Social Network Analysis:
Soccer market transfer Network Structure:
from 2009 to 2021

Giuseppe Pio Salcuni 1100090
Alessandro Tocco 1097880
Manuel Placella 1099701
Andrea Accornero 1097521
Master Degree in Computer Science

2023-2024

1 Introduction

The soccer market is the period when soccer teams can buy, sell or trade players. It is a crucial time for building and transforming teams, with strategies varying from team to team and league to league. A team's success or decline may prompt reinforcements or disposals to improve performance. Some clubs aim to win trophies in the short term, so they may invest heavily in experienced players, while others may focus on building a young and promising team for the future.
Teams competing for titles generally tend to seek immediate reinforcements and may be willing to pay more for established players. Some leagues may be attracted to players from around the world while others may have a greater concentration of local talent. Each team, as well as each league, is influenced by a myriad of factors.
Our work focuses on understanding just some of these factors and drawing conclusions regarding both clubs and leagues.

2 Problem and Motivation

The focus of the network analysis is to investigate the dynamics of influence within the football market, trying to identify the clubs that exert more or less impact on this scenario. A crucial aspect is to understand which teams made a significant number of purchases during the period under consideration and, conversely, to identify those that had a less pronounced presence in the transfer market.
Another goal is to explore the presence of communities or clusters within the network, distinguishing clubs that tend to collaborate through player transfers. This will allow us to reveal significant connections between particular teams or leagues, revealing any partnerships or preferential interactions.
Another challenge addressed in the analysis is the identification of clubs that act as key intermediaries in transfers, looking at those that are involved in both buying and selling the same

player more frequently. This aspect reveals peculiar market dynamics and negotiating relationships that may be crucial in the context of football negotiations.
Through this analysis we want to explore the presence of different groupings within the network. We want to identify both groups of teams that have frequently interacted with each other and clubs that share similar financial characteristics. The goal is to identify any subgroups within the transfer network to reveal whether some clubs have similar goals or approaches.
The analysis also extends to the general behavior of soccer leagues, aiming to understand a number of relevant dynamics. It is proposed to examine the influence exerted by different leagues in the context of player transfers by outlining the leagues that record a greater number of purchases/disposals. From a financial perspective, attention is also turned to leagues where there is higher spending in transfers, as well as leagues that tend to make more purchases abroad rather than within domestic borders. An attempt is then made to identify the pair of leagues between which player exchanges occur most frequently.
In subsequent analytical insights, we aim to investigate the tendency of similar nodes to associate in a network by examining the presence of a scale-free power law-based distribution in the network structure. In addition, we aim to identify structural similarities among clubs by considering connectivity and interaction patterns among nodes in the network. Similarly, we intend to explore the presence of groups of nodes that are fully interconnected, as well as to identify particularly cohesive areas through the identification of groups of nodes in which each element has a minimum number of connections with other nodes belonging to the same cluster. In sum, we aim to unveil the complex dynamics and significant structures within the network, thus contributing to an in-depth understanding of the interactions in the context under consideration.

3 Dataset and Network

3.1 Dataset

Considering the objectives and possible problems of the analysis, the following dataset was chosen on GitHub. The Dataset takes into account all market movements of the top 7 Euoropean soccer leagues over a time frame from 2009 to 2021.
The following records are removed from the dataset before network creation:

- Records with missing values
- Records related to retired players
- Records related to transfers due to end of loan *(loan_end = true)*
- Records related to released players *(Without Club)*

3.2 Network

Each **Node** *(Team)* in the network has the following characteristics:

- ID
- Team Name
- Country Code (*es GB1 for England*)

Each **Arch** *(Transfer)* in the network has the following characteristics:

- ID
- Player ID
- Player Name
- Season
- Market window *w o s*
- Transfer fee

Only clubs that have at least one outgoing arc (sale) and one incoming arc (purchase) are considered in the network, so that clubs that are not part of the 7 leagues considered are excluded from the network.
At the end of "Pre-Processing," our network consists of:

- 236 nodes
- 11648 arches

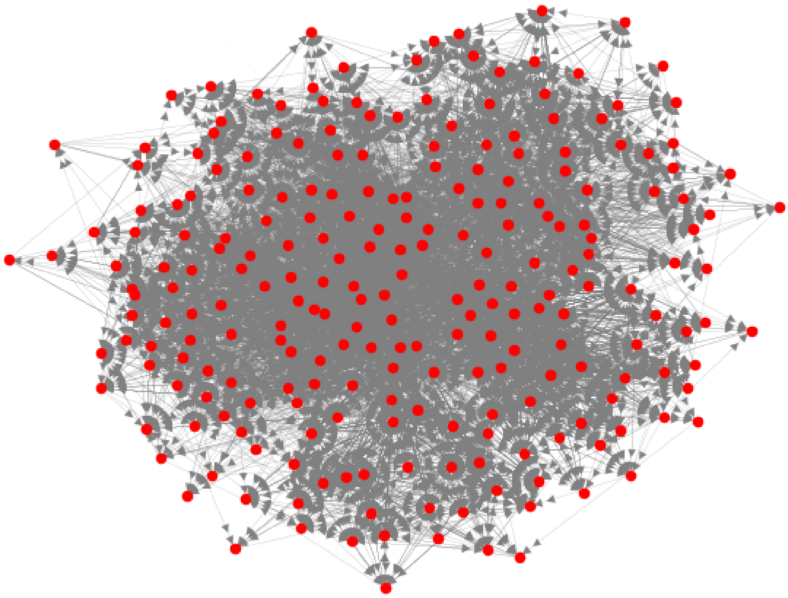

Figure 1: Network

4 Validity and Reliability

The dataset, carefully crafted by transfermarkt, emerges as a comprehensive and reliable source for data and statistics in the context of the soccer market. The accuracy and completeness of this resource are key pillars to ensure the robustness of our study. The careful selection of variables helped to accurately delineate the crucial dynamics of transfers in soccer. The representativeness of the dataset, spanning a wide range of clubs, players and transfer scenarios, provided us with a solid basis for extending our analysis to the entire soccer transfer landscape. The data cleaning phase was handled with transparency, clearly documenting the steps to address missing values and ensure the reliability of the analyses. The methodologies adopted in the data study, network analysis, and metrics calculation were described in detail, maintaining consistency to facilitate reproducibility of our approach. In addition, we paid special attention to the robustness of the results through sensitivity analyses, assessing the impact of variations in methodological choices on our results.

5 Measures

In the growing complex and interconnected context of modern soccer, network analysis emerges as an essential tool for understanding the subtle and intricate dynamics that characterize the transfer market and interactions between clubs. Through the application of key concepts such as degree centrality, proximity centrality, Kats centrality, modularity and other measures of relevance, we are able to probe deeply into the complexity of the relationships between the players in this world.

- **Centrality of degree**: indicates the number of direct connections a node will have. In simple terms, it represents how many "friends" or "connections" a node has in the network. A node with a high degree centrality is connected to many other nodes, while a node with a low degree centrality has fewer direct connections.
  In our context, it represents the number of transfers (purchases or disposals) that the club (or league), represented by the node, has made. This can provide insights into its activity in the transfer markets.

- **Closeness Centrality**: is a network centrality measure that reflects the average distance between a node and all other nodes in the network. Basically, it measures how "close" a node is to others in terms of shortest paths. A node with a high closeness centrality is more easily reached by other nodes in the network.
  In our case, the closeness centrality of a team indicates how accessible or close it is to other teams in terms of transfer possibilities. A team with high proximity centrality is more likely to be involved in negotiations with several teams because it is "close" to more potential transfer partners. High proximity centrality could also indicate an easy negotiation, extensive connectivity, or a liaison role for the team in question.

- **Kats centrality**: this type of Network centrality assigns a score to each node based on the sum of the paths between that node and all other nodes, with decreasing weight based on path length. In simpler terms, it reflects the importance of a node in relation to its connectivity and the quality of its connections in the network.
  The Katz centrality of a team thus indicates its importance in terms of connections in the network, taking into account both the quantity and quality of those connections. A

team with high Katz centrality might have quality connections with other key teams in the transfer network.

- **Modularity**: used to evaluate the subdivision of a network into distinct modules or communities. It represents the strength of the subdivision of the network into groups of highly connected nodes, with more sparse connections between groups. In other words, it measures the presence of internal structures or subnetworks within a larger network.
  In our context, it reveals the presence of "clusters" of clubs that interact more frequently with each other than with other teams in the network. These clusters could indicate communities of interest or collaboration within the soccer world.

- **K-Means Clustering**: in this case, it is an unsupervised machine learning algorithm used to group data into clusters so that the elements within each cluster are more similar to each other than elements in different clusters. Through the use of K-means clustering, clusters can be categorized according to different criteria. We decided to categorize them according to the number of transfers made in the period under consideration and according to the financial magnitude associated with player movements. The flexibility of this algorithm allows us to explore and identify patterns or groupings that may not be immediately apparent in the original data.
  To determine the optimal number of clusters to use in this analysis, we adopted the Elbow Method, a well-established and widely used procedure.

