# SPATIO-TEMPORAL COHESIVE NETWORKS FOR EVALUATING TEAM BEHAVIOR IN SOCCER

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMRE KÜLAH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

JULY 2019

Approval of the thesis:

**SPATIO-TEMPORAL COHESIVE NETWORKS FOR EVALUATING TEAM BEHAVIOR IN SOCCER**

submitted by **EMRE KÜLAH** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** _______________

Assist. Prof. Dr. Hande Alemdar
Supervisor, **Computer Engineering, METU** _______________

**Examining Committee Members:**

Assist. Prof. Dr. Elif Sürer
Graduate School of Informatics, METU _______________

Assist. Prof. Dr. Hande Alemdar
Computer Engineering, METU _______________

Assist. Prof. Dr. Bahri Atay Özgövde
Computer Engineering, Galatasaray University _______________

Date:

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Surname: Emre Külah

Signature :

# ABSTRACT

## SPATIO-TEMPORAL COHESIVE NETWORKS FOR EVALUATING TEAM BEHAVIOR IN SOCCER

Külah, Emre
M.S., Department of Computer Engineering
Supervisor: Assist. Prof. Dr. Hande Alemdar

July 2019, 71 pages

In this thesis, we analyze positional organizations of soccer teams during game transitions which end with an important result and individual sprinting performances of soccer players. Social networks and distance matrices of teams are used to obtain organizations of teams. Spatial features of pitch such as pitch value and pass probability value are used to evaluate sprint performances. Social networks that we call cohesion matrices are used as weights in both attacking and defending transitions. The norm of the weighted distance matrices forms team spread values. Cohesion matrices show player to player interactions and connections between clusters of teams on the pitch. The team spread values are used to characterize the behaviors of teams in a transition. The average team spread values show that top teams are more expansed while attacking and more tighter while defending. Moreover, the average team spread values confirms that teams are wider while attacking except when a transition ends with losing the possession of the ball. These results characterize organizations of teams in the Turkish Super League and the effects of individual players on those organizations. Sprint analysis results show that full-back and winger players have higher sprint value averages while midfielders have less. Also, teams that are focused on having the possession of the ball have less average sprint value than teams playing in counter-attack style.

Keywords: Soccer, Spatio-temporal data mining, Quantitative analysis, Sprint appraisal

# ÖZ

## FUTBOLDAKİ OYUNCULARIN MEKANSAL-GEÇİCİ KOHESİV AĞLARINI KULLANARAK TAKIM DAVRANIŞLARINI DEĞERLENDİRME

Külah, Emre
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi. Hande Alemdar

Temmuz 2019 , 71 sayfa

Bu tezde, takımların önemli bir olayla sonuçlanan oyun geçişleri sırasındaki konumsal organizasyonları ve oyuncuların bireysel sprint performansları analiz edilmiştir. Takımların sosyal ağ ve uzaklık matrisleri kullanılarak takım organizasyonları elde edilmiştir. Ayrıca saha değeri ve pas ihtimal değeri gibi, oyun sahasının mekansal özellikleri kullanılarak sprint performansları değerlendirilmiştir. Etkileşim matrisleri olarak adlandırdığımız sosyal ağlar hem atak hem de defans geçişleri sırasında ağırlık olarak kullanılmıştır. Ağırlıklı uzaklık matrislerinin normları takım dağılım değerlerini oluşturur. Takım dağılım değerleri, takımların geçişler sırasındaki davranışlarını karakterize etmek için kullanılır. Ortalama takım dağılım değerleri, takımların atak sırasında daha genişken, defans yaparken daha kompakt olduğunu gösterir. Ayrıca ataklar bir eylem ile sonuçlandığında takımların top kaybı yaptığı anlara göre daha yaygın olduğunu görürüz. Sprint analiz sonuçları, bek ve kanat oyuncuları yüksek sprint değer ortalamasına sahipken, orta saha oyuncularının sprint değer ortalamalarının düşük olduğunu ortaya koyar. Ayrıca topa sahip olmaya odaklı takımlar, kontra-atak

futbolu oynayan takımlara göre daha düşük sprint değer ortalamalarına sahiplerdir.

Anahtar Kelimeler: Futbol, Mekan-zamansal veri madenciliği, Sayısal analiz, Hızlı koşu değerlemesi

To my love Şeyma and to my dear family

## ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude and highest respect to my supervisor Assist. Prof. Dr. Hande Alemdar for her support and friendship. Without her guidance and contributions, it would not be possible to complete this journey successfully.

I wish to express my sincere gratitude to my committee members Assist. Prof. Dr. Elif Surer, Assist. Prof. Dr. Bahri Atay Özgövde for their helpful comments on this study.

I would also like to thank Mr. Serdar Alemdar and SentioSports team for their assistance in providing data, which is one of the most important parts of my thesis.

I owe a debt of gratitude to my mother, my father, my sister and my dear for their endless encouragement and love. They always be with me all the time. I also give my special thanks to Aslan and my nephew-to-be for their mental supports. I would like to thank my cousin Gürkan for his mental and professional support. In addition, my friends Görkem, Semih, Yusuf, Caner, Emre, Fırat, Gökhan, Kerem and Tevfik deserve a big thanks with their supports during my whole life.

**TABLE OF CONTENTS**

ABSTRACT . . . . . . . . . . v

ÖZ . . . . . . . . . . vii

ACKNOWLEDGMENTS . . . . . . . . . . x

TABLE OF CONTENTS . . . . . . . . . . xi

LIST OF TABLES . . . . . . . . . . xiv

LIST OF FIGURES . . . . . . . . . . xv

LIST OF ABBREVIATIONS . . . . . . . . . . xviii

CHAPTERS

1 INTRODUCTION . . . . . . . . . . 1

1.1 Motivation and Problem Definition . . . . . . . . . . 1

1.2 Contributions and Novelties . . . . . . . . . . 3

1.3 Data Collection . . . . . . . . . . 3

1.4 The Outline of the Thesis . . . . . . . . . . 4

2 RELATED WORKS . . . . . . . . . . 7

2.1 Pitch Value . . . . . . . . . . 7

2.2 Pass Probability . . . . . . . . . . 8

2.3 Team Spread . . . . . . . . . . 9

2.4 Social Networks . . . . . . . . . . 10

2.5 Sprint Value . . . 11

3 PITCH VALUE AND PASS INTERCEPTION VALUE . . . 13

3.1 Pitch Value . . . 13

3.1.1 Player Influence Area . . . 14

3.1.2 Pitch Value Calculation . . . 17

3.1.2.1 Effects of Defensive Players . . . 18

3.1.2.2 Effects of Ball and Goal . . . 19

3.1.2.3 Effect of Location . . . 19

3.2 Pass Interception Value . . . 21

3.2.1 Pass Interception Area . . . 22

4 TEAM SPREAD . . . 27

4.1 Constructing Cohesion Matrices . . . 28

4.1.1 The Closeness Matrix . . . 29

4.1.2 The Common Marking Count Matrix . . . 30

4.1.3 The Pass Count Matrix . . . 30

4.2 The Team Spread . . . 31

4.2.1 $\alpha - \beta$ Optimization . . . 32

4.3 Temporal Analysis . . . 32

4.4 Chapter Experiments . . . 34

4.4.1 Cohesion Matrices . . . 34

4.4.2 $\alpha - \beta$ Optimization . . . 37

4.4.3 Team Analysis . . . 38

4.5 Chapter Discussions . . . 46

4.5.1 Cohesion Matrices . . . . . . . . . . . . . . . . . . . . . . . 46

4.5.2 Team Analysis . . . . . . . . . . . . . . . . . . . . . . . . 48

5 QUANTIFYING SPRINT VALUE . . . . . . . . . . . . . . . . . . . . . 51

6 RESULTS AND DISCUSSION . . . . . . . . . . . . . . . . . . . . . . 57

6.1 Sprint Quantification . . . . . . . . . . . . . . . . . . . . . . . 57

6.2 League Level Analysis . . . . . . . . . . . . . . . . . . . . . . . 61

7 CONCLUSIONS AND FUTURE WORKS . . . . . . . . . . . . . . . . . . 65

REFERENCES . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67

# LIST OF TABLES

TABLES

Table 4.1 Statistical information of games which BJK played against BSK and KON. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

Table 4.2 Goal scores and average team spread values in the events which committed by the team specified in the columns. The teams are ordered according to their rank in the league. . . . . . . . . . . . . . . . . . . . . . 39

Table 4.3 Conceded goal counts and average team spread values in the events which suffered by the team specified in the columns. The teams are ordered according to their rank in the league. . . . . . . . . . . . . . . . . . 40

Table 4.4 Count of teams in the league (18 teams) who has smaller team spread in the condition specified. . . . . . . . . . . . . . . . . . . . . . . 41

Table 6.1 Top 20 players have most valuable sprint averages. . . . . . . . . . 62

Table 6.2 Bottom 20 players have worst valuable sprint averages. . . . . . . . 63

Table 6.3 Average sprint values of teams in the Turkish Super League 2018-2019 season. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

# LIST OF FIGURES

FIGURES

Figure 1.1 Rectangular Coordinate System used by SentioScope® . . . . . . 4

Figure 3.1 Comparison of two PLIAs with different speeds. . . . . . . . . . 15

Figure 3.2 Value of the whole pitch after applying the effect of the occupation of defensive players. . . . . . . . . . . . . . . . . . . . . . . . . . 18

Figure 3.3 Figure 3.3a shows pitch value when the effect of the ball added on the effect of the defensive players. The effect of the ball position added on the pitch value after as well and showed in the Figure 3.3b. . . 20

Figure 3.4 Final value of the whole pitch after applying the effect of the location. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

Figure 3.5 Comparison of two interception areas with different distance to the ball. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

Figure 3.6 Comparison of two PIAs with different speeds. . . . . . . . . . . 25

Figure 4.1 Conditions for closeness and common marking matrices. **(a)** $P1$ and $P2$ are close teammates **(b)** Teammates $P1$ and $P2$ are marking the same 2 opponents $P3$ and $P4$. . . . . . . . . . . . . . . . . . . . . . 29

Figure 4.2 Convolutional Neural Network structure to optimize $\alpha$ - $\beta$ factors. Two different weighted distance matrices are fed to the system. They are convolved and merged. Then the input becomes a vector with fully connected layer. . . . . . . . . . . . . . . . . . . . . . . . . . 33

Figure 4.3 Closeness Networks of Besiktas in games against Basaksehir **(a)** and Konyaspor **(b)**. Players are ordered according to formation of the team. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

Figure 4.4 Pass Networks of Besiktas in games against Basaksehir **(a)** and Konyaspor **(b)**. Players are ordered according to formation of the team. 35

Figure 4.5 Common Marking Networks of Besiktas in games against Basaksehir **(a)** and Konyaspor **(b)**. Players are ordered according to formation of the team. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

Figure 4.6 Two snapshots of formation of Besiktas from a game while has the possession of the ball and attacking to the right side of the pitch. . . 36

Figure 4.7 Offensive cohesive network of Besiktas **(a)** and difference matrix **(b)** showing $\mathbb{D} - \mathbb{D} \circ \mathbb{W}$. $\mathbb{D}$ is player-to-player distance matrix and $\mathbb{W}$ is offensive cohesive network. . . . . . . . . . . . . . . . . . . . . . 37

Figure 4.8 Changes in kernel values showing influence of pass and closeness networks over offensive cohesion matrix. . . . . . . . . . . . . . 38

Figure 4.9 Average team spread values of teams during events which finish with a **shot on target**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams. . . . . . . . . . . . . . . . . . 42

Figure 4.10 Average team spread values of teams during events which finish with **losing the ball**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams. . . . . . . . . . . . . . . . . . 43

Figure 4.11 Average team spread values of teams during events which finish with a **suffered shot on target**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams. . . . . . . . . . . . . . 44

Figure 4.12 Average team spread values of teams during events which finish with **gaining possession of the ball**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams. . . . . . . . . 45

Figure 4.13 Average team spread values of Turkish Super League for committed and suffered events. . . . . . . . . . . . . . . . . . . . . . . . . 46

Figure 5.1 Virtual organization of the attacking team around a circle with radius of 10 m. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

Figure 5.2 Pitch value distribution of the pitch at a game moment. . . . . . 54

Figure 5.3 Pass probability value distribution of the pitch at a game moment. 54

Figure 5.4 Final sprint value distribution of the pitch, including team spread effect, at a game moment. . . . . . . . . . . . . . . . . . . . . . . . . 55

Figure 6.1 Two insignificant sprints done by Adem Ljajic while attacking. . 59

Figure 6.2 Two valuable sprints done by Adem Ljajic while attacking. . . . 60

## LIST OF ABBREVIATIONS

| | |
|---|---|
| CNN | Convolutional Neural Network |
| FN | Frobenius Norm |
| PIA | Pass Interception Area |
| PLIA | Player Influence Area |
| TSL | Turkish Super League |

# CHAPTER 1

# INTRODUCTION

## 1.1 Motivation and Problem Definition

Soccer is the most popular sport in the world [1, 2]. This popularity makes it the most analyzed team sport as well. Soccer analysts work to examine players' performance in the match, both individually and as a team member [3, 4, 5, 6, 7, 8, 9, 10]. They try to find out beneficial indicators about both their teams and opposites. They analyze on-ball events, defending patterns and more. These analyzes are in need of spatio-temporal information players during games [11, 12, 13, 14]. Technology is being involved more and more in soccer analysis in parallel with these purposes.

Image processing [15, 16] and data mining [17] fields are two main technological benefits which we used in soccer analysis. Player-tracking systems have become imperative parts of soccer games. These tools are used to digitize players projections on the pitch during games. They collect that data to analyze individual player and team performances, the rights and wrongs, and the opposite teams' weaknesses. Although it is possible to do a visual inspection, technology helps us to find out more important indicators that are not easy to see with the naked eye.

There exists a lot of studies that analyze player and team performances. One of the main subjects which are deeply examined is pass patterns [18, 7]. Passes are crucial in soccer. Almost all top teams focused on keeping the possession of the ball. The easiest way to do this is passing. Pep Guardiola the manager of Manchester City once said "If there is not a sequence of 15 previous passes, a good transition between attack and defense is impossible. Impossible." [19]. To score a goal at the end of those transitions, teams try to find a gap between defenders by passing to each other. These sequences must end with a try which is committed by a player in a good location on the pitch [20]. This point is another key field of soccer for analyzing. Attacking players look for a space between defensive players all the time during games. There are different tactics to create and occupy those spaces. To utilize the spaces, attacking players sprints towards those spaces. For this reason, it is important to do a valuable sprint for soccer players during attacking moments.

This sprints towards those empty spaces generally remain inconclusive sometimes because of unsuccessful passes or good interceptions or sometimes because it is not possible to pass the ball to that point. When this situation happened, if the target player finds himself far far away from the rest of the team because of a vain sprint, helping the rest of the team for defending becomes impossible for this player. Besides, mean sprint time in a game is highly correlated with the distance covered by that player which means insignificant sprints cause less movement for that player [21]. Quantifying the quality of a sprint for a moment can help players to improve estimating a location while sprinting. While doing tactical training, analyzers can use these sprint values to improve sprint skills of attacking players. Furthermore, it is possible to use these values for defensive players. defensive players can observe possible high-value sprints for opponents.

A sprint is as valuable as the value of the point it targets. Value of a location on the pitch depends on some quantifiable conditions. Its location value is the first component of this quantity. Each point of the pitch has different value depends on the position of the ball and players [22]. This value is not enough for quantifying the sprint value. A location on the pitch may have a high value, but it may not be available to pass the ball to that location. For this reason, while quantifying the value of a sprint, we have to consider if it is possible to pass the ball to that location. Another condition we have to take into account is the team dispersion. A valuable location on the pitch may cause a player to be too far away from the rest of the team. If the pass to that location intercepted somehow, this player becomes nothing for defending tactics [23].

Taking into account these situations, it is possible to draw a heat map showing which points a team player can make more valuable sprints in the pitch. These heat maps help both attacking players to improve their sprinting skills and defensive players to improve their positioning and chasing skills.

## 1.2 Contributions and Novelties

First, we developed a physics-based model that exposes players' pass interception areas as distribution. Using this model, the pass probability values on the pitch can be examined instantly.

Weighted team spread model includes many novelties. We have created cohesion matrices that enable us to understand the organization of the teams and visualize them. The weighted team spread value created using these matrices allows us to see the teams' dispersion on the pitch, taking into account the interaction of the players.

Spatio-temporal sprint analysis enables us to quantify players' sprint performance, allowing us to see the sprint characteristics of both players individually and teams in general.

Sprint analysis and weighted team spread analysis can be used instantly as a useful tool for soccer player development and tactical deciding.

## 1.3 Data Collection

The state-of-art real-time two-camera player tracking system SentioScope® [16], developed by Sentio®, collects data from each Turkish Super League (TSL) game into a database. This database contains identifiers for games and players as well.

Using this database, we have created a separate data set for each game to analyze. For each second, position data of players of both teams and the ball in a rectangular coordinate system are saved in this data set $\mathcal{D}_{\mathcal{G}}$. $G$ represents a game in the whole league period. We identify the home team as $H$ and away team as $A$. The ball is labeled as $B$. The data set is constructed as follows:

$$
\begin{aligned}
\mathcal{D}_{\mathcal{G}}(t) &= \{t, X_{H_i}(t), Y_{H_i}(t) | i = [1, 11]\} \\
&+ \{t, X_{A_i}(t), Y_{A_i}(t) | i = [1, 11]\} \\
&+ \{t, X_B, Y_B\} \qquad (1.1)
\end{aligned}
$$

where $X$ and $Y$ shows coordinates and $X_{H_i}$ means $X$ position of $i^{th}$ player in the home team. $t$ is a moment representing a second of a game that lasts $T$ seconds.

