BARÇA INNOVATION HUB Universitas

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
School of Professional & Executive Development

Sports Analytics Postgraduate Final Project

Analysis of team passing networks depending on the possession outcome

Authors

Marc Estévez Inglada

Alex Marin Felices

Tutors

Javier Martín Buldú

Sergi Nadal Francesch

2022-2023

1. Introduction

Football is one of the most popular and widely played sports in the world. It is also a complex system that involves multiple agents (players, coaches, referees) interacting with each other under uncertain and dynamic conditions. The outcome of a football match depends on many factors such as individual skills, team tactics, opponent strategies, environmental conditions, and random events.

Traditionally, football analysis has relied on statistics such as goals scored or conceded, shots on target, possession percentage, or passes completed. However, these statistics do not capture all the aspects of the game that influence its result. For instance, they do not account for the quality or difficulty of the passes, the spatial distribution or movement of the players, or the temporal evolution or context of the match.

In recent years, network science has emerged as a powerful tool for analyzing football data from a new perspective. Network science is an interdisciplinary field that studies complex systems as networks or graphs, where entities are represented by nodes and interactions are represented by edges. Network science can be applied to various domains such as social sciences, biology, physics or computer science. In sports analytics, network science can be used to model and analyze team dynamics, strategies and performance.

The main advantage of network science applied in football is that it can capture and measure aspects of the game that are not easily observable or quantifiable by traditional statistics. Network science can reveal how players interact with each other on the pitch, how they create opportunities for scoring or defending goals, how they adapt to different situations or opponents, and how they contribute to the overall performance of the team. Moreover, network science can consider different levels of granularity and complexity of the game, such as its dynamics, space, time, and multilayer nature.

In this project, we review the current state of the field and some articles that apply network analysis methods to study the passing networks of football teams at different levels of granularity and complexity. We propose a new research project that aims to extend their work by incorporating a different approach and defining different networks based on how the play ended. We then highlight the main findings and contributions of this work and what could be done as future work.

The repository for this project can be accessed here:
https://github.com/alex11marin11/Analysis-of-team-passing-networks-depending-on-the-possession-outcome

2. Related Work

Network science is an interdisciplinary field that studies complex systems as networks of interacting elements (Newman, 2010). Network science has been applied to various domains, such as biology, sociology, physics, and computer science. In recent years, network science has also emerged as a powerful tool to analyze sports data, especially football (Gudmundsson and Horton, 2017).

Football is a complex sport that involves multiple players interacting dynamically in space and time to achieve a common goal. As we explained, traditional statistics based on individual performance or aggregated team measures may not capture the richness and complexity of the game. Network science offers a new perspective to study football by representing teams as passing networks, where nodes are players or regions of the pitch and links are passes between them. Passing networks can reveal the structure, dynamics, and evolution of team organization and performance during a match or a season.

Several studies have used network science to analyze football passing networks at different levels of granularity and abstraction. Some studies have focused on player passing networks, where nodes are players and links are weighted by the number of passes between them (Passos et al., 2011; Grund, 2012). These studies have investigated how network properties such as centrality, clustering, efficiency, modularity, and assortativity relate to team performance indicators such as ball possession, shots on goal, goals scored or conceded.

Other studies have considered pitch passing networks, where nodes are specific regions of the field connected through passes made by players occupying them (Cintia et al., 2015). These studies have explored how spatial patterns of passing vary across teams, positions, and match situations, and how they affect team outcomes such as scoring opportunities, defensive pressure, or territorial dominance.

A third type of passing network is the pitch-player passing network, where nodes are a combination of a player and its position at the moment of the pass (Cotta et al., 2013; Narizuka et al., 2014). These studies have examined how players move and distribute themselves on the pitch, and how their spatial positioning influences their passing behavior and network role.

In addition to these three types of passing networks, some studies have also incorporated other dimensions such as time and layers to capture the dynamic and multilayer nature of football passing networks. For instance, Buldú et al. (2018) proposed a framework to analyze football passing networks along four dimensions: dynamics, space, time, and layers. To do so, the authors construct different types of networks based on passes between players or regions of the pitch, and compute various network metrics to characterize their structure and evolution. They applied

this framework to study one season of the Spanish national league and compared different teams based on their network signatures.

One of the teams that stood out to these researchers was F.C. Barcelona coached by Pep Guardiola, which has been considered one of the best teams in football history. Buldú et al. (2019) further investigated this team and used data from 38 matches of the Spanish national league (La Liga) in the season 2008-2009, where Barcelona won with a record-breaking score of 99 points out of 114 possible. By using network science this research paper identified several features that characterized its style of play: high connectivity, clustering, entropy, centrality diversity, spatial diversity, temporal diversity, and high layer diversity and short distances and low modularity.

Some of the information that can be extracted from network science applied in football includes:

- The connectivity and cooperation among players: This can be measured by degree centrality, which indicates how many passes each player receives on average. A higher degree centrality means that a player is more involved in the game and has more options for passing.
- The cohesion and coordination among players: This can be measured by clustering coefficient, which indicates how likely it is that two players who share a common teammate also pass to each other. A higher clustering coefficient means that players tend to form triangles or cliques with their teammates.
- The structure and hierarchy of the team: This can be measured by betweenness centrality, which indicates how often a player acts as a bridge or intermediary between other players. A lower betweenness centrality means that there are fewer players who control the flow of passes in the team.
- The style and speed of play: This can be measured by efficiency, which indicates how quickly any node can reach any other node with fewer steps or passes. A higher efficiency means that a team plays faster and more directly.
- The homogeneity and integration of the team: This can be measured by modularity, which indicates how much the network is divided into subgroups or communities of players. A lower modularity means that a team is more homogeneous and integrated.

3. Problem statement

As mentioned in the section above there are a lot of different conclusions or findings that can be extracted from applying network science in sports analytics. In our case, we found it interesting to define 3 different types of networks referring to 3 different situations of the game that can occur. For any possession in football, there are basically 3 outcomes: (A) The action finishes but the team keeps the ball. (e.g. a player is fouled, or you get a corner kick). (B) The action ends up in a shot. (C) The action ends with the ball controlled by the opponent. For this study we decided to focus on finding a possible difference between the last 2 cases. Are there players who participate in more actions that end up in shots and some that participate more in the ones where their team loses the ball? This was the main focus for the project.

In the project, we wanted to use event datasets of football matches during a whole season to construct and analyze team passing networks. Our objective was to use a series of network metrics to understand how different teams organize their passing networks, if there is one or more key players and which are those key players. Next, we wanted to study how the passing networks change their structure depending on the final result of the action (two different networks, the first one with the passes where the action finishes with a shot, which is sometimes mentioned in the paper as Shot Network, and the second one with the passes that at the end the team loses the ball, which will be called Lost Network), and compare it to the complete network with all the passes in a match.

