**Media Engineering and Technology Faculty**
**German University in Cairo**

# Identifying Football Teams Tactics using Machine Learning

**Bachelor Thesis**

Author: Ahmed Osama
Supervisor: Assoc. Prof. Seif Eldawlatly
Submission Date: 30 July, 2020

**Media Engineering and Technology Faculty**
**German University in Cairo**

# Identifying Football Teams Tactics using Machine Learning

**Bachelor Thesis**

Author: Ahmed Osama
Supervisor: Assoc. Prof. Seif Eldawlatly
Submission Date: 30 July, 2020

This is to certify that:

(i) the thesis comprises only my original work toward the Bachelor Degree

(ii) due acknowlegement has been made in the text to all other material used

---

Ahmed Osama
30 July, 2020

# Acknowledgments

First of all, I want to express my deepest thanks and gratitude to my supervisor Assoc. Prof. Seif Eldawlatly for his guidance, his direction which was extremely important and productive, his great support and motivation. I would like to extend my gratitude to my family and friends for their support and encouragement especially my mother for her continuous motivation, her help in time of need and patience.

# Abstract

Football has always lagged behind other sports when it comes to Data analysis due to its complex style and nature. In this project we aim to identify football teams' tactics using unsupervised machine learning techniques using a dataset of soccer-logs collected from the 2017-2018 season from the 5 top European leagues (English, Spanish, German, Italian, French). Our workflow includes filtering the successful passing events, creating network of passes for every team during every match of the season, inputing the network to unsupervised clustering algorithms and then identifying the teams' tactical patterns from the machine learning output clusters. The resulting network provides a direct visual inspection of a teams' strategy. Our results show that top teams had special style of play with increased number of overall passes and dependence on short passes for attacking. Also, we show that the Premier League had more overall passes and depends on short balls more compared to other leagues. In addition, we do 2 case studies for Real Madrid and Liverpool teams (the UEFA champions League finalists this season) defensive play style. We found that teams that win against them cluster together mostly. Also, playing in home/away stadium and attacking from certain sides can increase/decrease the chances of winning against them.

# Contents

# Chapter 1

# Introduction

## 1.1 Motivation

One problem facing football analytics was the lack of accurate analytical data to assess the tactical actions of the teams. Until recently, tactical research was based on quantitative evidence in elite soccer. Fortunately, detailed game logs have been collected over the last few years through new monitoring technologies, and the largest open set of soccer-logs has been published, containing all the Spatio-temporal events (passes, goals, fouls, etc.) that occurred during each match of five reputable soccer competitions during the entire season [1]. Identifying football teams' tactics can help teams understand which tactical pattern they perform best with and which tactical pattern they perform worst with. It can also help managers to understand their opponents and create plans depending on their opponents' strengths and weaknesses. Also, for the fans it will help them understand their teams play style better as teams' analysis is widespread nowadays on the social media.

## 1.2 Aim of Work

It is not easy to detect the tactical behavior of a football team. Someone watching a team might classify it as offensive while another watching the same team might classify it as defensive, there is no standard for such a classification. It may be useful to develop an algorithm that does these classifications and this algorithm needs unsupervised machine learning techniques to do the job. In this thesis unsupervised clustering will be experimented to assign the different teams to different tactical clusters.

## 1.3 Structure of the Thesis

In Chapter 2, we discuss some definitions and the algorithms that are used in the implementation. Then we move on to the steps of the implementation in Chapter 3. Chapter 4 shows the results of our implementation and discussion. Concluding the thesis and adding recommendations for future work in Chapter 5.

# Chapter 2

# Background

## 2.1 Sports Analytics

Recently, scientists are using data science intensively in sports. Thanks to the recent technological development and widespread in the past few years, more data is now available for different sports. As a result, we are now understanding more about those sports, and the performing teams' tactical patterns, winning/losing conditions and points of strength or weakness which help guide for avoiding mistakes and capitalizing on the strengths to improve for either individual or team sports. For Cycling, Cintia et al. generated the first large scale data-driven study on data gathered from 30,000 cyclists that had a popular fitness application [2]. The authors used this data to build an optimal program for cyclists to improve. Also, In NBA basketball games Hollinger assesses the performance of players based on the Performance Efficiency Rating, which is calculated by combining the manifold type of data gathered during every game [3]. Moving on to tennis, Terroba et al. offer a pattern recognition algorithm to find common winning tactics [4]. Finally, Smith et al. offer a Baysian classifier to predict awards given for the best pitchers in Major League Baseball. The model achieved 80 percent success rate [5].

## 2.2 Football Analytics technologies

Despite the fact that Football (also known as Soccer in the United States of America) is one of the most popular sports in the world, automatically collected statistical information has not been available until recently, due to the uniqueness of football, with its continuous flow of the ball and relatively low scores. However, Football Analytics has attracted interest for a long time. Since the 1950s Charles Reep collected statistics by hand to create "The key to scoring goals is to transfer the ball as quickly as possible from back to front" concept [6]. leading to the start of the long-ball movement in English football [7]. In recent years, football statistics have developed due to the invention of new sensing technologies that provide different types of data about football matches:

### 2.2.1 Soccer-logs

They describe the events that occur during a match and collected through tagging devices automatically and reviewed for accuracy by specialists[8, 9, 10].

### 2.2.2 Video-tracking data

describe the trajectories of players during a match and are collected through video recorded matches [11, 12].

### 2.2.3 GPS data

describe the trajectories of players during training sessions and are collected through GPS devices worn by the players [1].

## 2.3 Machine Learning

Machine learning research is part of research on artificial intelligence, seeking to provide knowledge to computers through data, observations, and interacting with the world. That acquired knowledge allows computers to correctly generalize to new settings, or simply it is the science of getting computers to learn as well as humans do or better. Machine Learning algorithms can be classified into three categories: supervised learning, unsupervised learning and reinforcement learning [13].

### 2.3.1 Supervised Learning

In supervised learning, the desired output for each given input already exists in a set that can be referred to as the training set. Then the main target is to develop a model that can extract all necessarily features from the input and map them to the output to be benefited later to predict the most probable output to a given input outside this training set [13].

### 2.3.2 Unsupervised Learning

Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data [13]. That's why we use this method on our project.

### 2.3.3 Reinforcement Learning

Reinforcement learning is the training of machine learning models to make a sequence of decisions. The agent learns to achieve a goal in an uncertain, potentially complex environment. In reinforcement learning, and artificial intelligence faces a game-like situation. The computer employs trial and error to come up with a solution to the problem. To get the machine to do what the programmer wants, artificial intelligence gets either rewards or penalties for the actions it performs. Its goal is to maximize the total reward [13].

## 2.4 k-means Clustering

This clustering algorithm divides data into the best suited category based on already available knowledge. In k different clusters, which are typically chosen to be far enough apart from each other spatially, data is segregated in Eucledian Range, in order to be able to generate successful data mining results. Each cluster has a core called the centroid, and depending on how close the features are to the centroid, a data point is clustered into a certain cluster [14].
Iteratively, the k-means algorithm minimizes the distances between each data point and its centroid to find the most optimal solution for all data points.

1. $k$ random dataset points are chosen as centroids.
2. Then, it measures and stores distances between each data point and the $k$ centroids.
3. Points are assigned to the nearest cluster, based on distance calculations.
4. New centroid cluster positions are updated: equivalent to finding a mean in the locations of points
5. Unless the centroid positions have changed, the cycle continues from step 2, until the measured new center remains the same, signalling that the members and centroids of the clusters are now set.

The k-means algorithm defined above aims at minimizing an objective function, which in this case is the squared error function. The objective function for the k-means clustering algorithm is the squared error function:

$J = \sum_{i=1}^{k} \sum_{j=1}^{n} (\|x_i - v_j\|)^2 = 1$

where,

$\|x_i - v_j\|$ is the Eucledian distance between a point, $x_i$, and a centroid, $v_j$, iterated over all $k$ points in the $ith$ cluster, for n clusters.

In simpler terms, the objective function attempts to pick centroids that minimize the distance to all points belonging to its respective cluster so that the centroids are more symbolic of the surrounding cluster of data points.

## 2.5 Hierarchical Clustering

Hierarchical clustering is a general family of clustering algorithms which construct nested clusters by successively merging or dividing them. This cluster hierarchy is reproduced as a tree (or dendrogram). The tree's root is the unique cluster which collects all the samples, the leaves being the clusters with only one sample. There are two approaches for hierarchical clustering either bottom-up or top-down [15]. In this project we use bottom-up agglomerative clustering technique to cluster our data .

### 2.5.1 Agglomerative Clustering approach

The Agglomerative Clustering uses a bottom-up approach to perform a hierarchical clustering, each observation starts in its own cluster, and the clusters are merged successively. The metric used for the merge strategy is determined by the linkage criteria:

- Ward: minimizes the amount of squared differences across all clusters. It is a variance-minimizing approach, and is similar to the objective function of k-means in this sense, but tackled with an agglomerated hierarchical approach.
- Maximum or complete linkage: minimizes the maximum distance between pairs of cluster observations.
- Single linkage: minimizes the average of the distances between all observations of pairs of clusters.
- Average linkage: minimizes the distance between the closest observations of pairs of clusters.

We used Ward strategy because agglomerative clustering has a behavior of "rich getting richer" which leads to uneven cluster sizes. Single linkage is the worst strategy in this regard, and Ward gives the regularest sizes [16].

## 2.6 Principal Component Analysis

Given a set of points in two , three, or higher-dimensional space, a "best fit" line can be defined as one that minimizes the average squared distance from one point to the other. The next best fit line can be chosen similarly from perpendicular to the first direction.

Repeating this cycle provides an orthogonal basis in which different individual data measurements are uncorrelated. Such base vectors are called main components, and the main component analysis (PCA) of some similar procedures.[17].

Steps involved in PCA:

1. Standardization: Calculate the mean of all dataset measurements, except for the names. Scale the data so that every variable contributes to the analysis in equal measure. In the equation given, ($z = \frac{x-\mu}{\sigma}$), the scaled value is z, the initial value is x, and the mean and standard deviations are mu and sigma respectively.

2. Covariance Matrix Computation: We can compute the covariance of two variables X and Y using the following formula:

   $cov(X,Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{x})(Y_i - \bar{y})$

   Using the above formula, we can find the covariance matrix of A. Also, the result would be a square matrix of x*x dimensions.

3. Calculate Eigenvectors and corresponding Eigenvalues: In linear algebra, an own vector or characteristic vector of a linear transformation is a non-zero vector that changes by a scalar factor at most when applied to it. The proper value corresponding is the factor by which the own vector is scaled. In fact, a matrix A's own vector is the vector for which it carries the following:

   $A\vec{v} = \lambda\vec{v}$

   where lambda is a scalar value called the 'eigenvalue'. This means that the linear transformation is defined by lambda and the equation can be re-written as:

   $A\vec{v} - \lambda\vec{v} = 0$
   $\Rightarrow \vec{v}(A - \lambda I) = 0$

   where I is the identity matrix.
   It is important to note that both these individual vectors are unit own vectors, i.e. their lengths are both 1. These individual vectors give us the data patterns so that we can extract the most useful ones.

4. Choose k own vectors with the largest values of their own: Sort the own vectors with regard to their decreasing order of their own values, choosing k from them, where k is the number of dimensions you wish to have in the new dataset.

PCA is used in our research to help visualize the different passing networks into a two-dimensional graph.

## 2.7 Prior Work

Researchers have always been interested in searching for key performance indicators for elite association football. In the past, researchers used observational analysis to optimize the training and performance of the players and teams. With the development in high fidelity sensing technological equipment, researchers and practitioners are creating a system to benefit from this automatically generated data to find ways to fix the shortcomings and work towards optimal performances.

### 2.7.1 Grouping of soccer game records by multiscale comparison technique and rough clustering

To derive tactical patterns from match data [18]. Hirano and Tsumoto created a multiscale matching and rough clustering method based on temporal event data consisting of 168 time-series sequences from 64 games of the 2002 FIFA World Cup. Pass patterns such as side-attacks and zig-zag pass transactions leading to a goal could be automatically clustered.

### 2.7.2 Team activity recognition in association football using a bag-of-words-based method

In a sample of two annual Spanish La Liga matches played by four professional teams [19]. Montoliu et al. devised a Bag-of-Words-based approach to evaluate the most frequent movements. Typical team practices such as ball control, rapid attacks and set-piece play may be known for, among other potential uses, recognizing and assessing the strengths and weaknesses of one's team and the opposition team before, during and after a match.

### 2.7.3 Quantifying the relation between performance and success in soccer

A recent research involving a much broader dataset [20]. Including 395 games and 10 million events from the 2013-2014, 2014-2015 and 2015-2016 seasons of the Premier League, Seria A, La Liga, Bundesliga, Dutch Eredivisie, Ligue 1 used machine learning to measure the relationship between success (passes, kicks, shots, tackles, dribbles, clearances, goalkeeping acts, fouls, intercepts, aerial duels, scored goals). The model of logistic regression and classification could predict simulated team rankings near the actual rankings. The differential characteristics between top teams and bottom teams included making more passes and shots than rivals, and committing less fouls, tackles and goal-keeping behavior. Although the events derived from the match appear as standard notational analysis, the consistency of the playing of a team including passing accuracy and the spatial and temporal superiority of a team including average team position, speed and acceleration could also be calculated.

### 2.7.4 Using machine learning to draw inferences from pass location data in soccer

In addition, to pass quality and variability, the location of passes is also a determinant of a successful offense [21]. Providing context related to playing in critical pitch zones, Brooks et al. used a k-nearest neighbor approach to qualitatively assess the effect of passes traveling into and out of Zone 14, the zone located in the middle of the pitch immediately outside the opposing penalty area. Based on all passes from the 2012-2013 Spanish La Liga season, possession in Zone 14 often correlates to shooting opportunities. This is in support of previous studies using notational analysis, where Zone 14 was correlated with assists by making forward passes into the penalty area.

### 2.7.5 Artificial neural networks and player recruitment in professional soccer

In a study demonstrating the potential for machine learning to be used in the scouting and recruitment process in a professional football [22]. technical performance data were collected from 966 outfield players in the English Football League Championship during the 2008-2009 and 2009-2010 seasons. Key performance indicators that influence players' league status and accurately predict their future success in football were identified using quantifiable features, circumventing the subjective process and bias using traditional notational observation. Players most likely to end up in the English Premier League averaged the fewest unsuccessful first-time passes, had a higher mean number of possessions and averaged more passes to teammates in the penalty area.

### 2.7.6 An examination of expected goals and shot efficiency in soccer

In recent years, an expected goal value (xG) model has been developed to evaluate the offensive performance of players and teams [23]. The xG model assigns a value between 0 and 1 (with 1 being the maximum and representing a certain goal) to every attempt based on the quantity and quality (i.e. assist type, shot angle, and distance from goal, whether it was a headed shot, etc.) of shots taken. A variety of these models have proven valuable in predicting shooting outcomes and scouting for players with high conversion rates.

### 2.7.7 Automatic discovery of tactics in Spatio-temporal soccer match data

This research assesses the performance of a team by observing its behavior on the pitch as captured by football data extracted from games [24], utilizing a data-driven approach in contrast to what has been done historically of depending on past teams' performances. By modeling each football team as a complex system and infer a network whose nodes are players or zones on the pitch, and edges are movements of the ball between two nodes, and describing the performance of a team using three measurements: the team in a game, the variance of the degree of a network's nodes, a proxy for the volume of play expressed by the diversity of play.

### 2.7.8 Wide open spaces: A statistical technique for measuring space creation in professional soccer

Researchers in this paper utilize network theory to describe the strategy of football teams [25]. Using passing data during the FIFA 2010 World Cup, Researchers construct for each team a weighted and directed network in which nodes correspond to players and arrows to pass. The resulting network or graph provides a direct visual inspection of a team's strategy, from which researchers identify play patterns, determine hot-spots on the play, and localize potential weaknesses. Using different centrality measures, They can also determine the relative importance of each player in the game, the 'popularity' of a player, and the effect of removing players from the game .

# Chapter 3

# Methodology

This chapter is a discussion about the datasets used in this thesis; their usage, size, and partitioning. It also discusses the approach used to identify the tactical patterns of football teams in our project.

## 3.1 Datasets

To train the various systems in this thesis, an accurate dataset was required, i.e, a dataset that contains soccer-logs for different football matches for different leagues. So we settled on a dataset [1] that refers to season 2017/2018 of five national soccer competitions in Europe: Spanish first division, Italian first division, English first division, German first division, French first division. These competitions are the most important in Europe according to the UEFA country coefficient [26], which is used to rank the football associations of Europe and thus determine the number of clubs from an association that will participate in the UEFA Champions League and the UEFA Europa League. In total, five datasets corresponding to information about all competitions, matches, teams, players, events, referees, and coaches.

## 3.2 Tactics identification approach

The dataset we are using in this project provides data about matches that occurred in five major competitions. The data includes passes completed with the name of the pass's sender and his position and pass's recipient position but missing the recipient's name. To avoid this problem, we ignored the players and divided the football pitch into 5x3 zones as shown in figure 3.1 and assigned each pass position to its respective zone, and then we create the passing network based on the zone the ball was passed from to the zone, it was received, neglecting the players' names. This approach will help in generalizing the results for all teams in different leagues without the need to specify each teams' formation. A

workflow of those approaches is depicted in figure 3.2 below and a further explanation is given in the next sections.

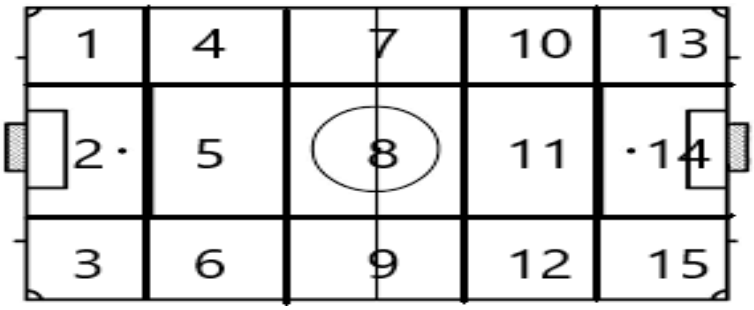



Figure 3.1: Football pitch divided into 15 zones

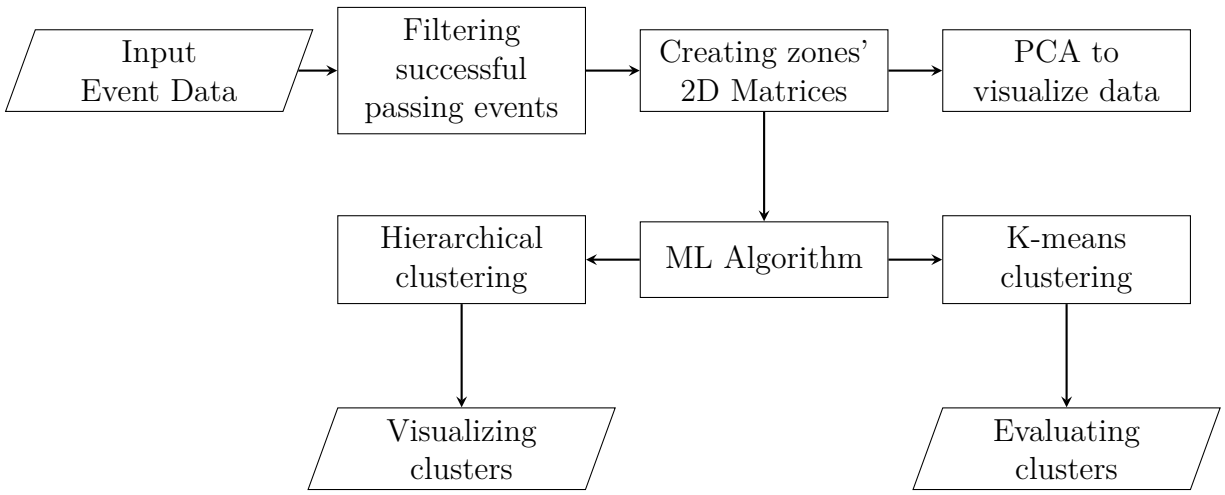



Figure 3.2: System workflow for clustering approaches

### 3.2.1 Input event data

The input (figure 3.3) is the events dataset describes all the events that occur during each match. Each event document contains the following information:

- **eventId:** the identifier of the event's type. Each eventId is associated with an event name (see next point);
- **eventName:** the name of the event's type. There are seven types of events (see Table 2): pass, foul, shot, duel, free kick, offside and touch;
- **subEventId:** the identifier of the subevent's type. Each subEventId is associated with a subevent name (see next point);

- **subEventName:** the name of the subevent's type. Each event type is associated with a different set of subevent types
- **tags:** a list of event tags, each describing additional information about the event (e.g., accurate). Each event type is associated with a different set of tags (see Table 2).
- **eventSec:** the time when the event occurs (in seconds since the beginning of the current half of the match);
- **id:** a unique identifier of the event
- **matchId:** the identifier of the match the event refers to. The identifier refers to the field "wyId" in a match document;
- **matchPeriod:** the period of the match. It can be "1H" (first half of the match), "2H" (second half of the match), "E1" (first extra time), "E2" (second extra time) or "P" (penalties time)
- **playerId:** the identifier of the player who generated the event. The identifier refers to the field "wyId" in a player document
- **positions:** the origin and destination positions associated with the event. Each position is a pair of coordinates (x, y). The x and y coordinates are always in the range [0, 100] and indicate the percentage of the field from the perspective of the attacking team. In particular, the value of the x coordinate indicates the event's nearness (in percentage) to the opponent's goal, while the value of the y coordinates indicates the event's nearness (in percentage) to the right side of the field;
- **teamId:** the identifier of the player's team. The identifier refers to the field "wyId" in a team document.

```
{"eventId": 8,
 "eventName": "Pass",
 "eventSec": 2.4175,
 "id": 253668302,
 "matchId": 2576335,
 "matchPeriod": "1H",
 "playerId": 3344,
 "positions":
 [{"x": 49, "y": 50}, {"x": 38, "y": 58}],
 "subEventId": 85,
 "subEventName": "Simple pass",
 "tags":
 [{"id": 1801}],
 "teamId": 3161}
```


Figure 3.3: shows an example of pass event (“eventId”: 8, “eventName”: “Pass”) generated by player 3344 (“playerId”: 3344) of team 3161 (“teamId”: 3161) in match 2576335 (“matchId”: 2576335) at second 2.41 of the first half of the match (“eventSec”: 2.4175, “matchPeriod”: “1H”). The pass started at position (49, 50) of the field and ended at position (38, 58) of the field (see field “positions”). Moreover, the pass was accurate as indicated by the presence of tag 1801 (field “tags”).