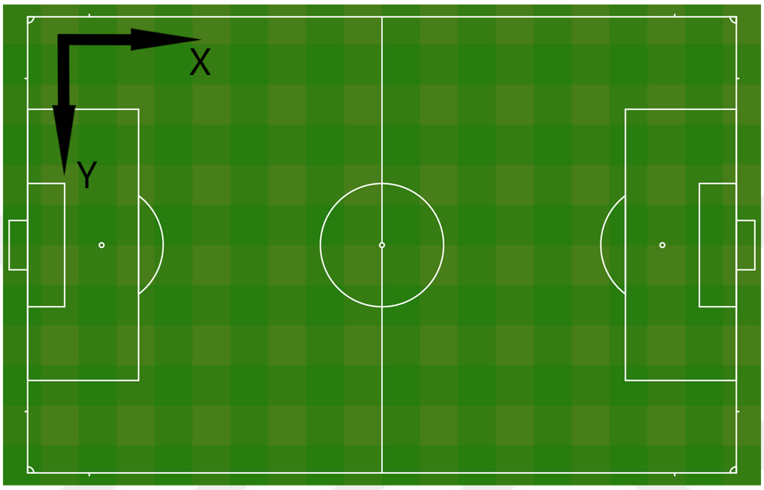



Figure 1.1: Rectangular Coordinate System used by SentioScope®

$$
\mathcal{D}_{\mathcal{G}} = \{\mathcal{D}_{\mathcal{G}}(t) | t = 0, ...., T\} \qquad (1.2)
$$

## 1.4 The Outline of the Thesis

2 chapters following Chapter 2 detail methodology of the thesis. Chapter 3 presents what pitch value is and how we calculate pitch value according to the position of the players and the position of the ball. In the second part of this chapter, the quantification of pass interception value is described which uses similar methods that are used in the first part of the chapter. Pitch value and pass interception value are two of the main parts of the quantification of sprint value. Chapter 4 explains what team spread is and how this value affects sprint value quantification. In Chapter 5, we created sprint value model using parts mentioned in Chapters 3, 4. Results are demonstrated

in the Chapter 6. In this chapter, various sprint value samples are visualized from moments of games in the form of heat maps on the pitch. The thesis is summarized in Chapter 7 and some future works are explained.

# CHAPTER 2

# RELATED WORKS

Related works chapter is divided into sections. Each section contains related works related to the main topics in the thesis.

## 2.1 Pitch Value

Fernandez [22] conducted a detailed space creation analysis in professional soccer. He calculates occupation and generation values for the spaces during games. To calculate these values, he created a model called the pitch control model. This model computes influences of defensive players on the pitch during a moment as a multivariate distribution where mean depends on speed vector of the player and covariance depends on both speed vector of the player and distance from the ball. This distribution is used in this study as a component for calculation of pitch value. He calculates space occupation and space generation values for players of F.C. Barcelona. He also visualized those values using heatmaps.

Spearman [24] quantifies the scoring probability of teams during an attacking transition in his study. He calculates next on-ball event probability and control probability for a position on the pitch. This control probability is calculated using a control model called the potential pitch control field. He uses a probability function which calculates the probability that a player can reach a location within some time. This function is used for each player of the team and the potential pitch control field is calculated. He uses this model as a component for the final calculation of the probability of scoring.

Seabra [25] approaches from another angle in his study. He creates a region that is occupied by the defenders and called as space of defensive occupation. He divides this region into zones which are classified according to their spatial features. These zones were examined and evaluated according to the actions that took place in these zones. The success rates of these actions are also calculated according to these zones.

### 2.2 Pass Probability

McHale [26] conducted a study which uses network analysis and pass difficulty to identify key players in teams. The main part of the study is estimating the key passers in teams. They created a model to predict pass probability using a generalized additive mixed model. They defined a couple of covariates. They defined each pass as a Bernoulli trial with the probability of a pass being successful depending on these covariates. They use an inverse function to convert pass probability into pass difficulty to calculate players' pass performances. They listed key players and their pass difficulty values in their results.

Spearman [27] calculates pass probability and pass value in his study. He uses Bernoulli method as well. He defines three values such as ball trajectory, time to intercept and time to control to use as parameters. He forms Bernoulli distribution and probability mass function for this distribution. He defines a likelihood and tries to find a set of parameters that maximizes this likelihood. He also defines pass value as a combination of pass probability, the benefit of pass to the attacking team and the benefit of unsuccessful pass to the opposite team.

Rein [28] analyzes pass effectiveness by evaluating changes in pitch control and outplayed defending player counts. They use the Voronoi-diagram approach to calculate pitch control values. They compare pass types towards attacking sides such as defense to midfield, midfield to attack in terms of pitch control changes and outplayed defender counts.

### 2.3 Team Spread

The team spread calculation has recently become an important metric used with team length or coverage area. Since each player's distance to each other is taken into account, it is very useful in explaining the team's current organization.

Bartlett [29] has a study to analyze team dispersion and team centroid changes during different events such as goals, shots on goal and tackles. They calculated several dispersion metrics such as surface area, stretch index, Frobenius norm, and team length in x and y directions during events listed above. In those events, they tried to find correlations between events and dispersion and centroid changes.

Moura [3] conducted an offensive and defensive analysis of 8 teams and 223 players as a preliminary and comprehensive study on this issue. In this work, Moura creates a distance-between-player matrix using 10 players without taking the goalkeepers into account. Then the Frobenius Norm of this matrix is calculated and the team spread value is formed. Moura's work shows how team spread increases when a team attacks and how they decrease when defend. Team spread values are not analyzed only in attacking and defending moments, but also in tackling and shot on target moments. In Moura's discussion section, it is mentioned that the coverage area might behave differently than the team spread. For example, a player who is too far away from the rest of the team will increase the coverage area very much, and the two values will vary since the effect will be less on the spread. This actually reveals the difference between the team spread and other geometric values.

Moura [11] in his another study, used team spread values to investigate the coordination between teams' dynamics. They used cross-correlation and vector coding techniques to find coherencies over a dataset which contains 257 players' trajectory data during 10 matches. Team spread values were calculated as functions of time and cross-correlation applied to these functions. Using the vector coding technique, they tried to identify coordination patterns during offensive sequences which ended with defensive tackles or shot on target. 4 different patterns are identified as anti-phase, in-phase, attacking team phase and defending team phase. Their results showed that although in-phase patterns are more commonly seen, there are anti-phase patterns during offensive sequences ending in a shot on target. Attacking teams are trying to behave conversely to increase shots on goal at the beginning of the attacking play. This study shows team spread value can be used in researches to investigate teams' dynamics and tactics.

### 2.4 Social Networks

Teams as a social organizations contains cooperative relations of individuals [5]. Social networks are used for the analysis of these structures [10]. This structure, which helps to find links within a community, can also work to find links between teammates [5]. With the tracking systems, the usage of these networks has increased considerably as a result of the digital tracking of player movements and passes. These networks can be created both for small groups of players and for the whole team. Since we created social networks for the whole team in our study, we reviewed studies that analyze the whole team.

Clemente [10] made a network analysis of a team's 5 matches from the Portuguese league. In this research, they formed a data set consisting of attack scenarios that started with the gaining the ball possession in 5 games and formed the pass networks in this set.

Grund [4] created pass networks for data that included 760 matches from the English Premier League. In this work, he used the most active 8 players of the team while building networks. This means goalkeepers and usually, strikers are excluded from these processes. Grund labeled passes made by a player as outgoing for him, while labeled passes received by him as incoming.

Goncalves [18] calculated the closeness and betweenness scores in addition to the pass networks. Closeness can be regarded as how much a player close to his teammate who has the ball. This score indicates the possibility of the player taking a pass. Betweenness is used as a score indicating the possibility that a player may be a bridge between passes. This study merges the positions of the whole team with the pass networks.

Wäsche [7] refers to two types of networks, which are called socio-centric and ego-centric. In Socio-centric networks, the connection between nodes is called cohesion. This has a similar logic to the cohesion matrices we have created. Cohesion can be considered as the strength of the link between the two players. The pass, closeness and marking networks we create are examples of these socio-centric networks.

### 2.5 Sprint Value

The sprint issue is generally addressed from a physical point of view. There are many studies on how training affects sprint speed, repetitive sprint recovery time, and sprint distance [30, 31, 32, 33]. Although these studies do not provide information about the valuation of the sprint, they show the impact of the sprint on the player during the match, allowing us to understand how bad the insignificant sprint is for the player. Spatio-temporal analysis of sprints in soccer is an uncharted area.

# CHAPTER 3

# PITCH VALUE AND PASS INTERCEPTION VALUE

This chapter contains pitch value and pass interception value calculations which are two main components of the sprint value quantification model.

## 3.1 Pitch Value

In soccer, while a player of a team that has the possession of the ball does not hold the ball, this player always seeks a location on the pitch which gives him an advantage. This advantage varies according to the duty of that player and position of the ball. If the player is an attacking player and the ball is in the second half of the pitch, then the player starts to seek a position to score a goal [34]. If this player plays in the middle of the field and team has just been trying to start a transition, then the player seeks a space between opponent players to take the ball [35]. For whatever reason, defensive players always try to occupy these locations. It is possible to say that spaces on the pitch which is occupied by defensive players more valuable. While calculating the value of a location on the pitch, all defensive players take effect according to their distance.

In addition to the value of occupation committed by opponents, the location itself has value as well. This value comes from the aim of the game. The aim of the game can be defined as controlling the ball and kicking a ball into the back of a net of an opponent [36]. So there are two important matters while calculating the value of a location. The first one is the distance from the ball and the second one is the distance from the goal. These two issues should be considered while calculating the value of a location.

### 3.1.1 Player Influence Area

Soccer players always are on the move on the pitch during the game, sometimes they walk slowly and sometimes they sprint. This speed changes the area in which they are effective. A study conducted by Javier Fernandez calls this area as player influence area (PLIA) [22]. If a player has zero speed, then the limits he can reach around him are the same. This means that the PLIA of that player can be shown as a perfect circle. However, if he has speed, then this PLIA becomes stretched in the direction of the speed vector of the player. Also, the center of the PLIA relocates toward speed vector as well. For a point $p$ on pitch, while calculating the PLIA of player $i$, we use standard multivariate normal distribution with mean $\mu_i(t)$ and covariance matrix $\Sigma_i(t)$. The velocity of the player is $\vec{s}$ and the angle is $\Theta$. Equation 3.1 shows how this influence is calculated. We use the same notation which Fernandez used in his study while defining PLIA.

$$f_i(p,t) = \frac{1}{\sqrt{(2\pi)^2 detCOV_i(t)}} exp(-\frac{1}{2}(p-\mu_i(\vec{s}_i(t)))^T COV_i(t)^{-1}(p-\mu_i(t))) \quad (3.1)$$

We calculate influence likelihood $I$ of a player on a point $p$ at time $t$ as Equation 3.2. This equation forms a degree of influence within a range $[0,1]$ for a location $p$ on the pitch.

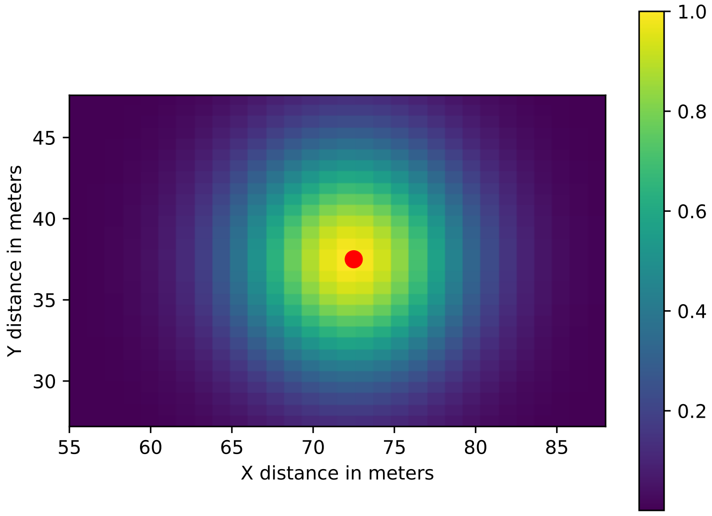
$$I_i(p,t) = \frac{f_i(p,t)}{f_i(p_i(t),t)} \quad (3.2)$$

The covariance matrix can be stated as a function of its characteristic vector and value using singular value decomposition algorithm [37]. $V$ and $L$ are corresponding eigenvector and eigenvalue matrices of covariance matrix respectively.

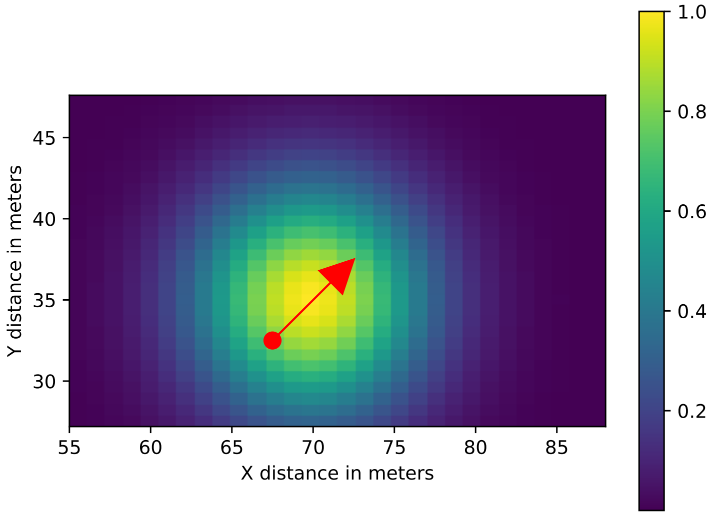
$$\Sigma = VLV^{-1} \quad (3.3)$$

We form rotation matrix $R$ as $R = V$ and scaling matrix $S$ as $S = \sqrt{L}$.

$$\Sigma = RSSR^{-1} \quad (3.4)$$




(a) PLIA of a player stands on the pitch.




(b) PLIA of a player moving in the direction of the red arrow with speed 7.07 m/s.

Figure 3.1: Comparison of two PLIAs with different speeds.

Those rotation matrix and scaling matrix can be expressed as Equation 3.5 and Equation 3.6 where $\Theta$ is the angle and $s_x$, $s_y$ are scaling factors in the corresponding directions.

$$R = \begin{bmatrix} cos(\Theta) & -sin(\Theta) \\ sin(\Theta) & cos(\Theta) \end{bmatrix} \tag{3.5}$$

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \tag{3.6}$$

To limit the radius of the influence area of the player according to the distance to the ball, we use an exponential function. Experts say that the influence area of a player can be within range [4, 10] meters. Equation 3.7 calculates influence area radius for a player where $d$ is the Euclidean distance function, $p_i(t)$ is the position of player $i$ and $p_b(t)$ is the position of the ball.

$$R_i(t) = min\left(\frac{d(p_i(t), p_b(t))^3}{1000} + 4, 10\right) \tag{3.7}$$

To expand the influence area in the speed vector direction and contract in the perpendicular direction we are going to increase the scaling value of x-direction and decrease the value of y-direction as defined in the Equation 3.9. $S_{rat_i}(s)$ is scaling function for the speed of the player according to possible maximum speed.

$$S_{rat_i}(s) = \frac{s^2}{13^2} \tag{3.8}$$

$$S_i(t) = \begin{bmatrix} \frac{R_i(t)+(R_i(t)S_{rat_i}(\vec{s_i}(t)))}{2} & 0 \\ 0 & \frac{R_i(t)-(R_i(t)S_{rat_i}(\vec{s_i}(t)))}{2} \end{bmatrix} \tag{3.9}$$

The rotation and the scaling matrices form covariance matrix as mentioned in the Equation 3.4. The mean value of this distribution is calculated by adding the half of his speed vector to his position to translate the location.

$$\mu_i(t) = p_i(t) + \vec{s_i}(t) \cdot 0.5 \tag{3.10}$$

Two visualizations in Figure 3.1 shows the PLIA of two different players. The first player stands on the pitch with zero speed while the other one has a speed which is shown in the figure as a white arrow. This speed creates the difference between their influence areas. The first one has a perfect circle around him with a radius depending on the distance to the ball. The second one also has a radius depending on the same condition but his speed expands the circle in the direction of his move and contracts in the perpendicular direction. Also, speed vector translates the mean of this area which defined in the Equation 3.10.

### 3.1.2 Pitch Value Calculation

The dominant part of the pitch value calculation comes from the areas occupied by the defensive players. We sum up the influence area values of all defensive players. This calculation gives us a good visualization of the valuable areas on the pitch. Figure 3.2 contains a heat map of the pitch which shows the value of the points on the pitch. Green colored players have the possession of the ball and attack towards the right side.

The influence area values of the defensive players are not enough to decide which location is valuable or not. Attacking team players always try to be close to the defending team's goal and this manner makes a point closer to the goal more valuable intrinsically.

Another important situation that makes a point more valuable is the distance to the ball. The only possible way to score a goal is to have the possession of the ball. For this reason, a point that is close to the ball becomes more valuable.

#### 3.1.2.1 Effects of Defensive Players

Each defensive player d has his PLIA $I_d$ which constructs valuable areas on the pitch. The first form of the value of a point on the pitch is calculated with Equation 3.11. We say pitch value of point $p_{k,l}$ at time t as $V_{k,l}(t)$. The position of the ball at time t is $p_b(t)$. Figure 3.2 shows the pitch value when Equation 3.11 calculated. A moment from a game selected and the pitch is transformed into $75 \times 75$ matrix. For each cell in the matrix, Equation 3.11 calculated and stored in that cell. Green circles show the attacking team which has the possession of the ball while red ones show the defending team. The bigger white circle shows the position of the ball. The attacking team attacks towards the right side of the pitch.

$$V_{k,l}(t) = \sum_{d} I_d(p_b(t), p_{k,l}(t)) \tag{3.11}$$

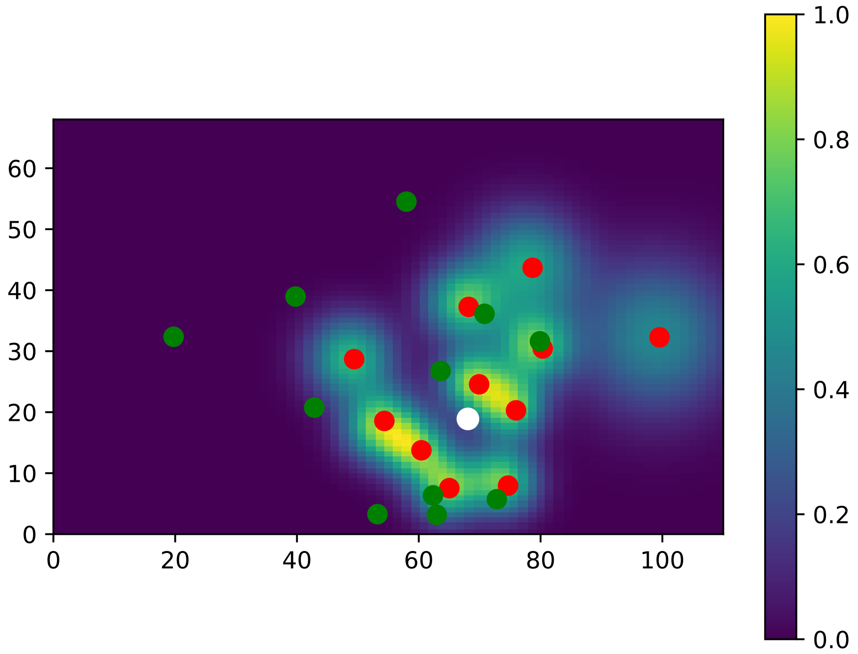



Figure 3.2: Value of the whole pitch after applying the effect of the occupation of defensive players.

