Identifying successful football teams in the European player transfer network

Tristan J. Dieles[1*], Carolina E. S. Mattsson[2] and Frank W. Takes[1]

*Correspondence:
tristandieles+liacs@gmail.com

[1] Leiden Institute of Advanced Computer Science, Leiden University, Niels Bohrweg, Leiden, The Netherlands
[2] CENTAI Institute, Turin, Italy

Abstract

This paper considers the European transfer market for professional football players as a network to study the relation between a team's position in this network and performance in its domestic league. Our analysis is centered on eight top European leagues. The market in each season is represented as a weighted directed network capturing the transfers of players to or from the teams in these leagues, and we also consider the cumulative network over the past 28 years. We find that the overall structure of this transfer market network has properties commonly observed in real-world networks, such as a skewed degree distribution, high clustering, and small-world characteristics. To assess football teams we first construct a measure of within-league performance that is comparable across leagues. Regression analysis is used to relate league performance with both the network position and level of engagement of the team in the transfer market, under two complimentary setups. Network position variables include, e.g., betweenness centrality, closeness centrality and node clustering coefficient, whereas market engagement variables capture a team's activity in the transfer market, e.g., total number of player transfers and total paid for players. For the season snapshots, the number of transfers correspond to weighted in- and out-degree. Our analysis first corroborates several recent findings relating aspects of market engagement with teams' league performance. A higher number of incoming transfers indicates worse performance and better resourced teams perform better. Then, and across specifications, we find that network position variables remain salient even when engagement variables are already considered. This substantiates the notion in the existing literature that a high degree corresponds to better team performance and suggests that network aspects of trading strategy may affect a team's success in their respective domestic league (or vice versa). In this sense, the approach and findings presented in this paper may in the future guide team's player acquisition policies.

Keywords: Network science, Football transfer market, Sport success, Football leagues

Introduction

Association football is Europe's most popular sport by any objective metric (Matheson 2003). Football teams compete against other teams in a national or regional league and are often defined on the basis of their performance in these leagues. The best performing teams of a league will promote to a higher league, and the worst teams relegate to a lower league. Teams competing in the higher leagues often attract the most interest of

fans, which is accompanied by more financial resources for the teams (Frick 2007). These social and financial stimuli contribute to the team's incentive to maximize match wins in the leagues they appear in. This makes that it is useful to assess what team characteristics contribute to a team's ability to win matches. Much research has been devoted to what in-game characteristics benefit this goal (Goes et al. 2020; Li and Zhao 2020; Fabíola Zambom-Ferraresi and Lera-López 2018). In addition, the composition of teams has been an identified contributor to a team's success (Gelade 2018). Furthermore, the exact relation between a match win and a team's league performance depends on the rating or ranking system the league applies. Different sports use different methods to rank its teams, see, e.g., the work by Langville and Meyer (2012). Moreover, as the performance of a team is relative to its competitors, the competitive profile of a league has shown to be a noteworthy determinant for a team's individual league performance (Vales et al. 2018).

Naturally, the quality of the players is an essential factor contributing to the team's performance. Therefore, the in- and out-flow of players significantly impacts the team's overall effectiveness. On this basis, Pantuso and Hvattum (2021) discuss tactical decisions to be applied to a club's transfer policy based on in-game data. Mourao (2016) describes these dynamics in closer detail and finds a significant effect of league performance on obtaining a higher number of incoming players. In connection to this, Mourao confirms the intuitive notion that if a team plays well, the club is likely to sell their players for a significantly higher price. Lu et al. observes a similar positive correlation, between the amount of transfer fee spent and received and the team's in-game performance (Liu et al. 2016). Furthermore, transfers of players play a major part in the financial health of clubs (Dobson and Goddard 2012). These factors all suggest an interrelation between, on one hand, a team's transfers and trading behaviour and, on the other, their league performance; a research topic suggested by Mourao as a direction of further research. Efforts directed at exposing a possible interaction between a team's sportive performance and their trading behaviour have been made (Liu et al. 2016; Matesanz et al. 2018) and underline the notion that football is a "money game". Aforementioned works have shown the presence of a correlation between variables describing the team's engagement in the transfer market and their league performance.