- **Statistical Homogeneity**: This type of study was conducted to evaluate the consistency of the clusters formed, allowing us to understand the degree of similarity between the nodes that make up each cluster. To quantify this similarity, we employed the following measure:

- **Coefficient of Assortativity**: this coefficient measures the propensity of similar nodes to associate with each other. A high assortativity coefficient indicates greater statistical homogeneity in clusters.

- **Scale-free**: a network is considered scale-free if it follows a degree distribution based on the power law, that is, if there are few highly connected nodes (hubs) and many nodes with few connections. Calculating the scale-free distribution can be useful in understanding the structure of the network and its organization. It is a probability distribution in which some entities have many connections and others have a few.

- **Structural Equivalence**: refers to the structural similarity between nodes within a network. Two nodes are considered structurally equivalent if they are involved in similar connectivity patterns. This concept helps to identify groups of nodes with similar roles or functionality in the network, which helps to better understand its overall structure and dynamics.

- **Clique**: a clique represents a subgraph in which each node is directly connected to all other nodes in the clique. In other words, this is a set of nodes in which each node is closely interconnected with all other nodes within that set. Cliques are important in that they identify strongly interconnected groups within a network, often indicating areas of dense activity or interaction. The size of a clique, defined as the number of interconnected nodes, can vary, but larger cliques play a significant role in the analysis because they reveal particularly tight connectivity structures and dense interactions between nodes.

- **Cores**: represent groups of nodes in which each node has a minimum number of connections with other nodes within the same group. These structures reveal particularly dense and cohesive areas within the network, allowing the most strongly interconnected nodes to be identified. Identifying cores helps to better understand the structure and dynamics of complex networks.

6 Results

After cleaning the dataset using the preprocessing discussed earlier, we proceeded to construct two networks using Python's NetworkX library. Specifically, we opted to create networks of type DiMultiGraph, a type of directed multigraph that allows for multiple arcs between the same nodes. This means that multiple directed arcs can exist from one node to another, each with distinctive characteristics.
Within our first network, nodes are represented by clubs and are identified by a unique id. Each node has two main attributes: the name of the club and the country to which it belongs. On the other hand, arcs in our network represent transfers and follow a specific direction from one node to another. Each arc has several attributes, including the transfer id, the id of the player involved, the name of the player, the age of the player at the time of the transfer, the season in which the transfer occurred, the market window, and the value of the transfer.
In the second network, nodes take on the role of representing leagues, while arcs reflect transfers between teams that are part of those leagues. The attributes associated with the arcs include the same information detailed in the previously described network.
In this section, all the results from the analyses conducted using the measures described in Section 5 are presented in detail, and the information that emerged from these measures will be examined.

6.1 Degree centrality

We performed degree centrality calculations on both networks in order to obtain meaningful insights. In the first case, the degree centrality analysis allowed us to identify the clubs that played a particularly influential role in the market during the 12 years between 2009 and 2021. In the second case, we focused on identifying the activity of the 7 leagues in our dataset in the market during the same period examined.

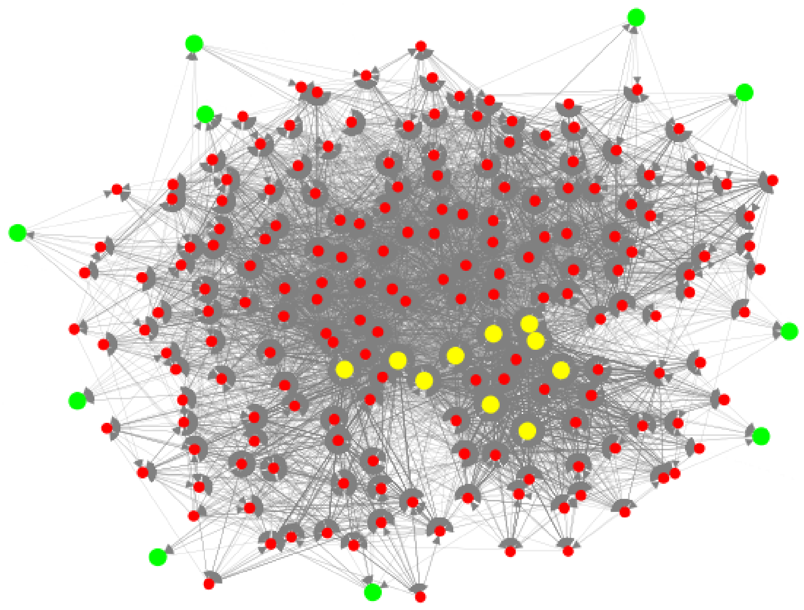

Figure 2: Degree centrality in the club network

In the Figure 2, nodes with a higher degree of centrality are highlighted in yellow, while those with a lower degree of centrality are marked in green. All other nodes are shown in red.