4. Data

4.1. Data description

The data used in this study consists of four different files for each of the 462 matches that were played in the season 2021/2022 from LaLiga Smartbank and it was extracted from Opta. The first files contain the identifier and the name of the teams that played each match. The teams that took part in the competition the season 2021/2022 can be seen in the Figure 1 with the final standings.

#	EQUIPO	PJ	G	E	P	G	PTS
1.	Almería	42	24	9	9	68:35	81
2.	Real Valladolid	42	24	9	9	71:43	81
3.	Eibar	42	23	11	8	61:45	80
4.	Las Palmas	42	19	13	10	57:47	70
5.	Tenerife	42	20	9	13	53:37	69
6.	Girona	42	20	8	14	57:42	68
7.	Real Oviedo	42	17	17	8	57:41	68
8.	Ponferradina	42	17	12	13	57:55	63
9.	Cartagena	42	18	6	18	63:57	60
10.	Real Zaragoza	42	12	20	10	39:46	56
11.	Burgos CF	42	15	10	17	41:41	55
12.	Leganés	42	13	15	14	50:51	54
13.	Huesca	42	13	15	14	49:44	54
14.	Mirandés	42	15	7	20	58:62	52
15.	UD Ibiza	42	12	16	14	53:59	52
16.	Lugo	42	10	20	12	46:52	50
17.	Sporting de Gijón	42	11	13	18	43:48	46
18.	Málaga	42	11	12	19	36:57	45
19.	Amorebieta	42	9	16	17	44:63	43
20.	R. Sociedad B	42	10	10	22	43:61	40
21.	Fuenlabrada	42	6	15	21	39:65	33
22.	Alcorcón	42	6	11	25	37:71	29

Figure 1: Standings of La Liga Smartbank 2021/2022

Source: flashscore.es

The second group of files contain information about the players that were in the squad list of both teams that played each match. There are 788 different players that were at least in one squad list during the whole competition. The information provided for each player is: the player identifier, the first and the last name, the name for what the player is known, the identifier of the team he belonged to, the number the player wore, the position where he played, if he started playing the match or not and the number of minutes he played. The possible values that the position variable

can have are: Goalkeeper, Defender, Midfielder, Striker and Substitute if the player was not in the starting line-up. So, there is just a position defined for the players that started playing the matches.

The third files consist of the passes that were made at each match. An average of over 897 passes were made between both teams in the matches played in LaLiga Smartbank the season 2021/2022. For each pass, it is included the match and the team identifier, an identifier for the pass, if it was made in the first or the second half, the minute and the second it was made, the identifiers of the players that made the pass and the one that received it, the outcome of the pass (1 if it succeeded and 0 if not), the origin and destination position of the pass and the number of the possession and the sequence the pass belonged to. Note that a new possession starts every time the team that has the ball changes, whereas there can be many sequences of passes in the same possession as the sequence change when the game is interrupted, but the team that has the ball can still be the same maintaining the possession. In the matches that are observed, there were on average 183 possessions and 285 sequences of passes.

The fourth group of files contain information about the shots that were made at each match. On average, more than 24 shots were made in the matches of that season of the second spanish division. For each shot made, the files have information of the identifier of the match where the shot was made, if the shot was done on the first or in the second half, the minute and the second it was made, how the shot ended, the position where the shot was made, the identifier, name, position and team from the player that made the shot, the team, identifier and name of the goalkeeper that receives the shot, the expected goal probability and some other details about the shots that are not used.

4.2. Data management

To be able to analyze how the teams pass the ball between their players and see if there are any key players, doing it also depending on the outcome of the pass sequences, we need to collect the information provided by the different files and create a database for each match with the passes that we are interested to study for each of the cases.

For the general case, we need to collect all the passes that are done in a match. To do so, we take the passes files for each match, containing the information mentioned before about all the passes that are done in that match, and we add the name of the team in possession of the ball from the teams file and the names of the players that are involved at each pass (the one that passes the ball and the one that receives it). Then, from all the passes we keep the ones that were successful, as we need to know who is the player receiving the ball in order to build later the passing networks for each team and match. Finally, the passes made by each team are filtered, obtaining for each team and for each match a database with the total passes made

with origin and destination player. On average, there are more than 315 passes per match and per team in the final databases.

Now, we want to obtain the sequence of passes made in a match that led to a shot in order to get those passes that ended in a positive action for the team having a chance to score. To do so, we take a copy of the total passes obtained previously and we add to them the shots made in the match from the shots files. Then, we order the actions (passes and shots) by the time of the match they were made and we keep the last sequence of passes that are made before each shot. As a result, filtering by the passes made by each team, we obtain the passes which led to a shot for each team at each match of the season, which are on average almost 40 passes.

Finally, we need to get the sequences of passes that ended with the team losing the ball. To do so, from the passes and shots database of each match we keep the sequences of passes where the team that possessed the ball in the following sequence of passes had changed and that did not come before a shot. Filtering for each of the teams that played each match we obtain the sequences of passes that led to losing the possession. There are on average more than 196 passes per team and per match from sequences of passes that end in the ball being lost.

At the end, from each match and team we obtain three different databases with information about the passes that are made (total passes, passes that led to a shot and passes that ended in losing the ball), which will be used to obtain the number of passes of each type made and received by any pair of players and to create three different networks of passes. From these networks we will compute the centrality values of its players, as it is explained in the methodology, getting for each player from each team and match the number of passes done and received and their centrality values in the different networks, as well as the minutes and the position they played that are obtained from the players files.

5. Methodology

5.1. Basic networks concepts

To analyze how the different teams from LaLiga Smartbank 2021/2022 performed when they were in possession of the ball and identify the key players when passing the ball depending on the outcome of the plays, three different passing networks were created for each team from each game of the season. The first group of networks contain all the passes done by each team in each of the matches they played, the second ones the passes of all the sequences that end up with a shot and the last networks contain the passes of the plays that end in losing the ball. Now, we define what a network is.

Definition 5.1.

A network N = (V, E) consists of a set V of vertices and a collection E containing the pairs of vertices that are connected called edges.

In our approach, the vertices from the different networks that are created consist of the players from the team that played that match and the edges connect two players if a pass was made between them during that match. In this case, we work with directed networks, which means that the edges are made of ordered pairs of vertices having an initial and a destination vertex, which gives importance to who is the player who gives the pass and who is the ball receiver, being a different edge if the pass goes in the other direction. In addition, the edges of the different networks have weights which represent for each edge the number of passes that are given from one player to the other. Finally, adjacency matrices are used in order to represent these networks.