#### 3.1.2.2 Effects of Ball and Goal

In addition to value comes from the occupation of defensive players, a location close to ball or goal is more valuable as well [38]. Especially attacking players always try to be close to the goal field. Besides, since the ball is the only target on the pitch to have the possession, it also needs to be in the calculation of the value of a point.

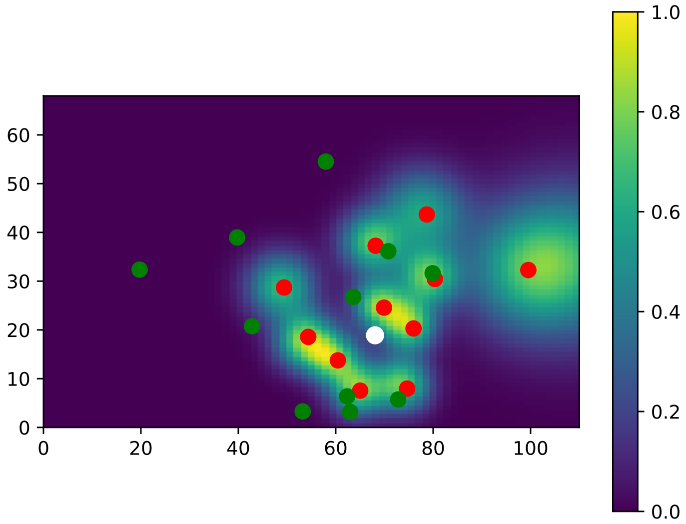
To form smooth effects around goal field and ball, we used a modified version of PLIA calculation. Since the goal field is stationary we can use Equation to increase the value of the area around the goal field. A point $p$ on the pitch gains the value of $I_g(p,t)$ at time $t$ according to the position of the goal field as expressed in the Equation 3.12. Similarly, the ball effect can be calculated using its position as defined in Equation 3.13. Speed values of both goal field and ball are assumed zero.

$$I_g(p,t) = \frac{f_g(p,t)}{f_g(p_g(t),t)} \tag{3.12}$$

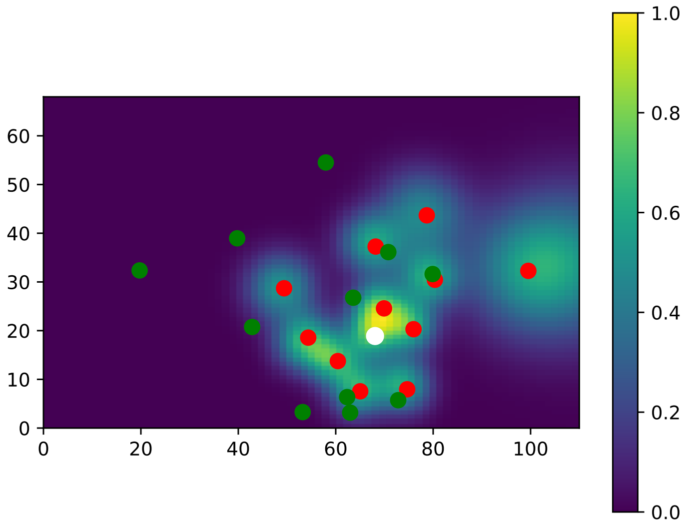
$$I_b(p,t) = \frac{f_b(p,t)}{f_b(p_b(t),t)} \tag{3.13}$$

#### 3.1.2.3 Effect of Location

For a team that has the possession of the ball and attacking to the goal post at the right, the value of the location increases while going to the right side of the pitch. For this reason, we decrease the value according to the horizontal distance to the right end of the pitch. Figure 3.4 shows the final pitch value visualization when the effect of location added.




(a) Pitch value when the effect of the goal field added.




(b) Pitch value when the effect of the ball position added.

Figure 3.3: Figure 3.3a shows pitch value when the effect of the ball added on the effect of the defensive players. The effect of the ball position added on the pitch value after as well and showed in the Figure 3.3b.

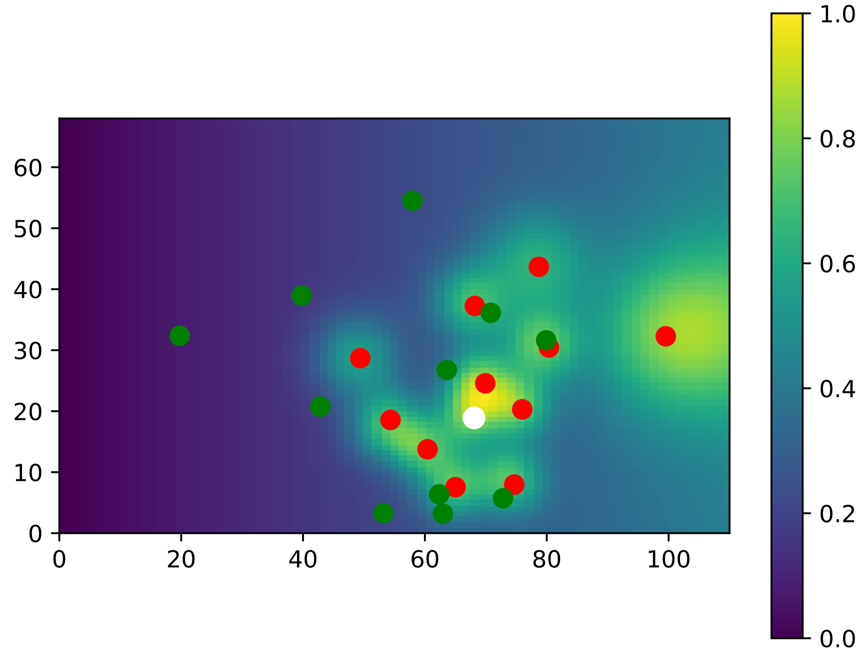



Figure 3.4: Final value of the whole pitch after applying the effect of the location.

### 3.2 Pass Interception Value

Soccer players do not hold the ball most of the time during a game [22]. Instead of that, they move and try to find a good location to increase their possibility of taking a good pass while their team has the possession of the ball. One of the most important considerations while finding a good location to ask for a pass is avoiding the opponents to intercept the pass. For this reason, the players try to avoid being behind the defenders and move to areas where there is no opponent between himself and the player holding the ball.

It is not always possible to avoid pass interception. Defensive players aim to intervene pass sequences of the attacking team. Each defensive player has his area around him which he can gain possession of the ball which we call pass interception area (PIA). Union of these PIAs of the defensive teams creates a collective troublesome area for the attacking team which decreases the value of that area for sprinting.

### 3.2.1 Pass Interception Area

PIA of a defensive player is calculated using a similar way which we used to calculate the PLIA in Chapter 3. The first difference between PLIA and PIA is the mean of the distribution. PLIA has a mean depends only on the speed of the player. However, PIA's mean and covariance matrix depends on the distance to the ball as well. If an opposing player is too close to the player who holds the ball, it is not possible to pass the ball into a great area behind the opponent. However, if the opponent is not too close (like 40-60m distance) then he only has a circle around him in which he has possibility to intercept the pass. If he is too far from the ball (>60m) then PIA shifts towards the ball since he has more time to move in this direction if the length of the pass is longer. To represent this manner in the calculation of the mean, we applied a closeness factor to the speed of the player. As defined in Equation 3.10, the mean value depends on the speed vector of the player. A function that exponentially decreases if the input value is less than 50 meters and exponentially increases if the input value is greater than 50 meters is used to transform the speed vector of the soccer player. This function $q$ returns a factor value for a defensive player at position $p_i(t)$ depending on distance to ball at position $p_b(t)$.

$$q(p_b(t), p_i(t)) = e^{\frac{1}{d(p_b(t),p_i(t))^{0.5}}} \tag{3.14}$$

We put a mean shifting vector $\vec{s_g}$, which we call general ball speed vector, between the ball and the player so that closeness factor $q$ shifts the mean of the distribution in the direction of these two objects. As mentioned in Equation 3.10 mean value depends on the speed vector of the player. Adding this $\vec{s_g}$ to the speed vector $\vec{s_i}$ of the player $i$ gives the needed shift of the mean. The speed of the player only shifts this direction a bit because of his inertia. Figure 3.5 shows how the distance between the ball and the player affects the mean of the PIA. Each figure contains a ball and a player who has zero speed. White circle shows the position of the ball $p_b(t)$ at time $t$ while white arrow shows $\vec{s_g}$ in the direction from ball to the player $i$. It's magnitude $v_g$ is a constant which is decided according to the opinions of soccer analysis experts. The red circle shows the position of the player.

Figure 3.6 shows the application of Equation 3.15 as well. It compares a player with zero speed and a player who has a distinct speed. Red arrow represents the speed vector $\vec{s_i}(t)$ of the player at time $t$. As can be seen in the Figure 3.6a, if the player does not move too fast, PIA falls right behind him, but if he has a certain speed, PIA shifts in that direction as shown in the Figure 3.6b.

$$\vec{s_i}(t) = \vec{s_i}(t) + \vec{s_g} * q(p_b(t), p_i(t)) \tag{3.15}$$

As mentioned at the top of this chapter, we use an equation similar to Equation 3.1. For a location $p$ on pitch, PIA value $h_i(p,t)$ for a defensive team player $i$ at time $t$ is a multivariate distribution as expressed in Equation 3.16.

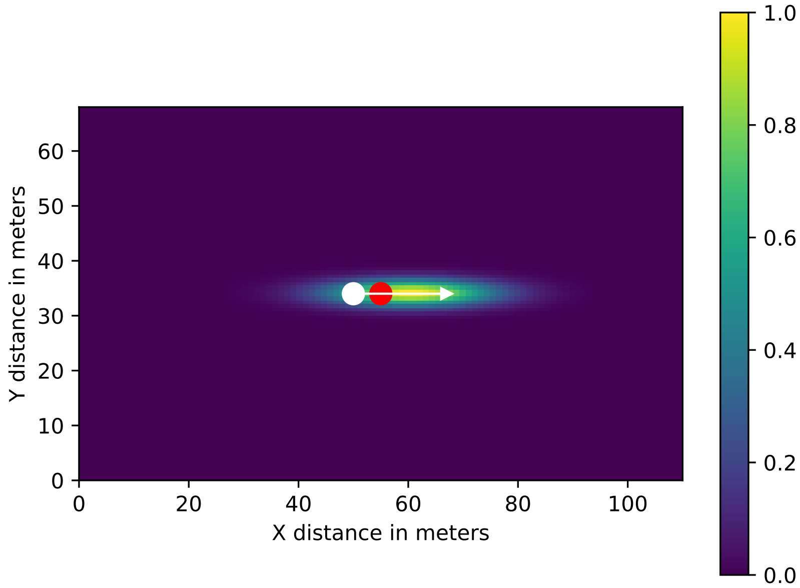
$$h_i(p,t) = \frac{1}{\sqrt{(2\pi)^2 detCOV_i(t)}} exp(-\frac{1}{2}(p - \mu_i(\vec{s_i}(t)))^T COV_i(t)^{-1}(p - \mu_i(t))) \tag{3.16}$$

We calculate pass interception likelihood $H$ of a player on a point $p$ at time $t$ as Equation 3.17. This equation forms a degree of interception value within a range $[0,1]$ for a location $p$ on pitch.

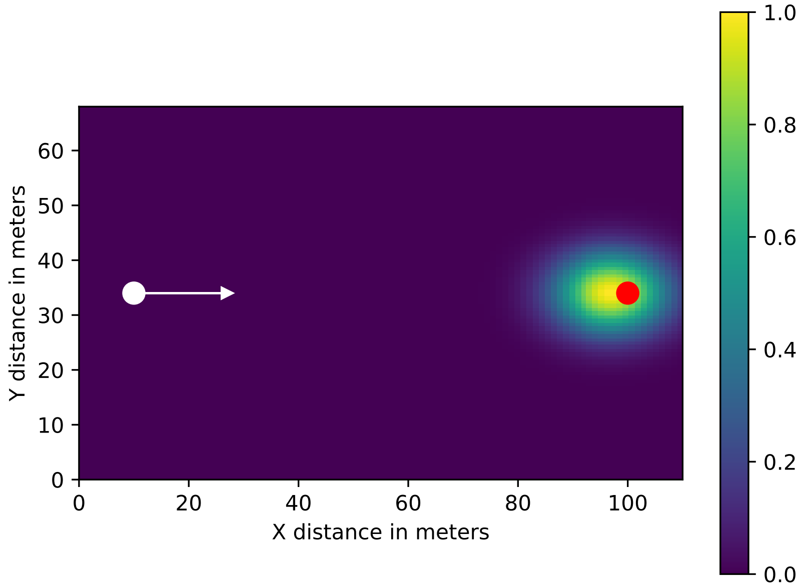
$$H_i(p,t) = \frac{h_i(p,t)}{h_i(p_i(t),t)} \tag{3.17}$$

Each defensive player d has his own PIA $H_d$ which constructs valuable areas on the pitch. The first form of the value of a point on the pitch is calculated with Equation 3.18. We say pass interception value of point $p_{k,l}$ at time t as $P_{k,l}(t)$. The position of the ball at time t is $p_b(t)$. For a moment from a game, the pitch is transformed into $75 \times 75$ matrix. For each cell in the matrix, Equation 3.18 calculated and stored in that cell. Figure 3.5 shows the heat map for that pass interception value matrix on the pitch. The yellow color means it is safe to pass the ball there, but purple colored areas have the possibility to be intercepted.

$$P_{k,l}(t) = \sum_d H_d(p_b(t), p_{k,l}(t)) \tag{3.18}$$

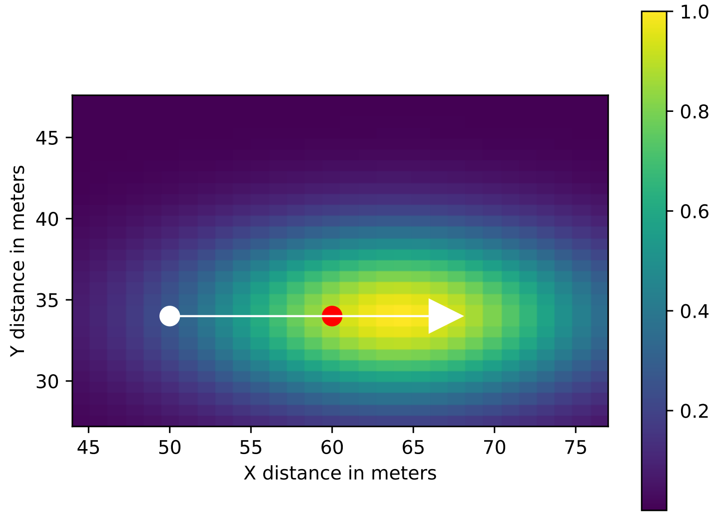



(a) PIA distribution if the opponent is too close to the ball.

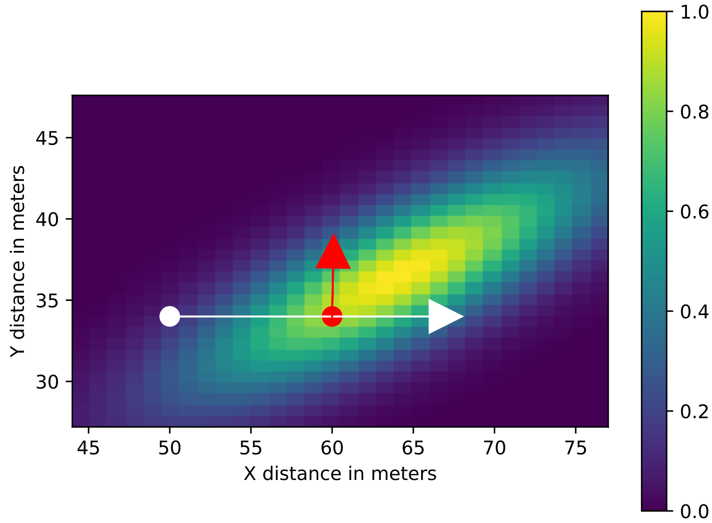



(b) PIA distribution if the opponent is too far to the ball.

Figure 3.5: Comparison of two interception areas with different distance to the ball.




(a) PIA of a player stands on the pitch.




(b) PIA of a player moving in the direction of the red arrow with speed 5 m/s.

Figure 3.6: Comparison of two PIAs with different speeds.

# CHAPTER 4

# TEAM SPREAD

Team spread is described as the sum of all players' distances from each other [3]. It gives useful information about the overall organization of the team since each player's position with respect to other teammates is taken into account. The change in the team spread value at the time of the attacking and the defending moments and during the soccer events such as shots, crosses, and final third entries show how the organization of the team has changed.

Another form of analyzes concerns the network analysis among the players. In competitive events like team sports, the performance also depends on the social relations between the players in the team [39, 5]. In particular, passing networks have been used in many studies and provided information about the interaction of players during attacking [4, 18, 7, 10]. However, other types of networks can be created alongside pass networks. For example, the proximity of the players during attacking and defensive moments can be shown in a network structure. A defensive network can also be created using the moments while the opposing team is attacking. This can be achieved by calculating the number of teammates marking the same players or pressing to the same player. There can be many different networks for attacking and defensive moments, but these networks do not mean much on their own [10]. As the number of structures to analyze increases, it becomes almost impossible to comprehend the big picture for the sports coaches. Given the limited duration between the successive games during the season, the time spent on the analysis should be as low as possible while making the most outcome.