Club	Degree Centrality
Genoa CFC	1.8553191489361702
FC Internazionale	1.4936170212765958
Juventus FC	1.4170212765957446
AS Roma	1.319148936170213
Parma Calcio 1913	1.319148936170213
AC Milan	1.2425531914893617
Udinese Calcio	1.2170212765957447
FC Porto	1.2127659574468086
SL Benfica	1.1787234042553192
UC Sampdoria	1.1702127659574468

Table 1: The 10 clubs with the highest degree of centrality

From the results of this first analysis, as highlighted in Table 1, a significant finding emerges: eight of the ten most active teams in the period under consideration are of Italian nationality while the other two are of Portuguese nationality.

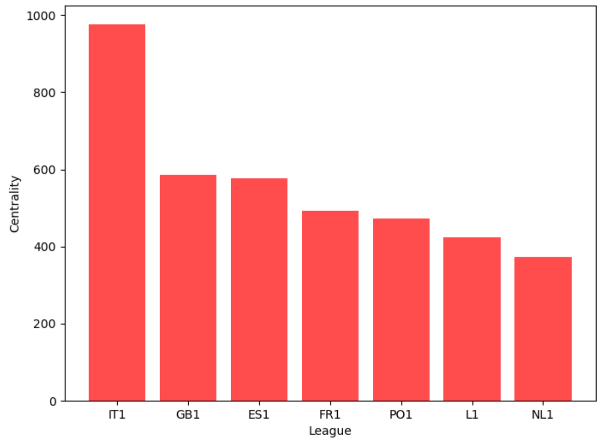

Figure 3: Degree centrality in the network of leagues

The result obtained previously is further confirmed through the calculation of degree centrality in the network incorporating the leagues and transfers between them (Figure 3). It is evident that league IT1 (Italian Serie A) shows a significantly higher degree centrality, almost twice as high as the second league in the rankings (which, in turn, shows relatively less distance from the centrality of the other six leagues). This discrepancy in degree centrality clearly suggests a prominent position of Serie A in the landscape of player transfers between clubs.
By deepening the network analysis by examining the degrees of entry and exit nodes, it is possible to clearly identify the clubs that stood out in terms of the number of purchases and disposals made.
The clubs that showed a significant propensity to make purchases were, in descending order: Genoa CFC, Parma Calcio 1913, UC Sampdoria, AC Milan, Juventus FC, AS Roma, FC Internazionale, ACF Fiorentina, Chievo Verona and Bologna FC 1909.
The clubs that made the most disposals, in descending order, were: FC Internazionale, Chelsea FC, Genoa CFC, FC Porto, SL Benfica, Juventus FC, AS Roma, Udinese Calcio, Manchester City and Sporting CP.
Even this simple analysis confirms the centrality of Italian clubs in the world of soccer marketing.
In the analysis of the network composed of leagues and transfers, the Italian, Spanish and English leagues stand out for a significant number of purchases and disposals, confirming the predominant activity of Italian clubs. Expenditure and revenue rankings see the English Premier League at the top in investment, while the Italian Serie A excels in revenue. In pairwise league comparisons, Spain and England stand out for the largest number of transfers, underscoring a strong interaction between the two nations' clubs in the international market.

6.2 Closeness Centrality

We next calculated the closeness centrality, previously illustrated. This value may reflect the ease of conducting negotiations, the presence of extensive connectivity, or a significant liaison role for the team in question. In general, however, closeness centrality provides a measure of how involved a team is in transfers with a larger number of clubs.

Club	Closeness Centrality
Genoa CFC	0.5759803921568627
Sevilla FC	0.573170731707317
Málaga CF	0.556872037914692
AS Monaco	0.556872037914692
Sporting CP	0.556872037914692
ACF Fiorentina	0.550351288056206
AS Saint-Étienne	0.5452436194895591
SL Benfica	0.5452436194895591
UC Sampdoria	0.5439814814814815
SS Lazio	0.5427251732101617

Table 2: The 10 clubs with highest proximity centrality

From the results obtained (shown in Table 2) it is clear that there are several new teams with a significant proximity centrality value. Among these teams, Genoa stands out, ranking at the top for both degree centrality and proximity centrality. This result suggests a role of considerable importance for Genoa within the transfer network, positioning itself at the center of connectivity dynamics and facilitating engagement with a large number of other teams. UC Sampdoria and SL Benfica are also present in both results.

6.3 Kats centrality

Next, we opted for Katz centrality analysis in order to obtain an evaluation that included both quantitative and qualitative aspects. The results of this analysis are detailed in Table 3.