Our work contributes to a growing literature considering the European football transfer market as a network. Network analysis allows for a robust and meaningful description of the network as a whole, and of its individual components. The toolbox of measures and metrics that network science offers to understand a system's components and their interaction is what makes network analysis a valuable method to tell the story of the football transfer market (Barabási 2016; Newman 2018). Centrality metrics, for example, can be used to capture the prominence of a team's position in a football transfer network (Liu et al. 2016; Bond 2020; Palazzo et al. 2023). We benchmark our findings against existing efforts to analyze multi-national multi-season football transfer markets from a network perspective (Liu et al. 2016; Matesanz et al. 2018; Wand 2022). Furthermore, Bond et al. described the European football loan system across multiple seasons (Bond 2020). Other relevant studies consider football transfer networks within a single season, within a single country, or both. Palazzo et al. (2023) identify teams with similar trading patterns (i.e., communities) considering both the network and team attributes. Clemente

(2023) identify communities on a global level, uncovering central leagues in the global transfer network. Turning the question around, Xu (2021) models link-creation among top teams in the summer transfer window in the 2019/2020 season.

In this paper we consider the complex interplay of different (network) variables derived from the structure of the underlying transfer market alongside simpler notions of engagement in the transfer market. In doing so we engage with findings in prior works describing football transfer networks per season (Matesanz et al. 2018; Wand 2022; Xu 2021) and over multiple seasons (Liu et al. 2016; Bond 2020), considering both average seasonal success and long-term cumulative success. Moreover, with respect to prior works, we focus less on financial flows and more on the movement of players. On this basis, we evaluate the relationship between team performance and team participation in the European transfer market at the highest levels of the sport over 28 years. We ask: How does domestic league performance of a top football club relate to the network position and the engagement of this club in the European player transfer market?

In answering abovementioned question, we make a number of contributions. First, we provide further description of the European transfer market network, investigate meaningful whole-network measures, and touch upon its community structure. Second, we corroborate several existing findings on the relationship between league competitiveness, team engagement in the transfer market, and team performance. Then, we show that network position measures such as degree or closeness centrality are positive indications of team performance, even taking into account the team's engagement with the market. Finally, we discuss a specific limitation that may be of relevance also more generally to network analysis of transfer markets for professional football players.

The remainder of this paper is organized as follows. First, the Data section discusses the data sets obtained and how the data was processed before analysis. Then, the Methodology section provides a number of preliminary definitions, before the Regression model section describes the layout of the statistical model and the setup of the experiments. After that, the Results and discussion section analyses the results, whereas finally the Conclusion section summarizes the main findings and provides suggestions for future work.

Data

In this section, we introduce the source of the data sets along with their properties. Subsequently, we will discuss the steps taken to preprocess the data.

Raw data

This research combines two raw data sets. The first is a transaction data set of player transfers involving teams competing in eight top European football leagues over 28 years. The second data set contains detailed results for all matches played within the relevant leagues over a large subset of the relevant years. A summary of the two data sets are displayed in Table 1, as well as the associated countries and country codes of the selected leagues.

As in Wand (2022); Xu (2021), we use a data set of player transfers scraped from https://www.transfermarkt.co.uk/. This data is available for download at https://github.com/ewenme/transfers and was produced by GitHub user `ewenme` on the 19th of April

