A Network Approach to Transfer Dynamics in European Soccer Leagues

Yunus Tan Kerestecioğlu

Abstract

This study investigates the dynamics of player transfers in the top European soccer leagues from the 1995/1996 season to the 2022/2023 season through a network science approach. Data on league standings were scraped from the publicly available website *transfermarkt.co.uk*. For constructing a weighted directed network, where clubs are nodes and transfer fees are edges, data from multiple CSV files were collected. This dataset enabled the analysis of various network quantities, such as degree distribution and the assortativity coefficient. An algorithm was developed to detect central nodes in the network by correlating node strength and betweenness centrality. Seasonal networks were also analyzed to identify active clubs for each season, and the correlation between transfer market activity and changes in league standings was examined. The findings reveal significant insights into the transfer behaviors of European clubs, highlighting key actors in the market and the impact of transfer activity on club performance. This study provides a deeper understanding of the structure and dynamics of the European soccer transfer market using a network science approach.

Keywords Directed network · Weighted network · Degree distribution

Assortativity · Central nodes · Transfer market activity

1 Introduction

1.1 Complex Networks

Networks have been studied to gain a better understanding of the overall behavior of large systems when traditional methods are insufficient. Networks are used in various areas such as the internet, social interactions, biological systems, and economic systems. In weighted networks, links represent more complex connections like socioeconomic flows. The transfer markets in football leagues, where teams often make big deals with each other rather than signing free agents, have also attracted research interest. This study looks at the European football transfer market as a weighted network to explore its detailed structure. By analyzing features such as which teams frequently trade players and analyzing the clubs with highest transfer market activity, this research provides new statistical insights for club management.

1.2 Soccer Transfer Market

European football is different from many American sports because it doesn't have a draft system to evenly distribute skilled players among teams. Instead, in European football, players work for their clubs and can be transferred to another club before their contract ends, as long as the new club pays a fee to the current club. The soccer transfer market operates during specific periods called transfer windows, occurring in the summer and winter. During these windows, clubs can buy, sell, or loan players. Transfer fees can vary widely, with top players often commanding multi-million dollar deals. Agents play a crucial role in negotiating transfers and player contracts. Additionally, factors like player performance, contract length, and market demand significantly influence transfer decisions and fees. The transfer market is a dynamic and essential part of soccer, impacting overall club strategy.

1.3 Data Sources

- **Transfer Market Data**

For the construction of transfer market data, csv files of most popular european soccer leagues including *Premier League(English league), Ligue 1(French league), Bundesliga(German league), Serie A(Italian league), Eredivisie(Dutch league), Liga Nos(Portuguese league), La Liga(Spanish league*) are found from *https://github.com/*, and then combined in single csv file. Then, the csv file is cleaned and filtered. Only transfers in the Summer Transfer Window for each season are filtered. Because, teams trade more important and valuable players in the Summer Transfer Window in order to better prepare for the next season. Also, free transfers and loan transfers are cleaned from the csv file in order to reduce the data redundancy, and assign the transfer fees as edge weights to the network. Also, multiple columns are filtered. A sample of transfer market data can be found in *Figure 1*.

	club_name	player_name	club_involved_name	fee	transfer_movement	transfer_period	league_name	year	season	fee_value
0	Arsenal FC	Dennis Bergkamp	Inter	€11.25m	in	Summer	Premier League	1995	1995/1996	11250000.0
1	Arsenal FC	David Platt	Sampdoria	€7.10m	in	Summer	Premier League	1995	1995/1996	7100000.0
6	Arsenal FC	Stefan Schwarz	Fiorentina	€2.20m	out	Summer	Premier League	1995	1995/1996	2200000.0
7	Arsenal FC	Jimmy Carter	Portsmouth	€0	out	Summer	Premier League	1995	1995/1996	0.0
11	Chelsea FC	Mark Hughes	Man Utd	€1.90m	in	Summer	Premier League	1995	1995/1996	1900000.0

Fig. 1 - Sample Transfer Market Data

- **Standings Data through Web Scraping**

For the related soccer leagues the standing tables are scraped from a publicly available website https://www.transfermarkt.co.uk/. Standing tables of leagues are collected to analyze the correlation between transfer market activity and ranking, points difference in consecutive

seasons. Then, a csv file having columns of rank, point, team_name, league_name is constructed. A python script is used In order to detect the mismatch team names between transfer market data and standings data, and the team names are matched manually.