Definition 5.2.

An adjacency matrix A of a directed weighted network is a $n \times n$ matrix, where n is the number of vertices, in which a_{ij} is the weight from the edge i to j. Note that if there is no edge from node i to j, then $a_{ij} = 0$.

Once we have represented the different networks for each team at each match they played by their adjacency matrix, we need to analyze these networks to understand which are the players that contribute more in the offensive phases by making more passes and how the teams distribute the passes they make between their players.

5.2. Eigenvector centrality

To extract relevant information about the importance of the players and be able to analyze the passing networks, the eigenvector centrality of each player in the different networks is calculated.

Definition 5.3.

The eigenvector centrality for node i is the i^{th} element of the eigenvector v in $Av = \lambda v$, where A is the adjacency matrix and λ is the largest eigenvalue of A.

The eigenvector centrality measures the influence that a node has in a network, considering also how important its neighbor nodes are in the network. This metric allows us to measure the relevance of each player in the passing network, indicating which ones are more central and the ones that take less part in the construction of the plays. The calculation of the eigenvector centralities has been done using the eigenvector_centrality function from the NetworkX library version 3.0 in Python.

Finally, with the passing networks and their eigenvector centralities computed, we analyze how the centralities of the players from all the teams are distributed and which were the most relevant players of the LaLiga Smartbank 2021/2022 for each type of network. In addition, the difference between the centralities in the passing networks which plays end in shot and the ones which ends in a lost ball is calculated to see the players that have more impact in positive plays than in passes whose outcome is negative for the team. Then, it is compared how heterogeneous (if the players from each team have similar centrality values or not) the total passing networks from the teams are by computing the standard deviations of the centralities of their players and at the end an example of a passing network with a very relevant player and another without any of note performance are shown.

6. Results

In this section we present the results obtained when calculating the eigenvector centralities for the passing networks of each team at each match played during the 2021/2022 season of LaLiga Smartbank. First of all, we analyze which are the distributions of the centralities of the players in the passing networks, which players did a great performance in only one match and which are the ones that performed better along the whole season for each of the different networks (total passes, sequences of passes that end in a shot and the ones in the ball being lost) and for the difference between the Shot centrality and the Lost centrality. Then, we study how heterogeneous are the total passing networks of the different teams by analyzing the variability of their players' centralities. Finally, we show two examples of passing networks, the first one which contains a really relevant player and the second one which the importance is more homogeneously distributed between their players.

6.1. Players performance

6.1.1. Centralities distributions and top performances in a match

First of all, we wanted to analyze the performance of the players in the passing networks and how the values of the eigenvector centralities were distributed in order to know which are common values for the players and which ones are standing out from the others being really relevant in a certain match or as per usual for their team. To do so, for all the centrality values obtained we do not consider the ones from the goalkeepers, as they are special players which can be a source of variability as teams use them in very different ways. In addition, as a first approach we wanted to compare just the centrality values from the players that had played the whole match, so that they are important players for their team and this way they have played the same number of minutes and had the same time to get involved in the passes of the team.

As a result, the distributions of the centralities of the outfield players that played the whole match for each of the three different passing networks are shown in Figure 2(A). It can be seen that the centrality distributions from the total passing networks and the ones that end in the ball being lost are really similar, with the distribution from the Shot passes Networks having a different pattern. This can be due to the fact that there are much more passes done in the Lost passes Networks than in the Shot passes Networks, as in matches there are always more lost balls than shots, being both networks a subset of the one with the total passes. In addition, it can be seen that most of the players have centrality values between 0.2 and 0.5 in the total and Lost passes Networks, whereas in the network of the passes that led to a shot there are more small and high values. This means that there are more players that are not

much involved in the sequences of passes that end in a shot and also more players that are relevant in these kinds of plays. This makes sense, as the shots are made from more advanced positions in the field and it is easier that offensive players are more involved in these passes and the defensive players do not participate that much.

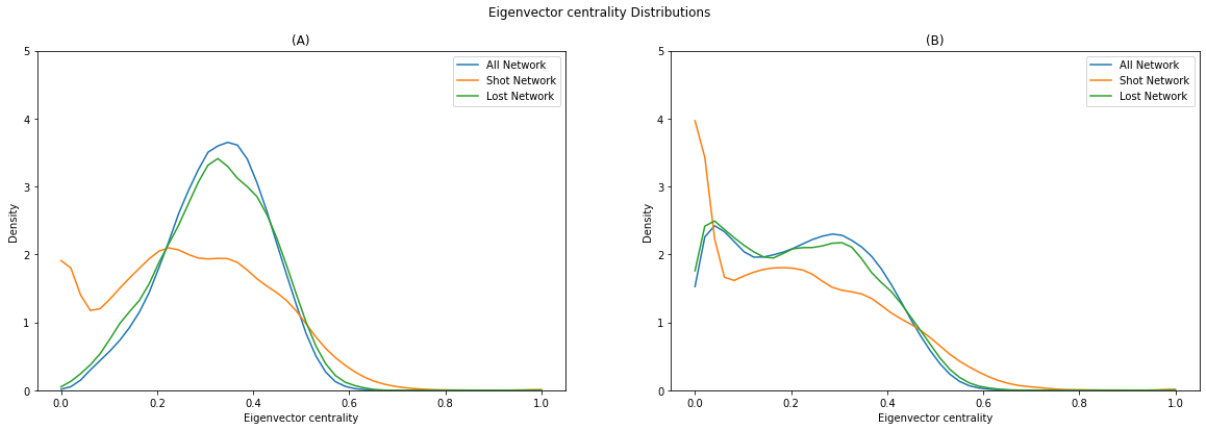

Figure 2: Distributions of the eigenvector centralities for the three types of networks. In (A) we removed goalkeepers and only kept the instances where a player played the whole match. In (B), we also removed goalkeepers but kept all players that played at least 360 minutes during the whole season.

This approach gives us a general idea of how the centralities of the different networks should be distributed in players that play the whole match. The problem with just considering the players that played the whole match is that we are losing many centrality values, some of them from players that were relevant in the passing networks and that played a large amount of minutes, as usually 5 substitutions per match are made, having only 6 remaining players per match that play all the minutes (5 if not considering the goalkeeper, as it is almost never substituted). In addition, most of the substitutions made involve midfielders or strikers, having much more defenders in the remaining centralities when filtering the players that played all the match.