### 3.2.2 Filtering passing events

After reading the input data, filtration is done on the raw event data to consider only the completed passing events including their sending/receiving positions.

### 3.2.3 Creating zones’ passing network

The zone’s passing network was created by dividing the football pitch into 15 zones (22 meters x 21 meters) depicted in figure 3.1. Then, we create 2D Matrices for every team in every league during every match to contain different passes made between different zones.

### 3.2.4 Principal Component Analysis

we utilized Principal Component Analysis to visualize the data points into a two dimensional graph.

### 3.2.5 Machine Learning algorithms

k-means and Hierarchical clustering algorithms were utilized as an unsupervised machine learning algorithms to cluster the different passing matrices. For hierarchical clustering we get the number of clusters from the created dendrogram. While for k-means, we chose the number of clusters to be 3, as it was the average number of clusters Hierarchical clustering produced, and to try narrow down teams’ playing style into few clusters to be able to evaluate them. Our analysis is done for every league separately, and we compare all leagues collectively at the end.

### 3.2.6 Clustering evaluation metrics

The evaluation metrics that are used to compare the different resulted clusters to search for differences in teams' tactical patterns:

- **Overall number of passes.**
- **The hot zones:** The football field is divided into 5 vertical zones. and the number of passes received in each vertical zone is calculated to decide the hot zones.
- **The dependence on short/long balls for attacking:** The percentage of the balls sent from the first 0-60%, and arrive at the front 80-100% of the field from the perspective of the attacking team.

# Chapter 4

# Results

This chapter discusses the results of running the k-means and hierarchical clustering algorithms to identify the tactical patterns of football teams. Two case studies for two teams (Real Madrid from La Liga and Liverpool from Premier League) to analyze their specific tactical behavior are included.

## 4.1 Premier League

### 4.1.1 Principal Component analysis

We use principal component analysis to visualize teams' 2D passing matrices into a two-dimensional graph, in figure 4.1.

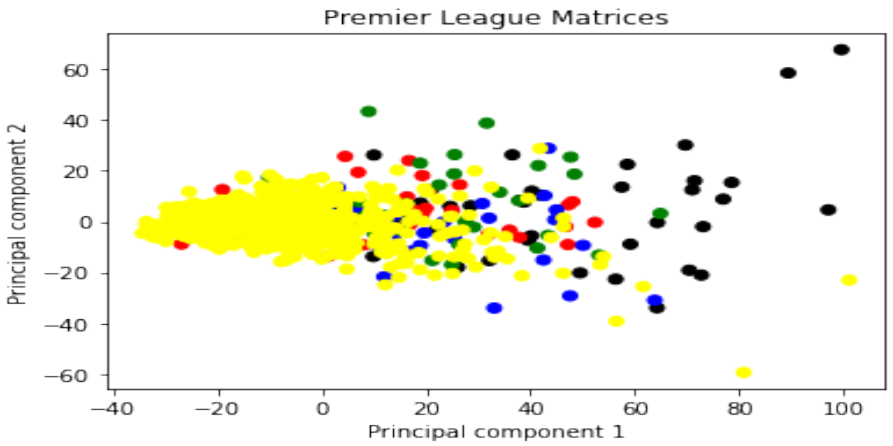
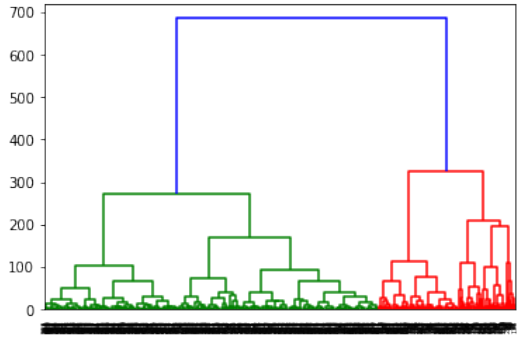


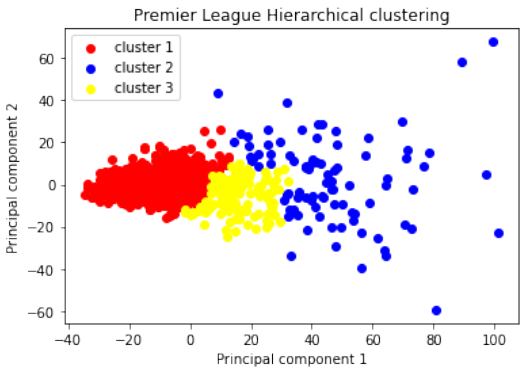

Figure 4.1: Premier League's 2D graph where each data point represent a passing network for a team during 2017/2018 season. Each team has 38 points corresponding to 38 matches. Points in Black represents Manchester City's passing matrices, Red for Manchester United, Blue for Tottenham, Green for Liverpool, and Yellow for the rest of the teams.

### 4.1.2 Hierarchical clustering

This subsection presents the results of running the hierarchical clustering algorithm on the Premier league test-set in figure 4.2 and table 4.1. We choose the number of clusters to be 3 according to the resulted dendrogram.




(a) Premier League's dendrogram




(b) Premier League's clusters created after running Hierarchical clustering algorithm on the data points.

Figure 4.2: Premier League's dendrogram and Hierarchical clustering

Table 4.1: shows clustering of Premier league teams using hierarchical clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Manchester City | 3 | 27 | 8 |
| Manchester United | 15 | 12 | 11 |
| Tottenham Hotspur | 12 | 9 | 17 |
| Liverpool | 11 | 15 | 12 |
| Chelsea | 17 | 6 | 15 |
| Arsenal | 12 | 14 | 12 |
| Burnley | 38 | 0 | 0 |
| Everton | 37 | 0 | 1 |
| Leicester City | 34 | 0 | 4 |
| Newcastle United | 35 | 1 | 2 |
| Crystal Palace | 33 | 0 | 5 |
| Bournemouth | 30 | 2 | 6 |
| West Ham United | 35 | 0 | 3 |
| Watford | 32 | 0 | 6 |
| Brighton | 37 | 0 | 1 |
| Huddersfield Town | 32 | 1 | 5 |
| Southampton | 26 | 3 | 9 |
| Swansea City | 32 | 0 | 6 |
| Stoke City | 34 | 0 | 4 |
| West Bromwich Albion | 35 | 0 | 3 |

### 4.1.3 k-means clustering

This subsection presents the results of running the k-means clustering algorithm on the Premier league test-set. Table 4.2 below shows the clustering of the Premier league teams' matches into 3 clusters. Then, we evaluate the clusters from the results, in the following table 4.6.

Table 4.2: shows clustering of premier leagues teams using k-means.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Manchester City | 12 | 26 | 0 |
| Manchester United | 21 | 8 | 9 |
| Tottenham Hotspur | 21 | 11 | 6 |
| Liverpool | 23 | 8 | 7 |
| Chelsea | 26 | 4 | 8 |
| Arsenal | 23 | 11 | 4 |
| Burnley | 1 | 0 | 37 |
| Everton | 9 | 0 | 29 |
| Leicester City | 11 | 0 | 27 |
| Newcastle United | 9 | 0 | 29 |
| Crystal Palace | 8 | 0 | 30 |
| Bournemouth | 11 | 2 | 25 |
| West Ham United | 9 | 0 | 29 |
| Watford | 8 | 0 | 30 |
| Brighton | 5 | 0 | 33 |
| Huddersfield Town | 10 | 1 | 27 |
| Southampton | 17 | 1 | 20 |
| Swansea City | 16 | 0 | 22 |
| Stoke City | 5 | 0 | 33 |
| West Bromwich Albion | 9 | 0 | 29 |

**Cluster 1**

Table 4.3: 2D Matrix showing the cluster 1 matches' passes of Premier League where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.4 | 1.2 | 0.1 | 2.4 | 0.6 | 0.2 | 0.7 | 0.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.2 | 1.3 | 1.3 | 3.7 | 3.2 | 3.2 | 0.8 | 0.7 | 0.8 | 0.6 | 0.6 | 0.5 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.4 | 1.3 | 0.3 | 0.7 | 2.4 | 0.0 | 0.2 | 0.7 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.6 | 2.7 | 0.1 | 13.1 | 7.8 | 2.0 | 8.5 | 2.5 | 0.7 | 2.4 | 0.4 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.6 | 0.3 | 8.1 | 9.0 | 7.9 | 4.3 | 4.8 | 4.6 | 1.0 | 0.6 | 0.9 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.9 | 1.5 | 1.8 | 7.6 | 13.0 | 0.6 | 2.8 | 8.1 | 0.2 | 0.6 | 2.5 | 0.0 | 0.0 | 0.3 |
| 7 | 0.1 | 0.3 | 0.0 | 5.1 | 2.7 | 0.5 | 11.0 | 5.6 | 1.3 | 9.7 | 2.0 | 0.8 | 0.6 | 0.2 | 0.3 |
| 8 | 0.0 | 0.2 | 0.0 | 1.6 | 2.9 | 1.4 | 5.9 | 7.9 | 5.6 | 4.5 | 4.0 | 4.7 | 0.6 | 0.2 | 0.5 |
| 9 | 0.0 | 0.2 | 0.1 | 0.3 | 2.8 | 5.4 | 1.1 | 5.8 | 11.3 | 0.7 | 2.2 | 9.4 | 0.2 | 0.2 | 0.9 |
| 10 | 0.0 | 0.0 | 0.0 | 0.8 | 0.2 | 0.1 | 5.8 | 2.2 | 0.2 | 22.4 | 6.7 | 1.0 | 6.6 | 1.4 | 0.6 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.9 | 1.8 | 0.9 | 5.7 | 9.8 | 5.8 | 2.3 | 1.8 | 2.4 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.8 | 0.3 | 2.4 | 6.2 | 0.9 | 6.4 | 21.4 | 0.6 | 1.7 | 6.6 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 3.1 | 0.6 | 0.1 | 4.3 | 2.3 | 0.4 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 0.4 | 1.2 | 0.4 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.5 | 3.1 | 0.3 | 2.4 | 3.8 |

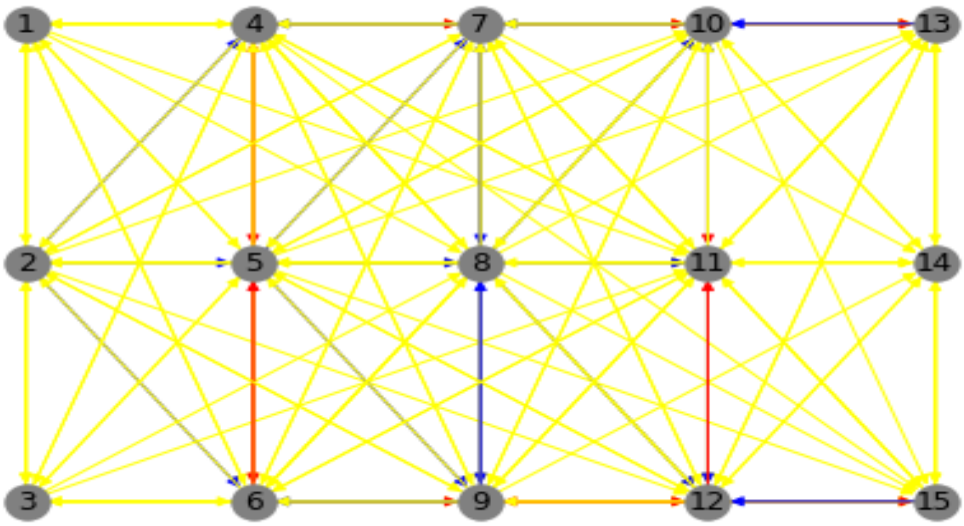



Figure 4.3: Cluster 1 average zones' passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

**Cluster 2**

Table 4.4: 2D Matrix showing the average passes of cluster 2 matches' in the Premier League where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.3 | 1.6 | 0.2 | 2.7 | 1.0 | 0.5 | 0.3 | 0.2 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.3 | 1.4 | 1.4 | 3.7 | 3.9 | 3.7 | 0.7 | 0.4 | 0.6 | 0.2 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 |
| 3 | 0.2 | 1.5 | 1.7 | 0.5 | 0.9 | 2.3 | 0.1 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 1.6 | 3.1 | 0.1 | 13.3 | 9.4 | 2.1 | 9.8 | 4.1 | 1.4 | 2.2 | 0.3 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.2 | 2.2 | 0.2 | 8.3 | 10.3 | 7.8 | 6.1 | 6.9 | 6.1 | 1.1 | 0.4 | 1.0 | 0.0 | 0.1 | 0.0 |
| 6 | 0.1 | 2.6 | 1.4 | 2.6 | 8.0 | 11.3 | 1.0 | 3.8 | 8.5 | 0.2 | 0.4 | 1.6 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 6.5 | 3.6 | 0.4 | 20.6 | 13.5 | 2.3 | 17.7 | 3.9 | 1.2 | 0.9 | 0.1 | 0.3 |
| 8 | 0.0 | 0.2 | 0.0 | 1.7 | 3.5 | 1.5 | 12.6 | 15.5 | 11.1 | 9.5 | 7.5 | 10.0 | 0.5 | 0.2 | 0.6 |
| 9 | 0.0 | 0.3 | 0.1 | 0.5 | 3.9 | 5.9 | 2.4 | 12.7 | 15.9 | 1.2 | 3.4 | 14.2 | 0.4 | 0.1 | 0.7 |
| 10 | 0.0 | 0.0 | 0.0 | 0.7 | 0.2 | 0.0 | 11.8 | 4.7 | 0.5 | 45.2 | 13.1 | 2.3 | 11.5 | 1.7 | 0.9 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 1.7 | 3.9 | 1.6 | 11.9 | 20.3 | 11.2 | 3.0 | 3.3 | 3.7 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.8 | 0.6 | 4.6 | 10.5 | 2.0 | 13.1 | 39.4 | 0.9 | 2.1 | 10.6 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 5.7 | 1.1 | 0.1 | 6.4 | 3.1 | 0.4 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.7 | 0.2 | 0.6 | 1.9 | 0.6 |
| 15 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.7 | 5.8 | 0.4 | 3.3 | 5.3 |

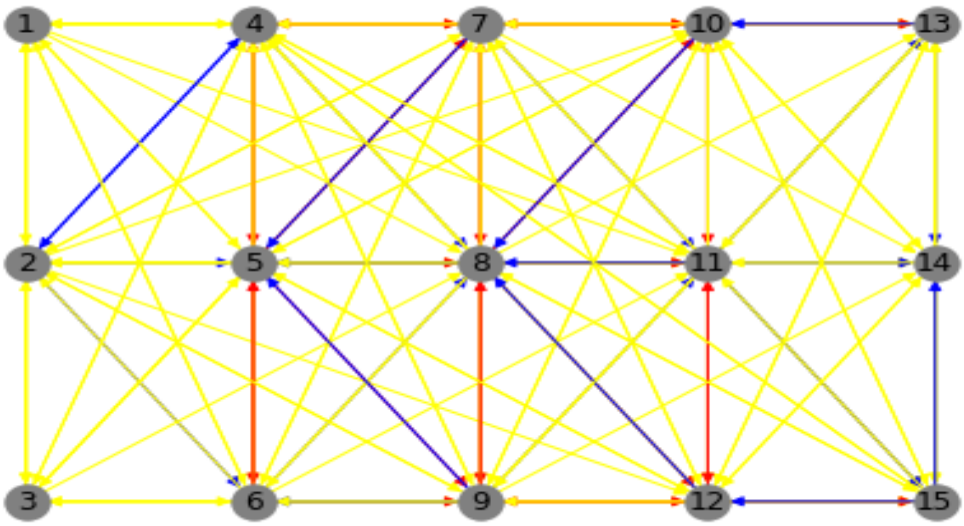



Figure 4.4: Cluster 2 average zones' passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

**Cluster 3**

Table 4.5: 2D Matrix showing the average passes of cluster 3 matches' in the Premier League where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.3 | 0.8 | 0.1 | 2.0 | 0.3 | 0.1 | 0.7 | 0.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 0.8 | 1.2 | 0.7 | 2.3 | 1.9 | 2.2 | 0.8 | 0.8 | 0.9 | 1.0 | 1.1 | 1.1 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 0.7 | 1.1 | 0.1 | 0.4 | 1.9 | 0.0 | 0.2 | 0.8 | 0.0 | 0.1 | 0.4 | 0.0 | 0.0 | 0.0 |
| 4 | 1.2 | 2.2 | 0.1 | 9.9 | 4.2 | 1.2 | 5.9 | 1.5 | 0.3 | 2.0 | 0.4 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.1 | 0.3 | 5.1 | 5.7 | 4.6 | 2.4 | 2.9 | 2.3 | 1.0 | 0.6 | 0.9 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.0 | 1.4 | 1.0 | 4.0 | 9.3 | 0.3 | 1.5 | 5.5 | 0.2 | 0.5 | 2.1 | 0.0 | 0.0 | 0.3 |
| 7 | 0.1 | 0.2 | 0.0 | 3.9 | 1.7 | 0.3 | 6.9 | 2.4 | 0.5 | 6.1 | 0.9 | 0.4 | 0.6 | 0.1 | 0.2 |
| 8 | 0.0 | 0.2 | 0.0 | 1.2 | 2.4 | 1.0 | 3.0 | 4.2 | 2.9 | 2.2 | 2.0 | 2.0 | 0.4 | 0.1 | 0.3 |
| 9 | 0.0 | 0.1 | 0.1 | 0.2 | 1.5 | 3.6 | 0.5 | 2.5 | 6.4 | 0.4 | 1.0 | 5.6 | 0.2 | 0.1 | 0.7 |
| 10 | 0.0 | 0.0 | 0.0 | 0.5 | 0.1 | 0.0 | 3.7 | 1.0 | 0.1 | 11.6 | 2.6 | 0.5 | 4.2 | 0.9 | 0.4 |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.5 | 1.0 | 0.4 | 2.5 | 3.7 | 2.4 | 1.0 | 0.9 | 1.1 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 1.1 | 3.4 | 0.4 | 2.5 | 11.2 | 0.4 | 1.0 | 3.9 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.7 | 0.3 | 0.1 | 2.8 | 1.5 | 0.2 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.2 | 0.7 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.5 | 0.2 | 1.6 | 2.6 |

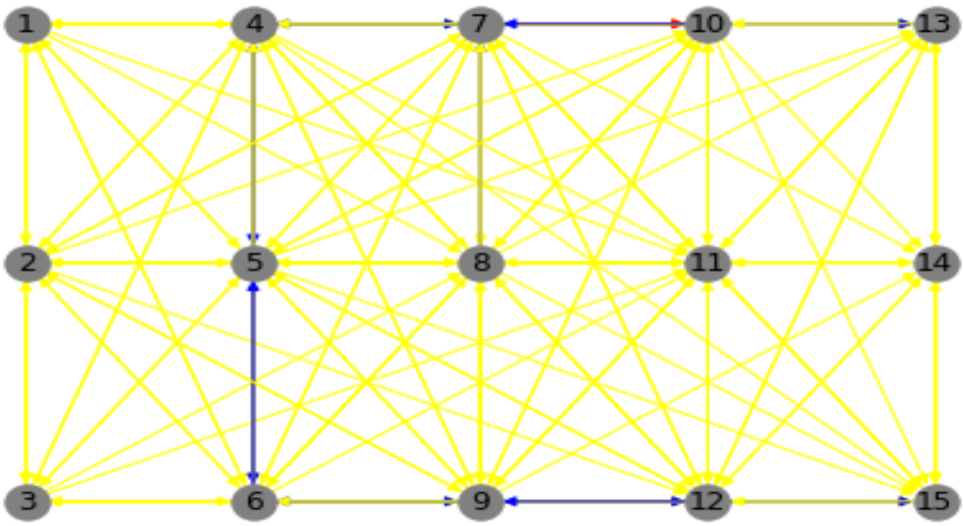



Figure 4.5: Cluster 3 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### 4.1.4 Examining results

From the shown below results from table 4.6 we can deduce that most of the top 6 teams' matches belong to cluster 1 and 2 which are characterized by the high number of passes, the domination of the fourth area of the football pitch (cluster 2), and the dependence on short balls for delivering the ball into the attacking area. We also notice that the league's winner (Manchester City) had the most matches in cluster 2.