Table 1 An overview of the data used in the research

League	Country	Country code	Transaction data	Performance data
English Premier League	United Kingdom	GB	93/94 – 20/21	93/94 – 20/21
French Ligue 1	France	FR	93/94 – 20/21	93/94 – 20/21
German 1.Bundesliga	Germany	DE	93/94 – 20/21	93/94 – 20/21
Italian Serie A	Italy	IT	93/94 – 20/21	93/94 – 20/21
Spanish La Liga	Spain	ES	93/94 – 20/21	93/94 – 20/21
Portugese Liga NOS	Portugal	PT	93/94 – 20/21	94/95 – 20/21
Dutch Eredivisie	Netherlands	NL	93/94 – 20/21	93/94 – 20/21
Russian Premier Liga	Russia	RU	93/94 – 20/21	12/13 – 20/21
Total			224	204

2022. Here, records are included on transfers to or from teams participating in the first divisions of Germany, Spain, Italy, France, Portugal, the Netherlands and Russia and the first and second divisions of England in a particular season. Teams outside these leagues (e.g., a Turkish club or a club from a different hierarchic level) are included as alters. However, transfers among teams outside these leagues are not included in the data. The downloaded data is a set of records that consist of *a player* that was transferred from a *team* to another *team* in a specified *period* of a certain *season* for a given *price*.

The league performance data stems from https://www.football-data.co.uk/data.php. This data set consists of match records of a *team* that played against another *team* which resulted in a *final score*. Combined with knowledge of the scoring rules for determining the final rank of the teams in the league at the end of the season, this data allows us to incorporate a league performance measure of the clubs (see League performance measure). Note that the league performance data set has incomplete overlap with the transaction data set, missing 19 seasons for the Russian Premier Liga and one season for the Portugese Liga NOS.

Preprocessing
This section sets forth the steps taken to guarantee a reliable, representative and reproducible configuration of the data.

1. Only the top tiers of every country will be considered. Therefore, transfers of clubs when appearing in the English Championship — which were originally present in the data set — have been excluded. The data set then includes 136,339 transactions in the default configuration.
2. All youth teams and non-first teams of a certain club have been aggregated into a seperate node, such that the influence of affiliated teams on the position and performance of the first team is interpreted collectively. This change of notation was applied in all 10,860 instances (which is 7.97% of the transactions). This did not affect the total number of transactions analysed.
3. Over the years, some clubs have changed names. Furthermore, some teams were noted under different names in different instances within the data set. The current or most recent name of a team was chosen as standard and the differing notations were adjusted. This did not lead to a change in the number of transactions.

4. Some transactions were present twice. This is the case when the player moved from a team present in the above competitions to another team present in the above competitions. Double notation of a transaction is also present when a player is loaned out. This transaction is taken into account twice: at the beginning of the loan and at the end. To account for the transactions that have been documented more than once, the first occurrence of every distinct transfer in the data set is used. 96,645 transactions are present after accounting for transfers which are documented more than once.

5. Moreover, some transfers existed where the player transferred to and from the same team. These transfers have probably not been documented properly and are therefore deleted from the configuration used in this research. This applied to 240 instances.

6. Lastly, transactions containing players that retired, moved to or from an unknown club, or that became a free agent have been disregarded. This was a total of 4,954 instances.

The final data set, after preprocessing, consists of 91,451 transactions and is provided in the Supplementary information. This data set is used to construct a network representation of the European football transfer market.

Methodology

This section describes the network approach and statistical methods used in this work. First, we introduce the measure used to quantify the in-game performance of teams within their leagues per season and overall. Second, we elaborate on the construction of the European transfer network, again per season and overall. Then, we present the engagement and network measures used in the analysis. Lastly, we discuss our statistical models for conducting regression analysis.

League performance measure

A season in football consists of a collection of matches that are played in one year between teams in the same league. To study team performance, the first step is to quantify a team u's performance or rank $R(u, y)$ in a given season y. A team's league performance is then defined as that collection of (in-game) performances over a season relative to its competitors — which is assumed to regress to the mean over time (Beck and Meyer 2011). We use the rank of a team in the final season standings. Since a lower rank implies better performance, this score is then inverted relative to the number of clubs in the competition. For example, if Team A finishes first (1st) out of twenty (20) clubs, Team A's score for that specific season will be 20. The raw rank is standardized to a format of sixteen teams. Teams within every league can gain at most sixteen (16) points and teams finishing in last place score one (1) point. This gives us the standardized performance indicator $P(u, y)$ which is comparable across leagues.

