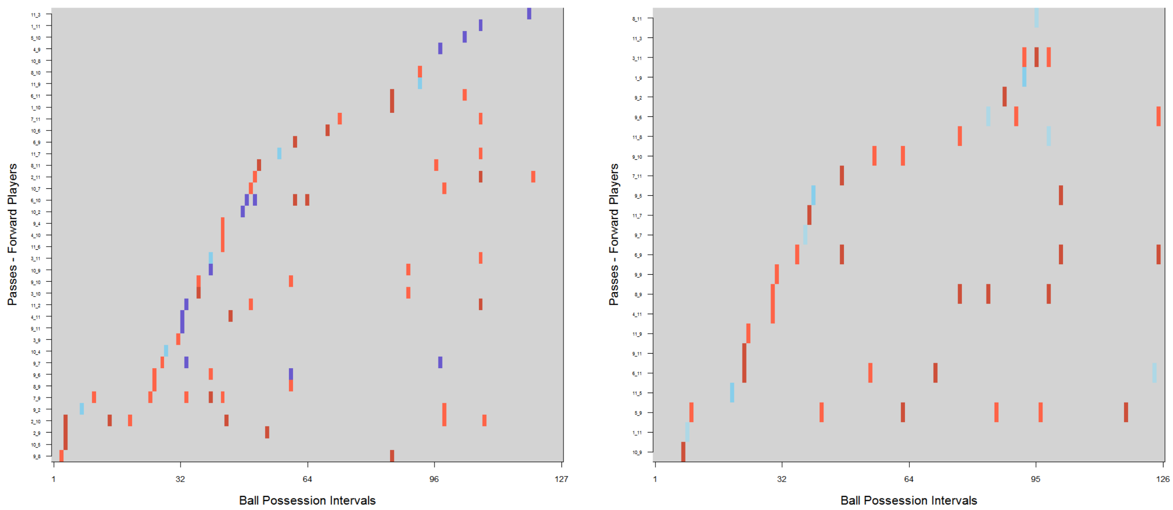
(a) Match 2 – Team A (b) Match 2 – Team C

Figure 4.14: Graph visual rhythms (ordered by players) encoding the patterns of passes of teams in **Match 2**.

This tool was developed in the R Shiny web application framework [20].

This tool allows loading data about soccer matches (information about players' location over time, and soccer match events), and encode them into graphs, depending on the type of analysis defined by the user (opponent aware or passes graphs). All complex network measures described in this thesis were implemented, so it is possible to visually analyze them by means of graph visual rhythms.

Figure 4.16 presents a typical usage example. In this case, we have graph visual rhythm images of two teams considering the diversity entropy measurement of the vertices, computed from opponent aware instant graphs. By clicking on the side-bar check boxes



(a) (b)

Figure 4.15: Graph visual rhythms encoding the patterns of passes of **forward players in Match 1**.

(area labeled with A), a user can highlight match events, such as goals and attacking moments, and also, can define a specific period of time in the match for zooming in targeting specific regions of the images (shown in region B). User can also view the corresponding opponent aware instant graphs by clicking anywhere on the graph visual rhythm image, or temporal graphs videos by selecting an area of interest in the image (field graph view in region C).

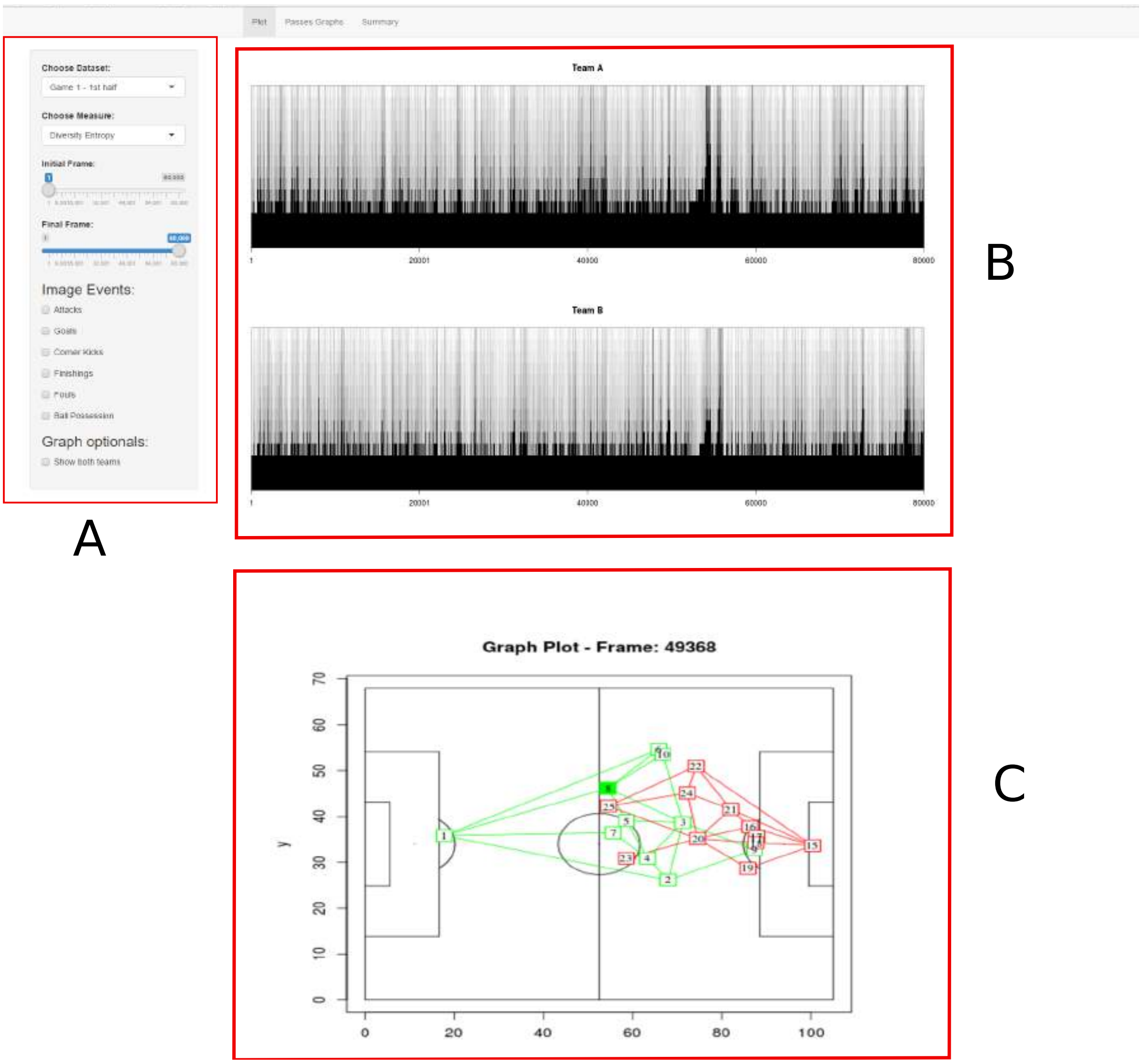


Figure 4.16: Screen shot of the soccer visual analytics tool developed.

Chapter 5

Network Measurements for Match Analysis

5.1 Introduction

One of the research questions addressed in this work refers to the investigation of appropriate complex network measurements for complex soccer match analysis. In Chapter 3, we proposed a novel method for modeling soccer matches, considering spatio-temporal data, the Opponent-Aware graphs. Using those graphs, we computed the Diversity entropy measure, applied the PCA of this measure for all players, and performed an ANOVA test, in order to analyze the relation between the measure and players' role. We also correlated this measure to passes performed by players, in order to verify the complexity of decision-making process of players. In Chapter 4, we propose a novel visualization tool for temporal graphs that allows to visually analyze the vertices behavior over time. In this chapter, we extend these previous studies, by proposing the characterization of the players' role according to multiple complex network measurements associated with them over time.

Some existing studies in the literature aim to characterize matches and team's performance [11, 22, 85], while others characterize passes [58] and goal attempts [69, 102, 103]. The studies consider players' trajectories, match events, and networks of flow of passes, but lack in considering simultaneously the spatio-temporal characteristics of the sport with the match events. In this chapter, we address this gap in the literature, extending the analysis performed in Chapter 3 by evaluating the effectiveness of seven different complex network measurements in the characterization of players according to their role in the match. Different from the existing studies in the literature, we exploit the spatio-temporal characteristics of the match as our analyzes are performed using temporal opponent-aware graphs.

5.2 Material and Methods

In this study, we have analyzed 10 soccer matches from the dataset, corresponding to 220 players, considering both teams of each match. To understand the behavior of players in different roles, we classified them into three categories: defenders, midfielders, and attacking. Players' role are determined automatically. We first computed the location distributions of team players on the pitch, and then estimated the relative position of each player given his team distribution. The team coverage area was divided into regions (bands), considering his location along the x-axis. Players whose most frequent locations were in bands 1 and 2 were labeled as defensive players. Those who spent most of their time in bands 3 and 4 were considered attacking players. The remaining players were labeled as midfielders.

We again use the same analysis framework defined in Section 3.4. Following this framework, we computed the opponent-aware graphs from each instant of time of each match, and extracted the following vertex complex network measurements: centrality, diversity entropy, pagerank, efficiency, vulnerability, degree, and eccentricity for each player. All the measures considered in this chapter are defined in Section 2.1. The extraction time of each measure, considering 45 minutes of play is about one hour. Once we have the measurement in a database, the training time of the algorithms is about a few seconds. In this way, once we have a trained model, it is possible to determine players'role in real time.

As the intention of this study is to investigate the behavior of players considering their measurement scores, we individually box-plotted these measurements for each team and match, considering the players grouped according to their roles. Next, we also compute the Principal Component Analysis (PCA) for all the measures, considering all matches. Finally, we evaluate the use of four classifiers – Nearest Neighbors, Support Vector Machines, Neural Networks, and Random Forest – to assign roles to players, according to their complex network measuremens, which we consider to be their features. All classification process is implemented using R Project libraries [90]. The Nearest Neighbors classifier is implemented using the 'class' package library [106]; Support Vector Machines implementation, in turn, considers the use of the 'e1071' package [73]; Random forests are implemented using the 'randomForest' package [17]; Neural Networks are implemented by the use of the 'neuralnet' package [49].

5.3 Results and Discussion

5.3.1 Boxplot and PCA Analysis

In the following, we present the analysis of each measure individually. First, we present a general boxplot of the measure, considering the 220 players, split into three categories (defensive, midfield, and attacking). Next, we analyze three matches (named Match 1, Match 2, and Match 3) as examples. We also plot the two principal component from PCA analysis. It is important to highlight that one of the teams is the same for all three matches, and will be called Team A, and the opponent teams in each match will be called, respectively: Team B, Team C, and Team D.