2 Soccer Transfer Market Network

2.1 Network Construction

In the soccer transfer network, nodes represent teams and edges represent the amount of total fee of transferred players between teams. The transfer weight of a player, w_{ij}(p), is defined by the transfer fee for the player p transferred from a team i to team j. The link weight between two teams is equal to the total weight of all traded players between them.

The edge direction of player weight is determined by using ‘club_name’, ’club_involved_name’, ’fee_value’, and ‘transfer_movement’ columns. It is constructed this way:

$if\ transfer\ movement\ =\ 'in'$:

$$w_{clubName-clubInvolvedName}\ +=\ 'fee'$$

For this case, the fee is added to the weight of direction from ‘club_involved_name’ to ‘club_name’. Meaning player is transferred from ‘club_involved_name’ to ‘club_name’.

$if\ transfer\ movement\ =\ 'out'$:

$$w_{clubInvolvedName-clubName}\ +=\ 'fee'$$

For this case, the fee is added to the weight of direction from ‘club_name’ to ‘club_involved_name’. Meaning player is transferred from ‘club_name’ to ‘club_involved_name’.

After the construction of the network, the weights are normalized in a smaller scale because, in a transfer network, there are high differences between the transfer fees which does not reflect the effects of transfers on the success of clubs exactly.

As a result, a global transfer network is constructed by analyzing transfer activity of 7 European leagues from season 1995 to 2022 mentioned in ‘Data Sources’. It has *3502* nodes and *27895* edges.

2.2 Degree Distribution

From degree distributions of the network, it can be deduced that degrees are not homogeneously distributed. Meaning that some teams were very active in the transfer market while some teams did not sell and did not buy too many players.