Table 4.6: shows Premier League evaluation metrics results

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Average number of passes | 437 | 659 | 260.8 |
| Overall matches | 33.5% | 9.5% | 57% |
| Top 6 teams | 55.3% | 29.9% | 14.8% |
| Passes received in the first 20 percent of the field | 22.5 (5.1%) | 23 (3.5%) | 17.3(6.8%) |
| Passes received in the second 20 percent of the field | 113.2 (25.9%) | 121.9 (18.5%) | 73.4(28.8%) |
| Passes received in the third 20 percent of the field | 118.4 (27%) | 197.7 (30%) | 67.8(26.6%) |
| Passes received in the fourth 20 percent of the field | 139 (31.8%) | 251.5 (38.1%) | 74.6(29.2%) |
| Passes received in the last 20 percent of the field | 44 (10%) | 64.9 (9.9%) | 27.6(10.8%) |
| Percentage of the long balls received in the attacking area from the total passes | 7.4% | 5% | 9.05% |

## 4.2 Serie A

### 4.2.1 Principal Component analysis

We use principal component analysis to visualize teams' 2D passing matrices into a two-dimensional graph in figure 4.6.

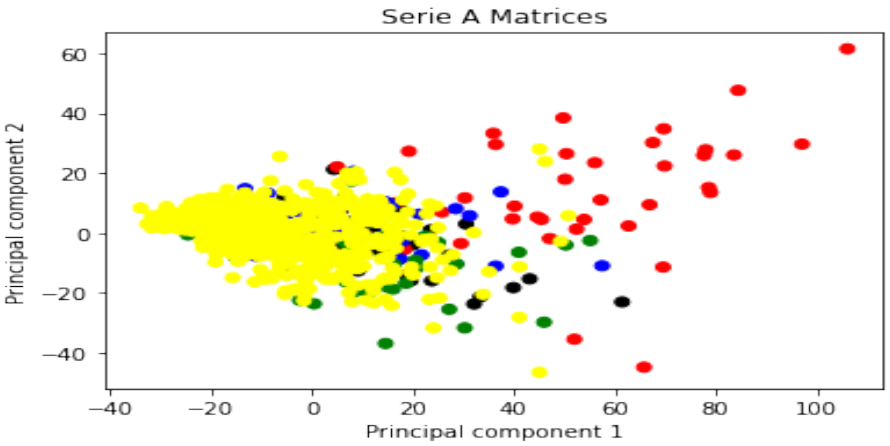
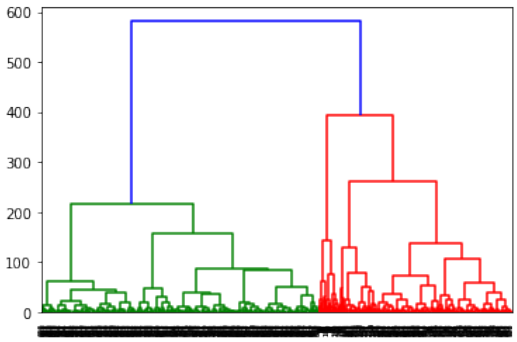


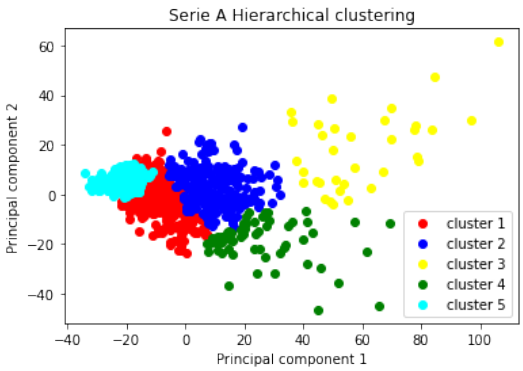

Figure 4.6: Serie A's 2D graph where each data point represent a passing network for a team during 2017/2018 season. Each team has 38 points corresponding to 38 matches. Points in black represents Juventus's passing matrices, Red for Napoli, Blue for Roma, Green for internazionale, and Yellow for the rest of the teams.

### 4.2.2 Hierarchical clustering

This subsection presents the results of running the hierarchical clustering algorithm on the Serie A's test-set in figure 4.7 and table 4.7. We choose the number of clusters to be 5 according to the resulted dendrogram.




(a) Serie A's dendrogram




(b) Serie A's clusters created after running Hierarchical clustering algorithm on the data points.

Figure 4.7: Serie A's dendrogram and hierarchical clustering

Table 4.7: shows clustering of Serie A teams using heirarchical clustering

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| Juventus | 6 | 21 | 0 | 9 | 2 |
| Napoli | 0 | 8 | 27 | 3 | 0 |
| Roma | 13 | 21 | 1 | 3 | 0 |
| Internazionale | 8 | 12 | 2 | 14 | 2 |
| Lazio | 7 | 23 | 2 | 0 | 6 |
| Milan | 14 | 14 | 1 | 9 | 0 |
| Atalanta | 7 | 25 | 1 | 3 | 2 |
| Fiorentina | 20 | 6 | 0 | 6 | 6 |
| Torino | 12 | 16 | 0 | 1 | 9 |
| Sampdoria | 16 | 14 | 0 | 5 | 3 |
| Sassuolo | 22 | 3 | 0 | 0 | 13 |
| Genoa | 19 | 10 | 0 | 1 | 8 |
| Chievo | 23 | 6 | 0 | 0 | 9 |
| Udinese | 22 | 8 | 0 | 0 | 8 |
| Bologna | 18 | 11 | 0 | 1 | 8 |
| Cagliari | 17 | 6 | 0 | 0 | 15 |
| SPAL | 24 | 4 | 0 | 1 | 9 |
| Crotone | 11 | 5 | 0 | 0 | 22 |
| Hellas Verona | 16 | 3 | 0 | 0 | 19 |
| Benevento | 19 | 4 | 0 | 2 | 13 |

### 4.2.3 k-means clustering

This subsection presents the results of running the k-means clustering algorithm on the Serie A test-set. Table 4.8 below shows the clustering of the Serie A' matches into 3 clusters. Then, we evaluate the clusters in table 4.12.

Table 4.8: shows clustering of Serie A teams using k-means clustering

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Juventus | 6 | 29 | 3 |
| Napoli | 0 | 8 | 30 |
| Roma | 11 | 23 | 4 |
| Internazionale | 4 | 30 | 4 |
| Lazio | 15 | 21 | 2 |
| Milan | 7 | 28 | 3 |
| Atalanta | 9 | 28 | 1 |
| Fiorentina | 18 | 19 | 1 |
| Torino | 22 | 16 | 0 |
| Sampdoria | 15 | 23 | 0 |
| Sassuolo | 33 | 5 | 0 |
| Genoa | 25 | 13 | 0 |
| Chievo | 27 | 11 | 0 |
| Udinese | 29 | 9 | 0 |
| Bologna | 26 | 12 | 0 |
| Cagliari | 31 | 7 | 0 |
| SPAL | 30 | 8 | 0 |
| Crotone | 33 | 5 | 0 |
| Hellas Verona | 32 | 6 | 0 |
| Benevento | 26 | 11 | 1 |

### Cluster 1

Table 4.9: 2D Matrix showing the average passes of cluster 1 teams' in the Serie A where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.8 | 1.2 | 0.1 | 2.7 | 0.5 | 0.2 | 0.6 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.4 | 1.4 | 1.3 | 3.2 | 2.5 | 3.0 | 0.9 | 0.8 | 0.9 | 0.6 | 0.6 | 0.4 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.2 | 1.8 | 0.2 | 0.5 | 2.8 | 0.1 | 0.2 | 0.8 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.6 | 2.3 | 0.1 | 11.4 | 5.0 | 1.4 | 6.3 | 1.7 | 0.4 | 2.0 | 0.4 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.1 | 0.3 | 5.8 | 6.8 | 5.5 | 2.7 | 3.3 | 2.7 | 1.0 | 0.6 | 0.8 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.2 | 1.6 | 1.3 | 4.6 | 10.9 | 0.4 | 1.6 | 6.4 | 0.2 | 0.6 | 1.9 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 4.1 | 1.6 | 0.3 | 6.9 | 2.8 | 0.6 | 5.9 | 1.2 | 0.4 | 0.7 | 0.1 | 0.2 |
| 8 | 0.0 | 0.1 | 0.0 | 1.1 | 2.7 | 1.1 | 3.2 | 5.1 | 3.2 | 2.3 | 2.0 | 2.3 | 0.4 | 0.1 | 0.4 |
| 9 | 0.0 | 0.1 | 0.1 | 0.2 | 1.6 | 3.5 | 0.5 | 2.7 | 6.8 | 0.4 | 1.3 | 6.0 | 0.2 | 0.1 | 0.6 |
| 10 | 0.0 | 0.0 | 0.0 | 0.5 | 0.1 | 0.0 | 3.4 | 0.9 | 0.1 | 10.4 | 2.7 | 0.5 | 3.9 | 0.9 | 0.4 |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.5 | 1.0 | 0.5 | 2.6 | 4.1 | 2.6 | 1.1 | 1.0 | 1.2 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.4 | 0.1 | 1.0 | 3.3 | 0.5 | 3.0 | 10.7 | 0.4 | 1.1 | 3.8 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.5 | 0.3 | 0.1 | 2.4 | 1.6 | 0.3 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.1 | 0.2 | 0.6 | 0.1 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 1.5 | 0.2 | 1.5 | 2.3 |

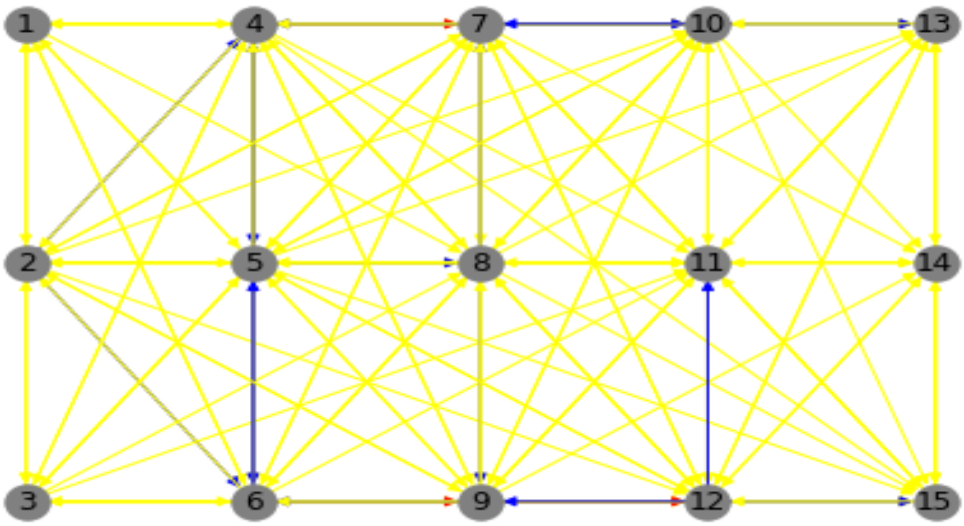



Figure 4.8: Cluster 1 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

**Cluster 2**

Table 4.10: 2D Matrix showing the average passes of cluster 2 teams' in the Serie A where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.7 | 1.4 | 0.1 | 2.9 | 0.7 | 0.3 | 0.6 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.5 | 1.7 | 1.5 | 3.9 | 3.5 | 3.8 | 0.7 | 0.7 | 0.6 | 0.2 | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.5 | 1.8 | 0.3 | 0.7 | 3.0 | 0.1 | 0.2 | 0.6 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| 4 | 1.5 | 2.5 | 0.1 | 13.6 | 7.9 | 2.1 | 8.6 | 2.9 | 0.8 | 2.3 | 0.4 | 0.2 | 0.2 | 0.0 | 0.0 |
| 5 | 0.3 | 2.4 | 0.4 | 8.3 | 9.3 | 7.8 | 4.9 | 5.1 | 4.6 | 1.0 | 0.5 | 0.9 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.5 | 1.7 | 2.1 | 7.6 | 13.7 | 0.7 | 2.7 | 8.6 | 0.3 | 0.5 | 2.2 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 5.5 | 2.9 | 0.4 | 11.9 | 6.5 | 1.3 | 11.0 | 2.1 | 0.7 | 1.0 | 0.1 | 0.2 |
| 8 | 0.0 | 0.1 | 0.0 | 1.4 | 3.3 | 1.4 | 6.6 | 10.2 | 6.6 | 5.3 | 4.2 | 5.1 | 0.5 | 0.2 | 0.6 |
| 9 | 0.0 | 0.2 | 0.1 | 0.5 | 2.9 | 4.8 | 1.3 | 6.1 | 11.6 | 0.7 | 2.1 | 10.5 | 0.2 | 0.2 | 1.0 |
| 10 | 0.0 | 0.0 | 0.0 | 0.7 | 0.3 | 0.0 | 6.8 | 2.5 | 0.3 | 21.4 | 5.9 | 1.1 | 6.1 | 1.7 | 0.6 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.9 | 2.2 | 0.9 | 5.1 | 8.7 | 5.1 | 1.8 | 1.7 | 2.2 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.7 | 0.3 | 2.2 | 5.9 | 1.1 | 5.7 | 21.0 | 0.6 | 1.8 | 6.6 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 2.5 | 0.5 | 0.1 | 3.3 | 2.3 | 0.3 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.1 | 0.3 | 1.1 | 0.3 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 2.6 | 0.4 | 2.5 | 3.6 |

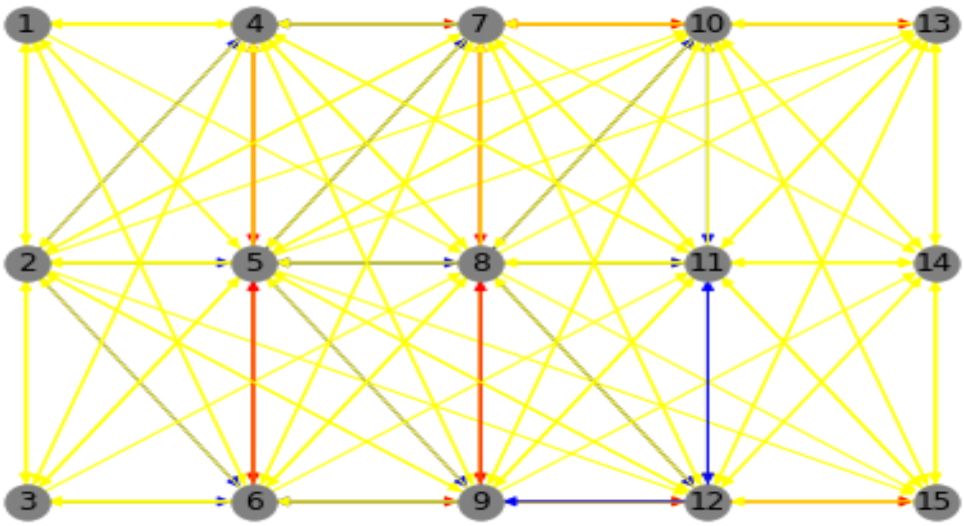



Figure 4.9: Cluster 2's average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### Cluster 3

Table 4.11: 2D Matrix showing the average passes of cluster 3 teams' in the Serie A where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.6 | 1.2 | 0.1 | 3.1 | 0.6 | 0.2 | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.4 | 1.2 | 1.0 | 4.2 | 3.5 | 3.4 | 0.5 | 0.3 | 0.7 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| 3 | 0.2 | 1.2 | 2.1 | 0.6 | 0.7 | 2.5 | 0.1 | 0.1 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 1.7 | 3.2 | 0.1 | 18.7 | 7.9 | 1.7 | 12.4 | 4.1 | 0.9 | 2.6 | 0.4 | 0.1 | 0.2 | 0.0 | 0.0 |
| 5 | 0.3 | 2.0 | 0.3 | 8.6 | 8.9 | 5.6 | 6.2 | 7.4 | 4.1 | 1.1 | 0.4 | 0.9 | 0.1 | 0.0 | 0.0 |
| 6 | 0.1 | 2.6 | 1.8 | 2.3 | 7.7 | 12.8 | 0.9 | 3.3 | 7.6 | 0.1 | 0.4 | 1.9 | 0.0 | 0.0 | 0.1 |
| 7 | 0.1 | 0.4 | 0.0 | 8.0 | 3.0 | 0.2 | 22.9 | 11.3 | 1.6 | 23.5 | 3.6 | 0.9 | 1.3 | 0.1 | 0.2 |
| 8 | 0.0 | 0.2 | 0.0 | 1.9 | 3.0 | 1.3 | 11.1 | 16.4 | 7.3 | 10.8 | 8.4 | 7.9 | 0.7 | 0.1 | 0.4 |
| 9 | 0.0 | 0.3 | 0.1 | 0.4 | 2.8 | 5.5 | 2.0 | 9.5 | 12.2 | 0.8 | 3.3 | 12.1 | 0.2 | 0.1 | 0.9 |
| 10 | 0.1 | 0.0 | 0.0 | 1.6 | 0.2 | 0.0 | 16.5 | 4.7 | 0.4 | 57.5 | 12.6 | 1.7 | 11.5 | 2.2 | 0.8 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 1.8 | 4.1 | 1.5 | 10.5 | 16.3 | 9.3 | 3.3 | 2.1 | 2.7 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.8 | 0.3 | 4.4 | 9.1 | 1.4 | 10.5 | 35.2 | 0.6 | 2.0 | 8.1 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.0 | 5.0 | 0.9 | 0.1 | 5.3 | 3.1 | 0.4 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 0.4 | 1.3 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.5 | 4.3 | 0.5 | 2.3 | 3.8 |

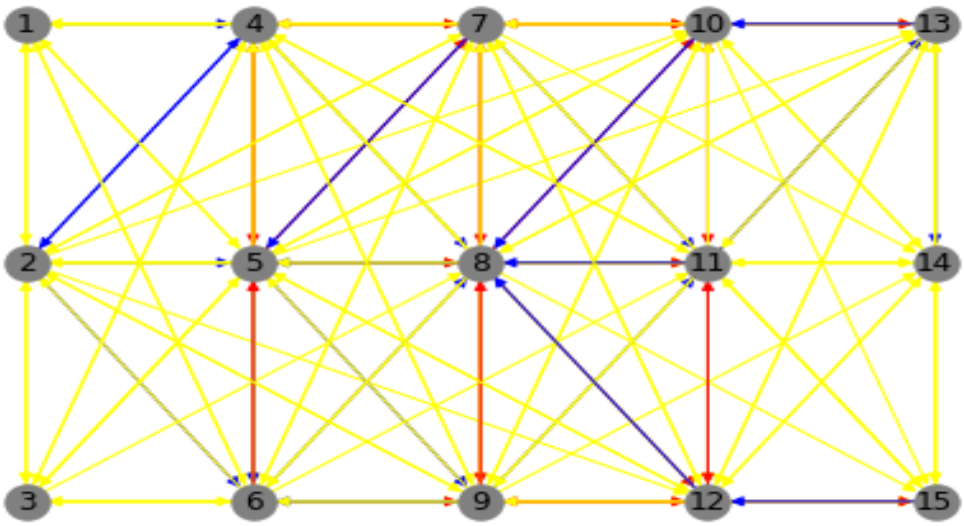



Figure 4.10: Cluster 3 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### 4.2.4 Examining results

From the shown below results in table 4.12 we can deduce that similarly to the Premier League, as we go higher in the league table, the number of matches that belong to the clusters (2, 3) with the higher number of overall passes and passes received in the fourth football field increase. While the number of long balls used for attacking decreases. As a special case "Napoli" team matches dominated cluster three with their special style of play of an increased average total number of passes, with the highest percentage of passes in the middle pitch area, their playstyle in the season can be considered by many to follow "Tiki-taka" football play style.