Centrality Analysis: The betweeness centrality measures the centrality of a vertex, considering the amount of shortest paths containing this vertex. In soccer context, this

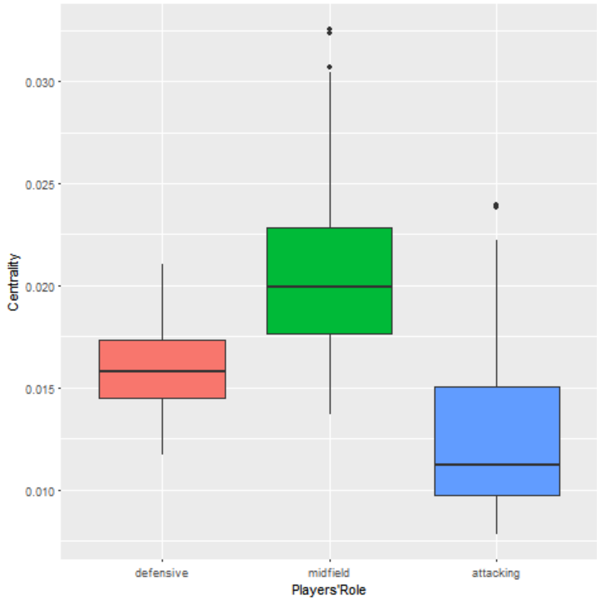
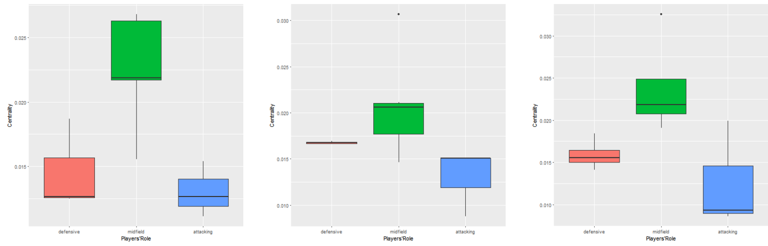


Figure 5.1: General Centrality boxplot, considering all the 220 players of the 10 matches analyzed.

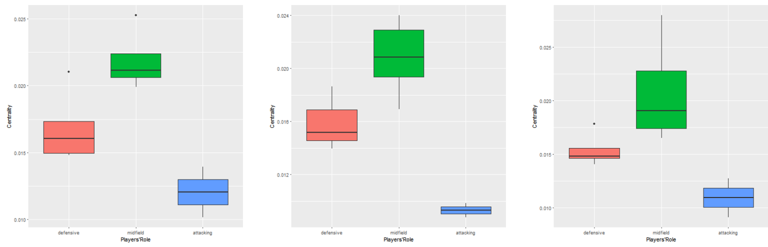
measure can indicate the players with better positioning in relation to their teammates. Figure 5.1 presents a complete boxplot of centrality considering all the 220 players of the 10 matches analyzed. In this plot, midfield players present higher centrality scores, which could mean they are better positioned on the pitch.

Figure 5.2 presents the betweeness centrality boxplots for Team A and its opponent teams in three matches. The individual match plots present the same structure as presented in the general one (Figure 5.1). As expected, midfield players seem to be well positioned, and present higher betweenness centrality scores, while attacking players, which in attacking moments usually have opponents nearby, have lower scores.

Figure 5.3 presents the PCA analysis for the same teams and matches. For Team A, we highlight that some midfield players behave very different from others given their distances in the PCA plot. For Match 1, Figure 5.3(a), almost all the midfield players are very distant from attacking and defensive players, except for Player 8. It explains the boxplot in Figure 5.2(a), where midfield players box is higher than others.

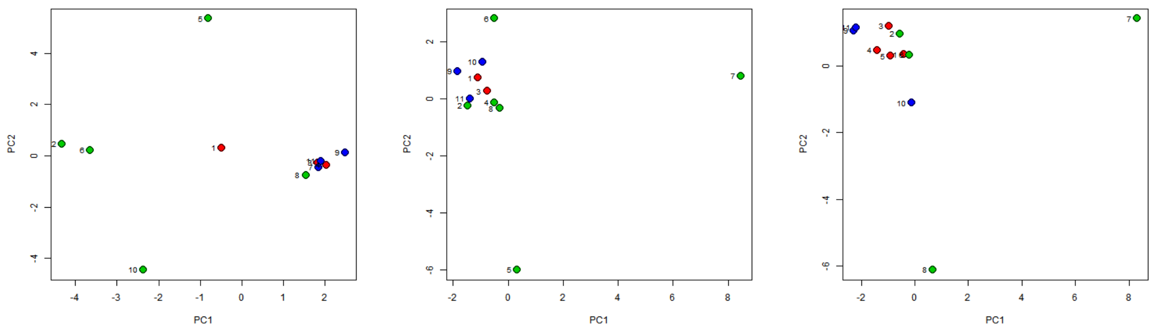


(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A

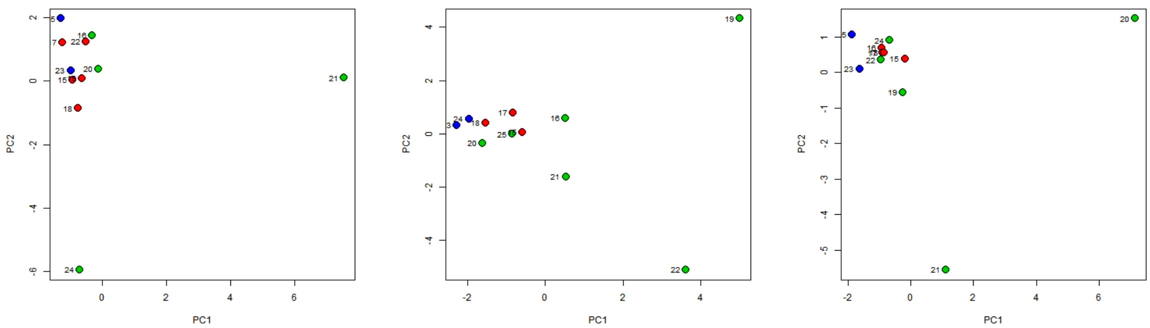


(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.2: Typical Centrality boxplots for three matches. Players were grouped into three classes, according to their role in the match.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A



(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.3: Typical centrality PCA for three matches. Players were colored in red, blue, and green, representing defensive, midfield, and attacking players, respectively.

Degree Analysis: The degree measurement quantifies the number of connections among vertices. In soccer analysis, it could be associated with a greater possibility of short and direct passes between players, and may indicate the player's availability for passes. Observing the plots in Figures 5.4 and 5.5, defensive players seem to have higher mean degree scores than midfield and attacking players. Also, Team A has similar scores of degree along the three matches. Analyzing the PCA plots, it seems that the degree scores organize players in categories, as suggested by the patterns in boxplots.

Comparing the degree and centrality scores of the teams, it is possible to conclude that, even though defensive players have higher degree scores, which means that they are connected to many teammates, they do not seem to be in strategic positions to make plays and ball passes, once their centrality scores are low.

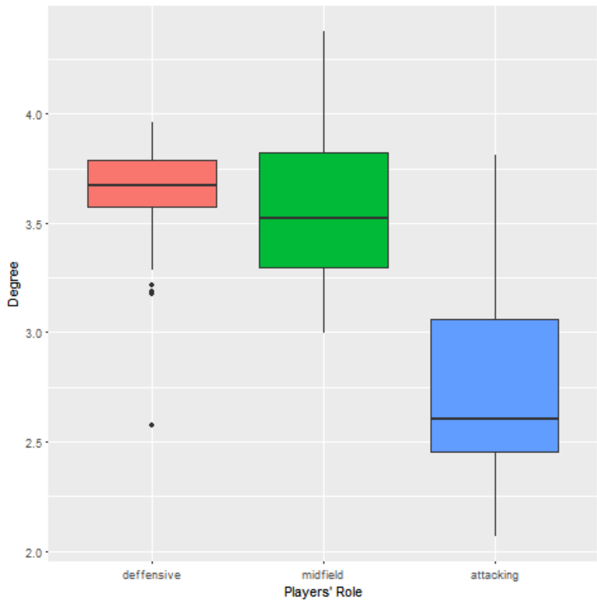
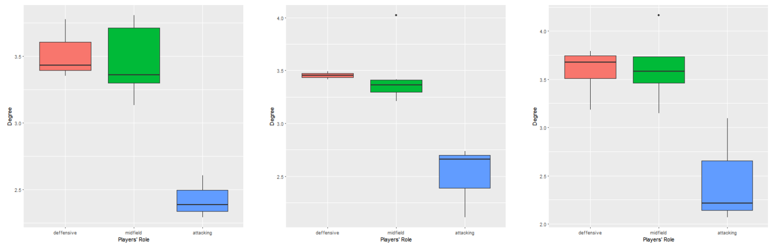
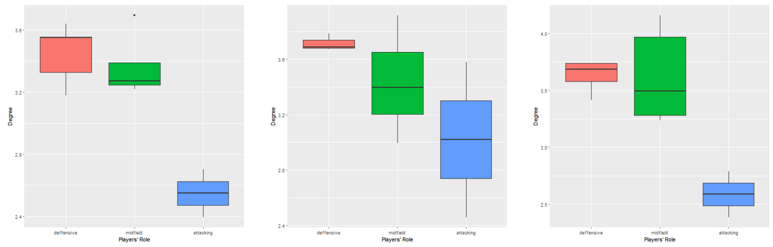


Figure 5.4: General Degree boxplot, considering all the 220 players of the 10 matches analyzed.

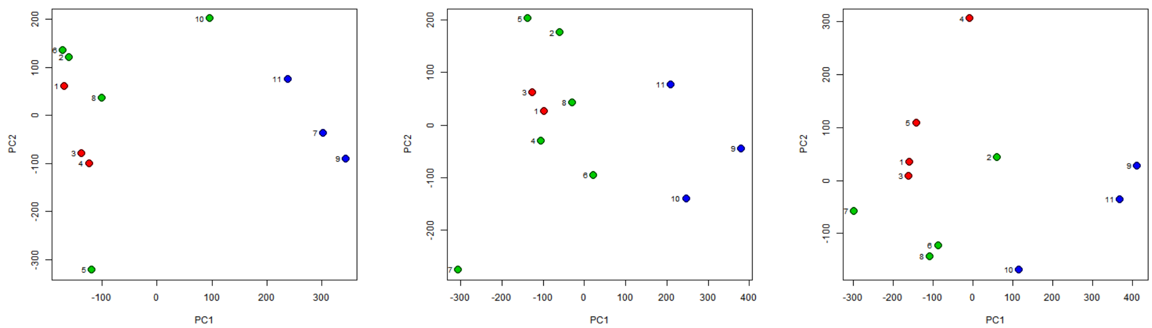


(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A

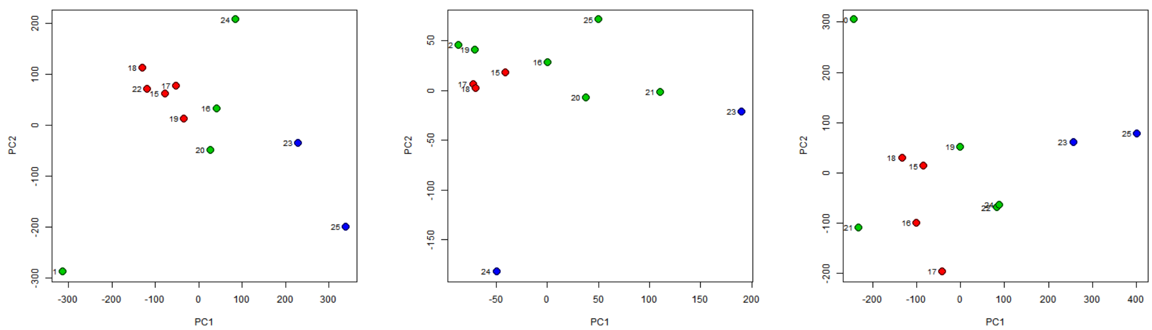


(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.5: Typical degree boxplots for three matches. Players were grouped into three classes, according to their role in the match.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A



(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.6: Typical Degree PCA for three matches. Players were colored in red, blue, and, green, representing defensive, midfield and attacking players, respectively.

Efficiency Analysis: The efficiency assesses the network's fault tolerance, by verifying the impact on local communication (direct neighbors) when a vertex and all its associated edges are removed. In soccer analysis, this can be associated with the players' importance for the local ball flow. Figures 5.7 and 5.8 present the Efficiency boxplots. The boxplots are very similar, for different teams and matches, with the one that takes into account all players. Similarly to what was observed in the Degree analysis, the Efficiency PCA graphs, Figure 5.9, provides well-defined groups of players, according to their classes.