To solve these issues, we consider again the centralities from all the players except the goalkeepers. This way we are obtaining the importance a player had in the passing network in a certain match no matter how many minutes he played. If a player played just a few minutes, then it is the time he had to contribute in the passing network of that match and it is considered as the relevance he had in the match. Nevertheless, we still want to keep the centrality values from the players that were important for their teams and not the ones that did not play enough minutes during the season or that just played one match with the team, which do not represent the usual way of playing from that team. For this reason, we keep the centrality values from the players that have played at least 360 minutes (4 whole matches, which accounts for approx 10% of the 42 total matches) with their team

which would also give the possibility to these players to have values of centralities from different matches so that a high mean centrality is not a result of just one outstanding performance. It is important to remark that a player could have played for more than one club, so the performances are separated and counted as if they were two different players.

The distributions of the centralities of these players (not considering again the goalkeepers) in the different networks are shown in Figure 2(B). It can be seen that, again, the distributions from the All passes Networks and the Lost passes Networks are really similar, whereas the Shot passes Networks have more really low and high values. In all the different networks there can be seen a greater amount of lower centrality values than using the previous approach due to the players that played a few minutes in the matches are expected to have low relevance in the passing networks.

Now, we analyze which had been the top performances from a player in a match in the different passing networks. In Table 1 there are shown the 10 highest values of the eigenvector centralities from the total passes network (eigen_cent_all). It can be seen that eight out of ten players that were more central in their passing network played as a defender, whereas the other two cases (Gorka Guruzeta, striker and Borja Sánchez, midfielder) come from matches where less passes were made. It is also relevant to mention that there are 3 values that come from Real Sociedad B matches, two of them belonging to Aritz Arambarri and being the only player that is repeated in the list. It can also be seen that the defenders centrality values of the Shot passes Networks (eigen_cent_shot) tend to be lower, whereas the ones from the Lost passes Networks (eigen_cent_lost) are more similar to the total ones, being the difference usually negative for them (*eigen_cent_diff = eigen_cent_shot - eigen_cent_lost*).

Top	Player	Team	Position	Minutes played	Passes done	Passes received	eigen_cent_all	eigen_cent_shot	eigen_cent_lost	eigen_cent_diff
1	Gorka Guruzeta	Amorebieta	Striker	89	8	32	0,7075	0,7266	0,6800	0,0466
2	Ignasi Miquel	Huesca	Defender	90	64	72	0,6330	0,4505	0,6207	-0,1702
3	Lluís López	Real Zaragoza	Defender	90	80	84	0,6149	0,3808	0,6312	-0,2504
4	Aritz Arambarri	Real Sociedad B	Defender	84	71	72	0,6139	0,5569	0,6196	-0,0628
5	Borja Sánchez	Real Oviedo	Midfielder	81	21	33	0,6093	0,0023	0,5463	-0,5440
6	Urko González de Zárate	Real Sociedad B	Defender	90	73	74	0,5935	0,3263	0,5873	-0,2610
7	Juanpe	Girona	Defender	90	78	85	0,5923	0,5687	0,6238	-0,0552
8	Pablo Vázquez	FC Cartagena	Defender	90	77	79	0,5889	0,5913	0,5739	0,0174
9	Jean-Sylvain Babin	Sporting de Gijón	Defender	90	50	51	0,5864	0,1997	0,5958	-0,3961
10	Aritz Arambarri	Real Sociedad B	Defender	90	63	70	0,5854	0,3363	0,5968	-0,2605

Table 1: Top 10 players by single match performance based on eigenvector centrality in the All passes Network.

In the tables A.1 and A.2 from the annex the same results are shown, but for the top 10 match performances in the Shot and in the Lost passes Networks respectively. First, it can be seen that the top centrality values from the Shot passes Networks have really high values (0.999), which happens as there are a few pass sequences in those matches that end with a shot and the player that is shooting does not make any pass of that type to any other player. There are 12 cases when the Shot passes Networks have a player with this centrality value, where three are from the Amorebieta and two of each from Burgos CF and FC Cartagena. What can be obtained from the overall top centralities from the Shot passes Networks is that there are more strikers and midfielders. Finally, the characteristics from the top performances in the Lost passes Networks are similar to the total passes networks, as there are many repeated players, nine out of ten are defenders and the team with more performances in the top ten is the Real Sociedad B again.

6.1.2. Top average performances

For the following analysis we decided to look at the performance of the players based on the whole season average. This way, we can find over the course of an entire season how important a player has been for its team. We have looked at this for each of the 3 different networks we have built to see the main differences, specially between the shot and the lost ones. We extracted the top 10 players (they at least need to have participated in 360 minutes during the season for their team), and we have ranked them by their average eigenvector centrality during the season. We also plotted the standard deviation to give us an idea of how consistent that player was; showing whether a player was consistently being an important player in the team, or having more ups and downs.

Overall, it can be seen by looking at the general structure of the 3 cases, that again, the All Network and Lost Network look very similar while the Shot Network looks quite different. The first two cases have players very close among them in terms of the mean value; being all of them between 0.4 and a bit under 0.5. In terms of the standard deviation, we can see the values are generally shorter than in the Shot Network case.

We can also look at it from the names perspective. In the (A) and (C) we can see the first 8 players are exactly the same but in a different order. This also shows that these 2 networks are very similar and coincides with the fact that most of the passes in a match end up in a ball lost instead of a shot which is a lot more rare. Here we can also see a difference in the position of the players. While most of the players in the All and Lost networks are defenders, in the Shot case we can find more Strikers and Midfielders.

Because of this, we also see that the plot in (B) looks quite different. We see that mostly, there is a larger standard deviation in this kind of network. We can also

observe that the biggest value is more or less the same but the other ones are smaller, with all of them being around 0.4 or lower.