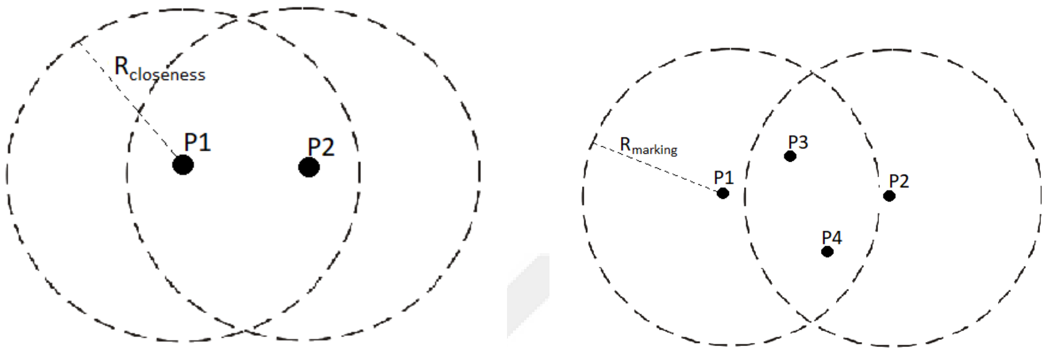
In order to address these issues, we propose a compact analysis method leveraging both the rich data obtained from the optical tracking technology and the power of compact visualization to give more insights to the coaches in a shorter amount of time. We use the attacking and defensive interactive behavior of the players as a weighting factor in the team's spread. We use a $N \times N$ distance matrix, which consists of the distance of the teammates to each other when making a team spread calculation. In soccer, $N = 11$ most of the time since red cards or severe injuries when the substitution is not possible is rare. We also use $N \times N$ cohesion matrices that contain the tactical behavior of the players together with the team spread to achieve a view that spans both the physical and the tactical aspects of the game. While the players who interact more with each other have a greater impact on the spread of the team, players who do not interact with each other will not affect this spread as much both when they are close to each other and when they are distant. In our analysis, distances between the goalkeeper and the offensive players do not affect the spread analysis much, therefore, we do not exclude them from the analysis like other studies do [3]. This allows us to create an inclusionary value and consider the team as a whole. Having a compact representation also allows us to focus on the temporal aspect of the change in team spread.

## 4.1 Constructing Cohesion Matrices

In soccer analysis, space occupancy related performance metrics are mostly calculated using the Euclidean distances between players. Rather than using the distances between players as the only metric, we propose using cohesion matrices a weighting factor to those distances. Then, using the weighted distance matrix to calculate the team spread for a given second. We define the team spread as the FN of the Euclidean distance matrix which is weighted using the cohesion matrices for a given second.

We construct cohesion matrices to weight the spread of a team at a given second. For the whole game, two distinct cohesion matrices are created per team. One matrix is for the times when the ball is in possession, $\mathbb{W}_{on}$, and the other one is for the times when the ball is in the opponent's possession, $\mathbb{W}_{off}$.

$\mathbb{W}_{on}$ is constructed using two other matrices: the closeness matrix and the pass count matrix. Similarly, $\mathbb{W}_{off}$ is constructed using two other matrices: the closeness matrix and the common marking count matrix. The rationale behind this construction is the following: when the team is in possession of the ball it should perform offensive by making passes and when the ball is in the opponent's possession, it should perform defensive behavior by marking the opponent players. In both situations, the formation of the team and the space occupancy is important so we use the closeness for both $\mathbb{W}_{on}$ and $\mathbb{W}_{off}$.




(a) Closeness Condition

(b) Common Marking Condition

Figure 4.1: Conditions for closeness and common marking matrices. **(a)** $P1$ and $P2$ are close teammates **(b)** Teammates $P1$ and $P2$ are marking the same 2 opponents $P3$ and $P4$.

### 4.1.1 The Closeness Matrix

The closeness matrix, $\mathbb{C}$, shows the number of seconds that a player is in the vicinity of some other teammates. If a player penetrates into a circle around a teammate that has a radius of $R_{closeness}$, we count it as being close. Each element $c_{ij}$ of $\mathbb{C}$ is calculated as follows:

$$c_{ij} = c_{ji} = \sum_{t=1}^{T} \mathbb{I}(d_{ij,t} < R_{closeness}) \tag{4.1}$$

where $t$ is a specific second in the game, $i$ and $j$ are the player indices, $R_{closeness}$ is the radius of a circle around players, $d_{ij,t}$ is the Euclidean distance between players $i$ and $j$ at time $t$, and $\mathbb{I}(.)$ is the indicator function.

### 4.1.2 The Common Marking Count Matrix

The common marking count matrix, $\mathbb{M}$, shows the number of common marking counts of teammates. At a given second in the game, if two teammates have the same opponents in their marking range, $R_{marking}$, we count them as common markings. In order to find the common markings, we first define the marking set of $i^{th}$ player at time t, $\mathcal{M}_{i,t}$ as follows:

$$\mathcal{M}_{i,t} = \{j \mid Q(i,j) = 0 \wedge d_{ij,t} < R_{marking}\} \tag{4.2}$$

where $R_{marking}$ is the radius of the marking circle around players. If an opponent penetrates in this circle, we count it as a marking. The $Q(.)$ function returns if two players are teammates or not.

$$Q(i,j) = \begin{cases} 1, & \text{if } team(Player_i) = team(Player_j) \\ 0, & \text{otherwise} \end{cases}$$

Each element $m_{ij}$ of $\mathbb{M}$ is calculated as follows:

$$m_{ij} = m_{ji} = \sum_{t=1}^{T} |\mathcal{M}_{i,t} \cap \mathcal{M}_{j,t}| \tag{4.3}$$

where | . | shows the set cardinality.

In Figure 4.1, we depict visually how we calculate the closeness and common marking conditions.

### 4.1.3 The Pass Count Matrix

The pass count matrix, $\mathbb{P}$, shows the number of passes between teammates. We count passes from $i^{th}$ player to $j^{th}$ player and from $j^{th}$ player to $i^{th}$ as same to make the matrix symmetric. Each element $p_{ij}$ of the matrix $\mathbb{P}$ is calculated as follows:

$$p_{ij} = p_{ji} = P(i,j) + P(j,i) \tag{4.4}$$

where $P(i, j)$ is the total pass count from $i^{th}$ player to $j^{th}$ player of a team.

### 4.2 The Team Spread

Having defined the ingredients of the cohesion matrices, we now define how we calculate the team spread. The team spread is the Frobenius norm of the weighted distances between the players and the weights are determined by the cohesion matrices $\mathbb{W}_{on}$ and $\mathbb{W}_{off}$ that are defined as follows:

$$\mathbb{W}_{on} = \alpha \cdot \mathbb{C} + \beta \cdot \mathbb{P} \tag{4.5}$$

$$\mathbb{W}_{off} = \alpha \cdot \mathbb{C} + \beta \cdot \mathbb{M} \tag{4.6}$$

We create these on-ball and off-ball matrices for both home and away teams. The numerical factors $\alpha$ and $\beta$ are optimized with a convolutional neural network. This optimization is detailed in Section 4.2.1.

We apply min-max normalization to cohesion matrices to scale data into $[0, 1]$ before adding matrices to each other to calculate weights. For each element $x \in W$, we replace the element with the normalized version $x'$ as follows:

$$x' = \frac{x - min(W)}{max(W) - min(W)} \tag{4.7}$$

where W is any matrix.

The team spread, $\mathbf{S_t}$, for a given second $t$ in the game is calculated as:

$$\mathbf{S_t} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij,t} d_{ij,t}} \tag{4.8}$$

where $i$ and $j$ are the player indices, $d_{ij,t}$ is the Euclidean distance between player $i$ and $j$ at time $t$. The weight $w_{ij,t}$ depends on the team's possession status. If the team is in possession of the ball at time $t$, then $w_{ij,t} = \mathbb{W}_{on}[i, j]$ and if the possession is in the opponent team, then $w_{ij,t} = \mathbb{W}_{off}[i, j]$.

### 4.2.1 $\alpha - \beta$ Optimization

The cohesion matrices have two weighting factors: $\alpha$, for the closeness matrix and $\beta$ for the pass count or common marking matrices depending on the possession of the ball. While these factors might be manually set to predefined values, it is better to optimize the values with a data-driven approach.

To optimize these factors, we propose using deep learning, which has been shown to work well in many real-world challenging artificial intelligence tasks. A convolutional neural network (CNN) is a type of deep neural network and it is most commonly used in computer vision since it allows us to learn the weight parameters of the convolutional kernels used for feature extraction in images [40]. We take a similar approach and we use CNNs to learn the weighting factors of the cohesion matrices by using the outcomes of the soccer events. We use shots and successful crosses as positive examples and ball losses as negative examples. Then we train the system using the well-known back-propagation algorithm which optimizes the $\alpha$ and $\beta$ factors.

In order to learn the factors, we constructed the CNN architecture in Figure 4.2. The left part of the structure until the merge layer simulates Equation (4.5) or Equation (4.6). $\alpha$ and $\beta$ filters takes the place of $\alpha$ and $\beta$ factors. The binary output layer represents the positive or negative outcomes of the soccer events. To train the network, we feed the CNN with a bag of soccer events of all 308 games in the league. There are a total of 7000 samples of successful shots, unsuccessful shots, successful crosses as positive labels and ball losses as negative examples.

## 4.3 Temporal Analysis

Although the team spread values can be calculated for any second of the game, in order to put it soccer context, we analyze the time sequences which ends with a concrete result. For this purpose, we use transitions which start with gaining the possession

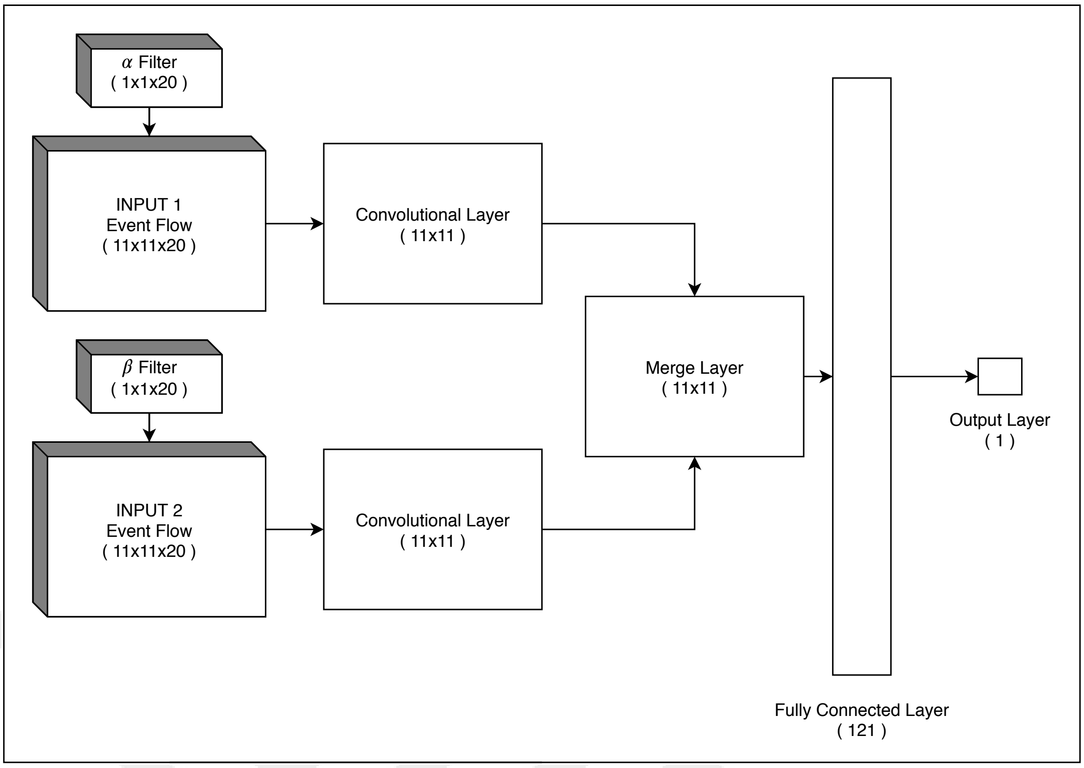



Figure 4.2: Convolutional Neural Network structure to optimize $\alpha$ - $\beta$ factors. Two different weighted distance matrices are fed to the system. They are convolved and merged. Then the input becomes a vector with fully connected layer.

of the ball and ends with one of the following events: shot on target, shot off target, successful cross and dispossession. Dispossession means losing possession of the ball.

These four events are labeled in the SentioScope database already but the exact beginning of the sequences are unknown. In order to find the beginning of each sequence, we go back in time and find the moment the team that committed the event gained the possession of the ball. If the resulting sequence is too short ($< 5$ secs) for analyzing, we ignore that sequence.

We observed that the majority of the soccer attacking sequences last around 20 seconds. Therefore, we scale all sequences to 20-time units in order to work with a comparable structure across teams and games. For the shorter sequences, we interpolate the data and for the longer ones, we shrink the sequences.

## 4.4 Chapter Experiments

To demonstrate the effectiveness of our proposed analysis using our spatio-temporal cohesive networks, we use the whole league data from the TSL.

### 4.4.1 Cohesion Matrices

While calculating the cohesive team spread values, we use cohesion matrices as weights. These cohesion matrices are not only weights of Frobenius Norm (FN) calculation, it also gives clues about the effectiveness of teammates. Each social network we used as cohesion matrices gives different information about players.

Pass networks are used in different studies to find pass patterns [18] and participation of players to offensive plays [10].

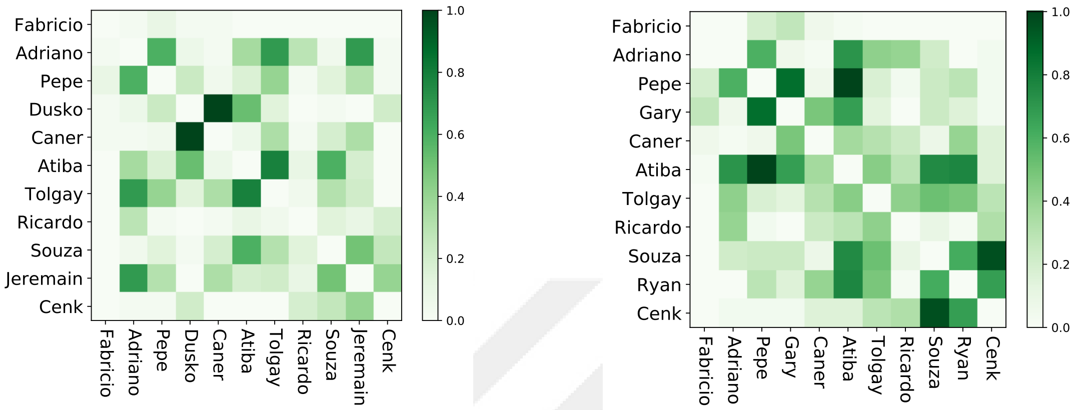
Closeness matrices show us which players become a possible bridge between pass sequences. We can see another good clustering using positional data instead of passes.

Common marking matrices work for defensive behaviors of the players. It is another network to link players using positional data of both their team and the opposite team.

To understand what cohesion matrices tell us, we compare two different games of Besiktas in the 2017-2018 season as a home team. The first game is against Basaksehir which is one of the teams aiming for the top. The other one is against Konyaspor stand in the $15^{th}$ position at the end of the season. Some statistical information is shown in Table 4.1.

Table 4.1: Statistical information of games which BJK played against BSK and KON.

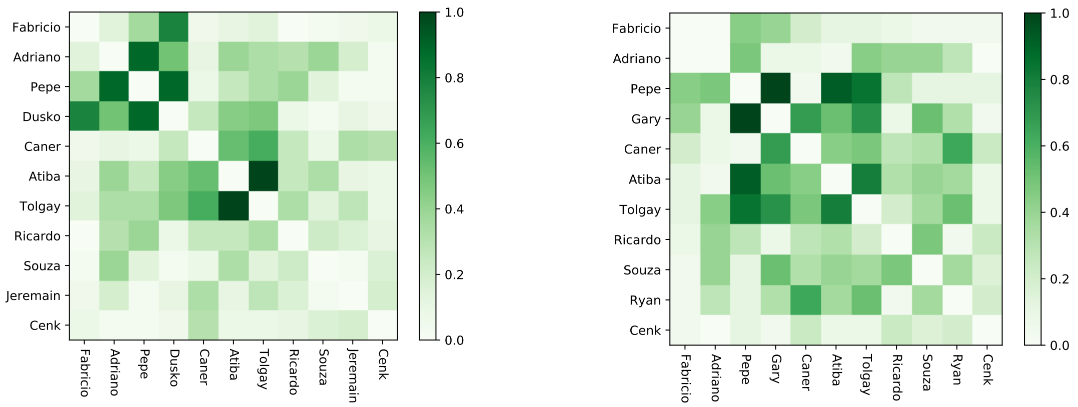
| | Score | Ball Possession | Shots on Target | Shots off Target | Pass |
|---|---|---|---|---|---|
| BJK-BSK | 1-1 | 49-51 | 3-3 | 14-9 | 467-487 |
| BJK-KON | 2-0 | 55-45 | 10-3 | 12-9 | 504-419 |




(a) Closeness of BJK (BJK-BSK)

(b) Closeness of BJK (BJK-KON)

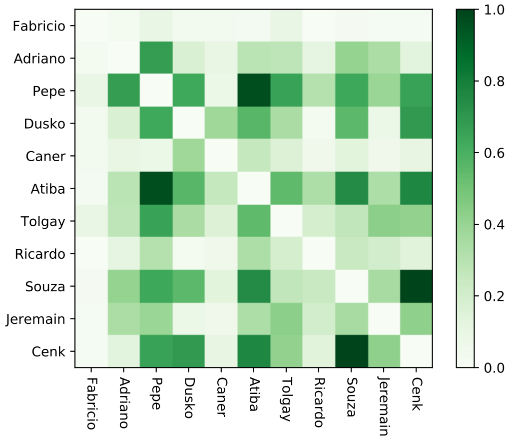
Figure 4.3: Closeness Networks of Besiktas in games against Basaksehir **(a)** and Konyaspor **(b)**. Players are ordered according to formation of the team.



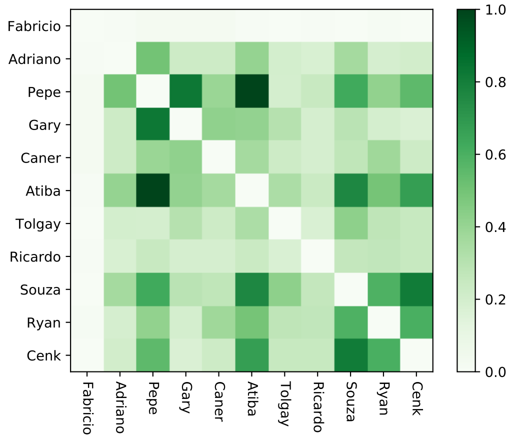

(a) Pass Network of BJK (BJK-KON)

(b) Pass Network of BJK (BJK-KON)

Figure 4.4: Pass Networks of Besiktas in games against Basaksehir **(a)** and Konyaspor **(b)**. Players are ordered according to formation of the team.