Club	Kats Centrality
Sevilla FC	0.10791031056980402
Chelsea FC	0.10714161470594029
Sporting CP	0.1062591922259398
AS Monaco	0.10257295424592182
SL Benfica	0.10089155347865199
FC Porto	0.10028351091920111
Manchester City	0.09986075077245195
AS Roma	0.09640167628419402
FC Internazionale	0.09560387419772876
Arsenal FC	0.09515917696981947

Table 3: The 10 clubs with highest Kats centrality

It can be seen that Sevilla FC and Chelsea FC emerge as the clubs with the highest Katz centrality in the network considered, so they are the clubs that from 2009 to 2021 made major market transactions in both quantity and quality (i.e., transactions with equally important teams in the network).
Benfica emerges as a unique exception, standing out as the only team present in the top 10 positions for all three measures of centrality. This fact significantly underscores the club's considerable importance and influence.

6.4 Modularity

Through modular analysis, we identified the presence of six distinctive communities (as can be seen in Figure 4) that show a higher frequency of interactions with each other. These communities also exhibit lower connectivity with clubs belonging to other communities.

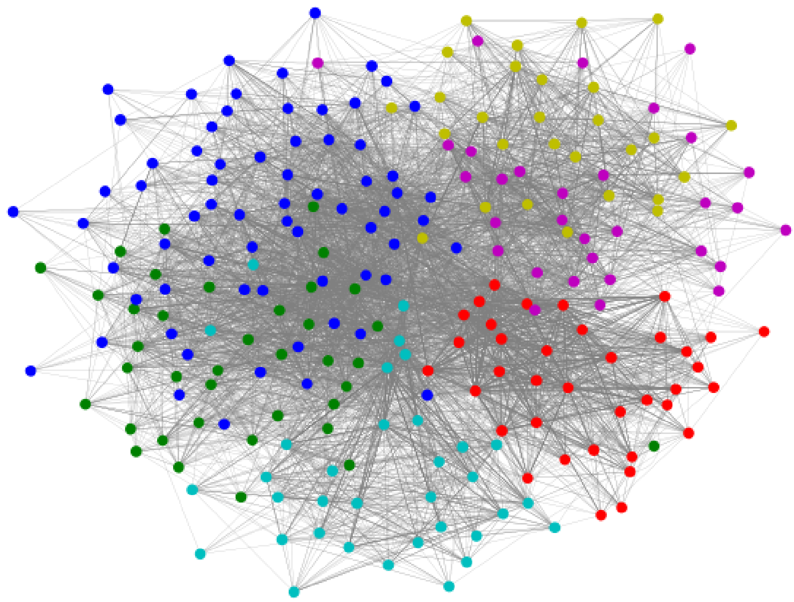

Figure 4: Communities in the network

Looking closely at the identified communities, Italian, French and German clubs are present in three specific communities, each comprising exclusively clubs from their respective nation. This suggests a propensity for Italian, French and German clubs to conduct market transactions primarily among themselves, strengthening ties and relationships within their respective national communities. Other communities reveal connections between clubs of different nationalities. In detail, one community consists of English and Spanish clubs, while another includes Portuguese clubs three Spanish clubs while another contains all Dutch clubs and three English clubs. These international ties indicate a broader market dynamic, with clubs of different nationalities engaging in transfer transactions and cooperation.
By calculating the modularity in the network formed by the leagues and transfers between them, as many communities spontaneously emerge as there are leagues in the data. This phenomenon is in line with expectations, since most transfers between clubs occur within the same nations. In other words, the communities identified reflect the geographical structure of the network, highlighting the closest ties between clubs within their respective nations.

6.5 K-Means Clustering

Concerning the application of the K-means method, we performed a comprehensive analysis considering both clubs and transfers between them. The determination of the optimal number of clusters was performed by using the Elbow Method. In Figure 5, the results of clustering based on the degree of the nodes are presented. As shown, three clusters emerged, each representing a distinctive grouping based on the number of operations. In the first cluster, representing the top clubs (many of them from Italy), the average number of operations was calculated to be 277. Next, a second cluster is identified characterized by an average number of transactions of 135, corresponding to about half of what was found in the previous cluster. Finally, in the third cluster, the average number of transactions stands at 50.