The aim of an overall league performance measure $P(u)$ is to indicate the dominance of a team u within their respective league *across seasons* in a way that the measure is comparable among leagues. Due to the dynamic nature of domestic club competitions (as a result of promotion and relegation), not all clubs will appear in one specific league for all seasons. Relegation to a lower league indicates worse performance. Therefore, $P(u, y) = 0$ when the team is playing in a lower league during season y. We then sum the

standardized performance indicator $P(u, y)$ for all y to obtain a team's cumulative score in the context of its specific league.

The standardized total score, the sum of the $P(u, y)$ values over all years, is then normalized to account for leagues of varying data coverage and size, as detailed in Table 1. This last transformation gives us the standardized league performance indicator $P(u)$ which is comparable across leagues. It should be noted that the measure is not intended to allow for cross-league comparisons of teams directly, were they to play one another. For example, Real Madrid has, on average, been as dominant in the Spanish La Liga as PSV Eindhoven has been in the Dutch Eredivisie (see Table 5).

Research by Vales et al. (2018) poses a comparative instrument for evaluating our league performance measure. Namely, the distribution of P gives an insight in the differing competitive profiles of the leagues and this question has also been studied in Vales et al. (2018). A lower standard deviation within a league implies that teams obtain a score that is, on average, closer to each other, suggesting a more competitive character as teams performances are relatively similar. Table 2 shows the resulting ranking of the European leagues by competitiveness. Considering only the teams also ranked by Vales et al., we obtain very similar results. The Dutch Eredivisie and the Russian Premier Liga were not incorporated in the research of Vales et al. Our findings suggest that these leagues have the highest deviations of league performance, thus implying they are the least competitive.

Network construction

The European football transfer network as used in this paper is constructed from the transaction data set introduced in the Preprocessing section. The data records transfers involving teams appearing in eight top European leagues $\mathcal{L} = \{L_{GB}, L_{FR}, L_{DU}, L_{IT}, L_{ES}, L_{PT}, L_{NL}, L_{RU}\}$ over a range of seasons. However, as noted above, not all clubs appear in one specific league for all seasons. The European football transfer network for season y is thus centered around the set of teams that competed in the top leagues that season $u \in \mathcal{L}_y$. The recorded transfers of players to or from these teams are used to construct the set of edges E_y. The set of nodes V_y is then the set of

Table 2 The competitive profiles of the leagues indicated by the distribution of P and a comparison of the results to research of Vales et al. (2018)

Symbol	League	Obs	Mean	Std. Dev.	Rank	Rank in Vales et al. (2018)
L_{GB}	English Premier League	49	3.48	4.10	3	3
L_{FR}	French Ligue 1	44	3.70	3.91	6	6
L_{DU}	German 1.Bundesliga	42	3.62	4.06	4	5
L_{IT}	Italian Serie A	51	3.06	4.01	5	4
L_{ES}	Spanish La Liga	48	3.56	4.17	1	1
L_{PT}	Portugese Liga NOS	43	3.32	4.10	2	2
L_{NL}	Dutch Eredivisie	30	5.07	4.66	(1)	n/a
L_{RU}	Russian Premier Liga	27	4.55	4.37	(2)	n/a
Total		334	3.68	4.14	-	-

teams that either appear in a top league or engage in a transfer with a top team that season, i.e., for all $u, v \in V$ it holds that $(u, v) \in E_y$ or $(v, u) \in E_y$, and $u \in \mathcal{L}_y$.