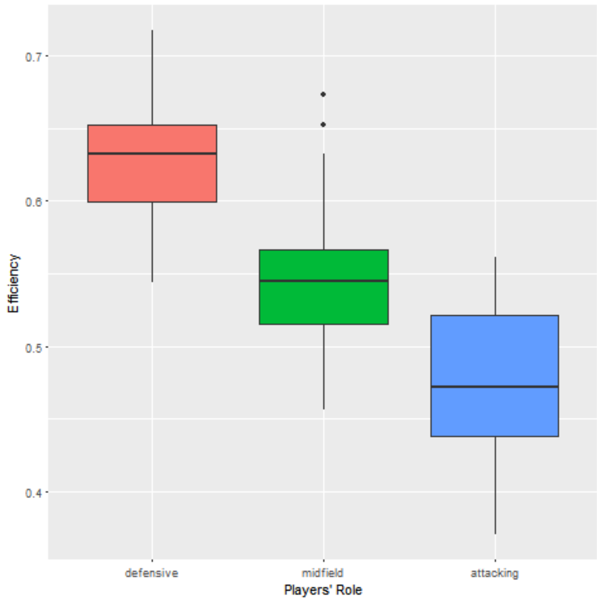
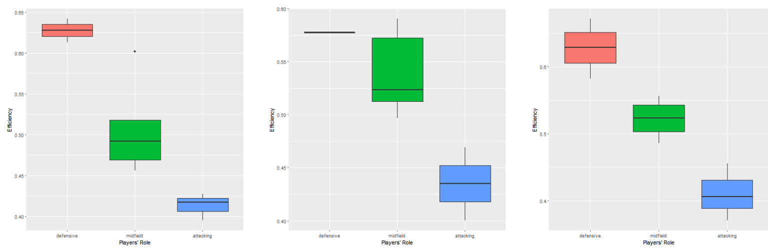
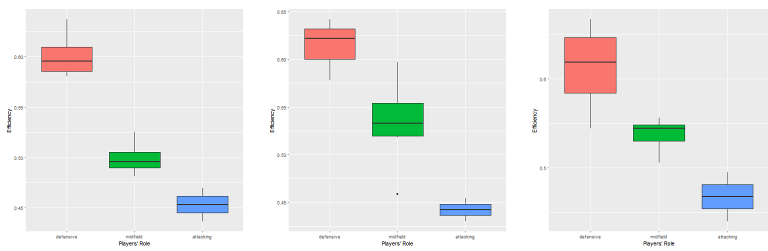


Figure 5.7: General Efficiency boxplot, considering all the 220 players of the 10 matches analyzed.

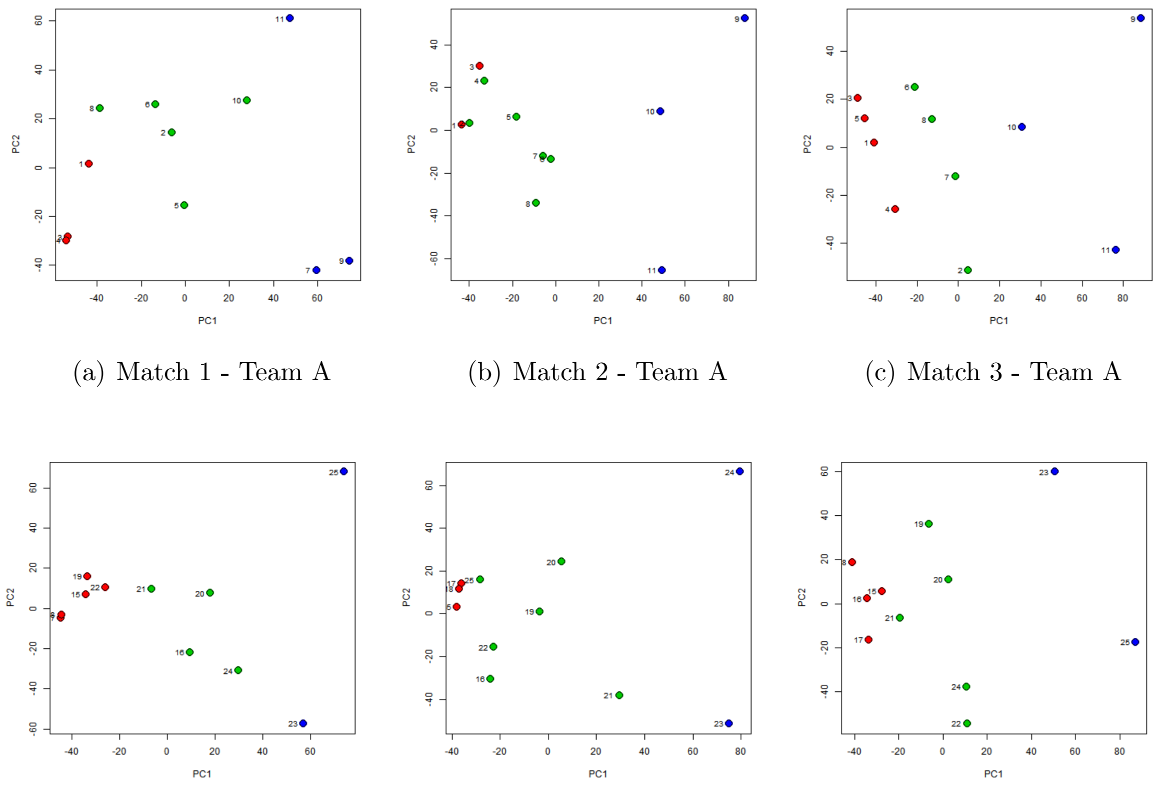


(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A



(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.8: Typical efficiency boxplots for three matches. Players were grouped into three classes, according to their role in the match.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A

(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.9: Typical Efficiency PCA for three matches. Players were colored in red, blue, and green, representing defensive, midfield, and attacking players, respectively.

PageRank Analysis: PageRank measures the prestige of a vertex based on the prestige of adjacent vertices. In soccer analysis, we can use this concept to verify if important players are connected within each other. By analyzing Figures 5.10 and 5.11, we can notice that for all teams and matches, defensive and midfield players present higher scores, meaning they have connections with important players. This is expected, once attacking players are usually isolated due to the presence of opponents nearby. The same phenomenon can be observed in the PCA graphics – Figure 5.12.

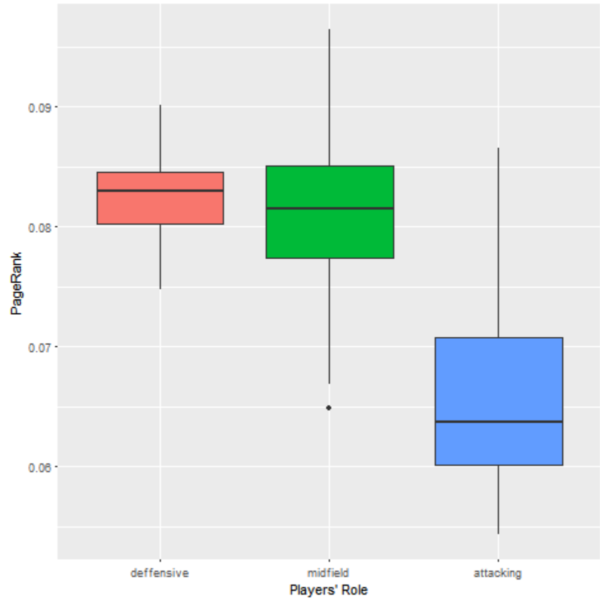
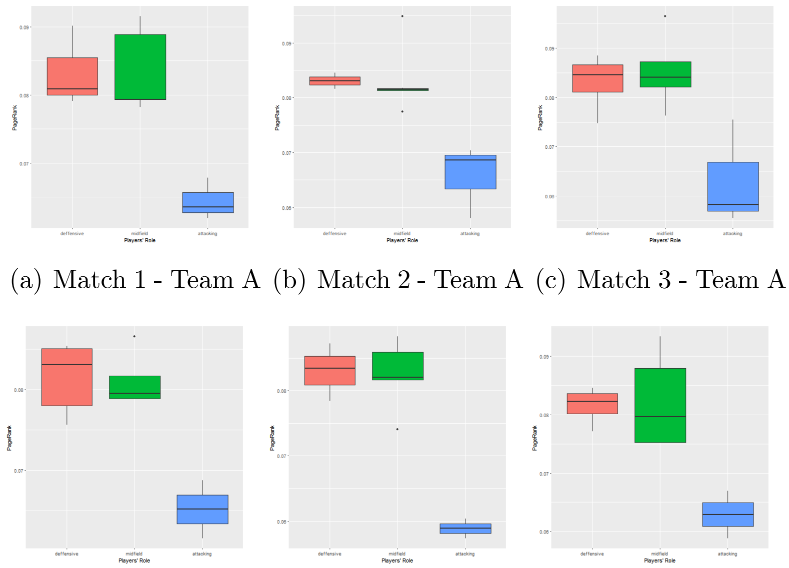


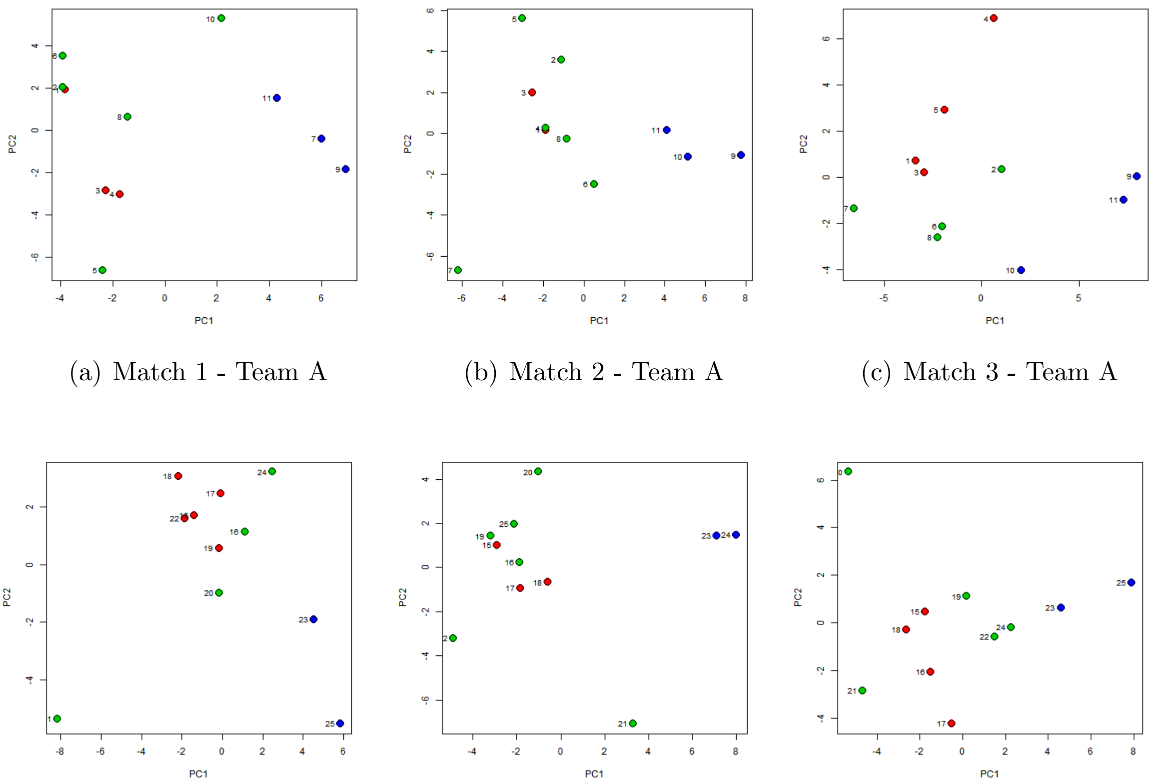
Figure 5.10: General PageRank boxplot, considering all the 220 players of the 10 matches analyzed.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A

(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.11: Typical PageRank boxplots for three matches. Players were grouped into three classes, according to their role in the match.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A

(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.12: Typical PageRank PCA for three matches. Players were colored in red, blue, and green, representing defensive, midfield, and attacking players, respectively

Vulnerability Analysis: The vulnerability of vertices quantifies the overall drop in global efficiency when one vertex is removed. In the soccer context, it can be used to measure the importance of a player considering the overall drop in possibility of passes, in case he is blocked by opponent players. Figures 5.13 and 5.14 show that, for all teams, on average, midfield players present higher values, meaning they play central roles in the team efficiency, considering the ball flow. In Figure 5.14, we observe that Team A presents boxplots with similar values across different matches. Since attacking players seem to have lower scores, in the PCA analysis shown in Figure 5.15, they appear very distant from other players.