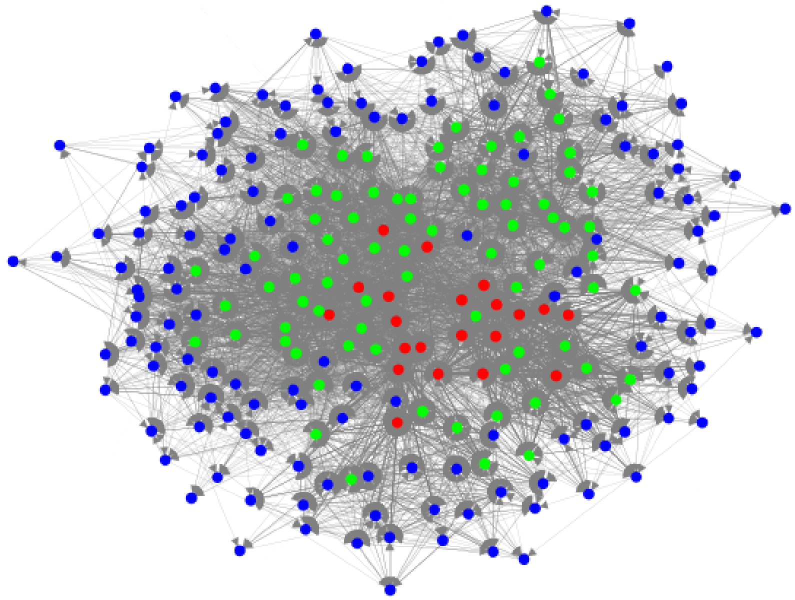

Figure 5: Cluster by grade

This analysis reinforced the idea that there are different levels of importance or relevance within the system under consideration.
In the next clustering step, we oriented our analysis on the financial magnitude of the clubs, taking into consideration the total amount of money spent and collected. As illustrated in Figure 6, we again came to the identification of three distinct clusters. The first cluster shows an average expenditure of approximately 66 million with as many as 186 clubs within it, the second shows a figure of approximately 460 million with 39 clubs, while the third cluster, consisting of 11 clubs, records a remarkable figure of approximately one billion. This observation further underscores and reinforces the presence of a limited number of clubs that clearly stand out as economically influential. Approximately 70% of the clubs are placed in the cluster with a lower financial magnitude highlighting the economic disparity compared to a limited number of entities that confirm economic dominance within the sports landscape.

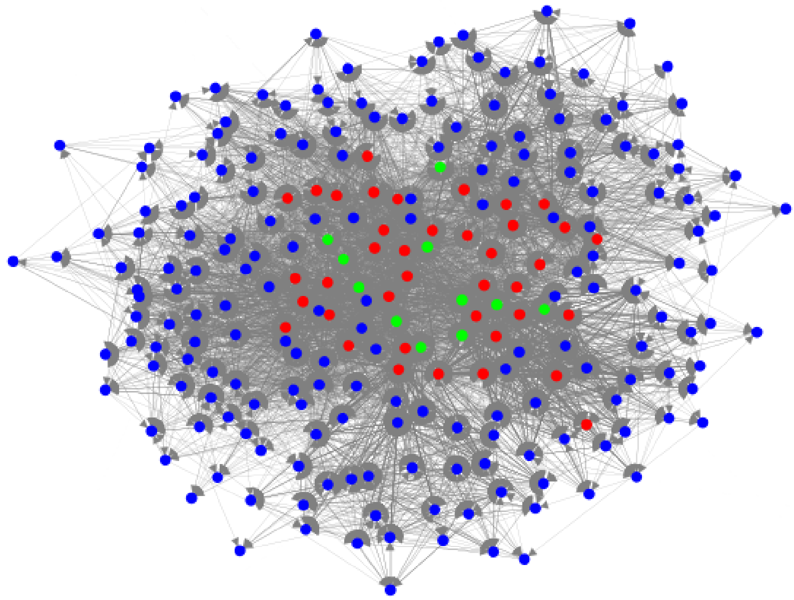

Figure 6: Cluster by financial entity

Analyzing the intersection of the results of the two clusters, further significant details emerge. In particular, it shows that among the clubs that stood out for high volumes of spending, there are three entities that, despite having invested conspicuously, made relatively few transfers. These clubs are Borussia, FC Barcelona and Real Madrid. This finding underscores the economic importance of the operations undertaken by these teams. On the other hand, we observe that a number of clubs, including Parma Calcio 1913, Genoa CF, Sporting CP and Manchester City, are in the category of those who have made numerous market transactions. However, it is interesting to note that these clubs are not among those that recorded the most impressive financial movements during the period under consideration. This dynamic suggests that although they actively participated in the transfer market, the economic size of their operations is in sharp contrast to the clubs mentioned above.

6.6 Statistical Homogeneity

Concerning the study of the above metric, the result obtained is 0.20144197958220117. The value, therefore, is not very high, but still positive. This suggests that there is a slight tendency for nodes with a certain degree to prefer connections with other nodes that have a similar degree. In more practical terms, we can say that in our network, teams with a certain number of inbound and outbound transfers tend to interact with other teams that have a similar number of transfers.