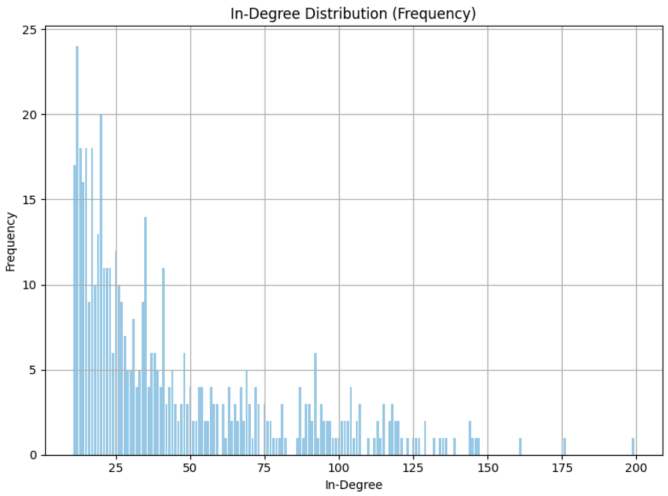

Fig. 2 - In Degree Distribution

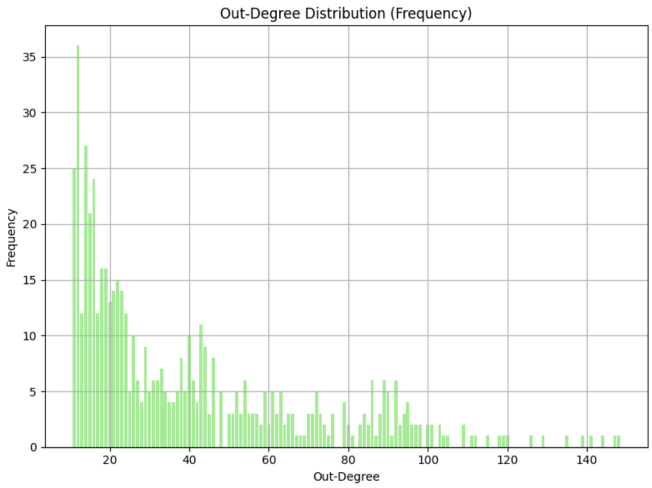

Fig. 3 - Out Degree Distribution

Even if the heterogeneity of degree distribution makes sense for the soccer transfer market, the difference does not reflect the reality exactly. Because, only transfer activities of 7 European leagues are analyzed and data includes transfers from these leagues to other leagues too. In order to reduce the redundancy caused by this situation, we defined a threshold that equals to *10*, and get the degrees greater than 10. *Figure 2* and *Figure 3* represent the degree distribution of degrees greater than this threshold.

3 Results

3.1 Centrality Measures

In this study, two key centrality measures are calculated for the clubs in the European soccer transfer network: Degree centrality and betweenness centrality. Degree centrality quantifies the number of direct transfer connections each club has, reflecting their overall activity in the transfer market. High degree centrality values indicate clubs that are heavily involved in player transactions, either as buyers or sellers. In contrast, betweenness centrality measures the extent to which a club acts as an intermediary in transfers between other clubs, highlighting its role as a bridge within the network. Clubs with high betweenness centrality play a crucial role in facilitating player movements across the network. Therefore, the betweenness centrality of a node is very important for the detection of active nodes in the transfer market.

As expected, the degree centrality distribution reveals a heterogeneous pattern, with a small number of clubs exhibiting exceptionally high centrality. Conversely, the majority of clubs have lower degree centrality, indicating less frequent participation in transfers.

The betweenness centrality distribution reveals the pivotal role of certain clubs, demonstrating that a few key actors are critical in maintaining the connectivity and flow of transfers within the network.

The distribution of betweenness centrality and the distribution of degree centrality can be found below. The figures highlight the heterogeneity of the network, and the hierarchical nature of the transfer market which is expected since some clubs aim to be successful by spending money and making transfers, while some clubs aim to be successful by training players from the infrastructure.

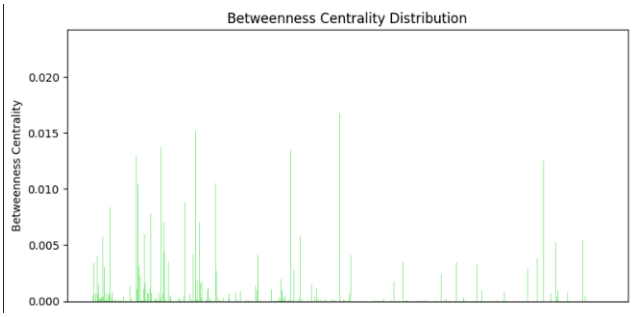

Fig.4 - Degree centrality distribution

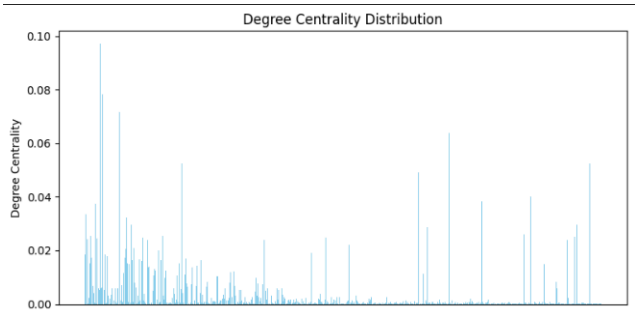

Fig.5 - Betweenness centrality distribution

3.2 Assortativity

To reveal the transfer interactions between low spending clubs and high spending clubs, the assortativity of the network is calculated. If the network is assortative, it means that high spending clubs tend to make transfers from other high spending clubs, and low spending clubs tend to make transfers from other low spending clubs. Briefly, if the network is assortative, clubs tend to make transfers from other clubs which are close to their own financial levels. However, in the European transfer network case, it is disassortative, meaning that high spending clubs may seek talented players from low spending clubs and low spending clubs may push their limit and try to make transfers from financially bigger clubs.