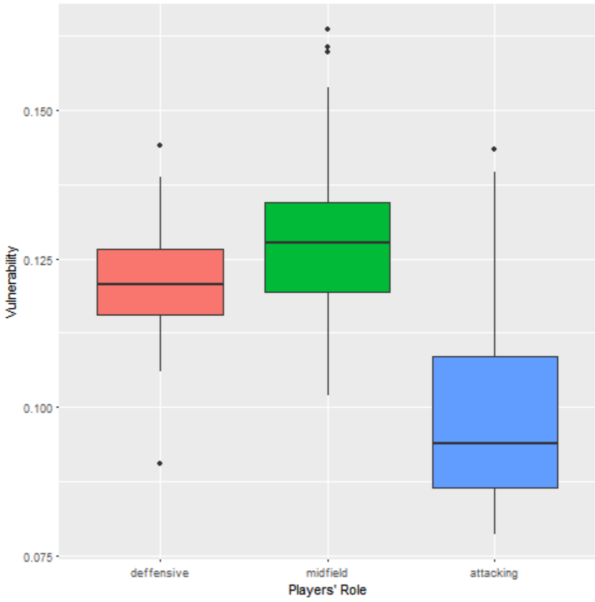
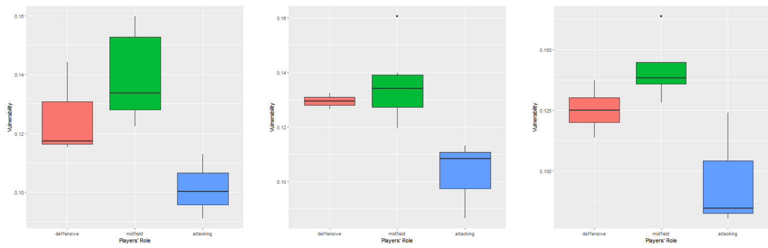
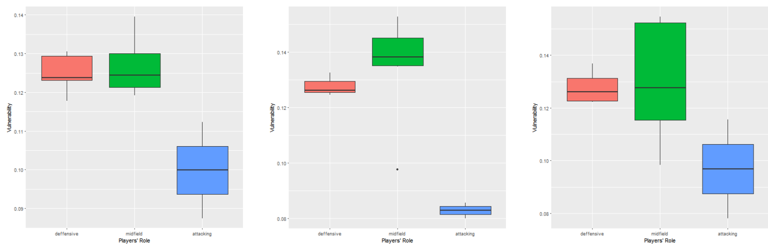


Figure 5.13: General Vulnerability boxplot, considering all the 220 players of the 10 matches analyzed.



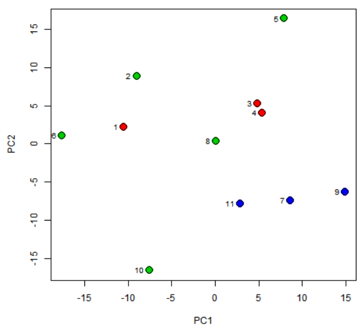
(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A



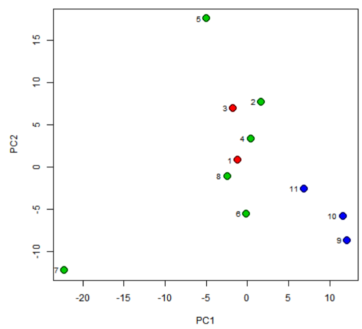
(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.14: Typical vulnerability boxplots for three matches. Players were grouped into three classes, according to their role in the match.

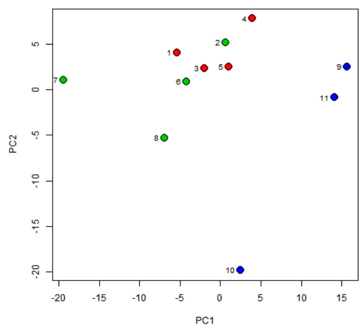
Eccentricity Analysis: The eccentricity measures how distant (considering connections) a node is from the most distant vertex of the graph. This concept can be used in soccer to measure the teams' spread on the pitch, considering the edges among players. Vertices with high eccentricity tends to be on the borders, in a non-central position. As expected, in Figures 5.16 and 5.17, it is possible to notice that midfield players present lower eccentricity scores, once they tend to be strategically positioned on the pitch in



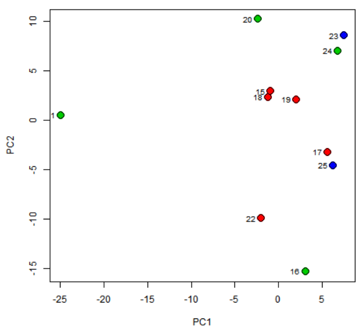
(a) Match 1 - Team A



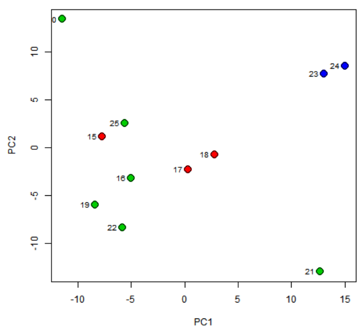
(b) Match 2 - Team A



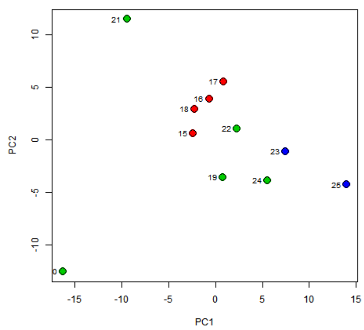
(c) Match 3 - Team A



(d) Match 1 - Team B



(e) Match 2 - Team C



(f) Match 3 - Team D

Figure 5.15: Typical Vulnerability PCA for three matches. Players were colored in red, blue, and green, representing defensive, midfield, and attacking players, respectively.

central positions acting in both defensive and attacking plays. As suspected by the degree/centrality analysis, defensive players seem to be in the borders, once they present high values of degree, associated with low centrality.

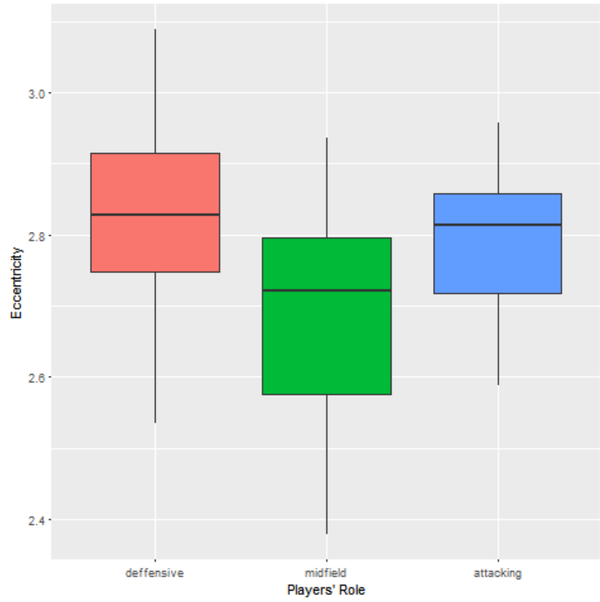
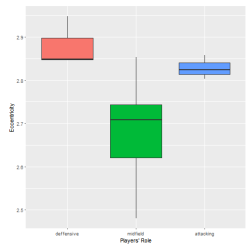
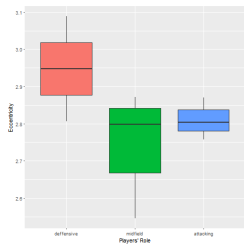


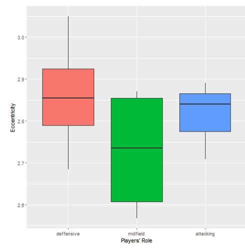
Figure 5.16: General Eccentricity boxplot, considering all the 220 players of the 10 matches analyzed.



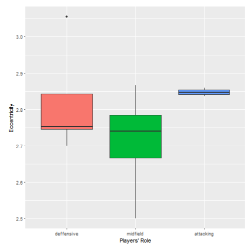
(a) Match 1 - Team A



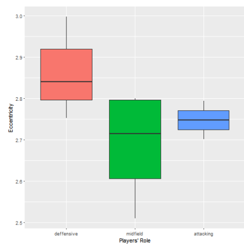
(b) Match 2 - Team A



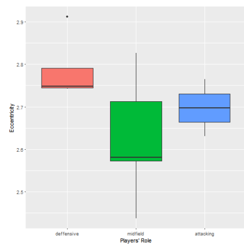
(c) Match 3 - Team A



(d) Match 1 - Team B

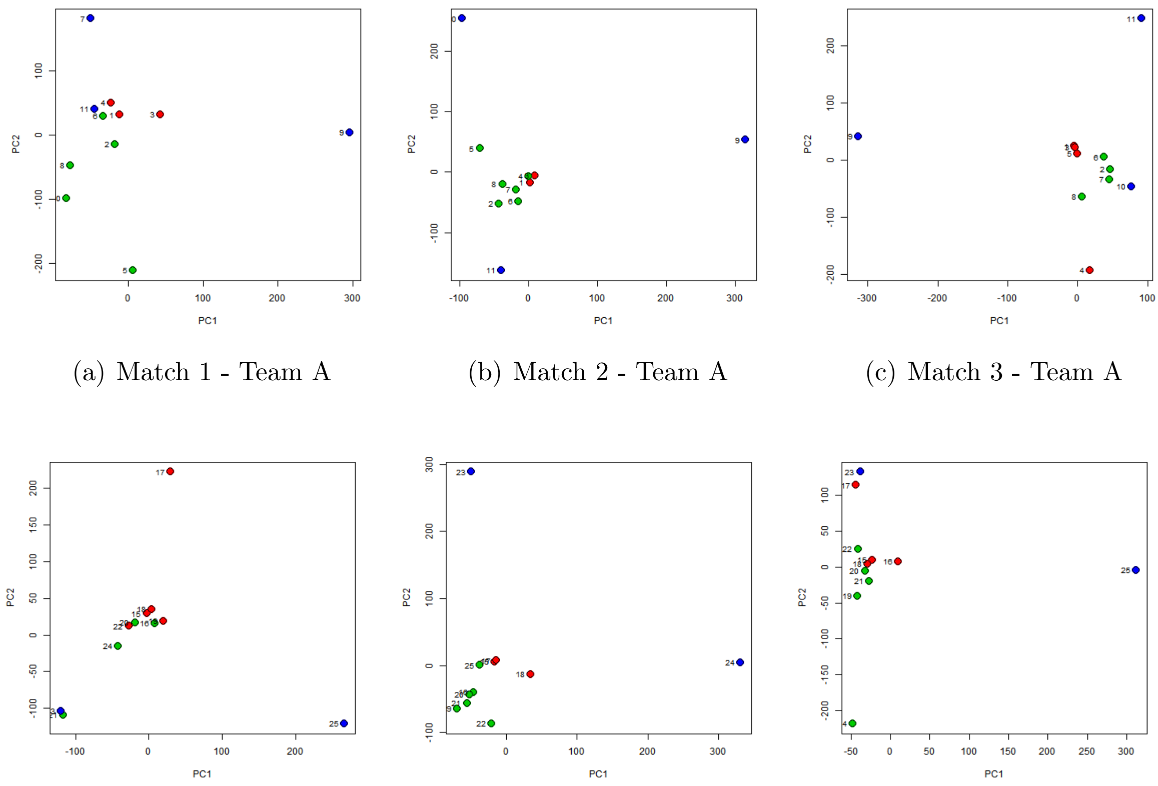


(e) Match 2 - Team C



(f) Match 3 - Team D

Figure 5.17: Typical Eccentricity boxplots for three matches. Players were grouped into three classes, according to their role in the match.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A

(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.18: Typical Eccentricity PCA for three matches. Players were colored in red, blue, and green, representing defensive, midfield, and attacking players, respectively.