Table 4.12: shows Serie A evaluation metrics results

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Average total number of passes | 279 | 441 | 634 |
| Overall matches | 54% | 39.6% | 6.4% |
| Top 6 teams | 19.8% | 60.1% | 20.1% |
| Passes received in the first 20 percent of the field | 21 (7%) | 23 (5%) | 24(3%) |
| Passes received in the second 20 percent of the field | 85 (30%) | 116 (26%) | 122(19%) |
| Passes received in the third 20 percent of the field | 72 (25%) | 126 (28%) | 186(19%) |
| Passes received in the fourth 20 percent of the field | 72 (26%) | 132 (30%) | 245(38%) |
| Passes received in the last 20 percent of the field | 26 (9%) | 41 (9%) | 55(8%) |
| Percentage of the long balls received in the attacking area from the total passes | 9.2% | 7.65% | 5.8% |

## 4.3 Ligue 1

### 4.3.1 Principal Component analysis

We use principal component analysis to visualize teams' 2D passing matrices into a two-dimensional graph in figure 4.11.

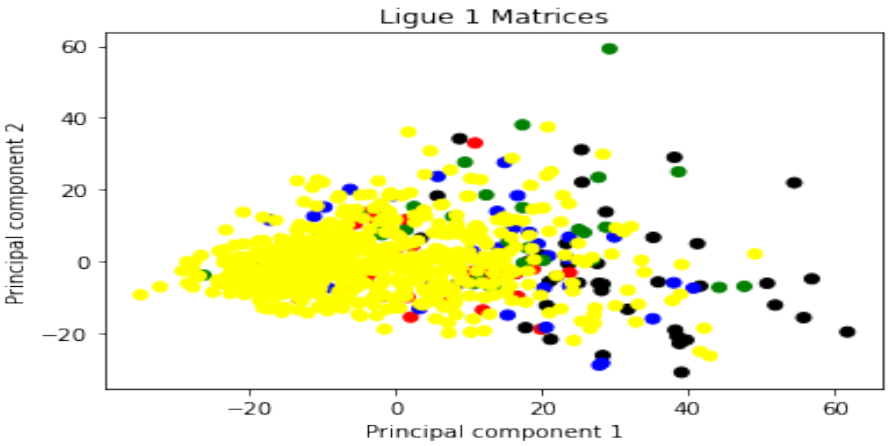
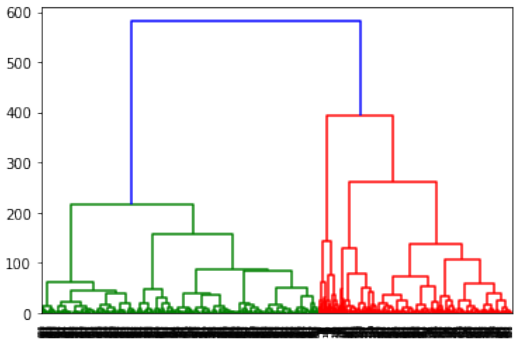


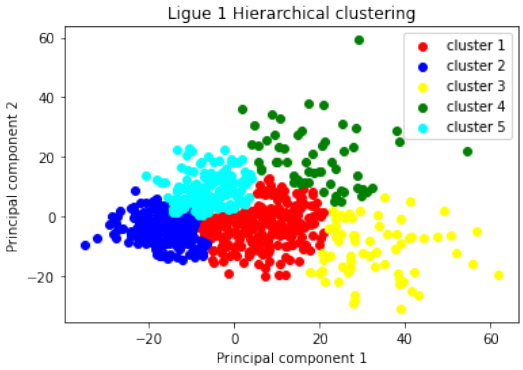

Figure 4.11: Ligue 1's 2D graph where each data point represent a passing network for a team during 2017/2018 season. Each team has 38 points corresponding to 38 matches. Points in Black represents Paris Saint-Germain's passing matrices, Red for Monaco, blue for Lyon, Green for Marseille, and Yellow for the rest of the teams.

### 4.3.2 Hierarchical clustering

This subsection presents the results of running the hierarchical clustering algorithm on the Ligue 1's test-set in figure 4.12 and table 4.13. We choose the number of clusters to be 5 according to the resulted dendrogram.




(a) Ligue 1's dendrogram




(b) Ligue 1's clusters created after running Hierarchical clustering algorithm on the data points.

Figure 4.12: Ligue 1's dendrogram and hierarchical clustering

Table 4.13: shows clustering of Ligue 1 teams using hierarchical clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| Paris Saint-Germain | 3 | 0 | 26 | 8 | 1 |
| Monaco | 21 | 8 | 2 | 1 | 6 |
| Lyon | 13 | 4 | 8 | 6 | 7 |
| Marseille | 13 | 4 | 3 | 10 | 8 |
| Rennes | 16 | 10 | 3 | 1 | 8 |
| Bordeaux | 15 | 7 | 2 | 3 | 11 |
| Saint-Étienne | 17 | 7 | 2 | 2 | 10 |
| Nice | 9 | 0 | 9 | 15 | 5 |
| Nantes | 12 | 19 | 0 | 1 | 6 |
| Montpellier | 14 | 14 | 2 | 1 | 7 |
| Dijon | 10 | 14 | 0 | 1 | 13 |
| Guingamp | 14 | 11 | 4 | 2 | 7 |
| Amiens | 8 | 19 | 0 | 0 | 11 |
| Angers | 10 | 12 | 1 | 2 | 13 |
| Strasbourg | 10 | 13 | 0 | 0 | 15 |
| Caen | 6 | 23 | 0 | 1 | 8 |
| Lille | 18 | 3 | 5 | 3 | 9 |
| Toulouse | 11 | 26 | 0 | 0 | 1 |
| Troyes | 11 | 11 | 1 | 0 | 15 |
| Metz | 18 | 16 | 0 | 0 | 4 |

### 4.3.3 k-means clustering

This section mentions the results of running the k-means clustering algorithm on the Ligue 1 test-set. Table 4.14 below shows the clustering of the Ligue 1 teams' matches into 3 clusters. Then, we evaluate the clusters in table 4.18.

Table 4.14: shows clustering of Ligue 1 teams using k-means.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Paris Saint-Germain | 0 | 31 | 7 |
| Monaco | 18 | 10 | 10 |
| Lyon | 10 | 13 | 15 |
| Marseille | 6 | 12 | 20 |
| Rennes | 24 | 6 | 8 |
| Bordeaux | 18 | 7 | 13 |
| Saint-Étienne | 16 | 10 | 12 |
| Nice | 3 | 12 | 23 |
| Nantes | 27 | 3 | 8 |
| Montpellier | 26 | 3 | 9 |
| Dijon | 30 | 4 | 4 |
| Guingamp | 24 | 5 | 9 |
| Amiens | 32 | 0 | 6 |
| Angers | 26 | 3 | 9 |
| Strasbourg | 26 | 1 | 11 |
| Caen | 32 | 1 | 5 |
| Lille | 12 | 8 | 18 |
| Toulouse | 33 | 3 | 2 |
| Troyes | 32 | 1 | 5 |
| Metz | 27 | 4 | 7 |

**Cluster 1**

Table 4.15: 2D Matrix showing the average passes of cluster 1 teams' in Ligue 1 where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.4 | 0.8 | 0.1 | 2.4 | 0.4 | 0.2 | 0.7 | 0.1 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 0.8 | 1.0 | 0.8 | 3.2 | 2.3 | 2.8 | 0.8 | 0.6 | 0.9 | 0.5 | 0.6 | 0.7 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 0.8 | 1.4 | 0.2 | 0.4 | 2.2 | 0.0 | 0.1 | 0.8 | 0.0 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 |
| 4 | 1.3 | 2.2 | 0.1 | 10.8 | 4.3 | 1.4 | 6.8 | 1.6 | 0.4 | 2.3 | 0.4 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.2 | 1.9 | 0.2 | 4.9 | 5.4 | 4.5 | 2.7 | 3.0 | 2.4 | 0.9 | 0.5 | 0.8 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.0 | 1.2 | 1.2 | 4.0 | 9.2 | 0.4 | 1.5 | 6.5 | 0.2 | 0.5 | 2.3 | 0.0 | 0.1 | 0.3 |
| 7 | 0.1 | 0.1 | 0.0 | 3.7 | 1.6 | 0.3 | 8.1 | 2.8 | 0.6 | 6.9 | 1.1 | 0.4 | 0.8 | 0.1 | 0.2 |
| 8 | 0.0 | 0.1 | 0.0 | 1.1 | 2.1 | 1.0 | 3.2 | 4.8 | 3.1 | 2.6 | 1.9 | 2.4 | 0.4 | 0.1 | 0.5 |
| 9 | 0.0 | 0.1 | 0.1 | 0.2 | 1.4 | 3.3 | 0.4 | 2.9 | 7.6 | 0.4 | 1.2 | 6.6 | 0.1 | 0.1 | 0.8 |
| 10 | 0.0 | 0.0 | 0.0 | 0.4 | 0.1 | 0.0 | 3.5 | 0.9 | 0.1 | 12.2 | 2.9 | 0.5 | 4.3 | 0.9 | 0.3 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.5 | 1.0 | 0.4 | 2.5 | 3.8 | 2.4 | 1.3 | 0.8 | 1.2 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 0.9 | 3.0 | 0.4 | 2.7 | 11.6 | 0.3 | 1.1 | 4.5 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.5 | 0.3 | 0.1 | 2.4 | 1.6 | 0.2 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.1 | 0.6 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.4 | 0.2 | 1.8 | 2.7 |

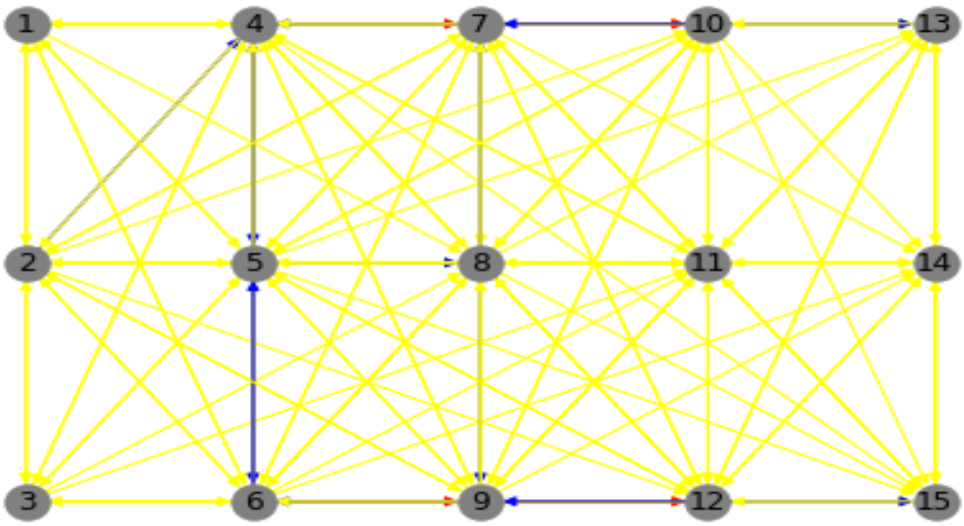



Figure 4.13: Cluster 1 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### Cluster 2

Table 4.16: 2D Matrix showing the average passes of cluster 2 teams' in Ligue 1 where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.7 | 1.4 | 0.1 | 2.6 | 0.8 | 0.4 | 0.5 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.2 | 1.2 | 1.0 | 3.3 | 3.3 | 3.5 | 0.5 | 0.5 | 0.6 | 0.1 | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.1 | 1.5 | 0.4 | 0.9 | 2.5 | 0.0 | 0.1 | 0.4 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| 4 | 1.5 | 2.1 | 0.0 | 12.5 | 8.0 | 2.3 | 9.8 | 3.2 | 1.0 | 2.2 | 0.2 | 0.1 | 0.1 | 0.0 | 0.1 |
| 5 | 0.2 | 1.7 | 0.2 | 7.2 | 9.7 | 7.6 | 5.5 | 6.2 | 5.7 | 1.0 | 0.4 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.2 | 1.3 | 2.3 | 8.3 | 12.2 | 1.0 | 3.3 | 9.3 | 0.2 | 0.4 | 1.9 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 6.1 | 3.0 | 0.6 | 16.2 | 8.2 | 1.8 | 12.9 | 2.2 | 0.7 | 1.0 | 0.1 | 0.3 |
| 8 | 0.1 | 0.1 | 0.0 | 1.6 | 3.2 | 1.6 | 7.9 | 11.8 | 8.4 | 5.5 | 4.7 | 5.9 | 0.6 | 0.2 | 0.6 |
| 9 | 0.0 | 0.2 | 0.1 | 0.5 | 3.2 | 5.7 | 1.9 | 8.3 | 16.9 | 0.8 | 2.2 | 12.7 | 0.2 | 0.2 | 1.2 |
| 10 | 0.0 | 0.0 | 0.0 | 0.5 | 0.2 | 0.1 | 7.4 | 2.4 | 0.3 | 23.5 | 6.3 | 1.1 | 7.0 | 1.4 | 0.6 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.8 | 2.2 | 1.0 | 4.9 | 10.2 | 5.6 | 2.1 | 1.5 | 2.4 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.6 | 0.3 | 2.4 | 7.2 | 0.8 | 6.4 | 26.0 | 0.6 | 1.6 | 7.9 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 2.9 | 0.4 | 0.1 | 3.6 | 2.2 | 0.3 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 0.3 | 1.1 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.4 | 2.9 | 0.4 | 2.8 | 4.3 |

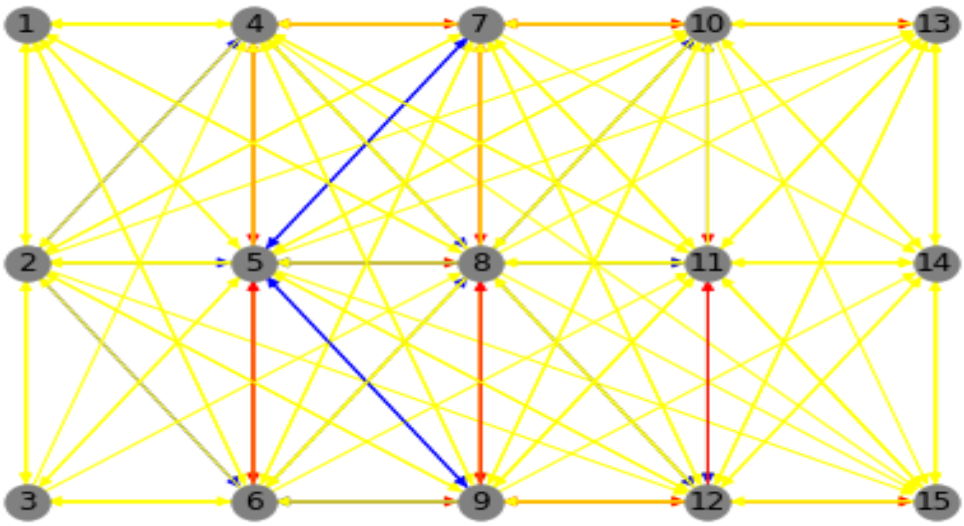



Figure 4.14: Cluster 2 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### Cluster 3

Table 4.17: 2D Matrix showing the average passes of cluster 3 teams' in Ligue 1 where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.8 | 1.6 | 0.2 | 3.3 | 0.8 | 0.3 | 0.6 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.6 | 1.4 | 1.5 | 4.1 | 3.3 | 4.3 | 0.7 | 0.5 | 0.8 | 0.3 | 0.3 | 0.5 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.4 | 2.1 | 0.4 | 0.7 | 3.3 | 0.0 | 0.1 | 0.6 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.8 | 3.0 | 0.1 | 20.2 | 10.5 | 3.4 | 10.3 | 2.9 | 0.7 | 3.1 | 0.4 | 0.2 | 0.3 | 0.0 | 0.0 |
| 5 | 0.3 | 2.3 | 0.4 | 11.0 | 12.2 | 10.3 | 4.6 | 5.2 | 4.4 | 1.1 | 0.5 | 1.2 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 3.1 | 1.9 | 3.0 | 10.4 | 19.2 | 0.7 | 2.9 | 9.5 | 0.2 | 0.5 | 3.1 | 0.1 | 0.1 | 0.4 |
| 7 | 0.1 | 0.2 | 0.0 | 6.3 | 2.6 | 0.4 | 9.9 | 4.3 | 0.9 | 7.9 | 1.6 | 0.6 | 0.8 | 0.1 | 0.1 |
| 8 | 0.0 | 0.1 | 0.0 | 1.8 | 3.3 | 1.8 | 4.1 | 7.1 | 4.3 | 3.4 | 2.9 | 3.3 | 0.4 | 0.1 | 0.6 |
| 9 | 0.0 | 0.2 | 0.0 | 0.4 | 2.6 | 5.6 | 0.7 | 4.1 | 10.1 | 0.6 | 1.4 | 9.0 | 0.2 | 0.2 | 0.8 |
| 10 | 0.0 | 0.0 | 0.0 | 0.8 | 0.1 | 0.0 | 4.6 | 1.5 | 0.2 | 14.7 | 3.5 | 0.6 | 4.5 | 0.9 | 0.4 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.6 | 1.2 | 0.7 | 3.1 | 5.3 | 3.2 | 1.2 | 0.8 | 1.6 |
| 12 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.7 | 0.2 | 1.3 | 4.9 | 0.5 | 3.9 | 16.7 | 0.5 | 1.0 | 5.6 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 1.6 | 0.3 | 0.1 | 2.5 | 1.7 | 0.2 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.1 | 0.6 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 2.1 | 0.3 | 2.1 | 3.0 |

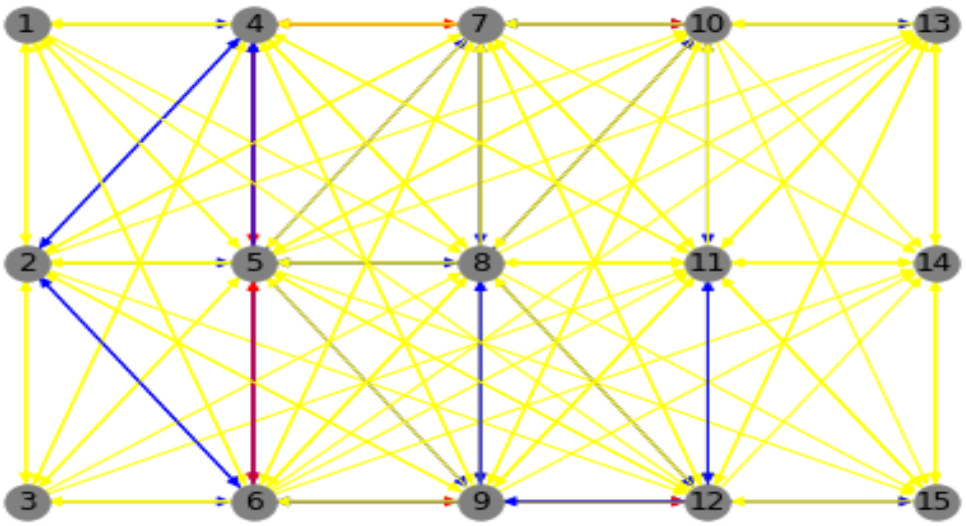



Figure 4.15: Cluster 3 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### 4.3.4 Examining results

From the shown below results in table 4.18, we can deduce that Ligue 1 is different than other leagues as it has lower average number of passes and depends more on long balls for attacking. Also, the top 6 teams were clustered evenly among the three clusters (except for Paris Saint Germain). Which infers that there are no major differences in tactical style between top and bottom teams in the French Ligue 1. This can the reason that Ligue 1 is considered the weakest league among the top 5 European leagues [26].