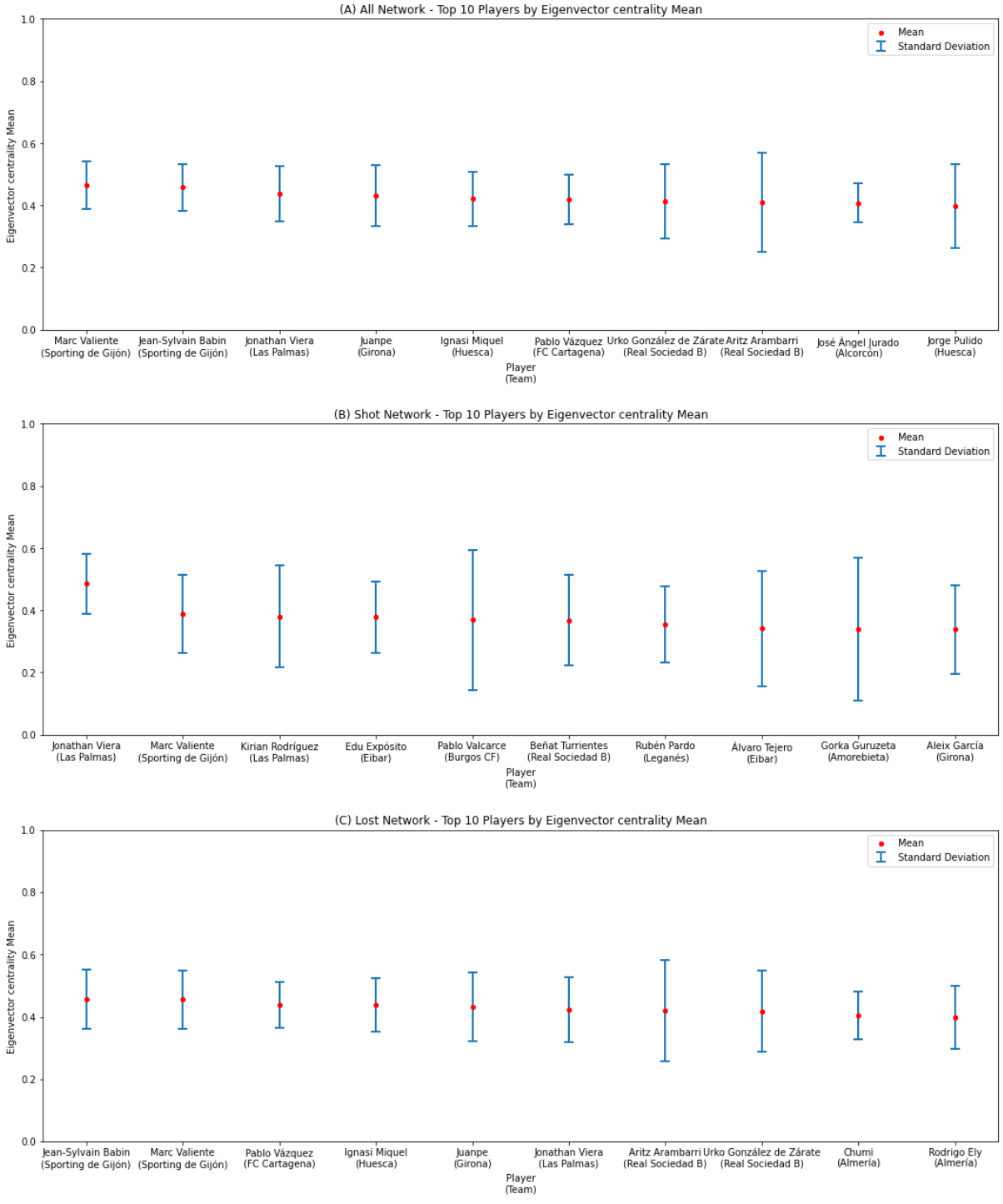

Figure 3: Top 10 players by the eigenvector centrality average score over the whole season. Standard deviation also shows the different variability of the values along the season Each subplot is based on a different network type; (A) All, (B) Shot and (C) Lost.

In the next section where we perform a team analysis we will see more about this, but we can see that some teams tend to appear a lot more on the top of the rankings. We see some appear in all 3 networks and in certain cases we can even see multiple players from the same team in the top 10. Sporting de Gijón has the top 2 in the All and Lost networks and the second one in Shots one. Real Sociedad B appears again a lot in the top spots, in all 3 network types with multiple players, and Girona, Huesca, Las Palmas and Cartagena also appear multiple times in the rankings. As a curious note, we can detect that Almería does not appear in the first 2 plots, but has 2 players in the top 10 of the Lost network.

6.1.3. Top average performances: Shot vs Lost networks

For the next analysis, we focus on one of the main topics and focus points of the projects. The goal of this part was to analyze if we could find players which participated and were more important in plays that ended up in shots, while some other players were more involved in plates where the ball was lost to the opponent.

In Figure 4, we can observe a similar plot to Figure 3. In this case, the plot displays the top 10, and also the bottom 10 players, and it shows the difference in the mean values between the Shot and Lost networks. It can be seen that there are some players who are definitely more involved in shot plays rather than in the ones where the team loses the ball. On the contrary, we can also find players which participate more in lost actions than in shot actions. If we look deeper into the players, we can find something quite interesting but probably expected. All the players in the top 10 are Strikers or Strikers and Midfielders, while all of the players in the bottom 10 are defenders.

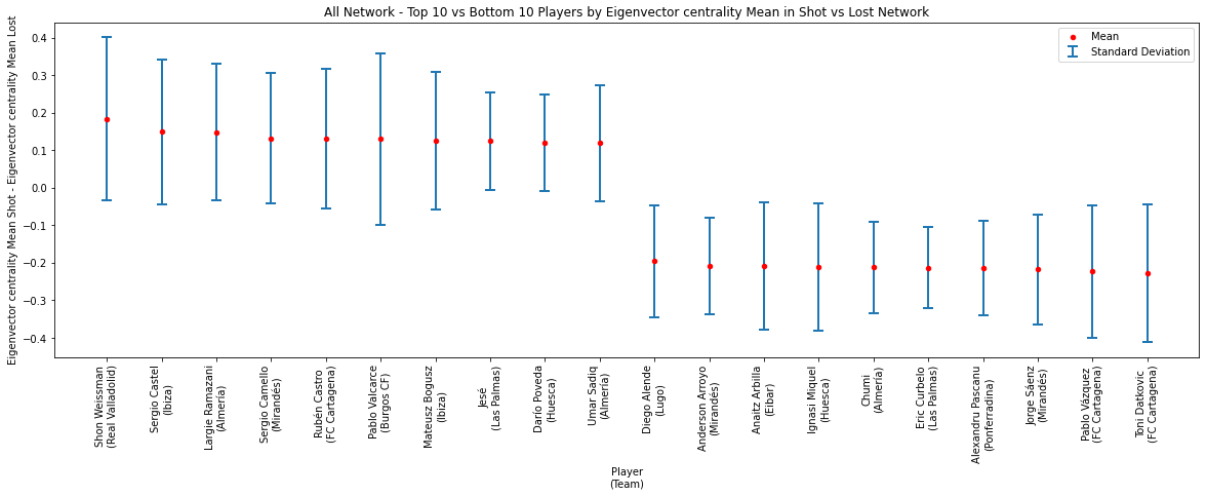

Figure 4: Top 10 and Bottom 10 players ranked by the difference in the eigenvector centrality mean in the Shot Network vs the Lost Network.
(*eigen_cent_diff = eigen_cent_shot - eigen_cent_lost)*

From the teams and concrete players we do not observe anything relevant or aligned with what we have observed in the previous sections. The teams and players that appear in this plot vary from the other top 10s from the Lost and Shot networks which also may show that these 2 networks are quite different.