Diversity Entropy Analysis: The diversity entropy quantifies the number of effectively accessible vertices at a given distance in steps. In soccer, we can use this to measure the accessibility of a player to his teammates, for ball passing purposes. As proposed in Chapter 3, diversity entropy highlights the difference among attacking players from other roles. Figures 5.19, 5.20, and 5.21 present results similar to those discussed in that chapter.

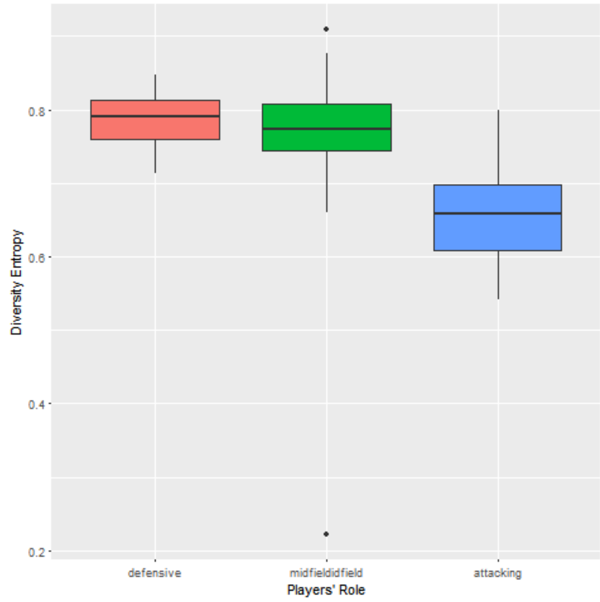
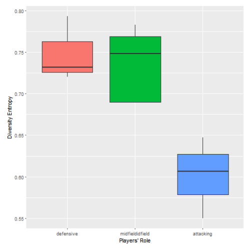
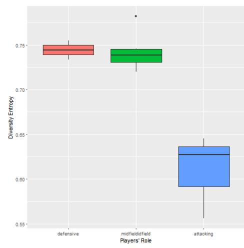
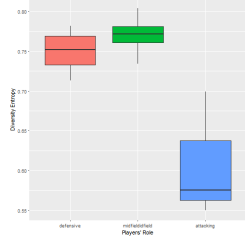
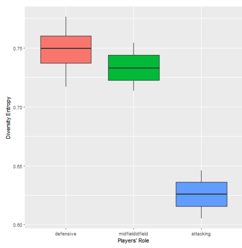


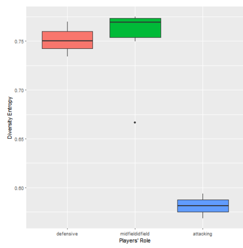
Figure 5.19: General Diversity entropy boxplot, considering all the 220 players of the 10 matches analyzed.


(a) Match 1 - Team A


(b) Match 2 - Team A


(c) Match 3 - Team A


(d) Match 1 - Team B


(e) Match 2 - Team C

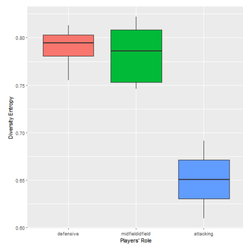
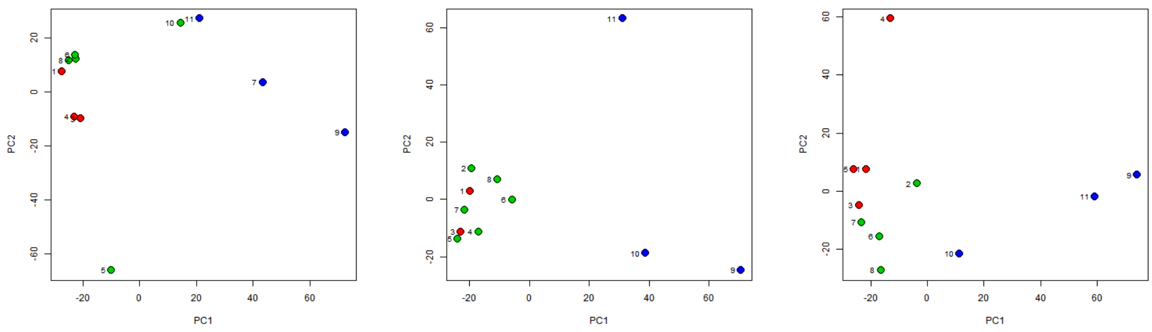
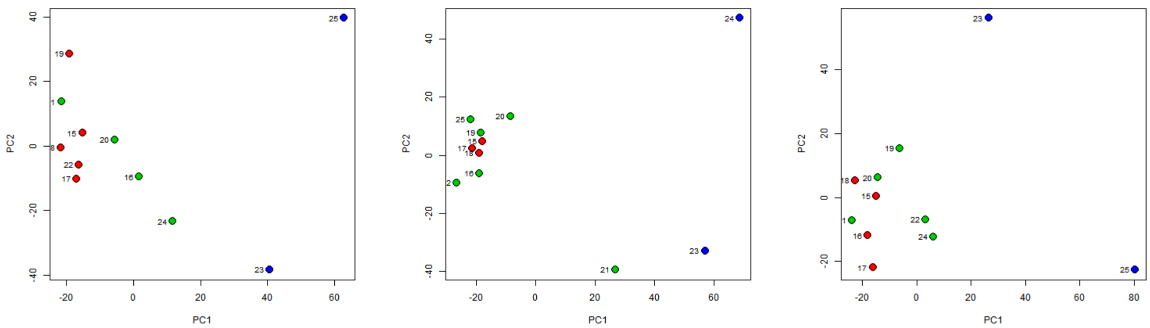

(f) Match 3 - Team D

Figure 5.20: Typical entropy boxplots for three matches. Players were grouped into three classes, according to their role in the match.



(a) Match 1 - Team A (b) Match 2 - Team A (c) Match 3 - Team A



(d) Match 1 - Team B (e) Match 2 - Team C (f) Match 3 - Team D

Figure 5.21: Typical diversity entropy PCA for three matches. Players were colored in red, blue, and green, representing defensive, midfield, and attacking players, respectively.

5.3.2 Classification

The boxplots images and PCA analysis show that some measurements seem to naturally lead to better classification of players according to their role in the match. Based on those results, we applied machine learning techniques, considering the mean of each measurement in a match as players' features. Each player is characterized by seven different features. For each one of the 10 matches, we considered both teams, resulting in 22 different players by match, and 220 players in all matches. These 220 players were automatically classified considering their positioning on the pitch along each match. Figure 5.22 presents an exploration view of the dataset, combining all the seven features in pairs. One relevant observation in this figure is the way players appear tightly grouped according to their classes, which could lead to an early suspicion that the players' classification according to those features is possible.

We evaluate the possibility of using these measurements for the classification of players. For this analysis, we compared the results of applying four classic machine learning algorithm commonly used in classification tasks: Nearest Neighbors (KNN), Support Vector Machines (SVM), Neural Networks, and Random Forest (RF). The algorithms were tuned to their best performance parameters.

We have performed a 10-fold cross validation to evaluate the algorithms results. The

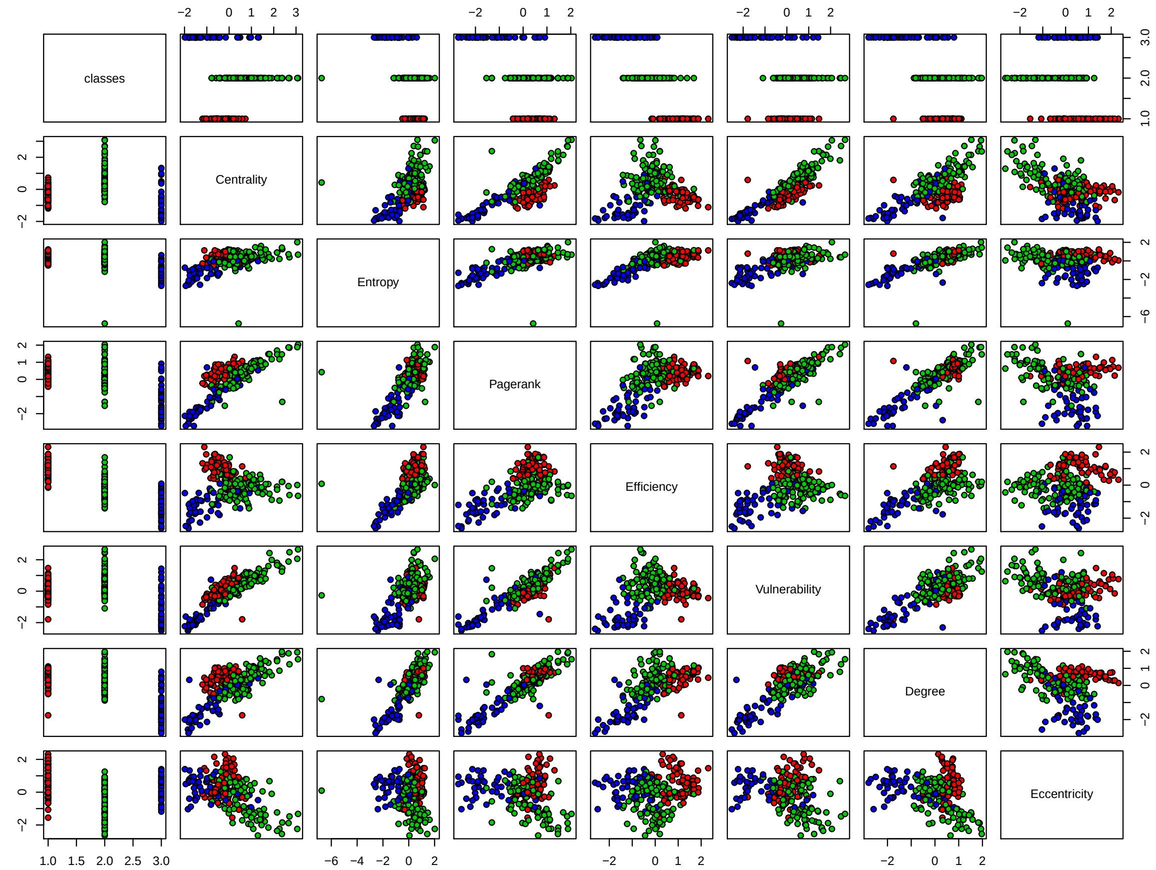


Figure 5.22: Dataset exploring plot, combining all the 7 features in pairs. The 220 players of 10 matches were colored according to their role as defensive (in red), midfield (in green), and attacking (in blue).