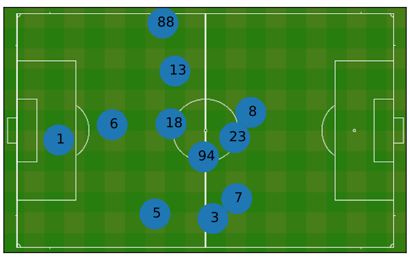



(a) Common Marking Network of Besiktas (Besiktas-Basaksehir)

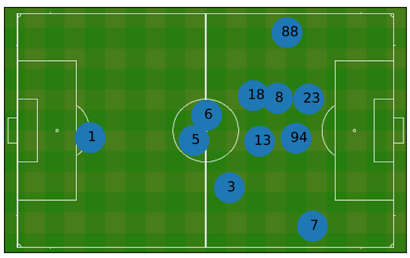



(b) Common Marking Network of Besiktas (Besiktas-Konyaspor)

Figure 4.5: Common Marking Networks of Besiktas in games against Basaksehir **(a)** and Konyaspor **(b)**. Players are ordered according to formation of the team.

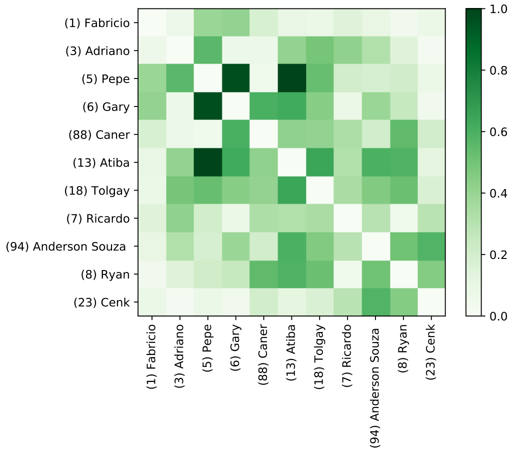



(a) A snapshot of formation of Besiktas while attacking. Team spread value without weight is $349.957m$ and weighted team spread value is $118.253m$.

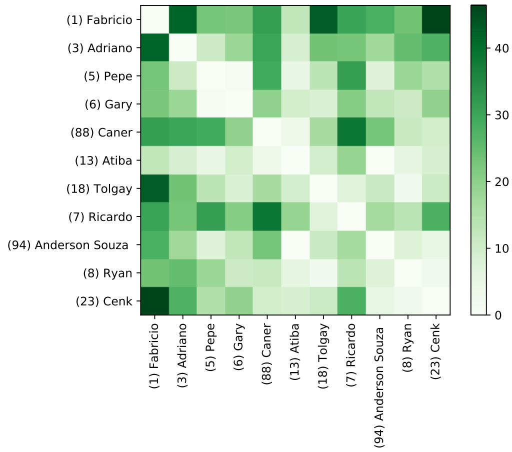



(b) A snapshot of formation of Besiktas while attacking. Team spread value without weight is $350.592m$ and weighted team spread value is $77.585m$.

Figure 4.6: Two snapshots of formation of Besiktas from a game while has the possession of the ball and attacking to the right side of the pitch.




(a) Offensive cohesive network of Besiktas.




(b) Matrix which shows the difference between team spread values without weight and with weight of Figure 4.6b.

Figure 4.7: Offensive cohesive network of Besiktas **(a)** and difference matrix **(b)** showing $\mathbb{D} - \mathbb{D} \circ \mathbb{W}$. $\mathbb{D}$ is player-to-player distance matrix and $\mathbb{W}$ is offensive cohesive network.

Another comparable area that cohesion matrices applied is weighting the team spread. 4.6 shows two different snapshots from a game of Besiktas while they had the possession of the ball.

In Figure 4.7a, we show the offensive cohesive weight of Besiktas in the game from which snapshots are taken. Figure 4.7b is the matrix that shows the difference between team spread values without weight and with weight which means which players influence team spread less and more.

### 4.4.2 $\alpha - \beta$ Optimization

We fed the deep learning model visualized in Figure 4.2 with the mentioned data in Section 4.2.1. Changes in kernel values are observed to analyze how pass and closeness networks affect the offensive cohesive weight matrix during a transition. We have two kernels of size $1 \times 20$ in the model as shown in the model figure. This $20$ stands for scaled transition sequence length. Figure 4.8 shows changes in kernels.

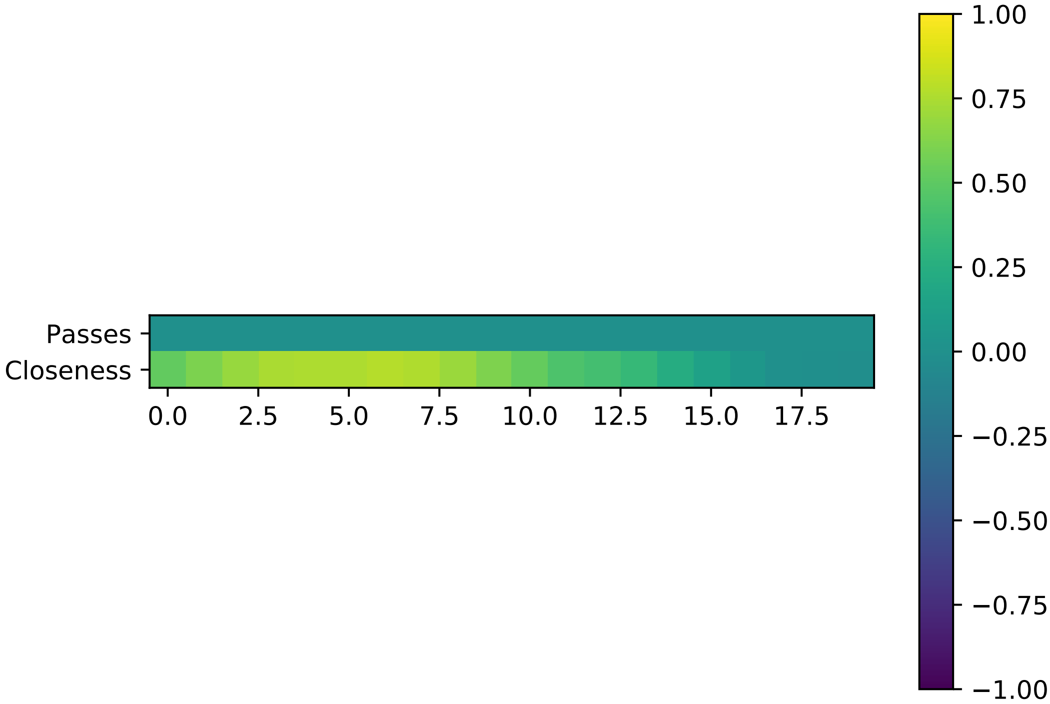



Figure 4.8: Changes in kernel values showing influence of pass and closeness networks over offensive cohesion matrix.

$1$ means that the matrix labeled is the only important network matrix for cohesion and $-1$ means the exact opposite. We observe that closeness is slightly more important than passes at the beginning however closeness becomes even more important than passes in the middle of the transition. At the end of the transition which the event occurs, we see that closeness loses its influence and they become equally effective.

### 4.4.3 Team Analysis

We make the analysis of each teams' averages and standard deviations for time-series of events which are listed in Section 3.4. Table 4.2 and Table 4.3 lists teams according to their rankings at the end of the 2017-2018 Turkish Super League season and show their statistics. The highest and lowest values are bold in the table.

Table 4.2: Goal scores and average team spread values in the events which committed by the team specified in the columns. The teams are ordered according to their rank in the league.

| Team | G. For | Shot on Target | Shot off Target | Succ. Cross | Ball Loss |
|---|---|---|---|---|---|
| Galatasaray | 75 | 100.09 ±15.5 | 94.970 ±19.6 | 97.490 ±18.1 | 84.468 ±20.2 |
| Fenerbahce | 78 | 103.376 ±17.4 | 99.034 ±13.8 | 93.351 ±17.1 | 88.752 ±20.2 |
| M.Basaksehir | 62 | 82.902 ±24.1 | 97.758 ±23.1 | 91.876 ±22.4 | 81.663 ±25.1 |
| Besiktas | 69 | 100.649 ±26.6 | 87.894 ±20.4 | 94.535 ±20.2 | 89.178 ±21.8 |
| Trabzonspor | 63 | 97.945 ±15.7 | 98.088 ±22.5 | 98.100 ±13.2 | 88.576 ±19.2 |
| Goztepe | 49 | 104.055 ±27.0 | 94.387 ±21.0 | 106.477 ±22.9 | 91.020 ±22.3 |
| Sivasspor | 45 | 74.103 ±20.9 | 99.922 ±18.2 | 100.960 ±15.2 | 86.752 ±20.3 |
| Kasimpasa | 57 | 93.927 ±20.1 | 101.575 ±19.5 | 99.727 ±23.7 | 88.177 ±20.8 |
| Kayserispor | 44 | 110.002 ±15.8 | 104.958 ±14.4 | 106.696 ±16.7 | 91.840 ±19.0 |
| Y.Malatyaspor | 38 | 102.537 ±26.1 | 96.997 ±24.0 | 93.243 ±20.2 | 88.745 ±22.5 |
| Akhisar B.Spor | 44 | 103.388 ±24.8 | **111.160 ±27.4** | **112.051 ±34.4** | 94.068 ±26.1 |
| Alanyaspor | 55 | 90.701 ±13.0 | 109.237 ±26.9 | 99.004 ±14.0 | 88.267 ±18.7 |
| Bursaspor | 43 | 95.188 ±23.2 | 95.255 ±24.8 | 99.089 ±24.8 | 83.518 ±22.4 |
| Antalyaspor | 40 | **110.974 ±48.9** | 93.966 ±23.3 | 105.531 ±23.8 | **107.535 ±39.9** |
| Konyaspor | 38 | 106.446 ±21.7 | 98.148 ±19.6 | **81.682 ±12.0** | 89.942 ±21.3 |
| Osmanlispor | 49 | 95.321 ±20.2 | 90.525 ±23.8 | 88.515 ±30.1 | 78.725 ±23.8 |
| Genclerbirligi | 37 | **69.823 ±20.0** | **88.842 ±23.0** | 92.994 ±22.9 | **72.929 ±16.3** |
| K.Karabukspor | 20 | 82.925 ±17.1 | 90.654 ±20.7 | 100.553 ±20.3 | 83.523 ±19.6 |

Table 4.3: Conceded goal counts and average team spread values in the events which suffered by the team specified in the columns. The teams are ordered according to their rank in the league.

| Team | G. Con. | Shot on Target | Shot off Target | Succ. Cross | Ball Loss |
|---|---|---|---|---|---|
| Galatasaray | 33 | 83.425 ±21.0 | 74.417 ±20.0 | 76.011 ±17.0 | 91.065 ±20.3 |
| Fenerbahce | 36 | 86.724 ±17.4 | 77.022 ±23.8 | **91.709 ±21.4** | 93.142 ±21.1 |
| M.Basaksehir | 34 | **97.542 ±24.1** | 89.405 ±27.1 | 68.365 ±27.8 | 91.116 ±25.6 |
| Besiktas | 30 | 80.748 ±24.7 | 94.783 ±24.6 | 80.501 ±22.1 | 96.818 ±22.1 |
| Trabzonspor | 51 | 94.486 ±29.3 | 87.721 ±24.7 | 85.996 ±15.2 | 94.817 ±20.4 |
| Goztepe | 50 | 96.568 ±15.0 | 86.978 ±22.9 | 87.659 ±19.0 | 99.128 ±25.0 |
| Sivasspor | 53 | 86.946 ±22.1 | 86.266 ±20.6 | 82.469 ±21.2 | 95.154 ±23.5 |
| Kasimpasa | 58 | 80.171 ±19.9 | 85.056 ±21.0 | 86.783 ±22.1 | 93.555 ±22.2 |
| Kayserispor | 55 | 94.381 ±18.8 | 84.992 ±21.5 | 89.512 ±25.2 | 96.579 ±20.8 |
| Y.Malatyaspor | 45 | 88.985 ±32.2 | 95.254 ±26.7 | 83.413 ±28.5 | 95.941 ±23.2 |
| Akhisar B.Spor | 53 | 89.969 ±30.9 | 87.120 ±27.2 | 80.289 ±21.3 | 99.776 ±27.9 |
| Alanyaspor | 59 | 78.411 ±17.5 | 78.536 ±18.4 | 91.372 ±19.8 | 94.602 ±20.0 |
| Bursaspor | 48 | 72.755 ±20.3 | 82.866 ±30.6 | 76.355 ±18.9 | 90.555 ±25.4 |
| Antalyaspor | 59 | 87.601 ±25.3 | **101.260 ±44.3** | 84.252 ±30.1 | **113.920 ±40.9** |
| A.Konyaspor | 42 | 84.787 ±23.4 | 76.407 ±17.3 | 80.280 ±17.5 | 97.567 ±23.5 |
| Osmanlispor | 60 | 75.281 ±28.4 | 74.374 ±20.2 | 75.860 ±20.1 | 80.715 ±23.1 |
| Genclerbirligi | 54 | **66.249 ±17.6** | **65.006 ±19.2** | **67.337 ±20.2** | **77.896 ±18.8** |
| K.Karabukspor | 86 | 86.401 ±27.1 | 73.366 ±24.2 | 79.623 ±24.3 | 89.228 ±21.2 |

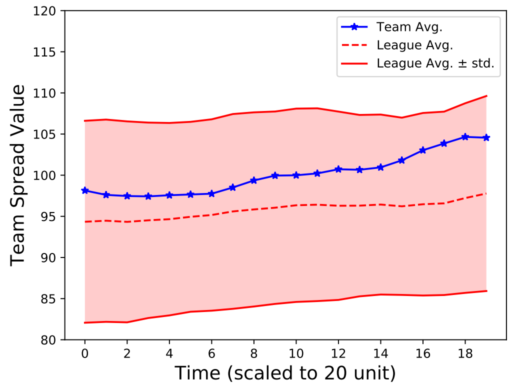
Table 4.4 shows some common behaviours for events. It compares committed and suffered events for each team. Rows are events we analyzed. We calculated the mean of the average team spreads of teams for each event separately. The first two columns are the comparison of those average team spreads of teams and the last two columns compare the average of the team spread values of the last second of the sequences.

In addition to team spread averages of teams during events, we analyzed team spread averages as a function of time for the whole transition which starts with gaining ball possession and ends with one of the events. By adding the league averages to the graphs, we show how a team differs from other teams. Also to compare teams, we show graphs of some of the top and bottom teams.

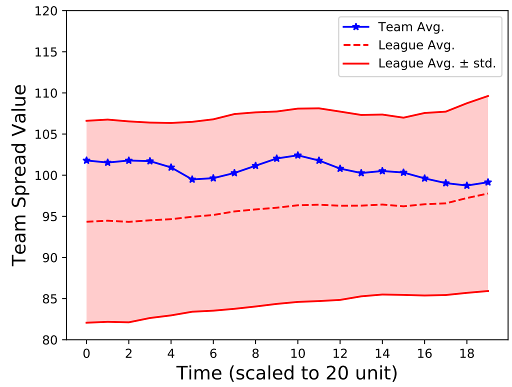
After all comparisons, league averages should be demonstrated as well. Figure 4.13 and Table 4.4 show league averages for both committed and suffered events.

Table 4.4: Count of teams in the league (18 teams) who has smaller team spread in the condition specified.

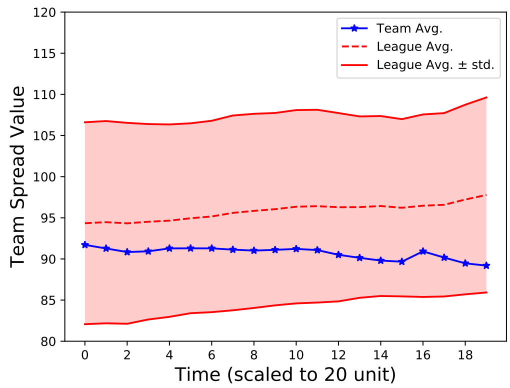
| | Committed Avg. | Suffered Avg. | Committed End | Suffered End |
|---|---|---|---|---|
| Shot on Target | 3 | 15 | 2 | 16 |
| Shot off Target | 2 | 16 | 2 | 16 |
| Succ. Cross | 0 | 18 | 0 | 18 |
| Dispossession | 18 | 0 | 18 | 0 |



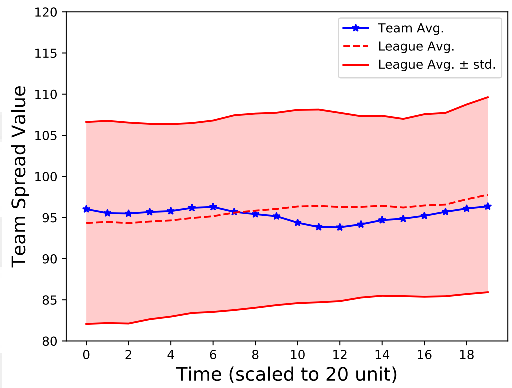

(a) Galatasaray




(b) Besiktas

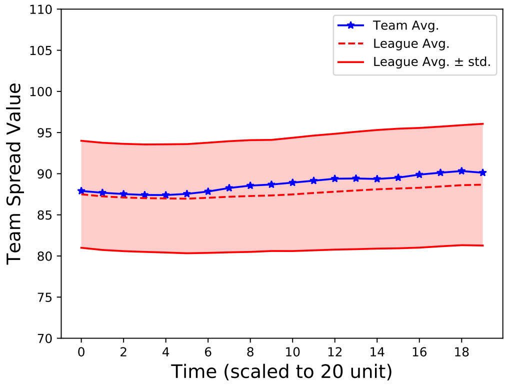



(c) Alanyaspor

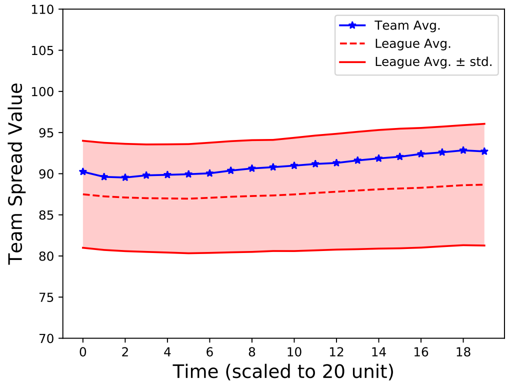



(d) Osmpanlispor

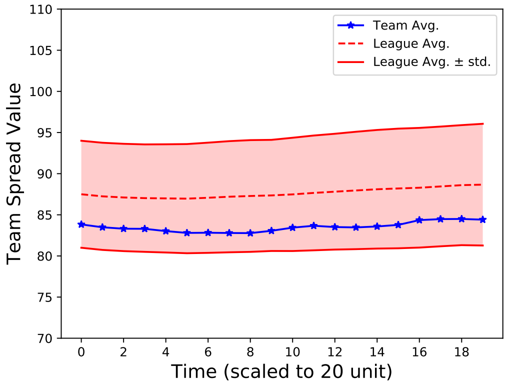
Figure 4.9: Average team spread values of teams during events which finish with a **shot on target**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams.