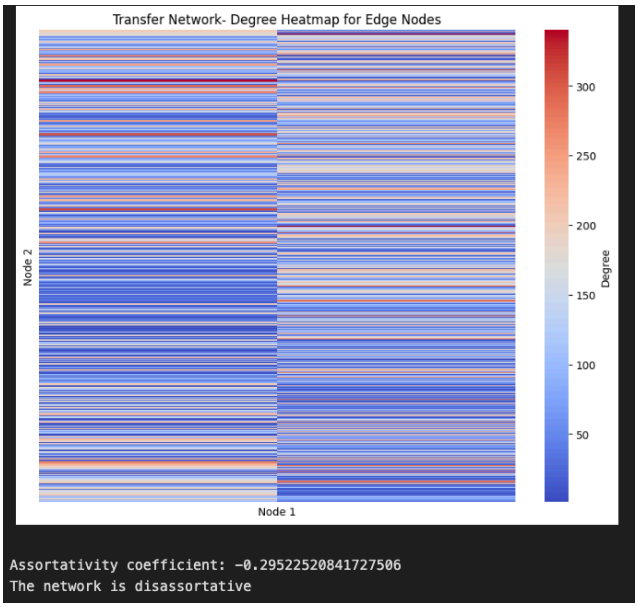

Fig. 6 - Degree Heatmap for Edge Nodes

3.3 Detecting Central Nodes

In this study, central nodes represent the active clubs in the European transfer market. A custom algorithm is designed to detect these central nodes. The steps of the algorithm are below:

1) Calculate betweenness centrality of nodes.
2) Calculate node strength of nodes. Node strength of a node is the sum of in-degree of the node and out-degree of the node.
3) Plot the scatter plot where x-axis is 'Node Strength' and y-axis is the 'Betweenness Centrality' of a node. *Figure 7* represents this plot, there is a positive correlation.
4) The nodes that are close to the fit line, meaning the nodes with a lower absolute residual value, are marked as central nodes.

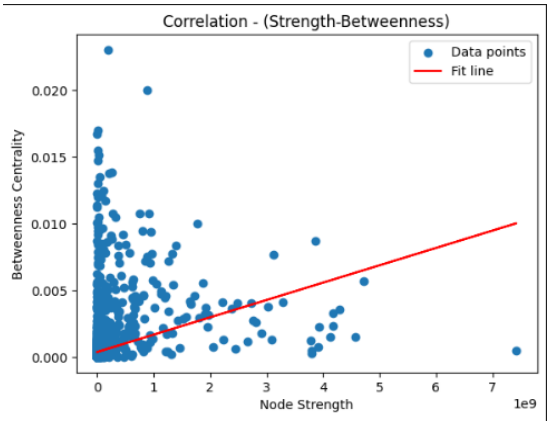

Fig. 7 - Node Strength-Betweenness Centrality Correlation

3.4 Success - Transfer Activity Correlation

In order to analyze the effect of transfer activity on success of clubs, a seasonal transfer network was constructed for each season from 1996 to 2022. For each seasonal transfer network, central nodes which are existing in the standings data fetched from website *'transfermarkt.co.uk'* are detected with the algorithm described in the *Section 3.3*. *Figure 8* represents the structure of standing data

position	team_name	points	league_name	season
1	Manchester United	82	Premier League	1995
2	Newcastle United	78	Premier League	1995
3	Liverpool FC	71	Premier League	1995
4	Aston Villa	63	Premier League	1995
5	Arsenal FC	63	Premier League	1995

Fig.8 - Standings Data

Standing data is formed of 5 columns: 'position', 'team_name', 'points', 'league_name', and 'season'. Two types of success are analyzed for the central nodes in this study, which are points difference of a club in the ending of current season and ending of previous season, and ranking difference of a club in the ending of current season and previous season. The line chart in *Figure 9* represents the average points difference, and average ranking difference of highly active clubs in the transfer market season by season.