There is a direction to the edges, as transferred players move from one club to another. Furthermore, the weight of an edge is defined by the number of players transferred in that direction in that season. This way we look primarily at the trading dynamics of teams. Previous literature has considered teams' financial activity and the evidence is clear that greater financial resources go hand-in-hand with better team performance (Frick 2007; Mourao 2016; Wand 2022; Liu et al. 2016; Matesanz et al. 2018). Price is considered in this particular research only as something of a control for the notion that "football is a money game" (Liu et al. 2016)

Each of the 28 seasons present in the transaction data set (93/94 - 20/21) is used to construct a directed weighted network $G_y = (V_y, E_y)$. Then, the season snapshots are combined into a single cumulative network $G = (V, E)$. There are $n = |V|$ teams in total and $m = |E|$ directed edges, as players move from one club to another.

Engagement measures

Specific measures are used to express a team's engagement in the European football transfer market, per season and overall. The main variables quantifying a team's engagement are the *total transfers in* and *total transfers out*. Following Wand (2022), we also consider the *total amount paid* and *total amount received* for transferred players. To study overall performance, we consider the average of these measures over the seasons wherein a team appeared in the top leagues.

Transfers This measure is the most straightforward; it captures the overall number of transfers involving a team. For season snapshots, we consider the number of players that a team has received from other teams $W_i(u, y)$ and the number of players that a team has transferred to other teams $W_o(u, y)$. Notably, these variables can also be expressed in the language of networks: our networks are weighted and directed by the number of transfer so they correspond to the weighted indegree and weighted outdegree of a team (also sometimes called in-strength and out-strength). For studying overall performance, we compute the average total transfers over the seasons where the team appeared in the top league because then the data set includes all its transfers. Specifically, we compute the average over the seasons where the league performance measure is greater than zero, i.e., $P(u, y) > 0$.

Transfer fees An alternative weighting is used to describe a second aspect of a team's engagement in the football transfer market. Specifically, the total fees paid for incoming players in a season $M_i(u, y)$ and total fees received for players transferred to other teams $M_o(u, y)$. Total transfer fees are the combined total $M(u, y) = M_i(u, y) + M_o(u, y)$. For studying overall performance, we compute the average over the seasons where the team appeared in the top league $P(u, y) > 0$.

Network position measures

A collection of measures is used to express a team's network position in the European football transfer market, per season and overall. We use a set of network measures to quantify a team's network position: *degree, betweenness centrality, closeness centrality*, and *clustering coefficient*. These definitions are derived from Barabási (2016), Newman

(2018), and Easley and Kleinberg (2010), and largely cover a set of network measures similar to what is also elaborated on from a sports analytics point of view in Palazzo et al. (2023).

Degree   The degree of a node $D(u)$ of a node $u \in V$ is the number of nodes it is directly connected to. In this context, regardless of the direction or frequency of the relationship, degree "helps us to understand the main market strategy of a team by considering the number of its partners" Palazzo et al. (2023). There is reason to expect a positive correlation between degree and league performance. Mourao (2016) describes the significant impact a team's league performance has on its number of trading partners.

Closeness   Player transfers are the links that connect teams in the European football transfer network, and we consider teams to be *indirectly* connected when there is a path between them via transfers with other teams. The closeness centrality of a node $C(u)$ describes the average distance, i.e., the average shortest path length, of one node to all other nodes in the networks. Whereas degree is a local network measure with a direct interpretation in terms of transfers, closeness captures a more global notion of network position. Teams with a higher $C(u)$ value are relatively well-connected with all other teams in the transfer market, directly and indirectly. In our case, closeness is computed on the undirected, unweighted version of the networks again disregarding the direction or frequency of the relationships. Previous literature has suggested a positive correlation with league performance (Liu et al. 2016).

Betweenness   The betweenness centrality $B(u)$ of a team quantifies how often the team helps connecting other (pairs of) teams via *shortest* paths. For this we use directed shortest paths, computed on the directed unweighted network. In this way, betweenness captures a second global notion of network position, albeit in substantially different way than closeness does. Teams with a higher $B(u)$ play an important role in brokering connections between parts of the extended transfer network that are otherwise relatively unconnected. Here again, previous literature has suggested a positive correlation with league performance (Liu et al. 2016).