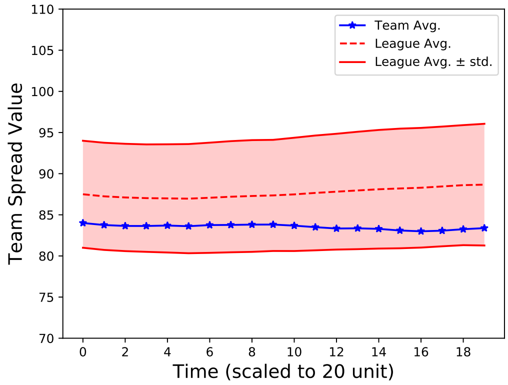

(a) Fenerbahce




(b) Goztepe

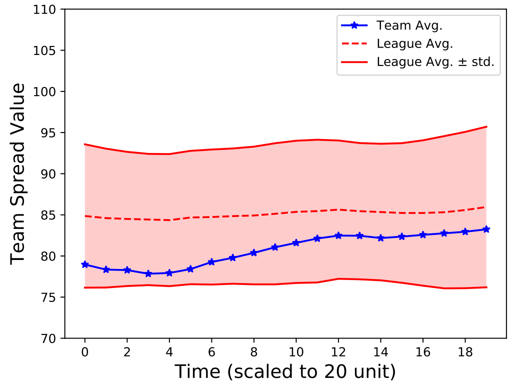



(c) Bursaspor

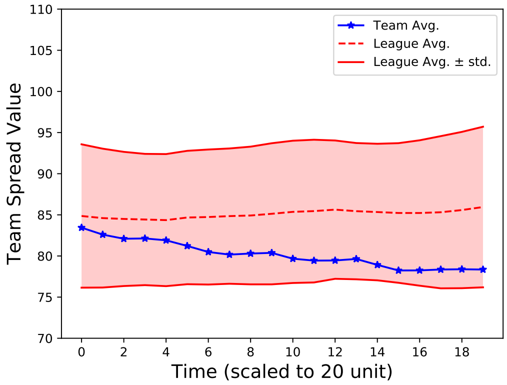



(d) KDC Karabukspor

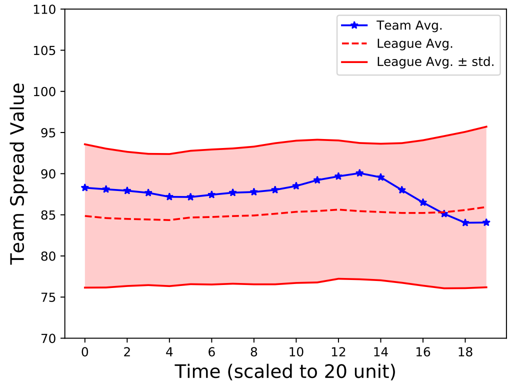
Figure 4.10: Average team spread values of teams during events which finish with **losing the ball**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams.



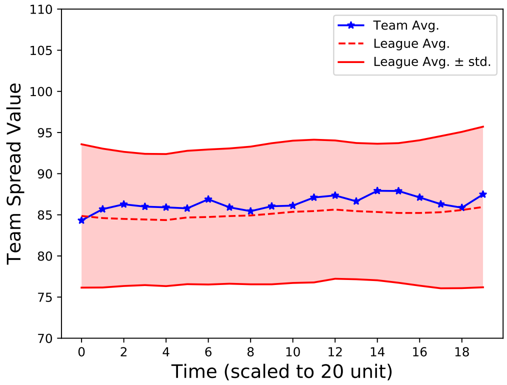

(a) Besiktas




(b) Kasimpasa

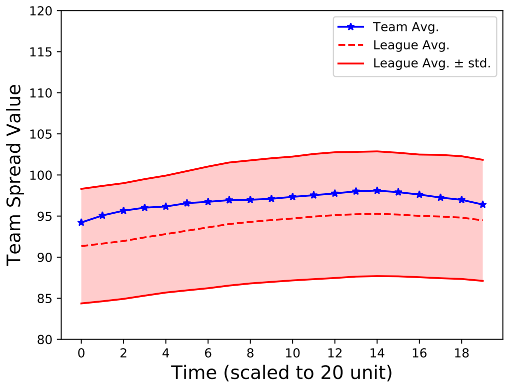



(c) Antalyaspor

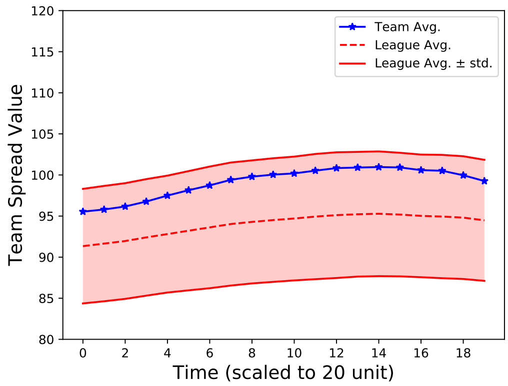



(d) KDC Karabukspor

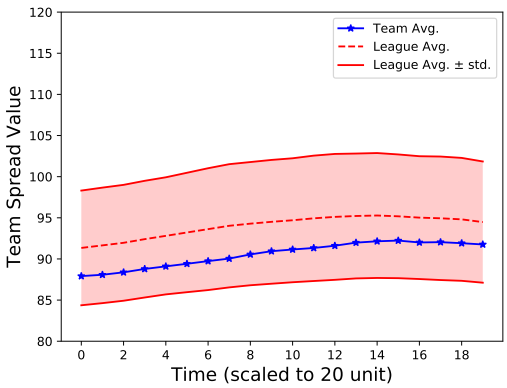
Figure 4.11: Average team spread values of teams during events which finish with a **suffered shot on target**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams.



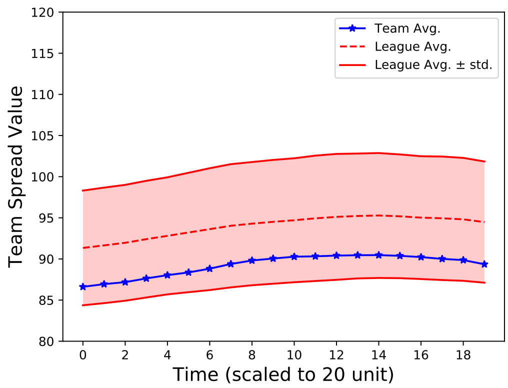

(a) Besiktas




(b) Goztepe




(c) Bursaspor




(d) KDC Karabukspor

Figure 4.12: Average team spread values of teams during events which finish with **gaining possession of the ball**. Two teams **(a)**, **(b)** from top six teams and two teams **(c)**, **(d)** from bottom seven teams.

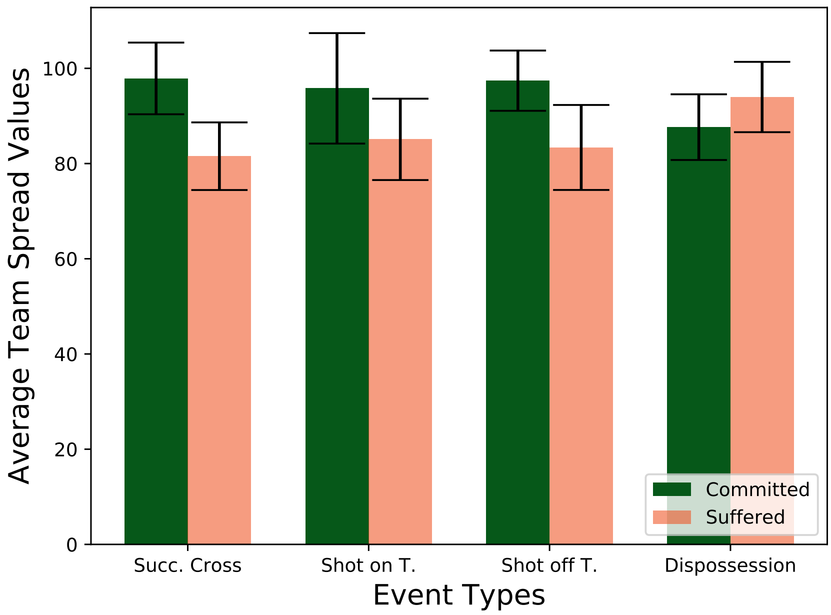



Figure 4.13: Average team spread values of Turkish Super League for committed and suffered events.

## 4.5 Chapter Discussions

### 4.5.1 Cohesion Matrices

Cohesion matrices as social networks give us clues about the behaviors of teammates in a game. Two games compared in Section 4.1 has fully different social networks.

The closeness matrix of the first game shows only defensive players were in contact during the whole game. Also especially attacking players are fully separated on the pitch. These problems end with a draw game. However, in the second game, all 3 parts of the team (defenders, midfielders, forwards) have more dense clusters which mean teammates in the clusters have more interaction between others in the same cluster. Also, connections between clusters are tighter in the second game when we compare with the first one. We can analyze wing back-defensive midfield links. Caner and Adriano are two full-backs of Besiktas and Atiba and Tolgay are defensive midfielders. Closeness and pass matrices show that in the second game links between full-backs and defensive midfielders are thicker. Furthermore, Souza (Anderson Souza Talisca) attacking midfielder of the team has thicker links with defensive midfielders and attacking players in both closeness and pass matrices in the second game.

When we compare common marking matrices, they look very similar. Also, shot on target and off target counts of two away teams are the same. It shows similar behaviors in defensive moments results with similar statistics.

These cohesion matrices give us information about the key players of the teams as well. When we look at the cohesion matrices' visualizations, we understand pass bridges, key midfielders. It is also possible to calculate closeness and betweenness centrality scores to quantify these key players in the teams.

While the FNs of the player-to-player distance matrices are a significant metric about the spread of the teams on the field, there are shortcomings in showing the team's compactness. The fact that the players who interact more with each other are more influential for the spread value of the team which is an example of a social community than the players who interact less with each other while they are distant from each other. Because of this reason, cohesive weights are crucial while calculating team spread value. There are two example snapshots that show the formation of the team from a game of Besiktas while they have the possession of the ball. Although the team spread values without weights are almost the same, their team spread values with weights are totally different. The first reason is the position of the goalkeeper. Unai Emery, manager of the Arsenal, once said: "When we are thinking in an attacking moment, I want the goalkeeper thinking, for that, he is the first. The same when we are thinking defensively." As part of both attacking and defending transitions, goalkeepers should be included in the team spread. However, since their interaction is very few with attacking players, the influence of the goalkeeper to the spread value is small. Dark colors in Figure 4.7b shows which player-to-player links influence team spread less. Ricardo Quaresma, Caner Erkin, Cenk Tosun, and Fabricio influence team spread less since their interaction with their teammates less.

### 4.5.2 Team Analysis

We calculated teams average spread values during both committed and suffered events showed in Table 4.2 and Table 4.3.

We find correlations between their ranks in the standings, their scored and suffered goal counts and their average team spread time-series. There are some outliers like Antalyaspor and Basaksehir. Table 4.2 shows that the average team spread values in attacking events are highly correlated with scored goal and rank of the teams.

Table 4.2 shows teams get wider when they ending their offensive play as a cross. 11 teams in the league have a greater average team spread in successful cross events. On the other hand, when teams cannot expand enough just after they gain the ball possession, more likely this offensive play ends with ball loss.

When defending, teams have more similar behaviors as shown in Table 4.3. 12 teams in the league have a smaller average team spread in suffered shots off target compared with shots on target. Moreover, 12 teams become more contracted when they defending crosses. All 18 teams of the league gain ball possession with a wider positional organization.

In addition to comparisons of events, Table 4.4 compares events themselves for committed and suffered sequences. Only 3 teams in the league have a smaller average team spread while committing shot on target and 2 teams while committing shot off goal. Although Sivasspor has a smaller average team spread during shot on target events, they became wider at the moment that they committed a shot on target.

Figure 4.9, 4.10, 4.11 and 4.12 compares average team spread values as a function of time. In each figure 2 teams from the top and 2 teams from the bottom part of the standing are compared for an event type.

In Figure 4.9, we compare average team spread values of teams for the shot on target sequences. As shown in the figure, Galatasaray and Besiktas are wider than the average of the league, while Alanyaspor and Osmanlispor are tighter. Also, Galatasaray which scored second most goal gets wider during whole sequences.

Figure 4.10 compares the average team spread values of sequences ends with losing the possession of the ball. Fenerbahce and Goztepe can expand from the beginning of the transition to the moment that they lose the ball. However, Bursaspor and KDC Karabukspor cannot become wide enough compared with the rest of the league. They are also two of the worst teams according to possession and successful pass rates in the season.

Figure 4.11 and 4.12 shows suffered events for teams. In figure 4.11, we compare suffered shot on targets. Besiktas, the team that conceded the least goals in the league has highly compact positional organizations while defending. Kasimpasa is another team has a tighter spread than the league average. Conversely, KDC Karabukspor and Antalyspor that first and third teams conceded the most goals in the league were very dispersed while defending.

Figure 4.12 includes average team spread values for teams. The most important inference of this figure is league averages. Almost all teams in the league expand in the first half of the transitions and in the second half they start to shrink and gain the possession of the ball.

League averages give general insights about the Turkish Super League. Figure 4.13 show the average team spread values for both committed and suffered events. The first noticeable point in that figure is that teams are tighter when they are defending except in events which end with a tackle. Results also show that teams apply more varying pitch organizations to make sequence eventuate in a shot on target. Moreover, teams are less variable than attacking sequences when they are defending.

# CHAPTER 5

# QUANTIFYING SPRINT VALUE

The quality of a sprint depends on how much advantage it is providing. It should target a valuable location on a pitch. It may be a space between defensive players or a good position behind them. It can target a location close to the ball as well. These beneficial purposes increase the value of that location on the pitch.

Whatever the purpose is, it should be a location in which it is possible to give the ball. If the player who holds the ball cannot pass the ball to that location, then there is no meaning to do a sprint to that location. There can be two types of pass interception. The first one is directly intercepting the pass between two players who try to pass each other. The second one is that a defensive player can reach the location that the ball goes, before the attacking player who sprints there. For this reason, pass interception areas of defensive players are important.

Since the pass interception area is something detrimental to sprint value, we are going to use its negative value while calculating sprint value and visualizing the pass interception value distribution. We call those areas as high pass probability areas and the metric as high pass probability value. High pass probability value is defined as in Equation 5.1. In Section 4.1.3 we were showing high interception value areas colored as yellow. However, in this chapter, we show high pass probability areas colored as yellow.

$$HP_{k,l}(t) = 1 - P_{k,l}(t) \tag{5.1}$$

Besides all these, while an attacking player sprints somewhere, he should consider the team spread. If he becomes lonely in a big range after a sprint, then he may not find a teammate to pass the ball. If he is a fullback (left or right back) and if he cannot get the ball which means pass is intercepted, then this sprint can be harmful because of a counter-attack. He will not be in the defensive organizations anymore. For these reasons, players should consider team spread as well while deciding a sprint.

These three matters construct the final value for a valuable sprint. Summing up the first two measurements are easy since both have distributions on the pitch. The third one is the challenging one since it is a scalar value. To transform this scalar into distribution, we created a circle centered to mean of the team at time $t$. The radius of the circle is decided using the weighted team spread value of the team. Figure 5.4 shows a moment from a game of TSL. Green circles show the player of the attacking team and red circle defenders. The ball is colored as white and a big blue circle shows the mean of the attacking team.

We are going to place a circle around the big blue circle. The radius of this circle will be proportional to the weighted team spread value of the green team. However, since the weighted team spread value is smaller than the physical spread of the team, we are going to use a multiplier to increase the weighted team spread value into a meaningful range. Generally, the unweighted team spread values are within the range $[350, 450]$ while weighted ones are within the range $[70, 120]$. For this reason, we decided a multiplier which is $4.3$. We create a virtual scenario in which all eleven attacking players including goalkeeper placed around a perfect circle where the radius is 10 meters. We calculate weighted team spread value of that organization by dividing unweighted team spread value to $4.3$ and we proportion real circle radius using these two virtual values.