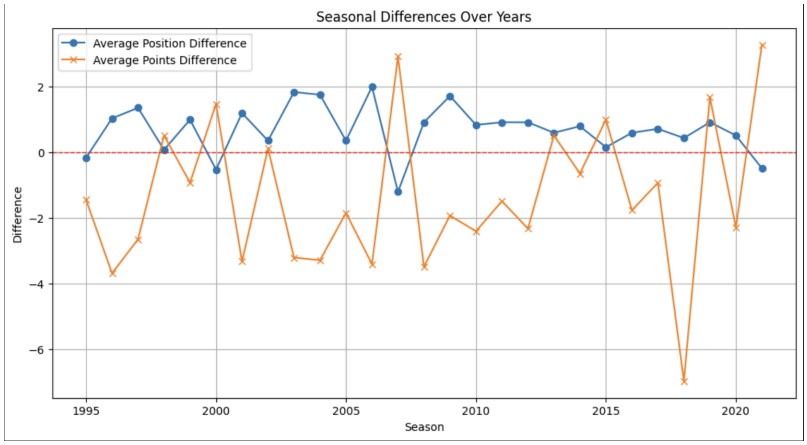

Fig.9 - Seasonal Success Differences of Central Teams over Years

From the findings, it can be deduced that there is a negative correlation between ranking difference and points difference of central clubs in each season. Meaning that, points collected by a club do not have too much effect on the ranking of the club at the end of the season. Generally, clubs that are active in the transfer market increased their rankings but the points

collected decreased. Therefore, the effect of transfer activity on the success of a club is open to interpretation.

4 Conclusion & Discussion

This study of the European soccer transfer network from the 1995/1996 season to the 2022/2023 season reveals several characteristics about transfer flow and club behavior.

Firstly, the findings indicate that player movement is driven by factors beyond club financial power or success. The calculated assortativity of the network shows a disassortative pattern, meaning that high-spending clubs do not necessarily buy players from other high-spending clubs. This suggests that transfers often occur from low-spending to high-spending clubs or vice versa, indicating that factors other than financial power, such as player potential and club needs, influence transfer decisions.

Secondly, we observed that clubs spending more on transfers and receiving higher fees tend to be more involved in transfer market interactions. This was analyzed by calculating the correlation between node strength and betweenness centrality. The positive correlation between these measures, both in the overall network and within seasonal networks, confirms that financially active clubs are central players in the transfer market, facilitating numerous transactions and acting as key intermediaries.

Lastly, we explored the hypothesis that heavy trader clubs become more successful in the following season. Our algorithm, which correlates node strength and betweenness centrality, combined with the analysis of points and position differences from scraped standings data, produced mixed results. While active clubs generally improved their rankings in the subsequent season, their points tended to decrease. This suggests that while transfer market activity may enhance a club's competitive position, it does not necessarily translate to a proportional increase in points, leaving the accuracy of this hypothesis open to interpretation.

Overall, this study offers an interesting look at the dynamics within the European soccer transfer market, highlighting the complex interaction between financial activity, market interactions, and club performance.

5 References

1. Lee, S., Hong, I., & Jung, W.-S. (Year). A Network Approach to the Transfer Market of European Football Leagues.

2. Tobias, W. (Year). Analysis of the Football Transfer Market Network.

3. Transfermarkt.co.uk. Retrieved from https://www.transfermarkt.co.uk

4. GitHub. Retrieved from https://github.com/