Table 4.18: shows Ligue 1 evaluation metrics results

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Average number of passes | 271 | 411 | 480 |
| Overall matches | 56.6% | 20% | 23.4% |
| Top 6 teams | 34.2% | 24.5% | 41.2% |
| Passes received in the first 20 percent of the field | 17 (6%) | 25 (6%) | 19(4%) |
| Passes received in the second 20 percent of the field | 75 (27%) | 148 (35%) | 113(23%) |
| Passes received in the third 20 percent of the field | 73 (26%) | 106 (25%) | 153(32%) |
| Passes received in the fourth 20 percent of the field | 76 (28%) | 100 (24%) | 148(30%) |
| Passes received in the last 20 percent of the field | 28 (10%) | 31 (7%) | 45(9%) |
| Percentage of the long balls received in the attacking area from the total passes | 10.5% | 10.12% | 8.11% |

## 4.4 Bundesliga

### 4.4.1 Principal Component analysis

We use principal component analysis to visualize teams' 2D passing matrices into a two-dimensional graph in figure 4.16.

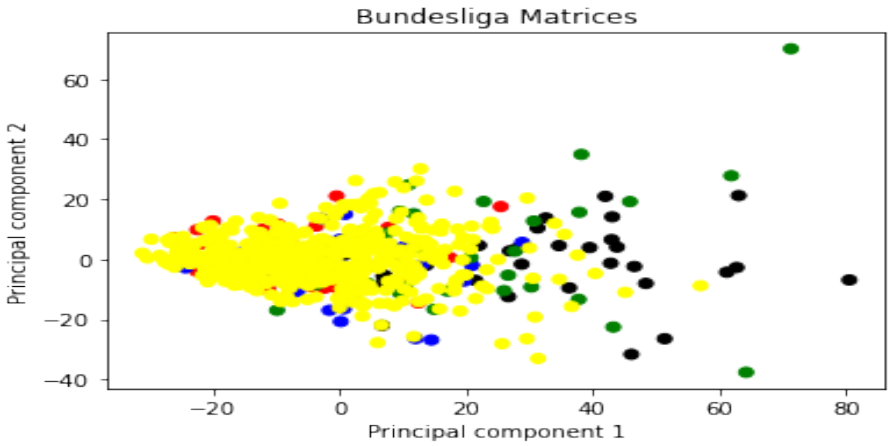
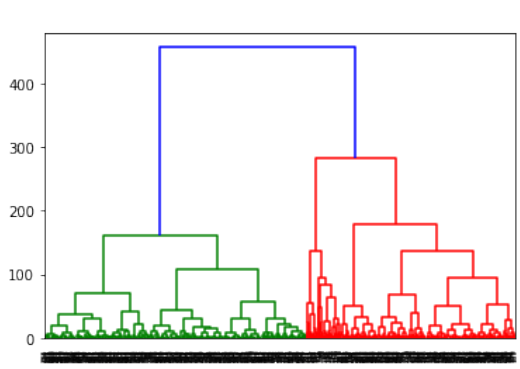


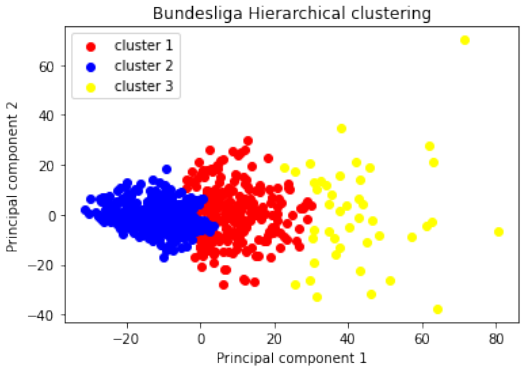

Figure 4.16: Bundesliga's 2D graph where each data point represent a passing network for a team during 2017/2018 season. Each team has 34 points corresponding to 34 matches. Points in Black represents Bayern Munich's passing matrices, Red for Schalke 04, Blue for 1899 Hoffenheim, Green for Borussia Dortmund, and Yellow for the rest of the teams.

### 4.4.2 Hierarchical Clustering

This subsection presents the results of running the hierarchical clustering algorithm on the Bundesliga test-set in figure 4.17 and table 4.19. We choose the number of clusters to be 3 according to the resulted dendrogram.




(a) Bundesliga's dendrogram




(b) Bundesliga's clusters created after running Hierarchical clustering algorithm on the data points.

Figure 4.17: Bundesliga's dendrogram and hierarchical clustering

Table 4.19: shows clustering of Bundesliga teams using hierarchical clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Bayern Munich | 15 | 1 | 18 |
| Schalke | 8 | 25 | 1 |
| 1899 Hoffenheim | 17 | 17 | 0 |
| Borussia Dortmund | 17 | 6 | 11 |
| Bayern Leverkusen | 18 | 12 | 4 |
| RB Leipzig | 21 | 10 | 3 |
| VfB Stuttgart | 10 | 22 | 2 |
| Eintracht Frankfurt | 11 | 23 | 0 |
| Borussia Mönchengladbach | 22 | 11 | 1 |
| Hertha BSC | 13 | 20 | 1 |
| Werder Bremen | 15 | 19 | 0 |
| FC Augsburg | 5 | 29 | 0 |
| Hannover | 7 | 27 | 0 |
| Mainz | 4 | 30 | 0 |
| SC Freiburg | 8 | 26 | 0 |
| VfL Wolfsburg | 14 | 18 | 2 |
| Hamburger SV | 8 | 25 | 1 |
| FC Köln | 14 | 20 | 0 |

### 4.4.3 k-means Clustering

This section mentions the results of running the k-means clustering algorithm on the Bundesliga test-set. Table 4.20 below shows the clustering of Bundesliga teams' matches into 3 clusters. Then, we evaluate the clusters in table 4.24.

Table 4.20: shows clustering of Bundesliga teams using k-means clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Bayern Munich | 22 | 0 | 12 |
| Schalke 04 | 1 | 19 | 14 |
| 1899 Hoffenheim | 1 | 13 | 20 |
| Borussia Dortmund | 14 | 6 | 14 |
| Bayern Leverkusen | 6 | 6 | 22 |
| RB Leipzig | 5 | 8 | 21 |
| VfB Stuttgart | 2 | 17 | 15 |
| Eintracht Frankfurt | 0 | 22 | 12 |
| Borussia Mönchengladbach | 5 | 9 | 20 |
| Hertha BSC | 1 | 18 | 15 |
| Werder Bremen | 1 | 17 | 16 |
| FC Augsburg | 0 | 27 | 7 |
| Hannover | 0 | 27 | 7 |
| Mainz | 0 | 27 | 7 |
| SC Freiburg | 0 | 27 | 7 |
| VfL Wolfsburg | 2 | 15 | 17 |
| Hamburger SV | 1 | 22 | 11 |
| FC Köln | 0 | 17 | 17 |

### Cluster 1

Table 4.21: 2D Matrix showing the average passes of cluster 1 teams' in the Bundesliga where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.3 | 1.5 | 0.2 | 2.7 | 1.1 | 0.3 | 0.4 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.0 | 1.3 | 1.0 | 3.7 | 3.3 | 3.4 | 0.7 | 0.7 | 1.0 | 0.3 | 0.3 | 0.2 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.3 | 1.2 | 0.2 | 0.9 | 1.9 | 0.0 | 0.2 | 0.5 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.6 | 3.0 | 0.1 | 12.1 | 8.6 | 2.9 | 10.1 | 3.9 | 1.6 | 2.4 | 0.2 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.7 | 0.3 | 8.6 | 8.9 | 8.3 | 6.8 | 5.8 | 6.1 | 1.0 | 0.5 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.3 | 1.3 | 2.8 | 7.9 | 10.4 | 1.1 | 3.5 | 8.9 | 0.1 | 0.3 | 2.6 | 0.0 | 0.1 | 0.2 |
| 7 | 0.2 | 0.3 | 0.1 | 6.4 | 3.9 | 0.7 | 18.7 | 11.2 | 3.1 | 16.7 | 3.2 | 1.1 | 1.0 | 0.2 | 0.4 |
| 8 | 0.0 | 0.1 | 0.0 | 1.5 | 3.3 | 1.6 | 10.1 | 11.1 | 10.5 | 7.5 | 6.6 | 7.5 | 0.6 | 0.3 | 0.8 |
| 9 | 0.0 | 0.3 | 0.1 | 0.5 | 3.8 | 6.4 | 3.2 | 10.1 | 15.9 | 1.3 | 3.3 | 15.0 | 0.2 | 0.3 | 1.1 |
| 10 | 0.1 | 0.0 | 0.0 | 0.9 | 0.2 | 0.2 | 11.2 | 4.3 | 0.6 | 35.8 | 10.4 | 1.8 | 9.3 | 1.7 | 0.9 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 1.5 | 2.9 | 1.7 | 9.4 | 14.8 | 10.3 | 3.0 | 2.4 | 3.5 |
| 12 | 0.2 | 0.1 | 0.0 | 0.2 | 0.6 | 0.8 | 0.4 | 3.5 | 9.9 | 2.1 | 10.2 | 31.2 | 0.6 | 1.9 | 8.5 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 4.1 | 0.8 | 0.1 | 4.7 | 3.1 | 0.5 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.4 | 0.2 | 0.3 | 1.7 | 0.4 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.7 | 3.9 | 0.5 | 3.3 | 4.3 |

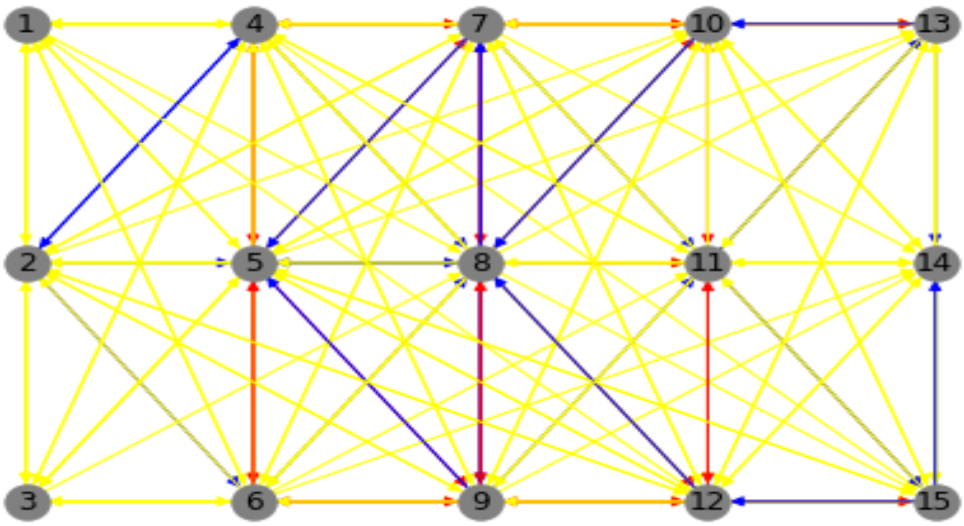



Figure 4.18: Cluster 1 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

**Cluster 2**

Table 4.22: 2D Matrix showing the average passes of cluster 2 teams' in the Bundesliga where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.4 | 1.0 | 0.1 | 2.3 | 0.5 | 0.1 | 0.8 | 0.2 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.0 | 1.1 | 1.2 | 3.2 | 2.5 | 3.1 | 0.9 | 1.0 | 1.1 | 0.7 | 1.0 | 0.9 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.0 | 1.2 | 0.2 | 0.5 | 2.1 | 0.0 | 0.2 | 0.9 | 0.0 | 0.2 | 0.4 | 0.0 | 0.0 | 0.0 |
| 4 | 1.7 | 2.8 | 0.1 | 9.4 | 5.0 | 1.3 | 6.2 | 1.7 | 0.4 | 2.2 | 0.7 | 0.2 | 0.2 | 0.1 | 0.1 |
| 5 | 0.3 | 2.4 | 0.3 | 5.9 | 6.3 | 5.6 | 2.5 | 3.0 | 2.5 | 1.0 | 0.7 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.5 | 1.4 | 1.1 | 4.8 | 8.7 | 0.3 | 1.6 | 5.7 | 0.2 | 0.7 | 2.3 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 3.6 | 1.7 | 0.3 | 6.3 | 2.3 | 0.5 | 5.6 | 1.4 | 0.3 | 0.6 | 0.2 | 0.2 |
| 8 | 0.0 | 0.2 | 0.0 | 1.2 | 2.2 | 1.2 | 2.8 | 4.5 | 2.6 | 2.3 | 2.1 | 2.1 | 0.3 | 0.1 | 0.4 |
| 9 | 0.0 | 0.2 | 0.1 | 0.2 | 1.5 | 2.9 | 0.5 | 2.4 | 5.7 | 0.4 | 1.3 | 5.7 | 0.1 | 0.1 | 0.6 |
| 10 | 0.0 | 0.0 | 0.0 | 0.5 | 0.1 | 0.0 | 2.9 | 0.8 | 0.1 | 9.5 | 2.5 | 0.4 | 3.5 | 0.9 | 0.3 |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.5 | 1.0 | 0.5 | 2.2 | 4.2 | 2.3 | 1.2 | 1.0 | 1.3 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.4 | 0.1 | 0.8 | 2.6 | 0.4 | 2.7 | 9.5 | 0.3 | 1.0 | 3.6 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.2 | 0.2 | 0.1 | 2.3 | 1.7 | 0.2 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.1 | 0.2 | 0.7 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.1 | 0.2 | 1.7 | 2.2 |

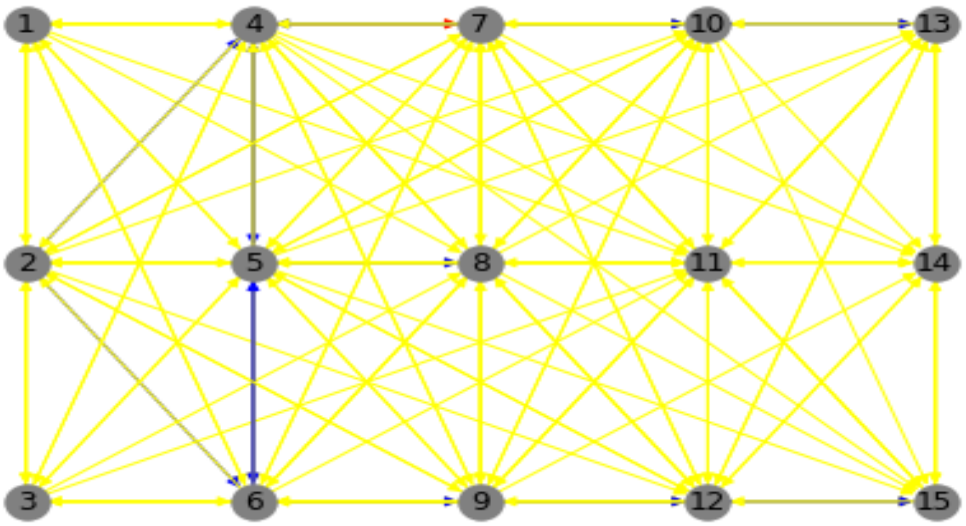



Figure 4.19: Cluster 2 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### Cluster 3

Table 4.23: 2D Matrix showing the average passes of cluster 3 teams' in the Bundesliga where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.3 | 1.4 | 0.1 | 2.6 | 0.6 | 0.3 | 0.6 | 0.2 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.5 | 1.4 | 1.3 | 4.2 | 3.6 | 4.2 | 0.9 | 0.9 | 1.1 | 0.4 | 0.6 | 0.7 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.4 | 1.4 | 0.2 | 0.7 | 2.6 | 0.0 | 0.2 | 0.8 | 0.0 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 |
| 4 | 1.7 | 3.4 | 0.2 | 14.4 | 8.6 | 2.1 | 9.3 | 2.8 | 0.8 | 2.8 | 0.5 | 0.2 | 0.2 | 0.0 | 0.0 |
| 5 | 0.4 | 3.1 | 0.4 | 9.0 | 8.6 | 8.6 | 4.7 | 4.6 | 4.7 | 1.1 | 0.7 | 1.1 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 3.4 | 1.7 | 2.2 | 7.9 | 12.3 | 0.6 | 2.4 | 8.3 | 0.2 | 0.6 | 2.7 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.3 | 0.0 | 5.8 | 3.1 | 0.4 | 11.2 | 5.2 | 1.2 | 9.5 | 1.9 | 0.7 | 0.7 | 0.1 | 0.2 |
| 8 | 0.0 | 0.2 | 0.0 | 1.8 | 3.3 | 1.5 | 5.4 | 7.2 | 5.2 | 4.0 | 3.1 | 3.9 | 0.5 | 0.1 | 0.6 |
| 9 | 0.0 | 0.2 | 0.1 | 0.4 | 3.0 | 5.4 | 1.1 | 4.9 | 11.1 | 0.6 | 2.0 | 9.8 | 0.2 | 0.1 | 0.9 |
| 10 | 0.0 | 0.0 | 0.0 | 0.8 | 0.3 | 0.0 | 5.4 | 1.8 | 0.3 | 17.5 | 4.4 | 0.7 | 5.5 | 1.3 | 0.6 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.1 | 0.7 | 1.7 | 0.7 | 3.8 | 6.0 | 3.8 | 1.6 | 1.4 | 1.9 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.7 | 0.3 | 1.9 | 5.1 | 0.8 | 4.5 | 18.3 | 0.4 | 1.4 | 5.8 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 2.1 | 0.3 | 0.1 | 2.8 | 2.2 | 0.3 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.4 | 0.1 | 0.3 | 1.1 | 0.3 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.3 | 2.0 | 0.3 | 2.1 | 3.0 |

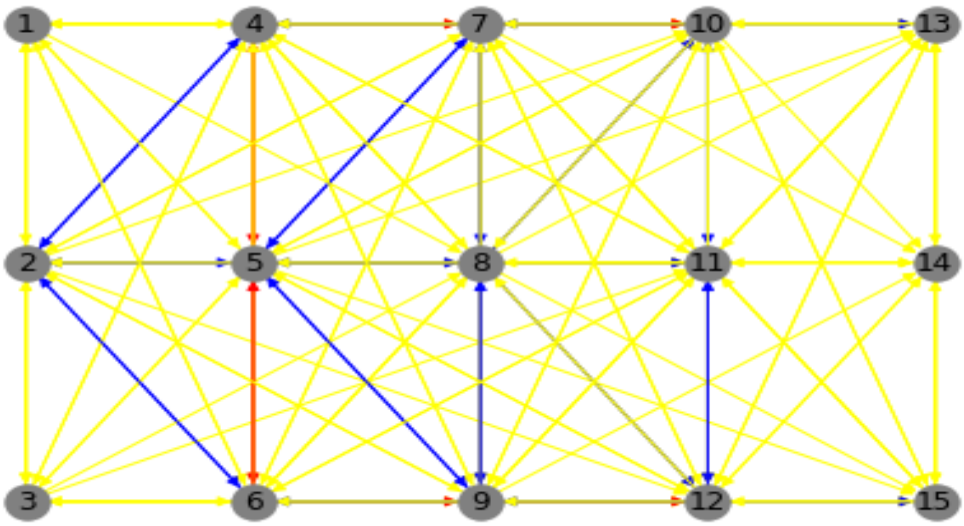



Figure 4.20: Cluster 3 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### 4.4.4 Examining results

From the shown below results in table 4.24, we can deduce that most of the top 6 teams depend less on long balls to attack and except for (Bayern Munich cluster 3 matches), they tend to pass the ball more in the middle (third) area of the pitch.