By calculating the virtual weighted team spread value $S_{v\_on}$, as defined in the Equation 5.2, for organization shown in the Figure 5.1, we obtain the proportion $e$ between virtual weighted team spread value and virtual radius 10 m. Then, we apply this proportion to optimum weighted team spread value $S_{opt\_on}$ and get an optimum radius $r_{opt}$ for the attacking team as in Equation 5.3. Using this circle, the optimum dispersion of the team is calculated for that moment $t$ and a distribution $S_{k,l}(t)$ which is $1$ inside the circle is placed on the pitch, whose value decreases as it moves away from

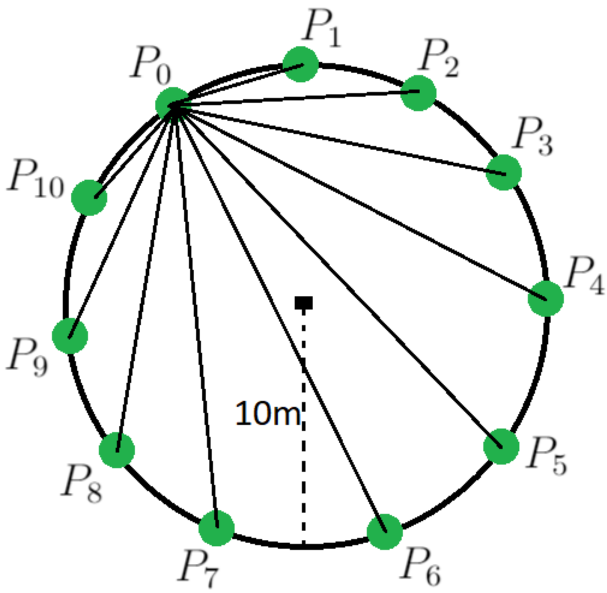



Figure 5.1: Virtual organization of the attacking team around a circle with radius of 10 m.

the circle.

$$e = \frac{S_{v\_on}}{4.3 \times 10} \tag{5.2}$$

$$r_{opt} = \frac{S_{opt\_on}}{e} \tag{5.3}$$

Finally, these three distributions forms final quantitative value $SV_{k,l}(t)$ of a sprint targets location $k, l$ at time $t$. As defined in Equation 5.4, the sprint value of that location $p$ is calculated by summing up the pitch value $V_{k,l}(t)$, the pass interception value $P_{k,l}(t)$ and optimum weighted team spread value $S_{k,l}(t)$ at time $t$. Figures 5.2, 5.3, 5.4 visualize this formation step by step. Green circles show attacking team players while red circles show defenders. White one shows the ball. Attacking team attacks towards the right side of the pitch.

$$SV_{k,l}(t) = V_{k,l}(t) + (1 - P_{k,l}(t)) + S_{k,l}(t) \tag{5.4}$$

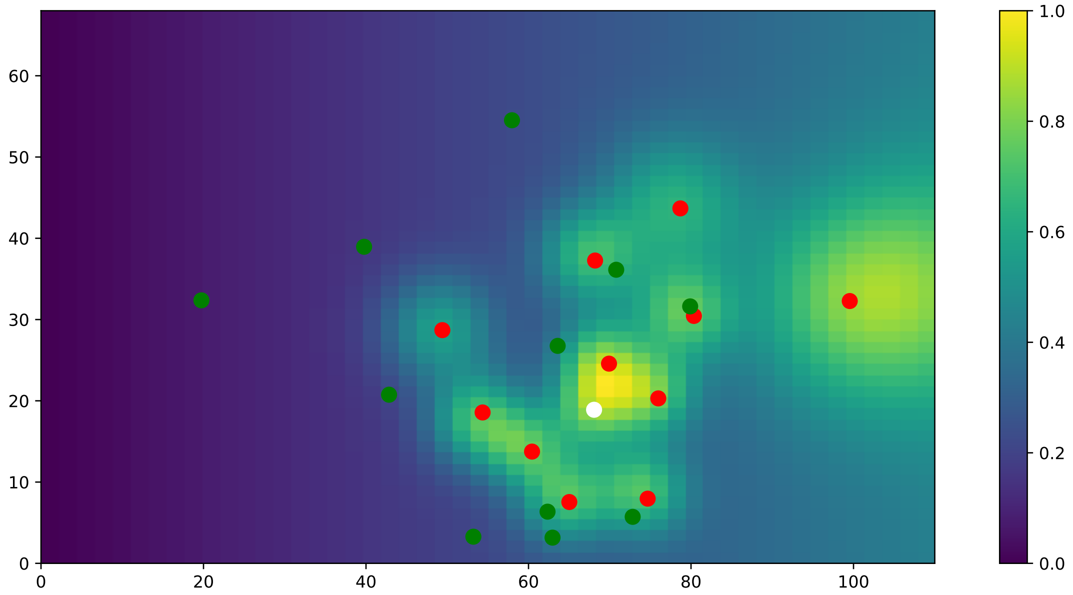



Figure 5.2: Pitch value distribution of the pitch at a game moment.

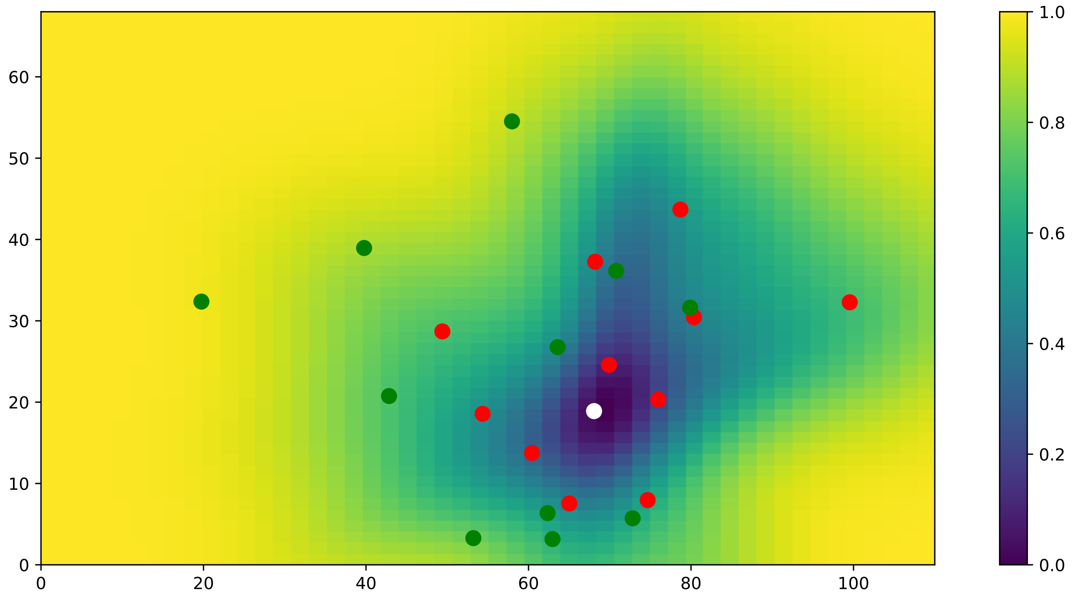



Figure 5.3: Pass probability value distribution of the pitch at a game moment.

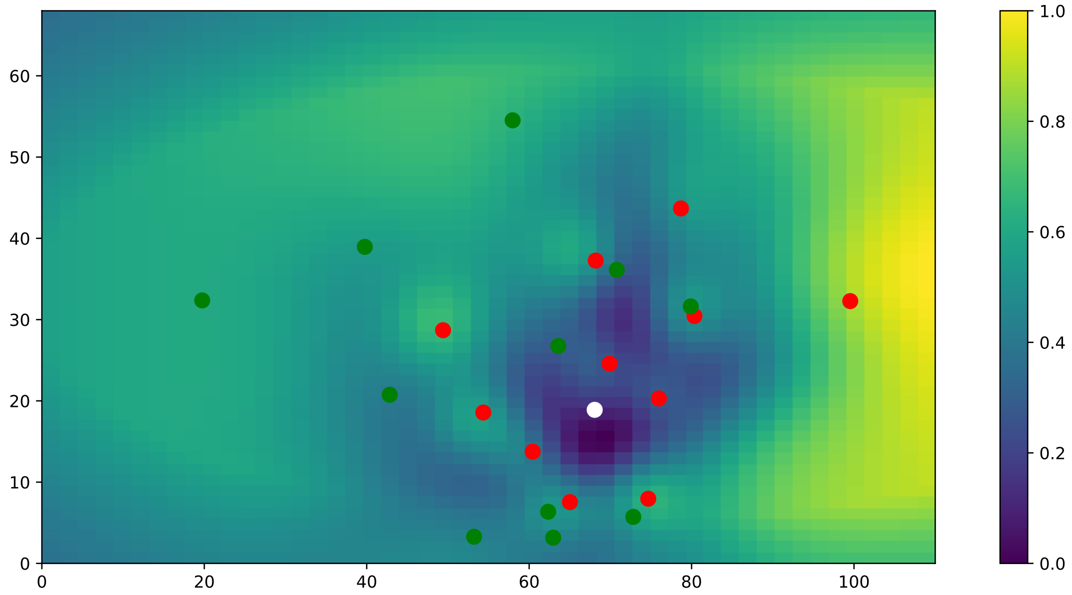



Figure 5.4: Final sprint value distribution of the pitch, including team spread effect, at a game moment.

# CHAPTER 6

# RESULTS AND DISCUSSION

This chapter split into two sections. The first section visualizes valuable and insignificant examples from real games. The second section includes tables listing best and worst sprinters and team averages.

## 6.1 Sprint Quantification

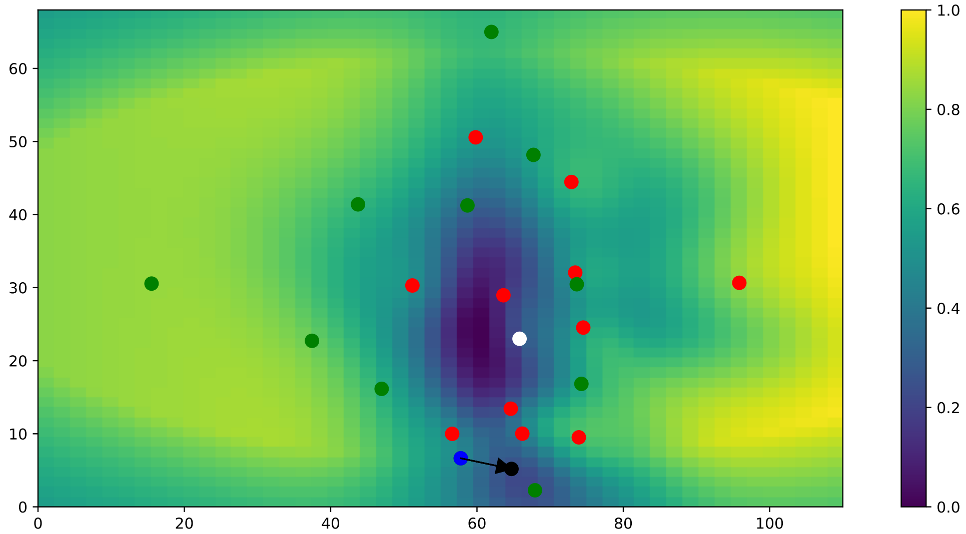
Quantifying sprint value and revealing a metric to show if it is a good sprint or not is something unimaginable if we do not picture it with real pitch moments. For this reason, we visualized two high and two low sprint value examples of real games. Heat maps are widely used in soccer analysis which is one of the most understandable visualization methods to show valuable and insignificant areas on the pitch that is represented as a matrix [41]. We used heat maps to show valuable sprint areas on the pitch for analyzed moments. This representation gives information about potential valuable sprint areas for the attacking team. Following examples analyze a selected player but it is possible to find out the good and bad sprint areas for all attacking players on that figure.

Green dots show the attacking team where red ones show defenders. There are one blue and one black dots on the pitch. Blue one represents the starting point of the sprinting player, Adem Ljajic attacking midfielder of Besiktas, while black one shows the ending point of the sprint. The black arrow also pictures the direction of the sprint. Besiktas attacks to the right side of the pitch. For these visuals, our goal in selecting the player in the attacking midfielder is to increase the likelihood of sprinting from anywhere in the field. The data can be biased as the forward players will often sprint towards the goal zone and the wing players will generally sprint towards the sideline. Figure 6.1 visualizes two insignificant sprints and Figure 6.2 pictures two valuable sprints done by Adem Ljajic.

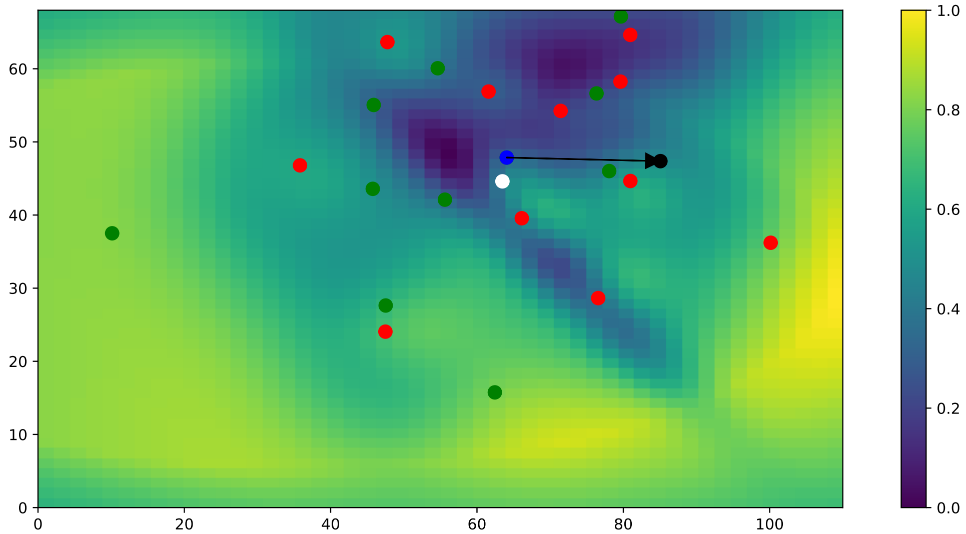
We listed two insignificant sprint moments in Figure 6.1. There can be different reasons for a sprint to be insignificant. In Figure 6.1a, the player shown with a blue dot, sprints next to sideline towards an empty space. This can be considered as a meaningful behavior. However, since there are two opponents who can intercept the pass aiming this sprinting player, the area involving the black dot becomes less valuable to sprint. The effects of defensive players around this area on pitch value are not more than the pass interception value of the same players. For this reason, this area becomes worthless.

In Figure 6.1b, we see another reason than the first example makes this sprint insignificant. The player who is shown with the blue dot sprints towards black dot which is represented with the black arrow as well. As can be seen in the figure, there is no opponent around that black arrow who can intercept the pass. However, the opponent at the southwest of the black dot controls the area containing the black dot. It is also possible to intercept the pass by the goalkeeper of the defending team who is at the rightmost side of the pitch shown with the red dot. This two situation makes this sprint insignificant. The heat map shows us that there are better options for that player to sprint.

Figure 6.2 contains two valuable sprint examples. As in Figure 6.1, there are different reasons which make a sprint valuable. In Figure 6.2a, we see that the player shown with blue dot sprints towards black dot which has a high pitch value since it is an empty space behind defending players. Another considerable situation that makes this sprint valuable is that there is no opponent who can intercept the pass directed to the black dot. On the other hand, in the second example pictured in Figure 6.2b, the main factor is being close to the goal spot.

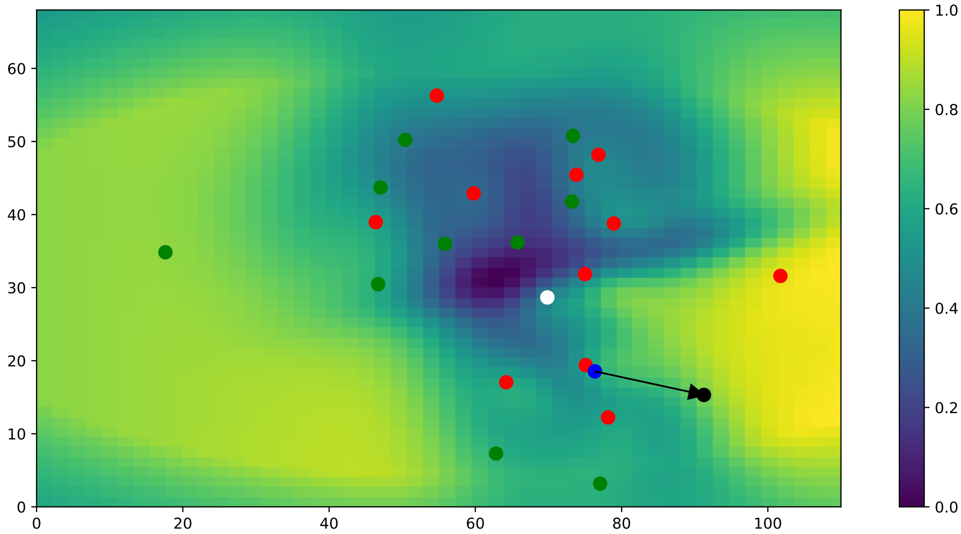



(a) An insignificant sprint which is measured as 0.30.

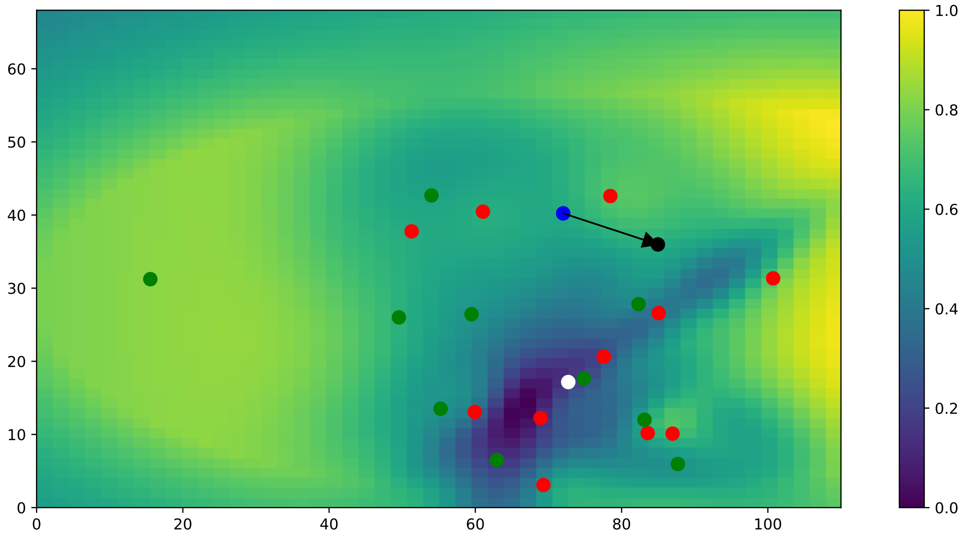



(b) An insignificant sprint which is measured as 0.37.

Figure 6.1: Two insignificant sprints done by Adem Ljajic while attacking.




(a) A valuable sprint which is measured as 0.81.




(b) A valuable sprint which is measured as 0.70.

Figure 6.2: Two valuable sprints done by Adem Ljajic while attacking.

## 6.2 League Level Analysis

We also analyzed the whole Turkish Super League 2018-2019 season. There were 306 games and 12074 sprints of 512 different players from 18 teams. Since not all players have adequate amount of sprints to get a consistent average and standard deviation value for him, we used players who have more than 30 sprints which involve 146 players. Table 6.1 lists the top 20 players with the highest average sprint value in the league, while Table 6.2 lists the 20 players with the lowest average. Table 6.3 lists team averages and standard deviations. We added their final league positions for the 2018-2019 season as a column to show how sprint behaviors of teams change according to their position.