6.2. Teams passing networks homogeneity

In this section we want to analyze the individual values obtained from the eigenvector centralities of the total passes network, but this time with a team perspective and compare the results obtained between the teams. To do so, we analyze the homogeneity of the passing networks from each team by seeing if their players have similar importance or, in contrast, they contribute in very different ways in the All passing Networks. This homogeneity is analyzed by computing the standard deviation of the eigenvector centrality values of all the players obtained in all the matches for each team. This time the goalkeepers are also excluded, as they are a source of extra variability that we are not interested in, as we want to analyze the difference of homogeneities between the performances of the outfield players of the teams. In addition, we compare the homogeneity of the passing networks of the teams with the maximum average centrality value from the players of the team, to see how having a key player (in terms of passes) or not affects the other values from the players.

It can be seen in Figure 5 the scatterplot of the maximum average centrality from the players for each team depending on the standard deviation of all its players' centralities and the regression line that best fits the relationship. The figure shows that the teams with more relevant players in the passing networks tend to have more variability between the centrality values of their players and the networks are more heterogeneous. Teams that are found in the upper right part of the plot, like Real Sociedad B, Sporting de Gijón and Girona, have one more relevant player in their total passing network and also their players' centrality values are more heterogeneous. In contrast, teams in the bottom left part of the plot do not have any really relevant player and their players distribute in a more homogenous way the passes they make. Eibar, Ponferradina, Amorebieta and Lugo are examples of teams with these characteristics. In addition, the teams located way over the regression line, like Las Palmas, have a more relevant player than the expected by the variability of the importance of their players in the passing networks, as opposed to the teams located way under the regression line like Burgos CF and Real Oviedo.

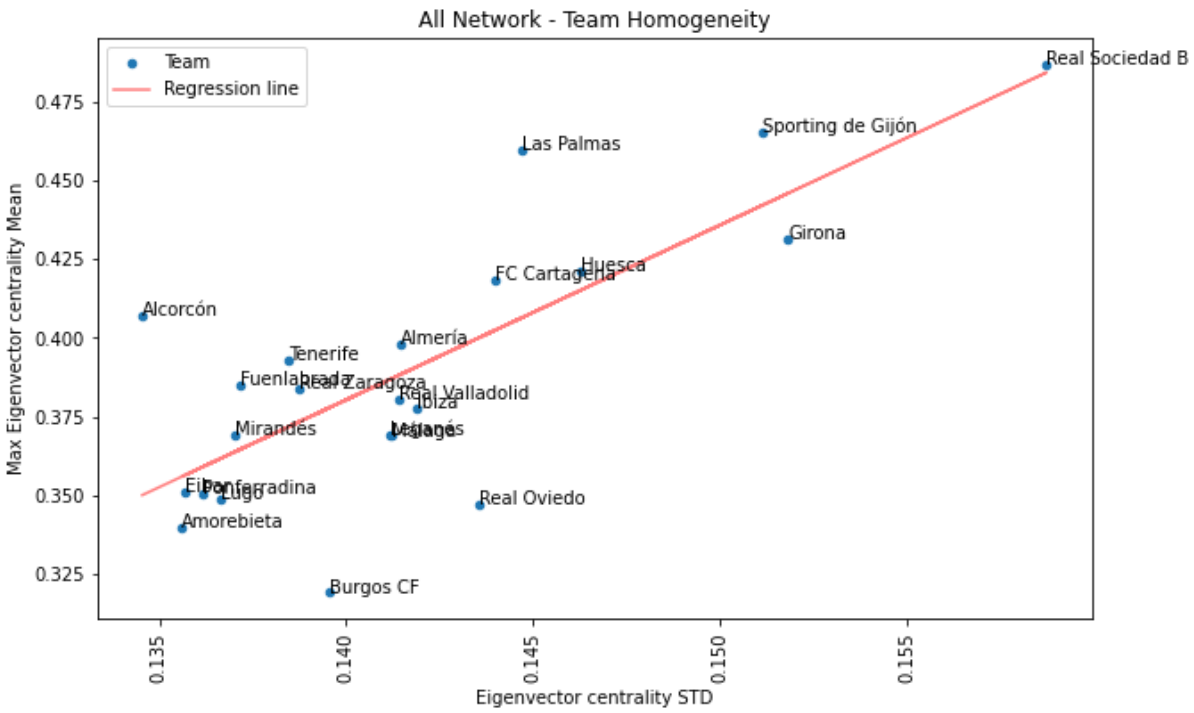

Figure 5: Scatterplot and regression line of the maximum average eigenvector centrality of the players depending on the centrality standard deviation by team.

6.3. Passing networks from two matches

Finally, we represent two different networks in order to show how a heterogeneous passing network works and how a network without any outstanding performance and where more players are highly involved is. The representation of the networks from Figure 6 and Figure 7 have been done with the code obtained from https://github.com/Friends-of-Tracking-Data-FoTD/passing-networks-in-python (note that both representations include the players that started from the bench and entered to play). With the objective to show a heterogeneous network with one relevant player involved, we have selected the match from Real Sociedad B where there was a player with higher eigenvector centrality, as we have seen that Real Sociedad B has more variability in their players centralities and also has two players between the top 10 most central players.

The total passing network from the Real Sociedad B's selected match can be seen in Figure 6. In this representation, we have that the size of the nodes depends on the value of the eigenvector centrality of the player in that network, the color of the node depends on the number of passes that the player has given and the color and size of the edges indicates the number of passes made between the players. It can be seen that there are four players that are really relevant in the network, whereas the other players are not much involved. The player with higher centrality on that match is Aritz Arambarri, who is one of the players with higher average eigenvector centrality in the whole competition, followed by Jon Ander Olasagasti and Urko González (which is

also at the top in the average eigenvector centralities), and the players that more passes did between them are Aritz Arambarri with Jon Ander Olasagasti and again the first one with Urko González. So, it can be seen that the most central players from the Real Sociedad B network are the defenders and also one midfielder.