Clustering   The clustering coefficient $CC(u)$ is a local network measure describing the extent to which the neighbours of a node are also directly linked. In this context, clustering coefficient expresses the extent to which trading partners of a team also trade with each other. This often happens for teams from the same league. The clustering coefficient captures the embeddedness of a top team in its local network neighbourhood and serves to identify closely-knit groups among the top teams. As it is difficult to meaningfully give real-world interpretation to measures of clustering that take weights into account, we use the directed unweighted clustering coefficient.

Cumulative network

When describing and interpreting the network structure of the European football transfer market as a whole in the Football transfer network section, several network level measures are reported on. The average degree $\overline{D}$ describes the average number of trading partners of each team in the network. The average weighted degree $\overline{W}$ is the average of the weighted degrees of all nodes, and in our case denotes the average number of transactions a team has made as or with a team in the top leagues. A connected component is a subset of nodes in which each node can reach all others via an undirected

path. We use the average distance $\bar{d}$, or average shortest path length, between teams to understand the connectivity within the network. The diameter d_{max} is the longest shortest path or maximal distance observed in the network. It gives the distance between the two teams that are furthest away from each other with respect to observed transfer relationships. A low diameter indicates a transfer network where even the least connected teams are well-integrated. The density ρ of a network is the ratio between the number of edges m actually present in a network and the total possible number of edges $n \cdot (n-1)$. The node clustering coefficient $CC(u)$ is a measure of the network density around individual nodes (defined above). The average clustering coefficient $\overline{CC}$ is the average $CC(u)$ value over all nodes $u \in V$. It indicates the average local embeddedness of teams in the transfer network. Real-world networks can often be partitioned into "network communities" where nodes in a community are more strongly connected to each other relative to the rest of the network. The modularity score Q quantifies the quality of a division of a network into communities obtained after applying a standard modularity maximization algorithm to find the communities (Blondel et al. 2008), and is generally higher if there is a profound community structure.

Control variables

League   The league of a team is used as a categorical variable to account for the differences among leagues that influence the sportive performance of the teams (Vales et al. 2018) and their financial performance (Liu et al. 2016) (e.g., the competitive profile of a league and correlated structural differences in participation in the European football transfer market). We denote the first league of country CO as L_{CO} and include the league L_{CO} of team u as a categorical control variable.

Prior Rank   In considering league performance per season, we control for the rank of the team in the previous season $P(u, y-1)$. This captures autocorrelation produced when teams continue to perform well from one season to the next. Recall that $P(u, y) = 0$ in years where the team does not appear in the top league.

Regression model

This research focuses on the performance of football teams that have appeared in any one of eight European domestic leagues in the available seasons (see Table 1). First, a linear regression model is used to predict league performance $P(u, y)$ of a team u in season y on the basis of its network position and level of engagement in the European football transfer market that season. There are 150 teams playing in the eight top leagues in a complete year. Each of these top teams has a league performance score $P(u, y)$ as defined in the League performance measure section. The network position of a team as defined in the Network position measures section is expressed by $D(u, y)$, $B(u, y)$, $C(u, y)$ and $CC(u, y)$. The engagement of a team as defined in the Engagement measures section is expressed by $W_i(u, y)$ and $W_o(u, y)$, as well as $M_i(u, y)$ and $M_o(u, y)$. The league L_{CO} of team u is included as a control, and the rank of the team in the prior year $P(u, y-1)$ is included to capture autocorrelation.

We then construct an analogous analysis relating the overall league performance of a team to its position in the cumulative transfer network. The engagement of a

team is captured by the average of $W_i(u, y)$ and $W_o(u, y)$ over the seasons that the team appears in the league. The network position of a team, on the other hand, is defined by the same metrics as defined in the Network position measures section computed for the cumulative transfer network $D(u)$, $B(u)$, $C(u)$ and $CC(u)$. Note that the interpretation of these metrics is not straightforward for teams that rarely promote into the top leagues, as relegated teams are not centered in the data collection. For this reason, we limit our regression analysis to the top 15 teams in each league by performance $P(u)$ and provide a quantitative discussion in the Limitations section.