Table 5.1: Accuracy of different classification algorithms.

Classifier	Accuracy	Std. Dev.
KNN	0.85	0.07
SVM	0.86	0.06
Neural Networks	0.71	0.07
Random Forest	0.85	0.05

dataset was split in 10 folds, and therefore, the algorithms were executed 10 times, which garantees that all folds were used for training, and each fold was used for validation. We then calculated the mean accuracy scores and the standard deviation of the 10 results for each classifier. Results are shown in Table 5.1. As we can observe, KNN, SVM, and RF yield comparable results, which are better than those observed for the Neural Network classifier.

Figure 5.23 presents the accuracy and purity of each variable in the dataset. It is in accordance with our previous boxplots and PCA analyses.

Considering the players misclassified by the algorithms, there is still a chance that

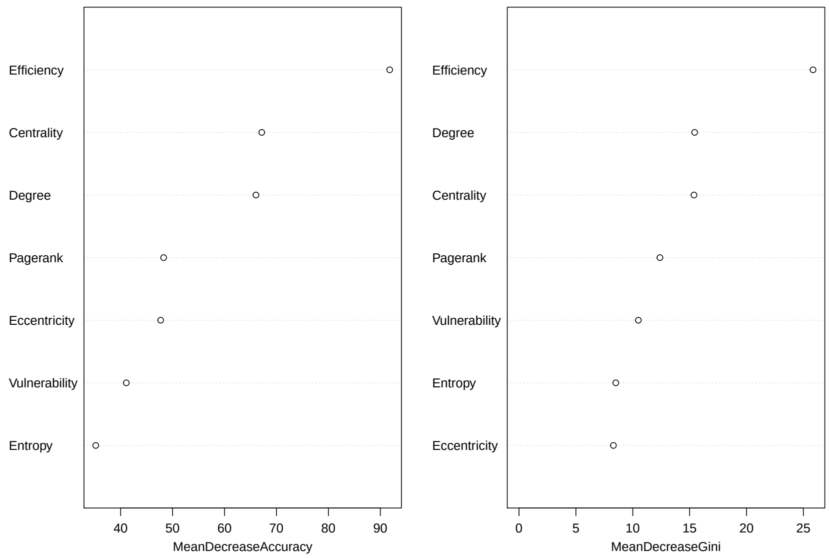


Figure 5.23: Random Forest results of importance of variables and purity of nodes in trees

Table 5.2: Feature vectors of Player 10, compared to mean scores for attacking and midfield players considering all matches.

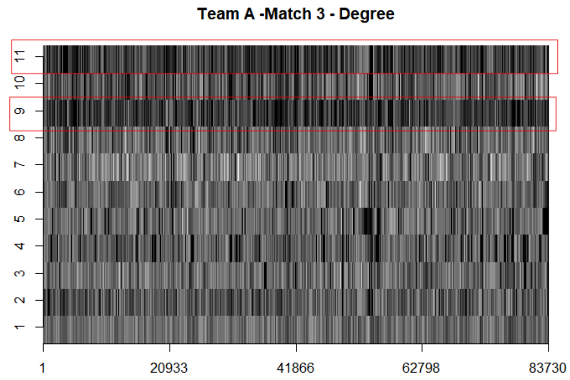
	Centrality	Degree	Efficiency	PageRank	Vulnerability	Eccentricity	Diversity Entropy
Player 10 - Team A Match 3	0.02	3.10	0.45	0.08	0.12	2.70	0.70
Attacking Mean Score	0.01	2.77	0.47	0.07	0.10	2.79	0.66
Midifield Mean Score	0.02	3.58	0.54	0.08	0.13	2.67	0.77

they were incorrectly labeled by the automatic classification algorithm initially used, or they changed their behavior along the match, according to events related to the success of the team, as goals scored. To analyze the first situation, we used as an example of misclassified player, Player 10 from Team A in Match 3. Originally, he was labeled as an attacking player, but all four algorithms classified him as a midfielder. By analyzing the PCA images described in Section 5.3.1 for all the measurements, we observe that either this player is located near midfield players, or he presents a very different behavior from the other attacking players, which indicates that in spite of being located in the attacking zone, he behaves as a midfield player. The feature vector for this player is presented in Table 5.2, and compared to the mean values found for attacking players considering all matches. For some measurements, as centrality, efficiency and pagerank, Player 10 presents scores that are more likely to be classified in midfield players group.

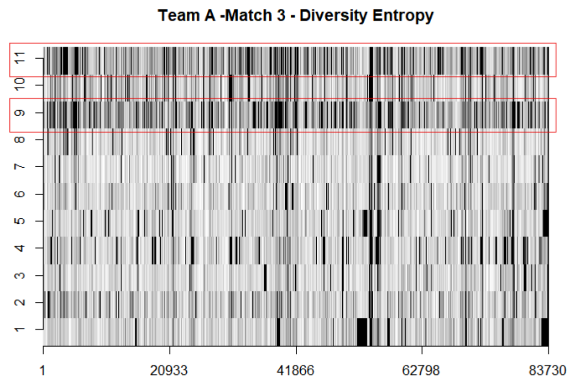
The different behavior of Player 10 along Match 3 can also be observed in Figure 5.24. This figure presents the graph visual rhythm images from Team A in Match 3 considering three measurements: degree, diversity entropy, and pagerank. The figure shows that the player's behavior is consistent along the match, as there are no breaks or color changes for the line related to this player, for each image. The figure also shows that his behavior is really different from attacking players, highlighted in red, which could explain his classification as a midfielder. These graph visual rhythm images allow a different analysis when compared to the table presented before. The measurement mean scores disregard the particularities of each match. Figure 5.24 allows to compare the behavior of Player 10 in relation to his teammates, which may be preferred in his performance analysis, since it privileges the spatio-temporal aspects of that particular match.

We summarize our findings from the conducted analysis as follows:

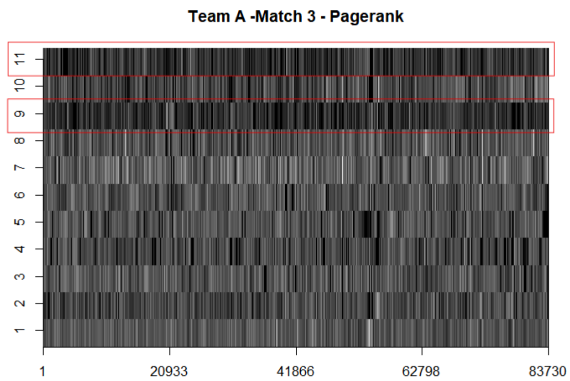
- Complex Network measurements are suitable for characterizing players' behavior according to their roles.
- By taking into account the results obtained by the machine learning algorithms, it is possible to infer that automatic classification of players according to measures of complex networks is feasible.
- Some measurements, such as efficiency, centrality, and degree seems to have better performance in classification algorithms, while others, like entropy, vulnerability, and pagerank seems to better characterize attacking players.
- Outlier players should be individually analyzed in order to find distinguished behaviors, when compared to teammates, considering the spatio-temporal aspects of a particular match.



(a)



(b)



(c)

Figure 5.24: Graph Visual Rhythm images for three measurements considering Team A of Match 3: (a) Degree, (b) Diversity Entropy, (c) PageRank. Lines highlighted in red refer to attacking players.

Chapter 6

Conclusions

In this chapter we summarize the main contributions of this thesis, as well as the possibilities of its extension.

6.1 Contributions

Sport analysis is a growing study area, which takes advantage of the big data made available through the use of automatic monitoring systems. This big data have been used to support the knowledge discovery process that might aid the definition of appropriate match strategies and training programs for several sports, including soccer, a sport that is very popular worldwide and that drives large volumes of money annually. In this analysis, the task of combining both spatial and temporal data has been challenging science to develop systems able to model, visualize, and analyze the dynamic nature of this sport. Among its contributions, this thesis has investigated the use of complex network measures for soccer match analysis based on the location of players and their opponents along the match, which means, considering the spatio-temporal dynamics of matches. The main goal was to support the knowledge discovery process by offering tools for modeling, analyzing, and visualizing this kind of information.

This study was developed according to the research questions proposed in Section 1.2. In the following, we address each research question raised, identifying scientific contributions developed:

- The first question was concerned with the investigation of **which graph model would better capture the spatio-temporal aspects of soccer matches based on players' location on the pitch**. We addressed this question by proposing a novel graph-based representation (Chapter 3), named *Opponent-Aware graph*, which takes into account the location of opponents in instant graphs. This representation considers both spatial and temporal information, since the physical locations of the players (vertices of the graph) determine the existence of edges between them. Since the edges represent the possibility of passes between teammate players, the existence of very close opponents decrease the chance of occurrence of passes, or make it very difficult to execute them. In this sense, we analyzed the complexity of making passes decisions for highly marked players, like the ones in forward role.

Performed analyses based on professional soccer matches demonstrated that the proposed graph representation is effective for match characterization. Following this approach, we also discussed the relationship among diversity entropy measure and the possibility of performing passes. This measurement was used to characterize the possibility of passes.

- Another research question concerned the identification of **which information visualization approach would be suitable for supporting the analysis of temporal changes in dynamic graphs**. To address this question, we introduced a novel visualization approach (Chapter 4), the *graph visual rhythm* representation, a compact visual structure to encode changes in temporal graphs, making it a suitable solution to handle large volumes of data. We demonstrated its applicability in several usage scenarios concerning the analysis of soccer matches, whose various dynamic aspects were encoded into temporal graphs. The graph visual rhythm analysis was developed through the implementation of a visual analytics tool.

- Those scenarios exploited in the Chapter 4 also addressed another proposed research question related to the identification of **which complex networks measurements better characterize events of interest in soccer matches**. Through the use of graph visual rhythm, we demonstrated that events of the match, attacking/defensive moments were evidenced.

- The remaining question, which refers to the investigation of **which complex networks measurements better characterize the players' role**. This question was addressed in Chapter 5. We presented the anlaysis of seven complex network measurements used as features vectors associated with players. We also proposed a classification system that considers players' feature vectors to determine their roles as defenders, midfield, and forwards. Classification results demonstrated that the use of measurements was quite effective in determining players' role, especially for the Nearest Neighbor, Support Vector Machines, and Random Forest classifiers. We believe that these findings open new opportunities related to the investigation of the use of complex network measures for characterizing multiple graphs over time.

In summary, the main contributions of this work are:

1. The proposal of framework for the analysis of soccer matches based on the use of graph-based representations and complex-network measurements.

2. The proposal of a graph-based representation, named opponent-aware graph, which takes into account the location of opponents. Performed analyses based on professional soccer matches demonstrated that the proposed graph representation is effective for match characterization. Furthermore, we were able to demonstrate that there exist a moderate correlation among the frequency of passes and the diversity entropy scores of the different players. We also demonstrated that diversity entropy scores are useful to characterize the roles of attacking players. We believe that these findings open new opportunities related to the investigation of the use of complex network measures for characterizing multiple graphs over time.

3. The proposal of the graph visual rhythm representation, a compact visual structure to encode changes in temporal graphs, making it a suitable solution to handle large volumes of data. We demonstrated its applicability in several usage scenarios concerning the analysis of soccer matches, whose several dynamic aspects are encoded into temporal graphs. We also proposed a soccer analytics visual tool intended to highlight aspects of the game.