Table 4.24: shows Bundesliga evaluation metrics results

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Average number of passes | 577 | 259 | 401 |
| Overall matches | 10% | 48.5% | 41.5% |
| Top 6 teams | 24% | 25.5% | 50.4% |
| Passes received in the first 20 percent of the field | 23 (10.2%) | 20 (7%) | 25(6%) |
| Passes received in the second 20 percent of the field | 122 (15.5%) | 77 (29%) | 119(29%) |
| Passes received in the third 20 percent of the field | 175 (36.1%) | 65 (25%) | 112(27%) |
| Passes received in the fourth 20 percent of the field | 201 (21.7%) | 70 (27%) | 109(27%) |
| Passes received in the last 20 percent of the field | 54 (16.8%) | 25 (9%) | 35(4%) |
| Percentage of the long balls received in the attacking area from the total passes | 7% | 10% | 7.82% |

## 4.5 La Liga

### 4.5.1 Principal Component analysis

We use principal component analysis to visualize teams' 2D passing matrices into a two-dimensional graph in figure 4.21.

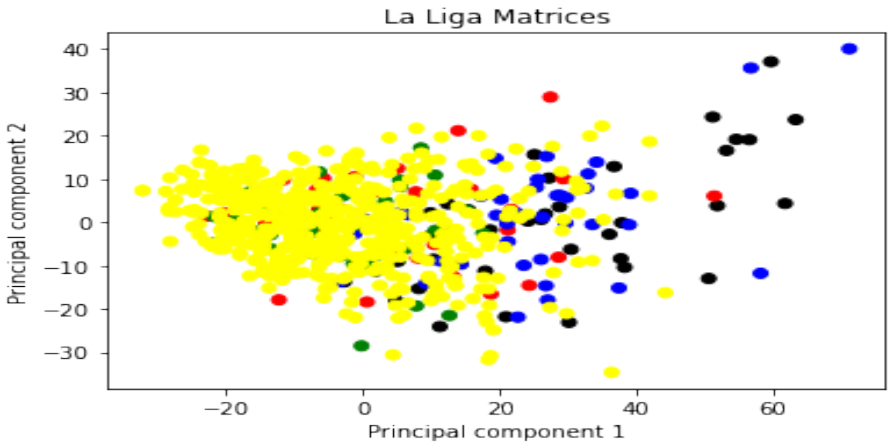
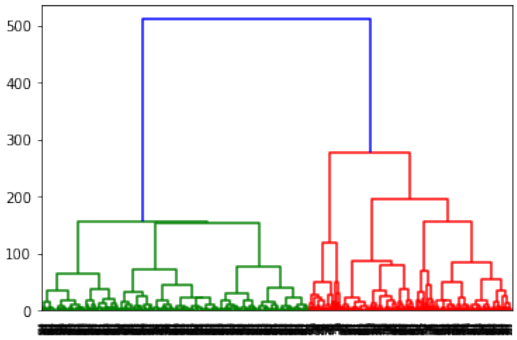


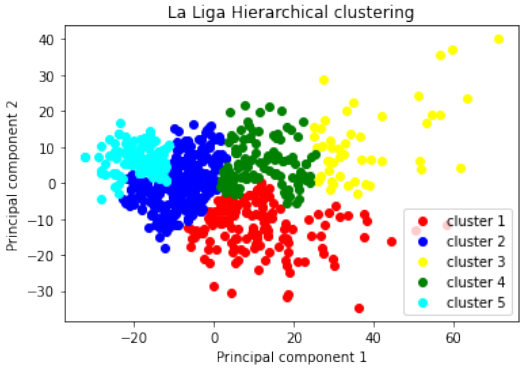

Figure 4.21: La Liga's 2D graph where each data point representss a passing network for a team during 2017/2018 season. Each team has 38 points corresponding to 38 matches. Points in Black represents Barcelona's passing matrices, Red for Atletico Madrid, Blue for Real Madrid, Green for Valencia, and Yellow for the rest of the teams.

### 4.5.2 Hierarchical clustering

This subsection presents the results of running the hierarchical clustering algorithm on the La Liga test-set in figure 4.22 and table 4.25. We choose the number of clusters to be 5 according to the resulted dendrogram.




(a) La Liga's dendrogram




(b) La Liga's clusters created after running Hierarchical clustering algorithm on the data points.

Figure 4.22: La Liga's dendrogram and hierarchical clustering

Table 4.25: shows clustering of La Liga teams using hierarchical clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| Barcelona | 16 | 0 | 16 | 6 | 0 |
| Atlético Madrid | 8 | 14 | 3 | 8 | 5 |
| Real Madrid | 11 | 2 | 15 | 10 | 0 |
| Valencia | 8 | 18 | 0 | 8 | 4 |
| Villarreal | 13 | 16 | 1 | 7 | 1 |
| Real Betis | 17 | 12 | 3 | 6 | 0 |
| Sevilla | 11 | 12 | 4 | 11 | 0 |
| Getafe | 5 | 25 | 0 | 4 | 4 |
| Eibar | 0 | 19 | 1 | 9 | 9 |
| Girona | 3 | 24 | 0 | 1 | 10 |
| Espanyol | 5 | 25 | 0 | 4 | 4 |
| Real Sociedad | 14 | 12 | 3 | 9 | 0 |
| Celta Vigo | 16 | 13 | 2 | 7 | 0 |
| Alavés | 0 | 14 | 0 | 1 | 23 |
| Levante | 2 | 16 | 0 | 2 | 18 |
| Leganés | 0 | 19 | 0 | 2 | 17 |
| Deportivo La Coruña | 2 | 17 | 0 | 9 | 10 |
| Las Palmas | 16 | 14 | 1 | 6 | 1 |
| Málaga | 1 | 21 | 0 | 4 | 12 |

### 4.5.3 k-means clustering

This section mentions the results of running the k-means clustering algorithm on the La Liga test-set. Table 4.26 below shows the clustering of La Liga teams' matches into 3 clusters. Then, we evaluate the resulted clusters in table 4.30.

Table 4.26: shows clustering of La Liga teams using k-means clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Barcelona | 14 | 0 | 24 |
| Atlético Madrid | 11 | 18 | 9 |
| Real Madrid | 10 | 1 | 27 |
| Valencia | 17 | 21 | 0 |
| Villarreal | 18 | 17 | 3 |
| Real Betis | 22 | 10 | 6 |
| Sevilla | 18 | 11 | 9 |
| Getafe | 11 | 27 | 0 |
| Eibar | 7 | 27 | 4 |
| Girona | 6 | 32 | 0 |
| Espanyol | 11 | 27 | 0 |
| Real Sociedad | 20 | 11 | 7 |
| Celta Vigo | 24 | 11 | 3 |
| Alavés | 2 | 36 | 0 |
| Levante | 5 | 33 | 0 |
| Leganés | 3 | 35 | 0 |
| Deportivo La Coruña | 12 | 25 | 1 |
| Las Palmas | 20 | 15 | 3 |
| Málaga | 8 | 30 | 0 |

**Cluster 1**

Table 4.27: 2D Matrix showing the average passes of cluster 1 teams' in La Liga where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.7 | 1.6 | 0.1 | 2.7 | 0.7 | 0.2 | 0.7 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.6 | 1.4 | 1.4 | 4.0 | 3.6 | 3.6 | 0.7 | 0.8 | 0.8 | 0.4 | 0.5 | 0.3 | 0.0 | 0.0 | 0.0 |
| 3 | 0.2 | 1.4 | 1.8 | 0.3 | 0.6 | 2.8 | 0.0 | 0.2 | 0.8 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.9 | 3.1 | 0.1 | 15.2 | 8.4 | 2.4 | 8.5 | 2.5 | 0.7 | 2.4 | 0.4 | 0.2 | 0.2 | 0.0 | 0.0 |
| 5 | 0.4 | 2.7 | 0.4 | 8.7 | 9.3 | 8.6 | 4.2 | 4.6 | 4.2 | 1.0 | 0.6 | 0.9 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 3.3 | 1.9 | 2.3 | 7.7 | 14.9 | 0.6 | 2.6 | 8.7 | 0.2 | 0.4 | 2.8 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 5.4 | 2.7 | 0.5 | 10.1 | 5.2 | 1.2 | 9.0 | 1.8 | 0.8 | 0.6 | 0.1 | 0.2 |
| 8 | 0.0 | 0.2 | 0.0 | 1.7 | 3.1 | 1.7 | 5.1 | 8.0 | 5.3 | 4.2 | 3.4 | 4.2 | 0.4 | 0.1 | 0.5 |
| 9 | 0.0 | 0.2 | 0.1 | 0.5 | 2.6 | 5.6 | 0.9 | 5.3 | 10.4 | 0.6 | 1.9 | 9.7 | 0.2 | 0.1 | 0.7 |
| 10 | 0.0 | 0.0 | 0.0 | 0.9 | 0.2 | 0.0 | 6.0 | 1.9 | 0.2 | 18.3 | 4.9 | 0.9 | 5.3 | 1.2 | 0.6 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.7 | 1.7 | 0.8 | 4.3 | 7.1 | 4.6 | 1.9 | 1.3 | 2.1 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.8 | 0.2 | 2.2 | 6.0 | 0.8 | 5.1 | 20.5 | 0.6 | 1.3 | 6.6 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 0.4 | 0.1 | 2.7 | 2.0 | 0.4 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.1 | 0.2 | 0.9 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.4 | 2.5 | 0.3 | 2.4 | 3.3 |

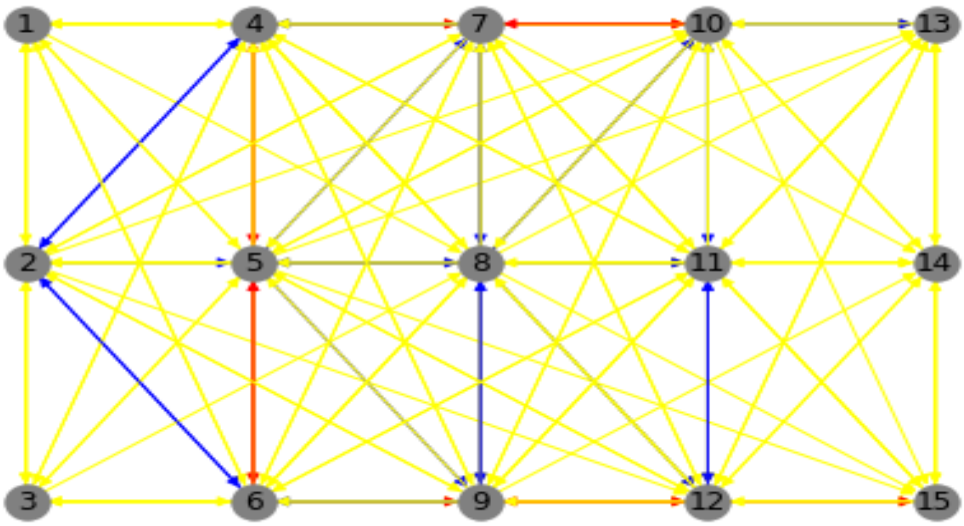



Figure 4.23: Cluster 1 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### Cluster 2

Table 4.28: 2D Matrix showing the average passes of cluster 2 teams' in La Liga where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.4 | 0.8 | 0.1 | 2.2 | 0.4 | 0.2 | 0.8 | 0.2 | 0.0 | 0.4 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 0.9 | 1.2 | 0.9 | 2.4 | 2.1 | 2.2 | 0.8 | 0.7 | 0.8 | 0.8 | 0.8 | 0.8 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 0.8 | 1.3 | 0.1 | 0.5 | 2.1 | 0.0 | 0.1 | 0.7 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.6 | 2.1 | 0.1 | 11.1 | 4.7 | 1.2 | 6.3 | 1.6 | 0.4 | 2.0 | 0.5 | 0.3 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.0 | 0.3 | 5.1 | 5.9 | 5.0 | 2.7 | 3.3 | 2.7 | 0.9 | 0.6 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 1.9 | 1.3 | 1.3 | 4.5 | 9.6 | 0.3 | 1.6 | 5.9 | 0.2 | 0.5 | 2.1 | 0.0 | 0.0 | 0.3 |
| 7 | 0.1 | 0.2 | 0.0 | 4.1 | 1.7 | 0.3 | 7.5 | 2.7 | 0.7 | 6.0 | 1.1 | 0.5 | 0.6 | 0.1 | 0.2 |
| 8 | 0.0 | 0.2 | 0.0 | 1.1 | 2.5 | 1.1 | 3.1 | 5.0 | 3.0 | 2.4 | 2.1 | 2.3 | 0.4 | 0.1 | 0.5 |
| 9 | 0.0 | 0.2 | 0.1 | 0.3 | 1.9 | 3.7 | 0.5 | 2.8 | 6.9 | 0.5 | 1.1 | 5.9 | 0.3 | 0.1 | 0.8 |
| 10 | 0.0 | 0.0 | 0.0 | 0.5 | 0.1 | 0.0 | 3.8 | 1.1 | 0.1 | 10.8 | 2.6 | 0.5 | 4.1 | 0.9 | 0.5 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.5 | 1.1 | 0.5 | 2.4 | 3.6 | 2.3 | 1.3 | 0.8 | 1.2 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.5 | 0.1 | 1.1 | 3.7 | 0.5 | 2.4 | 10.9 | 0.4 | 1.0 | 4.1 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 1.5 | 0.2 | 0.1 | 2.4 | 1.7 | 0.2 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.1 | 0.2 | 0.6 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.5 | 0.2 | 1.8 | 2.4 |

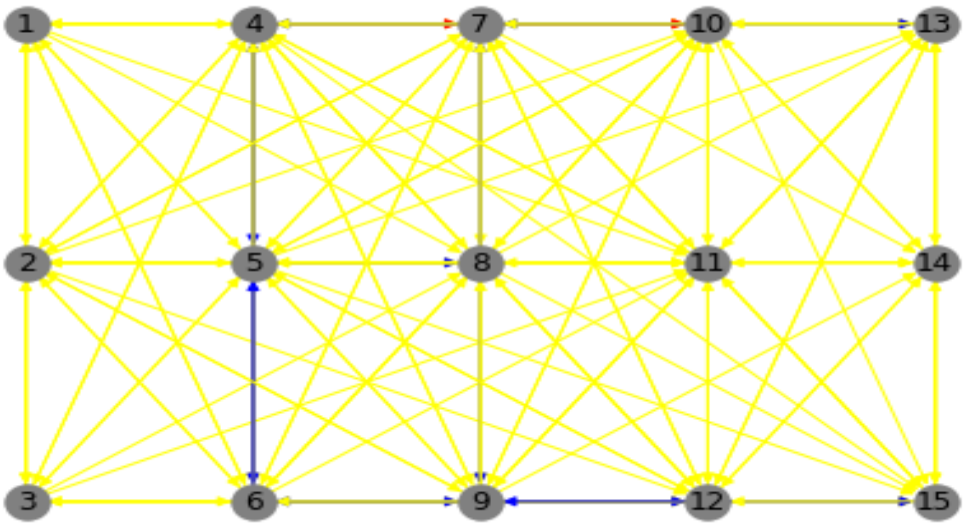



Figure 4.24: Cluster 2 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### Cluster 3

Table 4.29: 2D Matrix showing the average passes of cluster 3 teams' in La Liga where the rows are for the sender's zone and the columns are the receiver's zone.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.9 | 1.7 | 0.2 | 2.9 | 0.7 | 0.5 | 0.6 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.9 | 1.6 | 1.5 | 3.4 | 3.9 | 3.4 | 0.5 | 0.6 | 0.6 | 0.2 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| 3 | 0.2 | 1.6 | 1.7 | 0.4 | 0.7 | 2.7 | 0.0 | 0.1 | 0.5 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| 4 | 1.7 | 2.7 | 0.1 | 14.2 | 7.6 | 2.4 | 9.8 | 2.9 | 1.1 | 2.5 | 0.3 | 0.3 | 0.1 | 0.1 | 0.1 |
| 5 | 0.3 | 2.1 | 0.3 | 7.9 | 7.8 | 6.4 | 5.5 | 5.4 | 5.0 | 0.9 | 0.4 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.6 | 1.6 | 2.7 | 7.4 | 12.5 | 1.0 | 2.6 | 7.6 | 0.2 | 0.3 | 2.0 | 0.0 | 0.1 | 0.2 |
| 7 | 0.1 | 0.3 | 0.0 | 6.2 | 2.9 | 0.5 | 15.1 | 6.9 | 2.1 | 15.1 | 2.4 | 1.2 | 1.0 | 0.2 | 0.3 |
| 8 | 0.0 | 0.1 | 0.0 | 1.6 | 3.1 | 1.2 | 7.2 | 9.7 | 6.6 | 6.9 | 4.7 | 6.3 | 0.6 | 0.1 | 0.6 |
| 9 | 0.0 | 0.1 | 0.1 | 0.5 | 2.9 | 5.7 | 1.7 | 7.2 | 12.9 | 0.8 | 2.4 | 12.5 | 0.2 | 0.1 | 0.8 |
| 10 | 0.0 | 0.0 | 0.0 | 0.9 | 0.2 | 0.1 | 8.9 | 3.0 | 0.5 | 36.2 | 9.1 | 2.2 | 10.5 | 1.9 | 1.1 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 1.1 | 2.5 | 1.1 | 7.8 | 12.7 | 7.8 | 2.9 | 2.3 | 3.3 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.9 | 0.4 | 3.1 | 7.9 | 1.7 | 9.0 | 31.0 | 0.9 | 1.7 | 9.8 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 5.0 | 0.9 | 0.2 | 5.6 | 3.5 | 0.5 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.6 | 0.1 | 0.5 | 1.6 | 0.5 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.8 | 4.9 | 0.5 | 3.4 | 4.9 |

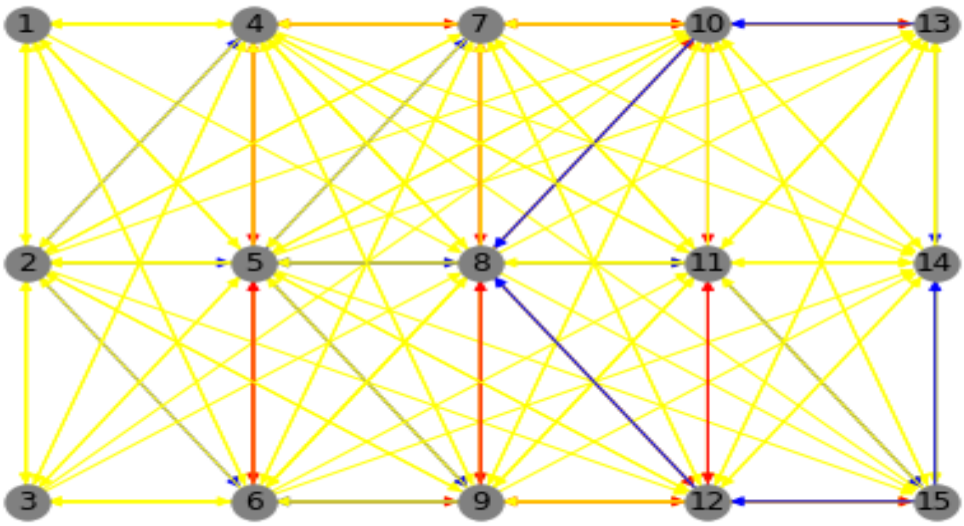



Figure 4.25: Cluster 3 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### 4.5.4 Examining results

From the shown below results from table 4.30, we can deduce that similarly to the previous leagues, as you go higher in the league table, the number of matches that belong to the cluster with the higher number of passes and the lower dependence on long balls for attacking increase. We didn't find significant differences in the percentage of passes in each fifth area of the field expect for cluster 3 (fourth area). Also, we noticed that Real Madrid and Barcelona play style was nearly similar in this season.