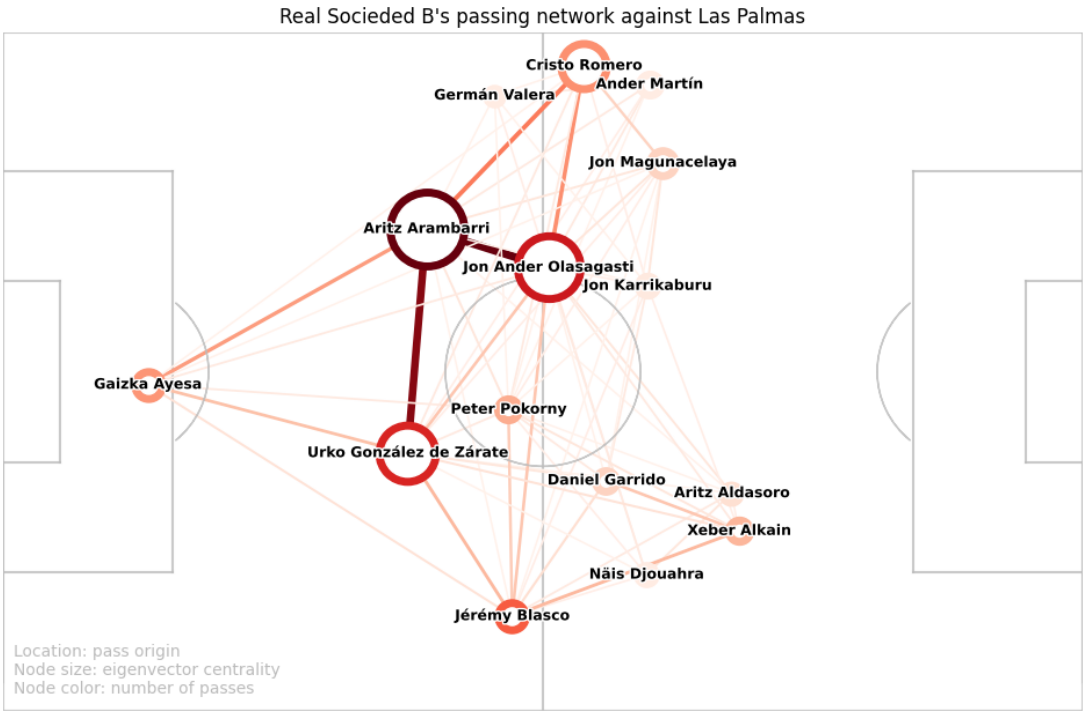

Figure 6: Passing network representation of Real Sociedad B against Las Palmas

To represent a passing network where there was not any key player and the passes were more distributed between all the players, we have selected the match that has a lower maximum eigenvector centrality. This way, we guarantee that the value of the most relevant player in the passing network of that match will be as low as possible. The resulting match is played by Real Zaragoza, who was between the teams with more homogeneous passing networks and that did not have a really outstanding player, and its representation network is shown in Figure 7. In this case, it can be seen that there are not that many differences between the centrality values of the players that started the game (the smallest nodes belong to the players that entered in the second half to play and also the one from Álvaro Giménez, who was a starter and did not participate too much) and that there are more players that were actively involved in the passing network. Again, the higher eigenvector centralities belong to the defenders, but in this case a midfielder in more advanced positions also played an important role. It also seems that more passes were made in the sidelines than in more centered positions, where the midfielders that were closer to the defenders participated less in the passes.

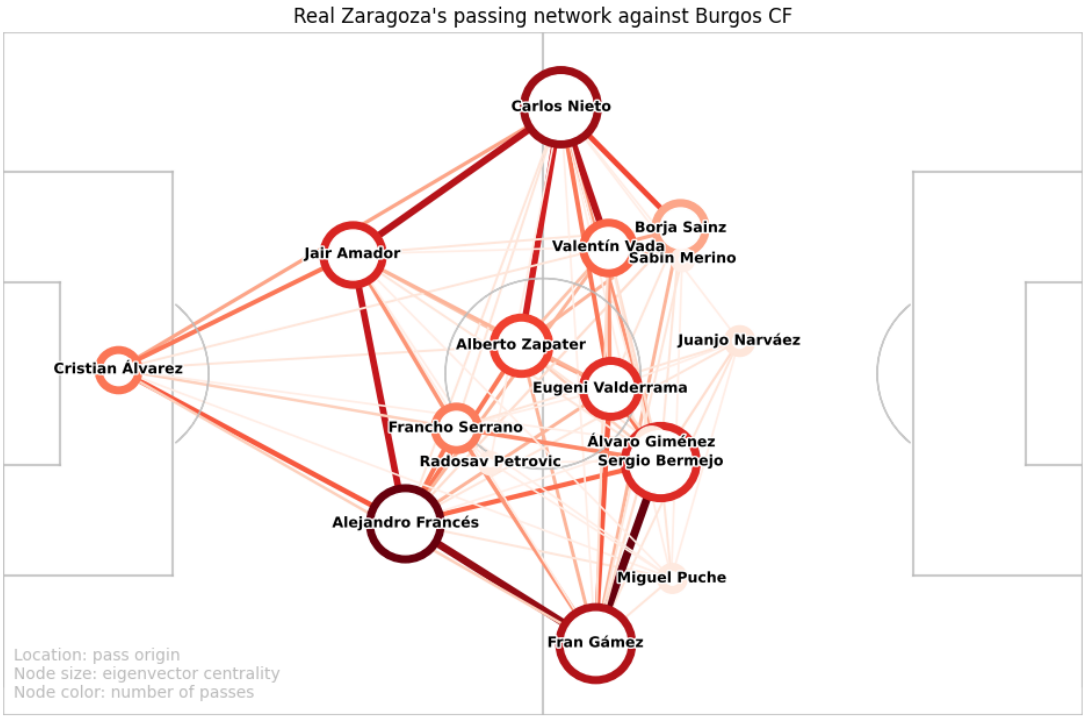

Figure 7: Passing network representation of Real Zaragoza against Burgos CF

7. Discussion and Future Work

As a first approach, we kept the centrality of the players that had played the whole match in order to analyze the relevance of the players in the passing networks that had the same time to get involved. This allowed us to observe how the values of the centralities from the players who played enough time should be distributed, but the condition of playing the whole match made not to consider too many players, being a big amount of the remaining values given by defenders. As a result, we considered those players who had at least played 360 minutes for their team in order to remove those that were not important in the team during the season or that they did not provide enough information to be considered, but giving the possibility of considering players with low amount of minutes in a match that could contribute less in the passing network. Nevertheless, this is considered to be the contribution of the player to that match passing network, although it has not played that much.

In the analysis we saw how similar the All and Lost Networks are, and that they are in fact quite different to the Shot ones. Shot ones tend to contain a lot less passes which cause some players to have higher values from a single match perspective. Nevertheless, overall the values tend to be more polarized, with either high or small values. We also saw defenders tend to appear more in All and Lost Networks while strikers and very offensive midfielders tend to have more presence in the Shot Networks. Moreover, we were able to see this fact in the difference between the Shot and Lost networks. In terms of the teams, we saw there are some teams which have a more homogeneous distribution of the passes while others tend to put more focus on one or two players in the squad. These last types of teams normally have players with very high eigenvector centralities.