6.7 Scale-free

In this case, the power exponent obtained is 4.76, a relatively high value. In a scale-free network, the power exponent is expected to be low, generally between 2 and 3. A lower exponent indicates a more strongly scale-free degree distribution. Our result suggests that the degree distribution may not adhere perfectly to a power law, or that there may be significant variation in the degree distribution. The graphical result of this measurement is shown in Figure 7.

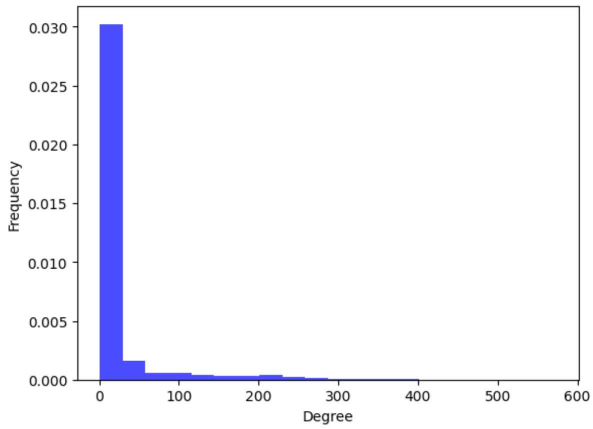

Figure 7: Degree distribution

6.8 Structural Equivalence

Analysis of cosine similarity between player transfers revealed interesting correlations among some soccer clubs. The top 10 clubs with the highest similarity coefficients, listed in a table, show an exceptionally high level of overlap in players traded over time. These clubs, including PSV Eindhoven, West Ham United, and FC Schalke 04, had a maximum similarity coefficient of 1.0000000000000007. This indicates complete overlap in player transfer patterns, suggesting strong similarities in market strategies or preferred types of players. The high cosine similarity among these clubs could be attributed to specific market dynamics, established relationships or common trends in player transfers.

Club	Cosine Similarity
PSV Eindhoven	1.0000000000000007
West Ham United	1.0000000000000007
West Bromwich Albion	1.0000000000000007
Swansea City	1.0000000000000007
FC Schalke 04	1.0000000000000007
FC Porto	1.0000000000000007
Liverpool FC	1.0000000000000007
Levante UD	1.0000000000000007
Hertha BSC	1.0000000000000007
Rayo Vallecano	1.0000000000000004

Table 4: The 10 clubs with greater cosine similarity

6.8.1 Jaccard's coefficient

Analysis through Jaccard's coefficient revealed interesting connections between the entities involved. The most closely connected clubs, represented in the top 10 nodes, indicate considerable overlap in the players exchanged. In particular, clubs such as SC Cambuur-Leeuwarden, FC Penafiel, and Leixões SC, all with a coefficient of 0.6, show significant correspondence in their transfer patterns, suggesting that these clubs may share similar market strategies or manifest a common preference in the types of players acquired or transferred. At the same time, other clubs, such as Excelsior Rotterdam and Vitória Setúbal FC, with Jaccard coefficients of 0.6 and 0.5869565217391305 respectively, reveal similarly significant links in their player movements.

Club	Jaccard Coefficient
SC Cambuur-Leeuwarden	0.6
FC Penafiel	0.6
Leixões SC	0.6
Excelsior Rotterdam	0.6
Vitória Setúbal FC	0.5869565217391305
Moreirense FC	0.5869565217391305
FC Arouca	0.5789473684210527
SC Beira-Mar	0.5789473684210527
PEC Zwolle	0.5652173913043478
Willem II Tilburg	0.5652173913043478

Table 5: The 10 clubs with the highest Jaccard coefficient

6.8.2 Pearson's Correlation Coefficient

Pearson's correlation coefficient of 0.232 shows a moderate positive correlation between patterns of player transfers between clubs. This value suggests the existence of a relationship, albeit not extremely strong, between player movements between clubs over time. The extremely low p-value (1.15e-142) indicates very high statistical significance, supporting the claim that the correlation found is not due to chance. This result may indicate that some clubs share, in part, similar strategies or trends in player transfers. However, it should be noted that the correlation, while significant, does not necessarily imply a cause-and-effect relationship and would require further analysis.

6.8.3 Hamming distance

The use of Hamming distance in examining player transfers between soccer clubs reveals peculiar aspects in the relationships between entities. The table of the top 10 nodes with highest Hamming distance highlights clubs such as Sporting CP, PSV Eindhoven and Chelsea FC, which are characterized by significant differences in their player transfer patterns compared to other clubs in the network. For example, Sporting CP has the highest Hamming distance of 113, indicating considerable dissimilarity in players acquired or transferred compared to other clubs. These results suggest marked divergence in market strategies or dynamics of player transfers, providing another key to understanding the complexity of relationships among soccer clubs.