Table 4.30: shows La Liga evaluation metrics results

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Average total number of passes | 415 | 271 | 532 |
| Overall matches | 33.1% | 53.6 % | 13.2% |
| Top 6 teams | 40.3% | 29.3% | 30.2% |
| Passes received in the first 20 percent of the field | 25 (6%) | 18 (6%) | 24(4%) |
| Passes received in the second 20 percent of the field | 122 (29%) | 78 (29%) | 114(21%) |
| Passes received in the third 20 percent of the field | 111 (26%) | 73 (27%) | 142(26%) |
| Passes received in the fourth 20 percent of the field | 118 (28%) | 73 (26%) | 191(35%) |
| Passes received in the last 20 percent of the field | 36 (8%) | 27 (10%) | 60(11%) |
| Percentage of the long balls received in the attacking area from the total passes | 6.8% | 9.7% | 5.9% |

## 4.6 All Leagues

### 4.6.1 Principal Component analysis

We use principal component analysis to visualize teams' 2D passing matrices into a two-dimensional graph in figure 4.26.

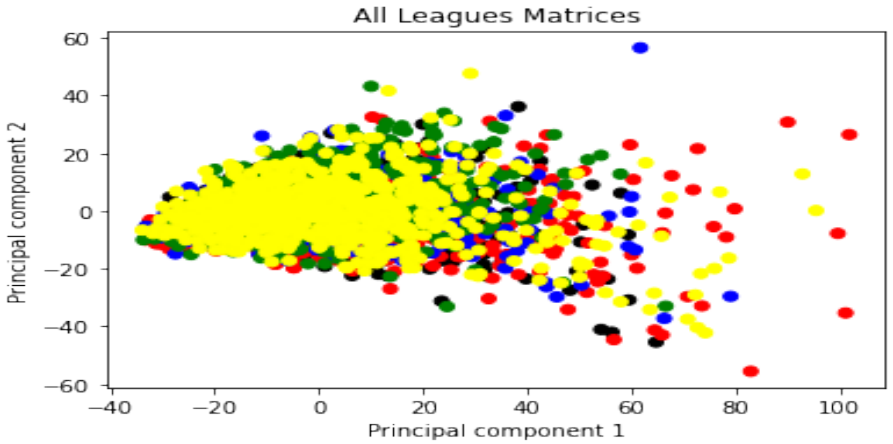
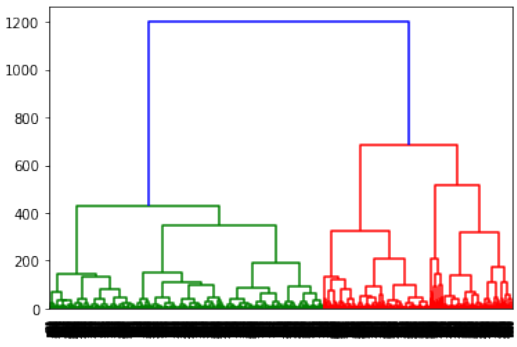


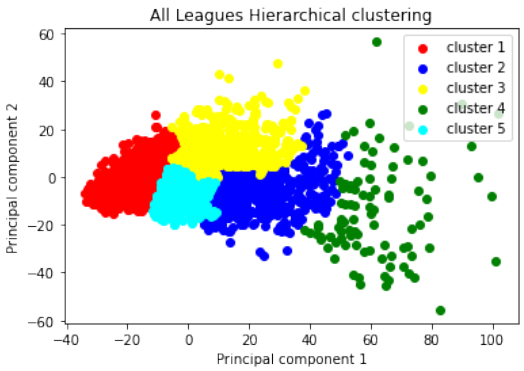

Figure 4.26: All leagues' 2D graph where each data point represent a passing network for a team during 2017/2018 season. Points in Black represents La Liga's passing matrices, Red for Premier League, Blue for Bundesliga, Green for Ligue 1, and Yellow for Serie A.

### 4.6.2 Hierarchical clustering

This subsection mentions the results of running hierarchical clustering algorithm on the test-set including the whole 5 leagues. Figure 4.27 and table 4.31 below shows the clustering of all leagues teams' matches into 5 clusters according to the generated dendrogram.




(a) All leagues' dendrogram




(b) All leagues's clusters created after running Hierarchical clustering algorithm on the data points.

Figure 4.27: All leagues's dendrogram and hierarchical clustering

Table 4.31: shows clustering of all leagues teams using hierarchical clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| La Liga | 296 | 115 | 160 | 13 | 138 |
| Premier League | 281 | 148 | 119 | 33 | 179 |
| Bundesliga | 277 | 68 | 154 | 12 | 101 |
| Ligue 1 | 323 | 93 | 215 | 5 | 124 |
| Serie A | 282 | 126 | 184 | 31 | 137 |

### 4.6.3 k-means clustering

This section mentions the results of running the k-means clustering algorithm on all leagues test-set. Table 4.32 below shows the clustering of all leagues teams' matches into 3 clusters.

Table 4.32: shows clustering of all leagues teams using k-means clustering.

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| La Liga | 385(53.3%) | 69(9.5%) | 268(37.1%) |
| Premier League | 406(53.4%) | 102(13.4%) | 252(33.1%) |
| Bundesliga | 335(54.7%) | 39(6.3%) | 238(38.8%) |
| Ligue 1 | 406(53.4%) | 51(6.7%) | 303(39.8%) |
| Serie A | 360(47.3%) | 74(9.7%) | 326(42.9%) |
| Overall Matches | 52.3% | 9.3% | 38.4% |

**Cluster 1**

Table 4.33: 2D Matrix showing the average passes of cluster 1 leagues where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.4 | 0.9 | 0.1 | 2.3 | 0.4 | 0.2 | 0.7 | 0.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.0 | 1.2 | 1.0 | 2.8 | 2.2 | 2.6 | 0.8 | 0.8 | 0.9 | 0.7 | 0.8 | 0.8 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 0.9 | 1.4 | 0.2 | 0.4 | 2.2 | 0.0 | 0.2 | 0.8 | 0.0 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 |
| 4 | 1.5 | 2.3 | 0.1 | 10.6 | 4.6 | 1.3 | 6.3 | 1.6 | 0.4 | 2.1 | 0.5 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.1 | 0.3 | 5.3 | 6.0 | 5.0 | 2.6 | 3.1 | 2.5 | 1.0 | 0.6 | 0.9 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.1 | 1.4 | 1.2 | 4.3 | 9.6 | 0.3 | 1.6 | 5.9 | 0.2 | 0.5 | 2.2 | 0.0 | 0.0 | 0.3 |
| 7 | 0.1 | 0.2 | 0.0 | 3.8 | 1.6 | 0.3 | 7.0 | 2.5 | 0.6 | 6.0 | 1.1 | 0.4 | 0.7 | 0.1 | 0.2 |
| 8 | 0.0 | 0.1 | 0.0 | 1.1 | 2.4 | 1.1 | 3.0 | 4.6 | 2.9 | 2.3 | 2.0 | 2.2 | 0.4 | 0.1 | 0.4 |
| 9 | 0.0 | 0.1 | 0.1 | 0.2 | 1.5 | 3.4 | 0.5 | 2.6 | 6.5 | 0.4 | 1.2 | 5.8 | 0.2 | 0.1 | 0.7 |
| 10 | 0.0 | 0.0 | 0.0 | 0.5 | 0.1 | 0.0 | 3.4 | 0.9 | 0.1 | 10.9 | 2.6 | 0.5 | 4.0 | 0.9 | 0.4 |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.5 | 1.0 | 0.5 | 2.4 | 3.8 | 2.4 | 1.2 | 0.9 | 1.2 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 1.0 | 3.1 | 0.4 | 2.6 | 10.6 | 0.4 | 1.0 | 3.9 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 1.5 | 0.3 | 0.1 | 2.5 | 1.6 | 0.2 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.2 | 0.6 | 0.2 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.4 | 0.2 | 1.6 | 2.4 |

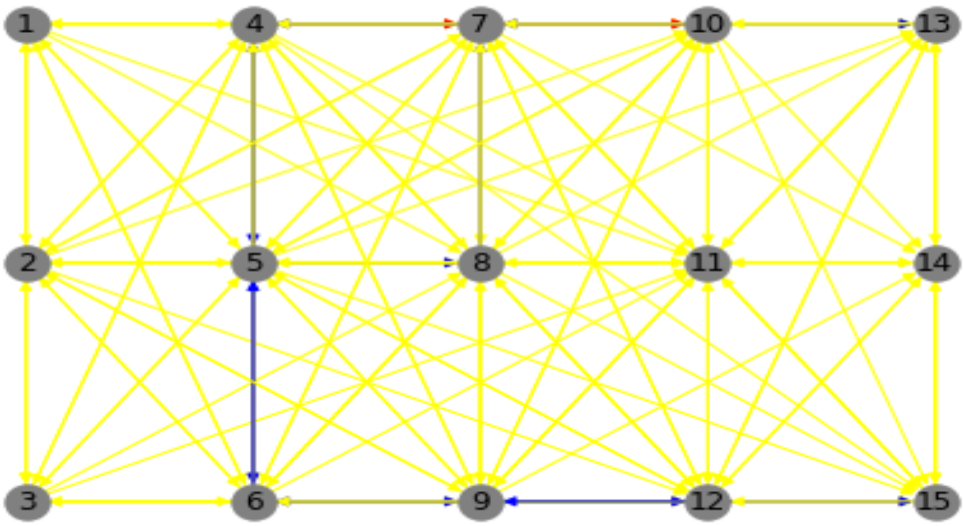



Figure 4.28: Cluster 1 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

**Cluster 2**

Table 4.34: 2D Matrix showing the average passes of cluster 2 leagues where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.8 | 1.5 | 0.1 | 2.8 | 0.9 | 0.4 | 0.3 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.4 | 1.4 | 1.3 | 3.7 | 3.9 | 3.5 | 0.6 | 0.4 | 0.6 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| 3 | 0.2 | 1.4 | 1.8 | 0.4 | 0.8 | 2.5 | 0.0 | 0.1 | 0.4 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| 4 | 1.6 | 2.8 | 0.1 | 14.0 | 8.6 | 2.2 | 9.9 | 3.7 | 1.1 | 2.3 | 0.3 | 0.2 | 0.1 | 0.0 | 0.1 |
| 5 | 0.3 | 2.1 | 0.3 | 8.2 | 9.7 | 7.0 | 6.2 | 6.7 | 5.6 | 1.0 | 0.4 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.5 | 1.5 | 2.4 | 8.1 | 12.1 | 1.0 | 3.5 | 8.3 | 0.2 | 0.3 | 1.9 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 6.4 | 3.3 | 0.4 | 18.4 | 10.9 | 2.3 | 16.9 | 3.3 | 1.1 | 1.0 | 0.1 | 0.2 |
| 8 | 0.0 | 0.1 | 0.0 | 1.7 | 3.4 | 1.5 | 10.3 | 14.2 | 9.4 | 8.5 | 6.7 | 8.0 | 0.6 | 0.2 | 0.6 |
| 9 | 0.0 | 0.2 | 0.1 | 0.5 | 3.3 | 5.8 | 2.3 | 10.4 | 15.0 | 1.0 | 3.0 | 13.5 | 0.2 | 0.1 | 0.9 |
| 10 | 0.0 | 0.0 | 0.0 | 0.9 | 0.2 | 0.1 | 11.0 | 4.0 | 0.5 | 40.6 | 11.2 | 1.9 | 10.2 | 1.8 | 0.9 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 1.4 | 3.4 | 1.4 | 9.5 | 16.6 | 9.4 | 3.1 | 2.6 | 3.3 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.8 | 0.4 | 3.9 | 9.0 | 1.7 | 10.7 | 33.9 | 0.8 | 2.0 | 9.0 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 4.8 | 0.9 | 0.1 | 5.3 | 3.1 | 0.4 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.6 | 0.1 | 0.5 | 1.6 | 0.5 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.7 | 4.5 | 0.4 | 3.2 | 4.5 |

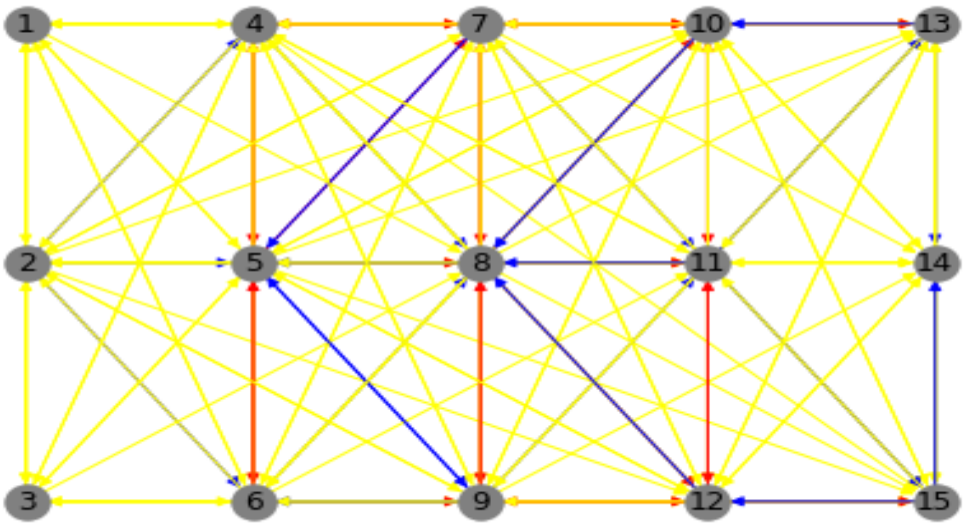



Figure 4.29: Cluster 2 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

**Cluster 3**

Table 4.35: 2D Matrix showing the average passes of cluster 3 leagues where the rows are for the sender's zone and the columns are the receiver's zone

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.5 | 1.4 | 0.1 | 2.7 | 0.7 | 0.3 | 0.6 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 1.4 | 1.4 | 1.3 | 3.9 | 3.4 | 3.7 | 0.7 | 0.7 | 0.8 | 0.4 | 0.4 | 0.4 | 0.0 | 0.0 | 0.0 |
| 3 | 0.1 | 1.4 | 1.6 | 0.3 | 0.7 | 2.7 | 0.0 | 0.2 | 0.7 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | 1.7 | 2.8 | 0.1 | 14.3 | 8.2 | 2.3 | 9.0 | 2.7 | 0.8 | 2.5 | 0.4 | 0.2 | 0.2 | 0.0 | 0.1 |
| 5 | 0.3 | 2.5 | 0.4 | 8.5 | 9.2 | 8.3 | 4.5 | 4.8 | 4.6 | 1.0 | 0.6 | 1.0 | 0.1 | 0.0 | 0.1 |
| 6 | 0.1 | 2.9 | 1.7 | 2.2 | 7.8 | 13.7 | 0.7 | 2.7 | 8.7 | 0.2 | 0.5 | 2.5 | 0.0 | 0.0 | 0.2 |
| 7 | 0.1 | 0.2 | 0.0 | 5.6 | 2.8 | 0.5 | 11.5 | 5.5 | 1.3 | 9.9 | 1.9 | 0.7 | 0.8 | 0.1 | 0.2 |
| 8 | 0.0 | 0.1 | 0.0 | 1.6 | 3.1 | 1.5 | 5.6 | 8.1 | 5.7 | 4.3 | 3.5 | 4.3 | 0.5 | 0.1 | 0.6 |
| 9 | 0.0 | 0.2 | 0.1 | 0.4 | 2.8 | 5.3 | 1.1 | 5.5 | 11.6 | 0.7 | 2.0 | 10.0 | 0.2 | 0.2 | 0.9 |
| 10 | 0.0 | 0.0 | 0.0 | 0.7 | 0.2 | 0.0 | 6.0 | 2.0 | 0.3 | 19.2 | 5.1 | 0.9 | 5.8 | 1.3 | 0.6 |
| 11 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.8 | 1.7 | 0.8 | 4.4 | 7.4 | 4.5 | 1.8 | 1.4 | 2.0 |
| 12 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.7 | 0.3 | 2.0 | 5.8 | 0.8 | 5.1 | 20.1 | 0.5 | 1.5 | 6.5 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 2.4 | 0.4 | 0.1 | 3.2 | 2.1 | 0.3 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.1 | 0.3 | 1.0 | 0.3 |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.4 | 2.5 | 0.3 | 2.4 | 3.5 |

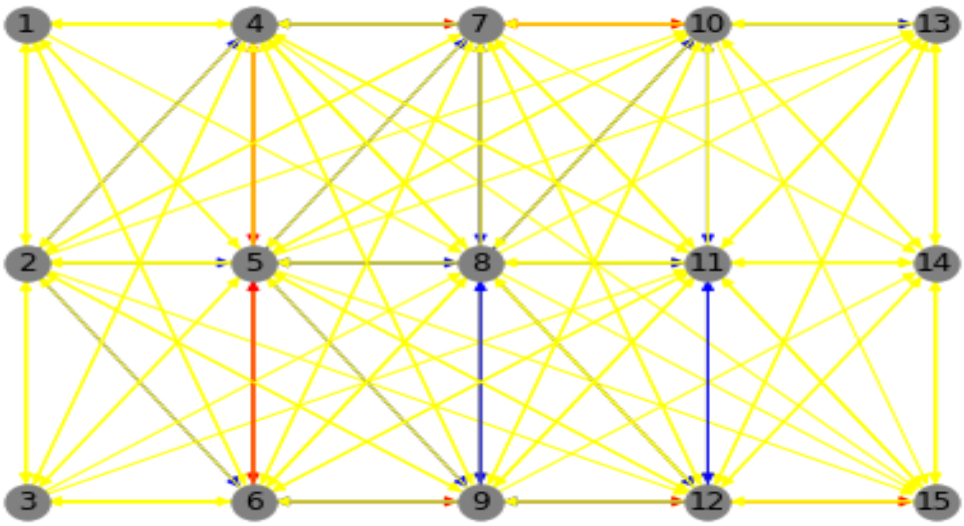



Figure 4.30: Cluster 3 average zone's passing network during 2017/2018 season where 1-3 are zones near teams' defense and goalkeeper. While, 13-15 are zones lying near to teams' attack, and a Red arrow means more than 6 passes between zones, Blue arrow means more than 3 passes and Yellow for less than 3.

### 4.6.4 Examining results

From the shown below results in table 4.36. We can notice that the Premier league has the highest percentage of played matches that belong to cluster 2 which is characterized by the highest number of passes overall, number of passes received in the fourth area and lowest in terms of dependence on long balls to attack. Followed by Serie A, La Liga, Ligue 1 and Bundesliga. We can also notice that Serie A is spread among clusters better than other leagues.

Table 4.36: shows All Leagues evaluation metrics results

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Average number of passes | 265 | 595 | 420 |
| Passes received in the first 20 percent of the field | 18 (7%) | 22 (3%) | 23(5%) |
| Passes received in the second 20 percent of the field | 77 (29%) | 120 (20%) | 118(28%) |
| Passes received in the third 20 percent of the field | 69 (26%) | 176 (29%) | 118(28%) |
| Passes received in the fourth 20 percent of the field | 72 (27%) | 217 (36%) | 121(28%) |
| Passes received in the last 20 percent of the field | 27 (10%) | 57 (5.7%) | 39(9%) |
| Percentage of the long balls received in the attacking area from the total passes | 10% | 5.7% | 7.7% |

## 4.7 Case studies

To dive into a deeper analysis of football tactical patterns, We chose two teams to analyze their opponents' tactical patterns and passing sequences. To try and understand the team's defensive weaknesses. The factors considered are:

- The cluster number that the opposing team belongs to
- The result of the match (Win/Lose/Draw)
- Whether the final standing of the opposing team in top 6 positions or not (qualifying into European major cup)
- Whether the match is played on the home team's stadium or away on the opposing team's stadium
- Total passes occurred during the match by the opposing team
- The percentage of long balls for the opposing team calculated by dividing the passes coming from the first 60% of the field (defensive zone) into the last (fifth) 20% of the field (attacking zone) over the total number of passes received on the attacking zone.
- The percentage of passes received from the opposing team in the left attacking side of the defending team divided by the total passes received in the attacking zone.
- The percentage of passes received from the opposing team in the middle attacking side of the defending team divided by the total passes received in the attacking zone.
- The percentage of passes received from the opposing team in the right attacking side of the defending team divided by the total passes received in the attacking zone.