Something we have learnt during this project is that the algorithm used to compute the eigenvector centrality gives more importance in directed networks to the receiving player rather than the passing player. We believe this causes some issues or unexpected results, specially in short possessions where there are not a lot of passes. The importance ends up mostly on the player at the end of the possession. This was seen previously in the top individual performances in the Shot Network where the players had a centrality of 0.99. In the Shot Network, for example, it is a problem because the shooter acts as a sink for the centrality of the network and accumulates it all. By having so few passes on those plays, they may not have any out passes and then the method does not work in the desired way.

We believe a possible solution for this would be to add a new connection between the last receiver and last passer in the opposite direction so that the last node is more connected. It would be a pass back to the player who sent the pass. That is to say, the assistance prior to any shot must have a pass in both directions.

As next steps, we think it would be interesting to do for example a clustering of the teams based on how homogeneous or heterogeneous they are to find which group

of teams are putting more importance into a single player and which rely in 2 main players, or which ones distributes it among even more players.

Since in this work we have mostly not used goalkeepers data, it is a possibility to do another study which focuses more on the goalkeepers. We believe it may be useful to someone who wants to study the implication of goalkeepers in the team's matches.

8. References

Balakrishnan, V. K. (1997). Graph Theory (1st ed.). McGraw-Hill.

Bonacich, P. (1987). Power and centrality: A family of measures. American journal of sociology, 92(5), 1170-1182.

Buldú, J. M., Busquets, J., Martínez, J. H., Herrera-Diestra, J. L., Echegoyen, I., Galeano, J., & Luque, J. (2018). Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game. Frontiers in psychology, 9, 1900.

Buldú, J. M., Busquets, J., Echegoyen, I., & Seirul. lo, F. (2019). Defining a historic football team: Using Network Science to analyze Guardiola's FC Barcelona. Scientific reports, 9(1), 13602.

Cintia, P., Rinzivillo, S., and Pappalardo, L. (2015). A network-based approach to evaluate the performance of football teams. In Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal.

Cotta, C., Mora, A.M., Merelo, J.J., and Merelo-Molina, C. (2013). A network analysis of the 2010 FIFA world cup champion team play. J Syst Sci Complex 26, 21.

Grund, T.U. (2012). Network structure and team performance: the case of English Premier League soccer teams. Soc. Netw. 34, 682-690.

J. Gudmundsson and M. Horton. (2017). Spatio-temporal analysis of team sports. ACM Comput. Surv. 50, 22.

Narizuka, T., Yamamoto, K., and Yamazaki, Y. (2014). Statistical properties of position-dependent ball-passing networks in football games. Physica A 412, 157-168.

Newman, M.E.J. (2010). Networks: An introduction. Oxford University Press, New York

Passos, P., K. Davids, K., Araújo, D., Paz, N., Mingueéns, J., and Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. J. Sci. Med. Sport, 14, 170-176.

9. Annex

Top	Player	Team	Position	Minutes played	Passes done All	Passes received All	Passes done Shot	Passes received Shot	Passes done Lost	Passes received Lost	eigen cent all	eigen cent shot	eigen cent lost	eigen cent diff
1	Josu Ozkoidi	Amorebieta	Defender	90	10	11	0	1	9	8	0,3564	0,99998	0,3530	0,64699
2	Viti	Real Oviedo	Striker	80	8	16	0	2	7	13	0,1669	0,99998	0,1688	0,83121
3	Matheus Aiás	Real Oviedo	Substitute	9	0	2	0	1	0	1	0,2134	0,99998	0,1643	0,83573
4	Oier Luengo	Amorebieta	Defender	90	21	14	0	1	15	8	0,2887	0,99998	0,2305	0,76949
5	Chris Ramos	Lugo	Striker	90	3	11	0	3	3	8	0,1601	0,99997	0,1420	0,85796
6	Pablo Valcarce	Burgos CF	Striker	88	5	16	0	1	5	11	0,2467	0,99997	0,2046	0,79535
7	Paris Adot	Ponferradina	Defender	90	14	13	0	2	13	9	0,3333	0,99997	0,2708	0,72918
8	Shon Weissman	Real Valladolid	Striker	38	5	9	0	1	3	6	0,1156	0,99997	0,8832	0,11679
9	Pablo Clavería	FC Cartagena	Midfielder	72	12	13	0	1	10	10	0,1839	0,99996	0,1605	0,83945
10	Gorka Guruzeta	Amorebieta	Midfielder	45	2	9	0	1	2	7	0,1272	0,99996	0,1166	0,88338

Table A.1: Top 10 players by single match performance based on eigenvector centrality in the Shot passes Network.

Top	Player	Team	Position	Minutes played	Passes done All	Passes received All	Passes done Shot	Passes received Shot	Passes done Lost	Passes received Lost	eigen cent all	eigen cent shot	eigen cent lost	eigen cent diff
1	Gorka Guruzeta	Amorebieta	Striker	89	8	32	1	4	7	23	0,7075	0,7266	0,6800	0,0466
2	Lluís López	Real Zaragoza	Defender	90	80	84	7	7	40	41	0,6149	0,3808	0,6312	-0,2504
3	Urko González de Zárate	Real Sociedad B	Defender	90	48	48	4	5	30	29	0,5652	0,5696	0,6246	-0,0549
4	Juanpe	Girona	Defender	90	78	85	10	8	48	57	0,5923	0,5687	0,6238	-0,0552
5	Jorge Pulido	Huesca	Defender	90	39	37	3	3	29	26	0,5722	0,5139	0,6227	-0,1088
6	Ignasi Miquel	Huesca	Defender	90	64	72	6	4	42	48	0,6330	0,4505	0,6207	-0,1702
7	Aritz Arambarri	Real Sociedad B	Defender	84	71	72	5	6	55	52	0,6139	0,5569	0,6196	-0,0628
8	Urko González de Zárate	Real Sociedad B	Defender	90	51	52	6	6	38	38	0,5820	0,3176	0,6115	-0,2939
9	Aritz Arambarri	Real Sociedad B	Defender	90	76	76	8	5	51	53	0,5808	0,3180	0,6094	-0,2914
10	Laure	Alcorcón	Defender	90	16	19	2	0	14	18	0,5044	0,7442	0,6069	0,1373

Table A.2: Top 10 players by single match performance based on eigenvector centrality in the Lost passes Network.