4. Identification of suitable complex-network measurements for the analysis of complex spatio-temporal patterns of soccer matches.

The main hypothesis that guided this study was: *Temporal graphs and associated complex-network measurements are effective to model the spatio-temporal dynamics of soccer matches and potentially improve soccer matches analysis.* We showed that the proposed *instant Opponent-Aware graphs* are suitable for the spatial representation of players on the pitch, and that the extraction of features (complex network measurements) from the graphs' vertices can accurately characterize players role and events of interest of the match. We believe, therefore, that the raised hypothesis was confirmed.

6.2 Research Limitations

Some limitations already identified of the conducted study are summarized in the following and may be addressed in future work:

- This study is based on the use of spatio-temporal information of soccer matches. Tests with sample rates between 20 Hz and 30 Hz were performed, although, sample rates lower than these values may impact the results obtained.

- In addition, the impact of replacing players during the matches were not considered. This analysis can provide interesting insights about the dynamics of games.

- All players were considered in the graph extraction process, including goalkeepers. The positioning of the goalkeeper can add bias in the analysis of the results, since most of the time, the goalkeeper does not participate actively in the construction of the plays.

- The use of other graph representation of even the use of different thresholds for the edge removal step in the Opponent-Aware Graph construction approach can impact the results of the complex-network measurements. All the measurements consider existing edges connected to vertices, which are directly dependent on the threshold value used for edge removal. The use of other graph representation can lead to different conclusions in the analysis of passes and performance of players as well.

6.3 Future Work

This research opens novel opportunities for investigation like:

- The use of several image processing algorithms to highlight important patterns in temporal graphs. We plan to follow this research venue in our future work. We also plan to incorporate matrix reordering methods [9] aiming to improve the identification of changing patterns in graph visual rhythm representations.
- The implementation of suitable visualization, considering graph visual rhythm images, to handle players' substitutions in a match.
- The evaluation of graph-based complex network measurements to characterize moments of interest like the attacking/defense transition, or defensive strategies poorly performed.
- By taking into account the characterization of players by means of a feature vector of complex network measurements, we propose, as future work, the analysis of how players change their roles in specific periods (e.g., with and without ball possession), verifying the impact of the opponent's tactical strategies in the team's performance.
- Another research venue refers to the mining of frequent subgraphs over time, with the objective of assessing the most used organization patterns on the pitch for each team during attack/defense actions.
- Given the analysis framework proposed in this study, some natural extensions are possible considering its different steps, such as:
  - The investigation of different graph models, based for example on the distances among players, or complete graphs as basis for edge removal, considering the opponents' position.
  - The evaluation of other complex network measures [34] in soccer match analysis based on players' locations over time.

6.4 Published Papers

This study has resulted in the following papers:

- The concept of Opponent-Aware graph computation, presented in Chapter 3, was presented as a poster, **Utilizando Redes complexas para Análise de Jogos de Futebol** [94] in the *XVII Congresso Brasileiro de Biomecânica*, Daniele Cristina Uchôa Maia Rodrigues, Felipe Arruda Moura, Sergio Augusto Cunha, and Ricardo da Silva Torres.

  This work was recipient of a honorable mention award.

- The content of Chapter 4 is based on the published full article: **Visualizing Temporal Graphs using Visual Rhythms – A Case Study in Soccer Match Analysis** [95]. Daniele Cristina Uchôa Maia Rodrigues, Felipe Arruda Moura, Sergio Augusto Cunha and Ricardo da Silva Torres. In the *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics*

Theory and Applications – Volume 3: IVAPP, (VISIGRAPP 2017), ISBN 978-989-758-228-8, pages 96-107. DOI: 10.5220/0006153000960107.

Bibliography

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

[2] Duarte Araújo and Keith Davids. Team synergies in sport: theory and measures. *Frontiers in Psychology*, 7:1449, 2016.

[3] Daniel Archambault, Helen Purchase, and Bruno Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552, 2011.

[4] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. Visualizing dynamic networks with matrix cubes. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pages 877–886. ACM, 2014.

[5] Natàlia Balague, Carlota Torrents, Robert Hristovski, Keith Davids, and Duarte Araújo. Overview of complex systems in sport. *Journal of Systems Science and Complexity*, 26(1):4–13, 2013.

[6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[7] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. The State of the Art in Visualizing Dynamic Graphs. In R. Borgo, R. Maciejewski, and I. Viola, editors, *EuroVis - STARs*. The Eurographics Association, 2014.

[8] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. A taxonomy and survey of dynamic graph visualization. In *Computer Graphics Forum*, volume 36, pages 133–159. Wiley Online Library, 2017.

[9] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. Matrix reordering methods for table and network visualization. In *Computer Graphics Forum*, volume 35, pages 693–716. Wiley Online Library, 2016.

[10] Francisco Nivando Bezerra and E Lima. Low cost soccer video summaries based on visual rhythm. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 71–78. ACM, 2006.

[11] Alina Bialkowski, Patrick Lucey, Peter Carr, Yisong Yue, Sridha Sridharan, and Iain Matthews. Identifying team style in soccer using formations learned from

spatiotemporal tracking data. In *2014 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 9–14. IEEE, 2014.

[12] Alina Bialkowski, Patrick Lucey, Peter Carr, Yisong Yue, Sridha Sridharan, and Iain Matthews. Large-scale analysis of soccer matches using spatiotemporal tracking data. In *2014 IEEE International Conference on Data Mining (ICDM)*, pages 725–730. IEEE, 2014.

[13] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175–308, 2006.

[14] John Adrian Bondy and Uppaluri Siva Ramachandra Murty. *Graph theory with applications*, volume 290. Elsevier Science Ltd., 1976.

[15] Ulrik Brandes and Steven R. Corman. Visual unrolling of network evolution and the analysis of dynamic discourse. *Information Visualization*, 2(1):40–50, 2003.

[16] Ulrik Brandes, Natalie Indlekofer, and Martin Mader. Visualization methods for longitudinal social networks and stochastic actor-oriented modeling. *Social Networks*, 34(3):291–308, 2012.

[17] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[18] Michael Burch, Markus Hoferlin, and Daniel Weiskopf. Layered timeradartrees. In *2011 15th International Conference on Information Visualisation (IV)*, pages 18–25. IEEE, 2011.

[19] Arnaud Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 27(5):387–408, 2012.

[20] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. *shiny: Web Application Framework for R*, 2016. R package version 0.13.2.9004.

[21] Seong Soo Chun, Hyeokman Kim, Kim Jung-Rim, Sangwook Oh, and Sanghoon Sull. Fast text caption localization on video using visual rhythm. In *International Conference on Advances in Visual Information Systems*, pages 259–268. Springer, 2002.

[22] Paolo Cintia, Fosca Giannotti, Luca Pappalardo, Dino Pedreschi, and Marco Malvaldi. The harsh rule of the goals: data-driven performance indicators for football teams. In *IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015. 36678 2015.*, pages 1–10. IEEE, 2015.

[23] Paolo Cintia, Salvatore Rinzivillo, and Luca Pappalardo. A network-based approach to evaluate the performance of football teams. In *Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal*, 2015.

[24] Filipe Manuel Clemente, Micael Santos Couceiro, Fernando Lourenço Martins, and Rui Mendes. An online tactical metrics applied to football game. *Research Journal of Applied Sciences, Engineering and Technology*, 5(5):1700–1719, 2013.

[25] Filipe Manuel Clemente, Micael Santos Couceiro, Fernando Manuel Lourenço Martins, and Rui Sousa Mendes. Using network metrics in soccer: A macro-analysis. *Journal of Human Kinetics*, 45(1):123–134, 2015.

[26] Filipe Manuel Clemente, Micael Santos Couceiro, Fernando Manuel Lourenço Martins, Rui Sousa Mendes, and António José Figueiredo. Practical implementation of computational tactical metrics for the football game. In *International Conference on Computational Science and Its Applications*, pages 712–727. Springer, 2014.

[27] Filipe Manuel Clemente and Fernando Manuel Lourenço Martins. Estudo da sequéncia de passes entre jogadores profissionais de futebol durante os jogos em casa ao longo de uma época desportiva: aplicabilidade das medidas de social network analysis. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 12(2):195–202, 2017.

[28] Filipe Manuel Clemente and Fernando Manuel Lourenço Martins. Who are the prominent players in the UEFA champions league: An approach based on network analysis. *Walailak Journal of Science and Technology (WJST)*, 14(8), 2017.

[29] Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, Micael Santos Couceiro, Rui Sousa Mendes, and António José Figueiredo. A network approach to characterize the teammates' interactions on football: A single match analysis. *Cuadernos de Psicología del Deporte*, 14(3):141–148, 2014.

[30] Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, Dimitris Kalamaras, Joana Oliveira, Patrícia Oliveira, and Rui Sousa Mendes. The social network analysis of switzerland football team on fifa world cup 2014. *Journal of Physical Education and Sport*, 15(1):136, 2015.

[31] Filipe Manuel Clemente, Micael Santos Couceiro, Fernando Lourenço Martins, Rui Sousa, and Antonio Figueiredo. Intelligent systems for analyzing soccer games: The weighted centroid. *Ingeniería e Investigación*, 34(3):70–75, 2014.

[32] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[33] Luciano da Fontoura Costa, Osvaldo Novais Oliveira Junior, Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana, and Luis Enrique Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.

[34] Luciano da Fontoura Costa, Francisco Aparecido Rodrigues, Gonzalo Travieso, and Paulino Ribeiro Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1):167–242, 2007.

[35] Carlos Cotta, Antonio Mora, Juan Julián Merelo, and Cecilia Merelo-Molina. A network analysis of the 2010 FIFA world cup champion team play. *Journal of Systems Science and Complexity*, 26(1):21–42, 2013.

[36] David Gutierrez Diaz del Campo, Sixto Gonzalez Villora, Luis Miguel Garcia Lopez, and Stephen Mitchell. Differences in decision-making development between expert and novice invasion game players. *Perceptual and Motor Skills*, 112(3):871–888, 2011.

[37] Paramita Dey, Maitreyee Ganguly, and Sarbani Roy. Network centrality based team formation: A case study on t-20 cricket. *Applied Computing and Informatics*, 2016.

[38] Jordi Duch, Joshua Waitzman, and Luís Nunes Amaral. Quantifying the performance of individual players in a team activity. *PLoS ONE*, 5(6):e10937, 2010.

[39] Michael Farrugia and Aaron Quigley. Effective temporal graph layout: A comparative study of animation versus static display methods. *Information Visualization*, 10(1):47–64, 2011.

[40] Jean-Daniel Fekete, Jarke Jan Van Wijk, John Thomas Stasko, and Chris North. The value of information visualization. In *Information Visualization*, pages 1–18. Springer, 2008.

[41] Tharindu Fernando, Xinyu Wei, Clinton Fookes, Sridha Sridharan, and Patrick Lucey. Discovering methods of scoring in soccer using tracking data. *Large-Scale Sports Analytics, Sidney*, 2015.

[42] Jennifer Fewell, Dieter Armbruster, John Ingraham, Alexander Petersen, and James Waters. Basketball teams as strategic networks. *PLoS ONE*, 7(11):e47445, 2012.

[43] Pascual Jovino Figueroa, Neucimar Jerônimo Leite, and Ricardo Machado Leite Barros. Background recovering in outdoor image sequences: An example of soccer players segmentation. *Image and Vision Computing*, 24(4):363–374, 2006.