### 4.7.1 Real Madrid

When investigating table 4.37, containing data about Real Madrid team's fixtures during the 2017/2018 La Liga season, we find those results:

1. Real Madrid lost to three cluster 1, two cluster 2, one cluster 3 teams. Taking into consideration the fewer teams belong to cluster 1 and 3, we can notice that the increased number of overall passes, the less use of long balls for attacking (characteristics of cluster 1 and 3), can increase the chance of winning when facing Real Madrid.

2. Teams that win when playing against Real Madrid depend on long balls in 6.8% of their attacks, while the average dependence on long passes for teams playing against Real Madrid is 8.71%.

3. Teams that win when playing against Real Madrid attack from the left-side of Real Madrid's defense with 44% of their attacks. while the average for all teams playing against Real Madrid is 38%.

4. Teams that win when playing against Real Madrid attack from the center of Real Madrid's defense with 24.55% of their attacks. while the average for all teams playing against Real Madrid is 24.88%.

5. Teams that win when playing against Real Madrid attack from the right-side of Real Madrid's defense with 31.5% of their attacks. while the average for all teams playing against Real Madrid is 37.1%.

6. Opposing teams complete 338.8 passes in the home stadium of Real Madrid. While on the away stadium it was 334.9 passes. This show the dominance of Real Madrid when playing outside their stadium.

7. The average dependence of opposing teams on long balls when facing Real Madrid on its home stadium was 10.4% of the attacks. While it was 7.62% on away stadiums.

8. The average total passes for the top 6 teams (excluding Real Madrid which have the third rank) when playing against Real Madrid was 358.5 passes, higher than average for all teams which were 336.8.

9. We calculated the average long passes done for the Top 6 in La Liga when facing Real Madrid it was 7.18%.

We can safely assume the following after checking the previous factors:

- Using long balls for attacks against Real Madrid can statistically decrease the chance of winning.
- Exploiting the left defensive side when playing against Real Madrid can increase the chance of winning [27].
- Home/Away doesn't affect the number of passes the opposing team makes when playing against Real Madrid.
- The top 6 teams excluding Real Madrid, do more total passes(358.5) than average(336.8) when facing Real Madrid, and they depend less on long balls 6.21%.
- Opposing teams depend more on long balls when playing against Real Madrid in its home stadium.

Table 4.37: shows clustering of La Liga teams.

| | Cluster | W/D/L | Top 6 | H/A | passes | LB | Left | Center | Right |
|---|---|---|---|---|---|---|---|---|---|
| Deportivo La Coruña | 2 | W | No | A | 237 | 17.9% | 48.8% | 20.9% | 30.2% |
| Valencia | 2 | D | Yes | H | 229 | 14.3% | 42.6% | 27.7% | 29.7% |
| Levante | 2 | D | No | H | 181 | 0% | 34.3% | 22.9% | 42.9% |
| Real Sociedad | 1 | W | No | A | 387 | 2.8% | 20.7% | 32.4% | 46.8% |
| Real Betis | 1 | L | Yes | H | 303 | 0.0% | 42.7% | 22.9% | 34.4% |
| Alaves | 2 | W | No | A | 206 | 18.2% | 58.2% | 17.9% | 23.9% |
| Espanyol | 2 | W | No | H | 280 | 22.2% | 44.2% | 23.3% | 32.6% |
| Getafe | 2 | W | No | A | 173 | 6.9% | 44.4% | 22.2% | 33.3% |
| Eibar | 2 | W | No | H | 358 | 5.1% | 47.8% | 21.4% | 30.8% |
| Girona | 2 | L | No | A | 267 | 11.4% | 53.0% | 20.0% | 27.0% |
| Las Palmas | 1 | W | No | H | 375 | 0.0% | 31.7% | 31.7% | 36.7% |
| Atletico Madrid | 2 | D | Yes | A | 268 | 7.9% | 44.4% | 15.9% | 39.7% |
| Malaga | 2 | W | No | H | 302 | 9.1% | 26.1% | 27.5% | 46.5% |
| Sevilla | 1 | W | No | H | 389 | 3.7% | 37.8% | 23.3% | 39.0% |
| Leganes | 2 | W | No | A | 320 | 30.0% | 41.4% | 19.8% | 38.8% |
| Barcelona | 3 | L | Yes | H | 500 | 5.4% | 31.8% | 31.2% | 37.0% |
| Celta de Vigo | 1 | D | No | A | 402 | 13.0% | 33.5% | 27.5% | 39.0% |
| Villarreal | 1 | L | Yes | H | 385 | 4.8% | 39.7% | 21.4% | 38.9% |
| Deportivo La Coruña | 2 | W | No | H | 283 | 3.7% | 29.6% | 23.5% | 47.0% |
| Valencia | 2 | W | Yes | A | 359 | 2.0% | 34.3% | 31.9% | 33.7% |
| Levante | 2 | D | No | A | 195 | 7.7% | 41.7% | 29.2% | 29.2% |
| Real Sociedad | 1 | W | No | H | 481 | 2.4% | 28.9% | 29.4% | 41.7% |
| Real Betis | 1 | W | Yes | A | 442 | 11.4% | 32.1% | 25.7% | 42.2% |
| Alaves | 2 | W | No | H | 248 | 10.0% | 43.9% | 28.1% | 28.1% |
| Espanyol | 2 | L | No | A | 301 | 3.8% | 34.6% | 32.3% | 33.1% |
| Getafe | 2 | W | No | H | 183 | 18.2% | 44.1% | 18.9% | 36.9% |
| Eibar | 2 | W | No | A | 309 | 12.2% | 39.0% | 16.5% | 44.5% |
| Girona | 2 | W | No | H | 266 | 5.9% | 61.9% | 13.4% | 24.7% |
| Las Palmas | 1 | W | No | A | 363 | 8.8% | 38.4% | 22.0% | 39.6% |
| Atletico Madrid | 2 | D | Yes | H | 264 | 18.2% | 27.7% | 33.9% | 38.4% |
| Malaga | 2 | W | No | A | 333 | 4.8% | 52.6% | 16.4% | 31.0% |
| Sevilla | 1 | L | No | A | 295 | 15.4% | 62.3% | 19.5% | 18.2% |
| Leganes | 2 | W | No | H | 324 | 7.3% | 34.9% | 18.8% | 46.3% |
| Barcelona | 3 | D | Yes | A | 469 | 6.5% | 34.3% | 32.9% | 32.9% |
| Celta de Vigo | 1 | W | No | H | 408 | 6.9% | 33.6% | 37.2% | 29.2% |
| Villarreal | 2 | D | Yes | A | 366 | 6.7% | 31.5% | 36.1% | 32.4% |

### 4.7.2 Liverpool

When investigating table 4.38 containing data about Liverpool team fixtures during the 2017/2018 Premier League season, we found those results:

1. Liverpool lost to three cluster 1, one cluster 2, one cluster 3 teams. Taking into consideration the fewer teams belong to cluster 1 and 2, we can notice that the increased number of overall passes, the less use of long balls for attacking (characteristics of cluster 1 and 2), can increase the chance of winning when facing Liverpool.

2. Teams that win when playing against Liverpool depend on long balls in 12.6% of their attacks. while the average dependence on long balls for teams playing against Liverpool is also 16.43%.

3. Teams that win when playing against Liverpool attack from the left side of Liverpool's defense with 41.8% of their attacks. while the average for all teams playing against Liverpool is 38.9%.

4. Teams that win when playing against Liverpool attack from the center of Liverpool's defense with 18.5% of their attacks. while the average for all teams playing against Liverpool is 22%.

5. Teams that win when playing against Liverpool attack from the right side of Liverpool's defense with 39.7% of their attacks. while the average for all teams playing against Liverpool is 39.1%.

6. When checking the total average passes that the opposing teams do in the home stadium of Liverpool it was 263.0 passes. While, on the away stadium it was 293.3 passes. Which shows the dominance of Liverpool team in its stadium "Anfield".

7. When checking the average dependence of opposing teams on long balls when facing Liverpool on its home stadium it was 18.5% of the attacks. While it was 12.78% on away stadiums.

8. When checking the average total passes for the Top 6 teams (excluding Liverpool which had the fourth rank) when playing against Liverpool it was 387.8 passes higher than average for all teams which were 278.1.

9. We calculated the average long balls done for the Top 6 in the Premier League when facing Liverpool it was 14.71%.

We can safely assume the following after checking the previous factors:

- Using long balls for attacks against Liverpool can statistically decreases the chance of winning.

- Exploiting the left-side defense when playing against Liverpool can also increase the chance of winning slightly.
- Teams playing against Liverpool tend to do fewer total passes (263) when playing in Liverpool's home stadium (293.2) and use long balls slightly more often to attack (18.5%) on Liverpool's stadium compared to playing in their home stadium (12.3%). Which shows the importance of playing on the home stadium for teams generally and for Liverpool specially.
- The top 6 teams excluding Liverpool, do more total passes (387.8) than average (278.1) when facing Liverpool, and they depend less on long balls 14.7% less than the average 15.7%.

Table 4.38: shows clustering of Premier League teams.

| | Cluster | W/D/L | Top 5 | H/A | passes | LB | Left | Center | Right |
|---|---|---|---|---|---|---|---|---|---|
| Watford | 3 | D | No | A | 290 | 24.1% | 20.0% | 18.3% | 61.7% |
| Crystal Palace | 3 | W | No | H | 182 | 35.7% | 37.7% | 28.6% | 33.8% |
| Arsenal | 1 | W | Yes | H | 432 | 22.6% | 42.9% | 19.3% | 37.9% |
| Manchester City | 2 | L | Yes | A | 648 | 5.1% | 46.2% | 15.4% | 38.5% |
| Burnley | 3 | D | No | H | 181 | 26.7% | 41.3% | 34.9% | 23.8% |
| Leicester City | 3 | W | No | A | 249 | 13.8% | 32.8% | 21.6% | 45.7% |
| Newcastle United | 3 | D | No | A | 205 | 17.6% | 30.4% | 26.1% | 43.5% |
| Manchester United | 1 | D | Yes | H | 263 | 28.6% | 16.9% | 21.7% | 61.4% |
| Tottenham Hotspur | 1 | L | Yes | A | 265 | 18.2% | 37.2% | 17.9% | 44.9% |
| Huddersfield Town | 3 | W | No | H | 162 | 42.9% | 24.3% | 27.0% | 48.6% |
| West Ham United | 3 | W | No | A | 344 | 10.7% | 34.8% | 29.5% | 35.6% |
| Southampton | 3 | W | No | H | 249 | 6.7% | 54.8% | 16.7% | 28.6% |
| Chelsea | 1 | D | Yes | H | 403 | 15.9% | 29.5% | 28.7% | 41.9% |
| Stoke City | 3 | W | No | A | 251 | 26.3% | 41.2% | 21.2% | 37.5% |
| Brighton | 3 | W | No | A | 400 | 0.0% | 53.2% | 14.4% | 32.4% |
| Everton | 3 | D | No | H | 94 | 28.6% | 31.0% | 19.0% | 50.0% |
| West Bromwich | 3 | D | No | H | 194 | 13.0% | 48.8% | 21.2% | 30.0% |
| Bournemouth | 3 | W | No | A | 402 | 10.7% | 37.9% | 21.0% | 41.1% |
| Arsenal | 1 | D | Yes | A | 387 | 10.3% | 37.3% | 25.3% | 37.3% |
| Swansea City | 3 | W | No | H | 410 | 17.9% | 41.0% | 22.9% | 36.2% |
| Leicester City | 3 | W | No | A | 246 | 13.3% | 46.4% | 21.7% | 31.9% |
| Burnley | 3 | W | No | A | 264 | 11.5% | 35.5% | 18.7% | 45.8% |
| Manchester City | 2 | W | Yes | H | 557 | 7.7% | 26.4% | 31.4% | 42.1% |
| Swansea City | 3 | L | No | A | 204 | 18.2% | 31.5% | 20.5% | 47.9% |
| Huddersfield | 3 | W | No | A | 211 | 7.7% | 30.6% | 19.4% | 50.0% |
| Tottenham Hotspur | 1 | D | Yes | H | 464 | 17.1% | 36.5% | 22.5% | 41.0% |
| Southampton | 3 | W | No | A | 436 | 5.9% | 46.8% | 16.7% | 36.5% |
| West Ham United | 3 | W | No | H | 244 | 19.2% | 54.2% | 12.7% | 33.1% |
| Newcastle United | 3 | W | No | H | 215 | 17.6% | 54.3% | 21.4% | 24.3% |
| Manchester United | 1 | L | Yes | A | 196 | 14.3% | 55.8% | 11.7% | 32.5% |
| Watford | 3 | W | No | H | 314 | 4.3% | 37.4% | 20.6% | 42.1% |
| Crystal Palace | 3 | W | No | A | 215 | 7.1% | 57.9% | 25.0% | 17.1% |
| Everton | 3 | D | No | A | 230 | 9.7% | 56.1% | 14.0% | 29.8% |
| Bournemouth | 3 | W | No | H | 295 | 4.2% | 39.3% | 17.9% | 42.9% |
| West Bromwich | 3 | D | No | A | 252 | 11.1% | 25.5% | 33.0% | 41.5% |
| Stoke City | 3 | D | No | H | 147 | 14.3% | 44.6% | 23.2% | 32.1% |
| Chelsea | 1 | L | Yes | A | 262 | 7.3% | 38.3% | 27.0% | 34.8% |
| Brighton | 3 | W | No | H | 194 | 28.6% | 21.6% | 27.5% | 51.0% |

# Chapter 5

# Conclusion and future work

## 5.1 Conclusion

In this thesis, the aim was to identify the tactical patterns of football teams using an unsupervised machine learning approach. We considered the passes that every team made as input to the k-means and Hierarchical clustering models to classify the teams into clusters. In the five top European leagues. The top 6 teams had most of their matches clustered in one cluster that was characterized by the high overall number of passes, the domination of the middle area of the football pitch, and the dependence on short passes to attack, Our results agree with [28]. For the Premier Legue, Ligue 1 and the Bundesliga the league winner had different tactical style than other teams. While in Serie A Napoli, the runner-up had this unique play style and for La Liga Barcelona (winner) and Real Madrid (third) had a similar play style.

When comparing leagues we found that the Premier League is the league with most passes and less dependence on long balls for attacking. While, Ligue 1 and Bundesliga had the least number of passes and increased dependence on long balls for attacking. This may clarify why Ligue 1 and Bundesliga are considered the weakest leagues among the top 5 [26].

Also, we did 2 case studies. One was for Spanish Real Madrid team and we found that using long balls for attacks and exploiting the left defensive side against Real Madrid can statistically increase the chance of winning [27], and we found similar results that "Real Madrid" as a top team depend less on long balls and do more passes than the average team. Teams playing against Real Madrid in their home stadium didn't show increased number of passes, which shows the dominance of Real Madrid team outside its field.

As for the English team "Liverpool", we found that teams playing against Liverpool tend to do fewer total passes (263) when playing in Liverpool's home stadium than the average (293.2) and use long balls slightly more often to attack (18.5%) on Liverpool's stadium compared to playing in their home stadium(12.78%). In addition, we found that exploiting the left-side defense when playing against Liverpool can also increase the chance of winning slightly. And we found similar results that "Liverpool" as a top team depend less on long balls and do more passes than the average team.

## 5.2 Future work

Additional improvements could be made to identify the football teams' tactical patterns, approaches considered by this thesis:

- Adding teams' starting formation to the dataset can help in better identifying the tactical patterns.
- Increasing the dataset size to include more seasons instead of only one. To confirm the accuracy of the results. and study the development of teams' tactical patterns during different seasons
- Defining labeled dataset for football teams' tactics to open the possibility of the use of supervised machine learning algorithms.
- Analyzing the resulted networks using network analysis techniques may help understand the teams' play style better.
- Considering other events (shots, fouls, duels, free kicks, offsides) can help in better identifying the football teams' tactics.
- Examining the passing sequences that lead to goals can help identify strengths and weaknesses in teams.

# Appendix

# Appendix A

# Lists

# List of Figures

# Bibliography


[1] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):1–15, 2019.

[2] Paolo Cintia, Luca Pappalardo, and Dino Pedreschi. " engine matters": A first large scale data driven study on cyclists' performance. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 147–153. IEEE, 2013.

[3] John Hollinger. The player efficiency rating, 2009.

[4] Antonio Terroba, Walter A Kosters, and Jonathan K Vis. Tactical analysis modeling through data mining. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*. Citeseer, 2010.

[5] Lloyd Smith, Bret Lipscomb, and Adam Simkins. Data mining in sports: Predicting cy young award winners. *Journal of Computing Sciences in Colleges*, 22(4):115–121, 2007.

[6] Charles Reep and Bernard Benjamin. Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, 131(4):581–585, 1968.

[7] Joe Sykes and Neil Paine. How one man's bad math helped ruin decades of english soccer. *Five Thirty Eight*, 2016.

[8] Tom Decroos, Jan Van Haaren, and Jesse Davis. Automatic discovery of tactics in spatio-temporal soccer match data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 223–232, 2018.

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

[10] Joel Brooks, Matthew Kerr, and John Guttag. Developing a data-driven player ranking in soccer using predictive model weights. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 49–55, 2016.

[11] Javier Fernandez and Luke Bornn. Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *Sloan Sports Analytics Conference*, volume 2018, 2018.

[12] Xinyu Wei, Long Sha, Patrick Lucey, Stuart Morgan, and Sridha Sridharan. Large-scale analysis of formations in soccer. In *2013 international conference on digital image computing: techniques and applications (DICTA)*, pages 1–8. IEEE, 2013.

[13] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[14] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003.

[15] Vincent Cohen-Addad, Varun Kanade, Frederik Mallmann-Trenn, and Claire Mathieu. Hierarchical clustering: Objective functions and algorithms. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 378–397. SIAM, 2018.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[18] Shoji Hirano and Shusaku Tsumoto. Grouping of soccer game records by multiscale comparison technique and rough clustering. In *Fifth International Conference on Hybrid Intelligent Systems (HIS'05)*, pages 6–pp. IEEE, 2005.

[19] Raúl Montoliu, Raúl Martín-Félez, Joaquín Torres-Sospedra, and Adolfo Martínez-Usó. Team activity recognition in association football using a bag-of-words-based method. *Human movement science*, 41:165–178, 2015.

[20] Luca Pappalardo and Paolo Cintia. Quantifying the relation between performance and success in soccer. *Advances in Complex Systems*, 21(03n04):1750014, 2018.

[21] Joel Brooks, Matthew Kerr, and John Guttag. Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(5):338–349, 2016.

[22] Donald Barron, Graham Ball, Matthew Robins, and Caroline Sunderland. Artificial neural networks and player recruitment in professional soccer. *PloS one*, 13(10), 2018.

[23] Alex Rathke. An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport and Exercise*, 12(2):514–529, 2017.

[24] Paolo Cintia, Salvatore Rinzivillo, and Luca Pappalardo. Network-based measures for predicting the outcomes of football games. In *MLSA@ PKDD/ECML*, pages 46–54, 2015.

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

[26] UEFA. Uefa country coefficients. https://www.uefa.com/memberassociations/uefarankings/country/#/yr/2018, July 2018.

[27] as. Real madrid defensive weakness. https://en.as.com/en/2018/05/24/opinion/1527152316_826233.html, May 2018.

[28] Joel Oberstone. Differentiating the top english premier league football clubs from the rest of the pack: Identifying the keys to success. *Journal of Quantitative Analysis in Sports*, 5(3), 2009.