Tables 6.1, 6.2 tells us that role of the player is one of the most important determinant for player sprint quality value. 13 of the top 20 players in Table 6.1 have the role of fullback and 4 players have the role of the winger. Teams use different tactics and different formations during games but fullbacks and wingers are main sprinter players in the squad [42]. They are more likely to find empty corridors, which are generally next to sidelines, than other players in other roles. It is also possible to overlap behind the defenders for these roles. These capabilities enable players in these roles to have higher averages of sprint values. On the other hand, both the sprint numbers of the central midfielders and the high probability of intercepting the pass in this area cause the midfielders to have lower sprint averages. Especially defensive midfielders are the most troubled players in sprint value. The only thing that makes their sprints valuable is that they usually play close to the ball. As a result, 12 players in the list with the lowest sprint average are midfielders and 9 of them are predominantly defensive.

If we look at the team averages, here we see that the overall tactics of the teams are effective. Fullbacks and wingers have high sprint averages for teams playing possession football, which means that teams aim to keep the possession of the ball with passes, but the overall average is lower. However, teams that play more counter-attack football have higher sprint averages. The reason for this is that, as their opponents mostly try to play in the attacking third which is one-third of the field attacked by the team with the ball, their attackers sprint behind the defenders of the opposing team when they gain the possession of the ball.

Table 6.1: Top 20 players have most valuable sprint averages.

| Sprint Avg. Position | Name | Average | Std. Deviation |
|---|---|---|---|
| 1. | Vahid Amiri | 0.777 | 0.085 |
| 2. | Yusuf Erdoğan | 0.770 | 0.131 |
| 3. | Iasmin Latovlevici | 0.767 | 0.092 |
| 4. | Berkan Emir | 0.765 | 0.089 |
| 5. | Aly Cissokho | 0.765 | 0.114 |
| 6. | Umut Meras | 0.764 | 0.096 |
| 7. | Adis Jahovic | 0.764 | 0.104 |
| 8. | Mayele Fabrice N'Sakala | 0.763 | 0.102 |
| 9. | Emrah Bassan | 0.759 | 0.109 |
| 10. | Anthony Nnaduzor Nwakaeme | 0.758 | 0.119 |
| 11. | Karim Ramadan Seifeldin | 0.757 | 0.097 |
| 12. | Onur Ayık | 0.752 | 0.135 |
| 13. | Yuto Nagatomo | 0.751 | 0.097 |
| 14. | Tyler Boyd | 0.750 | 0.123 |
| 15. | Hasan Ali Kaldırım | 0.750 | 0.106 |
| 16. | Diafra Sakho | 0.750 | 0.096 |
| 17. | Caner Erkin | 0.747 | 0.118 |
| 18. | Uğur Çiftçi | 0.747 | 0.104 |
| 19. | Atila Turan | 0.746 | 0.095 |
| 20. | Bilal Başakçıoğlu | 0.746 | 0.126 |

Table 6.2: Bottom 20 players have worst valuable sprint averages.

| Sprint Avg. Position | Name | Average | Std. Deviation |
|---|---|---|---|
| 1. | Jose Ernesto Sosa | 0.650 | 0.124 |
| 2. | Efecan Karaca | 0.665 | 0.130 |
| 3. | Andre Ayew | 0.668 | 0.118 |
| 4. | Mossoro Da Costa | 0.669 | 0.141 |
| 5. | Adem Büyük | 0.670 | 0.144 |
| 6. | Dorukhan Toköz | 0.676 | 0.115 |
| 7. | Henri Gregoire Saivet | 0.677 | 0.132 |
| 8. | David Pavelka | 0.679 | 0.123 |
| 9. | Aytaç Kara | 0.679 | 0.107 |
| 10. | Andre Castro | 0.680 | 0.115 |
| 11. | Sergiy Rybalka | 0.681 | 0.126 |
| 12. | Papiss Demba Cisse | 0.681 | 0.119 |
| 13. | Mahmut Tekdemir | 0.682 | 0.128 |
| 14. | Irfan Can Kahveci | 0.683 | 0.118 |
| 15. | Nejc Skubic | 0.686 | 0.148 |
| 16. | Younès Belhanda | 0.686 | 0.127 |
| 17. | Bengali Fode Koita | 0.686 | 0.138 |
| 18. | Loret Sadiku | 0.686 | 0.131 |
| 19. | Arouna Kone | 0.688 | 0.126 |
| 20. | Tunay Torun | 0.688 | 0.113 |

Table 6.3: Average sprint values of teams in the Turkish Super League 2018-2019 season.

| | League Pos. | Team Name | Average | Std. Deviation | Pass Counts |
|---|---|---|---|---|---|
| 1. | 16. | Bursaspor | 0.727 | 0.116 | 361 |
| 2. | 11. | Çaykur Rizespor | 0.723 | 0.121 | 377 |
| 3. | 7. | Antalyaspor | 0.722 | 0.120 | 315 |
| 4. | 15. | Göztepe | 0.721 | 0.115 | 353 |
| 5. | 10. | Kayserispor | 0.718 | 0.121 | 340 |
| 6. | 12. | Sivasspor | 0.717 | 0.118 | 372 |
| 7. | 5. | Yeni Malatyaspor | 0.715 | 0.121 | 298 |
| 8. | 14. | Kasımpaşa | 0.715 | 0.127 | 347 |
| 9. | 18. | Akhisar Bld.Spor | 0.713 | 0.122 | 345 |
| 10. | 8. | Atiker Konyaspor | 0.712 | 0.129 | 344 |
| 11. | 17. | Erzurum BŞB | 0.711 | 0.124 | 345 |
| 12. | 13. | MKE Ankaragücü | 0.710 | 0.122 | 324 |
| 13. | 3. | Beşiktaş | 0.709 | 0.119 | 420 |
| 14. | 9. | Alanyaspor | 0.706 | 0.124 | 327 |
| 15. | 4. | Trabzonspor | 0.705 | 0.120 | 406 |
| 16. | 1. | Galatasaray | 0.705 | 0.125 | 431 |
| 17. | 2. | Medipol Başakşehir | 0.704 | 0.127 | 486 |
| 18. | 6. | Fenerbahçe | 0.696 | 0.123 | 431 |

# CHAPTER 7

# CONCLUSIONS AND FUTURE WORKS

Sprints are one of the actions that make a difference in a match. Players usually move during the match without having the ball. At this point, sprints become very important. Sprints are also one of the most decisive factors in player continuity. Therefore, the successful sprint rate is very important for the players. This thesis aims to quantify the quality of the sprints done by attacking team players using spatio-temporal features of those players during the game.

There are three multivariate distributions construct sprint quantification model. These distributions are pitch value, pass probability value and the team spread effect. Pitch value distribution calculates a value for locations on the pitch based on defenders' positions, distance to the ball, distance to the goal and horizontal coordination of that location. This value is the most important criterion to understand the benefit of the sprinting area for that player. Pass interception value predicts the probability of interception of the pass towards the sprinting player. We use the complement of this interception value as a pass probability value. If the pass towards the sprinting player is unlikely to be successful, it doesn't make sense that the sprinting area is valuable. Teams have a certain team spread both when attacking and defending. This is very important for the organization of the team. In case of a possible pass interception, the sprint should not increase the team spread too much so that the defensive organization will not cause any problems. Therefore, we use a circle that shows the optimum team spread. As we move away from this circle, the value of that place becomes less and less worthless. These three components form the final sprint distribution on the pitch.

The results show that players' roles are an important criterion for sprint values. Average sprint values are higher in roles such as full-back or winger, where sprinting is more important. While the majority of players with the highest sprint average are players in these roles, midfielder roles based on having the possession of the ball and center-back roles with a more static playing style have lower sprint value averages. Besides, the game tactics of the teams affect the overall sprint value averages of the teams. The teams that play with possession football style which based on passing to each other have lower averages, whereas the teams that play in counter-attack style have higher averages because they are based on the sprinting towards the back of the opponent's defense.

The sprint value model is a good indicator for both players' and teams' style of play as it demonstrates the sprint value averages of players and teams. Also, since the sprint value distribution can be visualized on the pitch, it can be used as a tool for both the development of attacking players' sprint skills and defending players' chasing skills using the opponent's sprint model.

Cohesion matrices are very good insights to understand the organizational tactics of teams. Therefore, it is possible to form various cohesion matrices. Interactions such as common movements, pass similarities, space generation-occupation match-ups can be analyzed. In addition, techniques such as Hidden Markov Model can be tried to create cohesion weights of transitions. Sprint value analysis shows that players' roles affect their averages, but it can be an obstacle to recognizing players who differ in their roles. Therefore, it is possible to make improvements in the impact of roles. In addition, the physical characteristics of the players affect these values. Players' heights, maximum sprint speeds can be used to make these distributions more precise. Besides, this model can be used to improve sprinting and chasing skills more effectively by using augmented virtual reality devices during training, allowing players to instantly see valuable sprint areas, to get sprint values and get feedbacks after their sprints. In addition, the weighted team spread model can be used in other team sports, such as basketball, handball, and football, as it examines the spatio-temporal organization of teams. It can also be used as a tool for analysis in military offensive and defense exercises that require organizational action as a team.

## REFERENCES


[1] I. K. Koerte, M. Mayinger, M. Muehlmann, D. Kaufmann, A. P. Lin, D. Steffinger, B. Fisch, B.-S. Rauchmann, S. Immler, S. Karch, *et al.*, "Cortical thinning in former professional soccer players," *Brain imaging and behavior*, vol. 10, no. 3, pp. 792–798, 2016.

[2] A. Junge, K. Cheung, T. Edwards, and J. Dvorak, "Injuries in youth amateur soccer and rugby players—comparison of incidence and characteristics," *British journal of sports medicine*, vol. 38, no. 2, pp. 168–172, 2004.

[3] F. A. Moura, L. E. B. Martins, R. D. O. Anido, R. M. L. De Barros, and S. A. Cunha, "Quantitative analysis of brazilian football players' organisation on the pitch," *Sports biomechanics*, vol. 11, no. 1, pp. 85–96, 2012.

[4] T. U. Grund, "Network structure and team performance: The case of English Premier League soccer teams," *Social Networks*, vol. 34, no. 4, pp. 682–690, 2012.

[5] R. Duarte, D. Arajo, V. Correia, and K. Davids, "Sports teams as superorganisms: Implications of sociobiological models of behaviour for research and practice in team sports performance analysis," *Sports Medicine*, vol. 42, no. 8, pp. 633–642, 2012.

[6] F. M. Clemente, M. S. Couceiro, F. M. L. Martins, and R. S. Mendes, "Using network metrics to investigate football team players' connections: A pilot study," *Motriz. Revista de Educacao Fisica*, vol. 20, no. 3, pp. 262–271, 2014.

[7] H. Wäsche, G. Dickson, A. Woll, and U. Brandes, "Social network analysis in sport research: an emerging paradigm," *European Journal for Sport and Society*, vol. 14, no. 2, pp. 138–165, 2017.

[8] D. Araújo, P. Silva, and K. Davids, "Capturing group tactical behaviors in expert team players," *Handbook of Sport Expertise*, no. JANUARY, pp. 209–220, 2015.

[9] E. Arriaza-Ardiles, J. M. Martín-González, M. D. Zuniga, J. Sánchez-Flores, Y. de Saa, and J. M. García-Manso, "Applying graphs and complex networks to football metric interpretation," *Human movement science*, vol. 57, no. September 2016, pp. 236–243, 2018.

[10] F. M. Clemente, M. S. Couceiro, F. M. L. Martins, and R. S. Mendes, "Using Network Metrics in Soccer: A Macro-Analysis," *Journal of Human Kinetics*, vol. 45, pp. 123–134, 2015.

[11] F. A. Moura, R. E. van Emmerik, J. E. Santana, L. E. B. Martins, R. M. L. de Barros, and S. A. Cunha, "Coordination analysis of players' distribution in football using cross-correlation and vector coding techniques," *Journal of Sports Sciences*, vol. 34, no. 24, pp. 2224–2232, 2016.

[12] Y. Wu, X. Xie, J. Wang, D. Deng, H. Liang, H. Zhang, S. Cheng, and W. Chen, "ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 65–75, 2018.

[13] J. Hobbs, P. Power, L. Sha, and P. Lucey, "Quantifying the value of transitions in soccer via spatiotemporal trajectory clustering," MIT Sloan Sports Analytics Conference, 2018.

[14] N. Mehrasa, Y. Zhong, F. Tung, L. Bornn, and G. Mori, "Learning person trajectory representations for team activity analysis," *arXiv preprint arXiv:1706.00893*, 2017.

[15] S. Iwase and H. Saito, "Parallel tracking of all soccer players by integrating detected positions in multiple view images," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 4, pp. 751–754, IEEE, 2004.

[16] S. Baysal and P. Duygulu, "Sentioscope: a soccer player tracking system using model field particles," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1350–1362, 2016

.

[17] S.-C. Chen, M.-L. Shyu, M. Chen, and C. Zhang, "A decision tree-based multimodal data mining framework for soccer goal detection," in *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*, vol. 1, pp. 265–268, IEEE, 2004.

[18] B. Gonçalves, D. Coutinho, S. Santos, C. Lago-Penas, S. Jiménez, and J. Sampaio, "Exploring team passing networks and player movement dynamics in youth association football," *PLoS ONE*, vol. 12, no. 1, pp. 1–13, 2017.

[19] M. Perarnau, *Pep Confidential: Inside Pep Guardiola's First Season at Bayern Munich*. Birlinn, 2014.

[20] R. Kannekens, M. T. Elferink-Gemser, and C. Visscher, "Positioning and deciding: Key factors for talent development in soccer," *Scandinavian Journal of Medicine and Science in Sports*, vol. 21, no. 6, pp. 846–852, 2011.

[21] M. Buchheit, A. Mendez-Villanueva, B. Simpson, and P. Bourdon, "Repeated-sprint sequences during youth soccer matches," *International journal of sports medicine*, vol. 31, no. 10, pp. 709–716, 2010.

[22] J. Fernandez and L. Bornn, "Wide Open Spaces : A statistical technique for measuring space creation in professional soccer," *MIT Sloan Sports Analytics Conferece*, pp. 1–19, 2018.

[23] T. Frias and R. Duarte, "Man-to-man or zone defense? measuring team dispersion behaviors in small-sided soccer games," *Trends in Sport Sciences*, vol. 21, no. 3, p. 135, 2014.

[24] W. Spearman, "Beyond Expected Goals," *MIT Sloan Sports Analytics Conference*, pp. 1–17, 2018.

[25] F. Seabra and L. E. Dantas, "Space definition for match analysis in soccer.," *International Journal of Performance Analysis in Sport*, vol. 6, no. 2, pp. 97–113, 2006.

[26] I. G. McHale and S. D. Relton, "Identifying key players in soccer teams using network analysis and pass difficulty," *European Journal of Operational Research*, vol. 268, no. 1, pp. 339–347, 2018.

[27] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop, “Physics—based modeling of pass probabilities in soccer,” in *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017.

[28] R. Rein, D. Raabe, and D. Memmert, ““which pass is better?” novel approaches to assess passing effectiveness in elite soccer,” *Human movement science*, vol. 55, pp. 172–181, 2017.

[29] R. Bartlett, C. Button, M. Robins, A. Dutt-Mazumder, and G. Kennedy, “Analysing team coordination patterns from player movement trajectories in soccer: methodological considerations,” *International Journal of Performance Analysis in Sport*, vol. 12, no. 2, pp. 398–424, 2012.

[30] T. Reilly, A. M. Williams, A. Nevill, and A. Franks, “A multidisciplinary approach to talent identification in soccer,” *Journal of sports sciences*, vol. 18, no. 9, pp. 695–702, 2000.

[31] L. M. Romer, A. K. McConnell, and D. A. Jones, “Effects of inspiratory muscle training upon recovery time during high intensity, repetitive sprint activity,” *International journal of sports medicine*, vol. 23, no. 05, pp. 353–360, 2002.

[32] A. L. Owen, D. P. Wong, D. Paul, and A. Dellal, “Effects of a periodized small-sided game training intervention on physical performance in elite professional soccer,” *The Journal of Strength & Conditioning Research*, vol. 26, no. 10, pp. 2748–2754, 2012.

[33] A. Haghighi, M. Moghadasi, A. Nikseresht, A. Torkfar, and M. Haghighi, “Effects of plyometric versus resistance training on sprint and skill performance in young soccer players,” *European Journal of Experimental Biology*, vol. 2, no. 6, pp. 2348–2351, 2012.

[34] W. Frencken, K. Lemmink, N. Delleman, and C. Visscher, “Oscillations of centroid position and surface area of soccer teams in small-sided games,” *European Journal of Sport Science*, vol. 11, no. 4, pp. 215–223, 2011.

[35] E. Olsen and O. Larsen, “Use of match analysis by coaches,” *Science and football III*, vol. 3, pp. 209–220, 1997.

[36] I. S. Lee and D. H. Nam, “Method and apparatus for manipulating character of soccer game,” May 12 2015. US Patent 9,028,313.

[37] P. Hall, D. Marshall, and R. Martin, “Adding and subtracting eigenspaces with eigenvalue decomposition and singular value decomposition,” *Image and Vision Computing*, vol. 20, no. 13-14, pp. 1009–1016, 2002.

[38] L. Vilar, D. Araújo, B. Travassos, and K. Davids, “Coordination tendencies are shaped by attacker and defender interactions with the goal and the ball in futsal,” *Human Movement Science*, vol. 33, pp. 14–24, 2014.

[39] J. A. Morales, S. Sánchez, J. Flores, C. Pineda, C. Gershenson, G. Cocho, J. Zizumbo, and G. Iñiguez, “Universal temporal features of rankings in competitive sports and games,” *EPJ Data Sci.*, 2016.

[40] S. Liu, J. Tang, Z. Zhang, and J.-L. Gaudiot, “Computer architectures for autonomous driving,” *Computer*, vol. 50, no. 8, pp. 18–25, 2017.

[41] S. Zhao, Y. Guo, Q. Sheng, and Y. Shyr, “Advanced heat map and clustering analysis using heatmap3,” *BioMed research international*, vol. 2014, 2014.

[42] M. Buchheit, B. Simpson, and A. Mendez-Villanueva, “Repeated high-speed activities during youth soccer games in relation to changes in maximal sprinting and aerobic speeds,” *International journal of sports medicine*, vol. 34, no. 01, pp. 40–48, 2013.