[44] Pascual Jovino Figueroa, Neucimar Jerônimo Leite, and Ricardo Machado Leite Barros. Tracking soccer players aiming their kinematical motion analysis. *Computer Vision and Image Understanding*, 101(2):122–135, 2006.

[45] Iztok Fister, Karin Ljubič, Ponnuthurai Nagaratnam Suganthan, and Matjaž Perc. Computational intelligence in sports: challenges and opportunities within a new research domain. *Applied Mathematics and Computation*, 262:178–186, 2015.

[46] Preparata Franco and Michael Ian Preparata Shamos. *Computational geometry: an introduction*. Springer Science & Business Media, 2012.

[47] Wouter Frencken and Koen Lemmink. Team kinematics of small-sided soccer games: A systematic approach. *Science and football VI*, pages 161–166, 2008.

[48] Wouter Frencken, Koen Lemmink, Nico Delleman, and Chris Visscher. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science*, 11(4):215–223, 2011.

[49] Stefan Fritsch and Frauke Guenther. *neuralnet: Training of Neural Networks*, 2016. R package version 1.33.

[50] José Gama, Micael Couceiro, Gonçalo Dias, and Vasco Vaz. Small-world networks in professional football: conceptual model and data. *European Journal of Human Movement*, 35:85–113, 2015.

[51] Lise Getoor. Link mining: a new data mining challenge. *ACM SIGKDD Explorations Newsletter*, 5(1):84–89, 2003.

[52] Thomas Grund. Network structure and team performance: The case of english premier league soccer teams. *Social Networks*, 34(4):682–690, 2012.

[53] Joachim Gudmundsson and Michael Horton. Spatio-temporal analysis of team sports. *ACM Computing Survey*, 50(2):22:1–22:34, April 2017.

[54] Silvio Jamil Ferzoli Guimarães, Michel Couprie, Arnaldo de Albuquerque Araújo, and Neucimar Jerônimo Leite. Video segmentation based on 2d image analysis. *Pattern Recognition Letters*, 24(7):947–957, 2003.

[55] Laszlo Gyarmati, Haewoon Kwak, and Pablo Rodriguez. Searching for a unique style in soccer. *arXiv preprint arXiv:1409.0308*, 2014.

[56] Adam Hewitt, Grace Greenham, and Kevin Norton. Game style in soccer: what is it and can we quantify it? *International Journal of Performance Analysis in Sport*, 16(1):355–372, 2016.

[57] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.

[58] Michael Horton, Joachim Gudmundsson, Sanjay Chawla, and Joël Estephan. Automated classification of passing in football. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 319–330. Springer, 2015.

[59] Mike Hughes, Stephen Mark Cooper, and Alan Nevill. Analysis of notation data: reliability. *Notational Analysis of Sport*, 2:189–204, 2004.

[60] Christophe Hurter, Ozan Ersoy, Sara Irina Fabrikant, Tijmen R. Klein, and Alexandru Telea. Bundled visualization of dynamicgraph and trail data. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1141–1157, 2014.

[61] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Gorg, Jorn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. *Lecture Notes in Computer Science*, 4950:154–176, 2008.

[62] Vassilis Kostakos. Temporal graphs. *Physica A: Statistical Mechanics and its Applications*, 388(6):1007–1023, 2009.

[63] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and evolution of online social networks. In *Link mining: Models, Algorithms, and Applications*, pages 337–357. Springer, 2010.

[64] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.

[65] Paul Larkin, Christopher Mesagno, Jason Berry, Michael Spittle, and Jack Harvey. Video-based training to improve perceptual-cognitive decision-making performance of australian football umpires. *Journal of Sports Sciences*, pages 1–8, 2017.

[66] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87(19):198701, 2001.

[67] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM, 2005.

[68] Heiko Lex, Kai Essig, Andreas Knoblauch, and Thomas Schack. Cognitive representations and cognitive processing of team-specific tactics in soccer. *PLoS ONE*, 10(2):e0118219, 2015.

[69] Patrick Lucey, Alina Bialkowski, Mathew Monfort, Peter Carr, and Iain Matthews. quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th Annual MIT Sloan Sports Analytics Conference*, pages 1–9, 2014.

[70] Pedro Malta and Bruno Travassos. Caraterização da transição defesa-ataque de uma equipa de futebol/characterization of the defense-attack transition of a soccer team. *Motricidade*, 10(1):27, 2014.

[71] Kenneth O. McGraw and Seok P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30, 1996.

[72] Daniel Memmert, Koen Lemmink, and Jaime Sampaio. Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 47(1):1–10, 2017.

[73] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2017. R package version 1.6-8.

[74] Stephen Mitchell. Improving invasion game performance. *Journal of Physical Education, Recreation & Dance*, 67(2):30–33, 1996.

[75] Felipe Arruda Moura. *Análise quantitativa da distribuição de jogadores de futebol em campo durante jogos oficiais.* PhD thesis, Universidade Estadual de Campinas, Campinas - SP, 2011.

[76] Felipe Arruda Moura, Luiz Eduardo Barreto Martins, Ricardo De Oliveira Anido, Ricardo Machado Leite De Barros, and Sergio Augusto Cunha. Quantitative analysis of brazilian football players' organisation on the pitch. *Sports Biomechanics*, 11(1):85–96, 2012.

[77] Felipe Arruda Moura, Luiz Eduardo Barreto Martins, Ricardo de Oliveira Anido, Paulo Régis C Ruffino, Ricardo Machado Leite Barros, and Sergio Augusto Cunha. A spectral analysis of team dynamics and tactics in brazilian football. *Journal of Sports Sciences*, 31(14):1568–1577, 2013.

[78] Felipe Arruda Moura, Luiz Eduardo Barreto Martins, and Sergio Augusto Cunha. Analysis of football game-related statistics using multivariate techniques. *Journal of Sports sciences*, 32(20):1881–1887, 2014.

[79] Felipe Arruda Moura, Richard Van Emmerik, Juliana Exel Santana, Luiz Eduardo Barreto Martins, Ricardo Machado Leite de Barros, and Sergio Augusto Cunha. Coordination analysis of players' distribution in football using cross-correlation and vector coding techniques. *Journal of Sports Sciences*, 34(24):2224–2232, 2016.

[80] Jovan Nahman and Dragoslav Perić. Path-set based optimal planning of new urban distribution networks. *International Journal of Electrical Power & Energy Systems*, 85:42–49, 2017.

[81] Mark E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[82] Chong-Wah Ngo, Ting-Chuen Pong, and Roland Chin. Detection of gradual transitions through temporal slice analysis. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999*, volume 1, pages 36–41. IEEE, 1999.

[83] Min-hwan Oh, Suraj Keshri, and Garud Iyengar. Graphical model for baskeball match simulation. In *Proceddings of the 2015 MIT Sloan Sports Analytics Conference, Boston, MA, USA*, volume 2728, 2015.

[84] Judy Oslin and Stephen Mitchell. Game-centered approaches to teaching physical education. *The handbook of Physical education*, pages 627–651, 2006.

[85] Luca Pappalardo and Paolo Cintia. Quantifying the relation between performance and success in soccer. *arXiv preprint arXiv:1705.00885*, 2017.

[86] Pedro Passos, Keith Davids, Duarte Araujo, N. Paz, J. Minguéns, and Jose Mendes. 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, 2011.

[87] Javier López Peña and Hugo Touchette. A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*, 2012.

[88] Nicola Perra and Santo Fortunato. Spectral centrality measures in complex networks. *Physical Review E*, 78(3):036107, 2008.

[89] Allan Pinto, William Robson Schwartz, Helio Pedrini, and Anderson de Rezende Rocha. Using visual rhythms for detecting video-based facial spoof attacks. *IEEE Transactions on Information Forensics and Security*, 10(5):1025–1038, 2015.

[90] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[91] Robert Rein and Daniel Memmert. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1):1410, 2016.

[92] Alfréd Rényi and Paul Erdős. On random graphs. *Publicationes Mathematicae*, 6(290-297):5, 1959.

[93] João Ribeiro, Pedro Silva, Ricardo Duarte, Keith Davids, and Júlio Garganta. Team sports performance analysed through the lens of social network theory: implications for research and practice. *Sports Medicine*, pages 1–8, 2017.

[94] Daniele Cristina Uchôa Maia Rodrigues, Felipe Arruda Moura, Sergio Augusto Cunha, and Ricardo da Silva Torres. Utilizando redes complexas para análise de jogos de futebol. In *XVII Congresso Brasileiro de Biomecânica*, pages 632–633, 2017.

[95] Daniele Cristina Uchôa Maia Rodrigues, Felipe Arruda Moura, Sergio Augusto Cunha, and Ricardo da Silva Torres. Visualizing temporal graphs using visual rhythms - a case study in soccer match analysis. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: IVAPP, (VISIGRAPP 2017)*, pages 96–107. INSTICC, SciTePress, 2017.

[96] Francisco Aparecido Rodrigues. *Caracterização, classificação e análise de redes complexas*. PhD thesis, 2007.

[97] Nicola Santoro, Walter Quattrociocchi, Paola Flocchini, Arnaud Casteigts, and Frederic Amblard. Time-varying graphs and social network analysis: Temporal indicators and metrics. *arXiv preprint arXiv:1102.0629*, 2011.

[98] Hugo Sarmento, Rui Marcelino, Maria Teresa Anguera, Jorge Campaniço, Nuno Matos, and José Carlos Leitão. Match analysis in football: a systematic review. *Journal of Sports Sciences*, 32(20):1831–1843, 2014.

[99] Malte Siegle and Martin Lames. Modeling soccer by means of relative phase. *Journal of Systems Science and Complexity*, 26(1):14–20, 2013.

[100] Bruno Augusto Nassif Travençolo and Luciano da Fontoura Costa. Accessibility in complex networks. *Physics Letters A*, 373(1):89–95, 2008.

[101] Bruno Augusto Nassif Travençolo, Matheus Palhares Viana, and Luciano da Fontoura Costa. Border detection in complex networks. *New Journal of Physics*, 11(6):063019, 2009.

[102] Jan Van Haaren, Vladimir Dzyuba, Siebe Hannosset, and Jesse Davis. Automatically discovering offensive patterns in soccer match data. In *International Symposium on Intelligent Data Analysis*, pages 286–297. Springer, 2015.

[103] Jan Van Haaren, Siebe Hannosset, and Jesse Davis. Strategy discovery in professional soccer match data. In *Proceedings of the KDD-16 Workshop on Large-Scale Sports Analytics*, 2016.

[104] Jarke Jan Van Wijk. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 79–86. IEEE, 2005.

[105] Corinna Vehlow, Michael Burch, Hansjorg Schmauder, and Daniel Weiskopf. Radial layered matrix visualization of dynamic graphs. In *Information Visualisation (IV), 2013 17th International Conference*, pages 51–58. IEEE, 2013.

[106] William Venables and Brian David Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.

[107] Luís Vilar, Duarte Araújo, Keith Davids, and Yaneer Bar-Yam. Science of winning soccer: Emergent pattern-forming dynamics in association football. *Journal of Systems Science and Complexity*, 26(1):73–84, 2013.

[108] Duncan James Watts and Steven Henry Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[109] Douglas Brent West. *Introduction to Graph Theory*. Prentice Hall Upper Saddle River, 2001.

[110] Carl Woods, Annette Raynor, Lyndell Bruce, Zane McDonald, and Sam Robertson. The application of a multi-dimensional assessment approach to talent identification in australian football. *Journal of Sports Sciences*, 34(14):1340–1345, 2016.

[111] Feng Zhao and Anthony Tung. Large scale cohesive subgraphs discovery for social network visual analysis. In *Proceedings of the VLDB Endowment*, volume 6, pages 85–96. VLDB Endowment, 2012.