Club	Hamming distance
Sporting CP	113
PSV Eindhoven	113
Chelsea FC	108
SL Benfica	108
FC Metz	108
Bayer 04 Leverkusen	107
SC Braga	106
FC Porto	106
Vitória Guimarães SC	106
Eintracht Frankfurt	104

Table 6: The 10 clubs with the greatest Hamming distance

6.9 Cliques

The survey of cliques within the network revealed an interesting structure of connections concentrated around several prominent clubs. The top 10 clubs with the most connected cliques are listed in the table below:

Club	Number of cliques
Bologna FC 1909	10
AS Roma	10
SSC Napoli	10
ACF Fiorentina	10
AC Milan	10
FC Internazionale	10
Torino FC	10
Atalanta BC	10
Juventus FC	10
Chievo Verona	10

Table 7: The 10 clubs with major cliques

These clubs emerge as central nodes within player transfer relationships, each involved in a significant number of cliques. The presence of 10 cliques linked to each club suggests a robust interconnectedness, indicative of relevant exchanges and relationships within the soccer landscape. Clique analysis then provides insight into the specific dynamics involving these clubs, underscoring their centrality in player transfer market interactions.

6.10 Cores

The assignment of Cores Numbers to football clubs, shown in the table, represents the degree of centrality within the player transfer network. Clubs with high Cores Numbers, such as Stoke City, 1. FC Köln, AC Milan, and others, emerge as central nodes in the network structure. This suggests that such clubs are deeply interconnected with others, actively participating in numerous player transfers with a wide range of partners. The centrality of these clubs could be indicative of a strong presence in the transfer market and significant influence on player exchange dynamics. In addition, a Cores Number of 30 for all these clubs implies high cohesion within the network, reflecting a strong connection between adjacent clubs. These results highlight the strategic importance of these clubs in the player transfer landscape and suggest a dynamic and dense network among the entities involved.

Club	Core Numbers
Stoke City	30
FC Köln	30
AC Milan	30
Arsenal FC	30
AS Roma	30
Atlético de Madrid	30
UC Sampdoria	30
Borussia Dortmund	30
Olympique Lyon	30
Bayer 04 Leverkusen	30

Table 8: The 10 clubs with higher core numbers

7 Conclusion

The present study provided in-depth insight into the dynamics of player transfers in the world of soccer, focusing on the relationships between clubs, leagues and the centrality of certain teams. The results obtained through the various analyses of centrality, modularity, clustering and other network metrics helped to map out the interactions in the transfer market.
In particular, the degree centrality analysis highlighted the prominent position of the Italian Serie A, emphasizing its centrality in player transfers between clubs. Clubs such as Genoa CFC, Parma Calcio 1913 and UC Sampdoria stood out for the significant number of purchases, confirming the predominant activity of Italian clubs.
Proximity centrality identified clubs such as Genoa with a notable role in the transfer network, positioning itself at the center of connectivity dynamics and facilitating engagement with a large number of other teams.
Katz's centrality analysis identified clubs such as Sevilla FC and Chelsea FC as those with significant market transactions in both quantity and quality during the period under consideration.
The modular analysis revealed the presence of distinctive communities, with Italian, French, and German clubs tending to carry out market transactions primarily among themselves, while other communities reflect connections between clubs of different nationalities.
The application of the K-means method revealed the presence of different categories of clubs based on the number of transactions and financial magnitude, highlighting the presence of a limited number of clubs that are clearly distinguished by their economic influence.
Similarity analyses through Jaccard's coefficient, Pearson's correlation and Hamming's distance provided further details on the relationships between clubs, highlighting overlaps in transfer patterns, moderate correlations and significant differences in market strategies.
Clique and cores analysis emphasized the centrality of some clubs in player exchange dynamics, highlighting their robust interconnectedness and high cohesion within the network.
In short, this study provides a comprehensive overview of interactions in the player transfer market in soccer, contributing to a greater understanding of the dynamics involving clubs, leagues, and players in this fascinating sporting world.

8 Critique

In the course of our project, we focused on the analysis of transfers during the period under consideration. However, it might be interesting to extend our investigation through a temporal analysis in order to reveal trends and evolving dynamics in transfer behavior over different seasons. At the same time, we consider it intriguing
to enrich our data by integrating them with the performance of clubs before and after transfers, thus allowing for a more in-depth exploration of how such movements affect performance both individually and collectively.
An additional perspective, which could bring significant value to our work, is the development of advanced predictive models. These models, based on past patterns and key parameters that emerged from our analysis, could prove to be valuable tools for anticipating possible future transfers in the soccer landscape. In this way, we would not only contribute to understanding past dynamics, but also pave the way for a predictive perspective that could be of crucial importance to clubs in developing market strategies and managing sporting resources.