An overview of the variables in the model and the associated node measures can be found in Table 3. As the explanatory variables do not share a comparable scale, in the regression, the coefficients are standardized. This enables us to compare the coefficients of the variables on an equivalent scale.

Experimental setup

The Data section details how the data was acquired and processed before analysis. After preprocessing, we used the software Gephi (https://gephi.org) for its visualisation tools and for the calculation of the whole-network measures reported in the Network construction section. The league performance measure was calculated for all teams through the method described in the League performance measure section. Network and engagement variables were computed using the *networkx* package in Python (https://networkx.org). The data set with engagement and network variables was combined with the data set containing the league performance variable. This enabled us to implement the model proposed in the Regression model section using the *pandas*, *statsmodels* and *patsy* packages in Python (https://pandas.pydata.

Table 3 An overview of the variables per season and overall as used in the regression model

Symbol	Node measure	Obs	Mean	Std. Dev.	Min.	Max.
$P(y)$	League performance measure	3814	8.36	4.57	1.00	17.6
$D(y)$	Degree	3814	21.31	9.55	1	138
$B(y)$	Betweenness centrality	3814	0.010	0.007	0	0.076
$C(y)$	Closeness centrality	3814	0.298	0.031	0.173	0.380
$CC(y)$	Clustering coefficient	3814	0.035	0.031	0	0.500
$W_i(y)$	Transfers in	3814	12.21	5.58	0	115
$W_o(y)$	Transfers out	3814	14.81	10.23	0	225
$M_i(y)$	Fees paid (mil. £)	3814	15.86	29.64	0	342.09
$M_o(y)$	Fees received (mil. £)	3814	12.63	23.39	0	299.04
P	League performance measure	120	8.30	3.53	2.81	15.46
D	Degree	120	220.1	64.6	45	440
B	Betweenness centrality	120	0.0062	0.0036	0.0004	0.0222
C	Closeness centrality	120	0.418	0.017	0.373	0.471
CC	Clustering coefficient	120	0.141	0.054	0.051	0.445
$\langle W_i \rangle$	Transfers in (avg)	120	11.70	3.08	4.46	25.67
$\langle W_o \rangle$	Transfers out (avg)	120	15.04	6.84	3.15	43.75
$\langle M_i \rangle$	Fees paid (mil. £, avg)	120	17.96	19.68	0.01	86.79
$\langle M_o \rangle$	Fees received (mil. £, avg)	120	14.55	12.08	0.19	48.12

org, https://www.statsmodels.org, https://patsy.readthedocs.io). Replication materials are provided in the Supplementary information.

Results and discussion

In this section, we describe the characteristics of the European football transfer network from 1993/94 through 2020/21. We then explore the relationship between a team's network position, market engagement, and league performance.

Football transfer network

In this section we investigate the characteristics of the European football transfer market. This section builds on the definition of the overall network as described in the Methodology section. The values for each measure are presented in Table 4 and a network visualisation is presented in Fig. 1.

The European football transfer network possesses many characteristics that mark other real-world networks. It is a sparse graph as only a small fraction of the possible pairs of teams in the network are directly connected by the transfer of a player. Expressed by the low density, this reinforces the notion that trading players happens on a non-incidental basis. Even so, all the nodes belong to the same connected component. This means that transfers in the European football transfer market *indirectly* connect all these teams together. Moreover, the average distance between nodes is 3.18. Such a low value implies that this European football player transfer network is a "small-world" where most teams are only a few former-teammates-of-former-teammates away from any other team. Notably, this appears to be an emergent property of the football transfer market over multiple seasons. Prior analyses of multi-season football transfer networks have found the same (Liu et al. 2016; Matesanz et al. 2018). Season snapshots, however, do not necessarily exhibit the "small-world" property (Wand 2022).

Interestingly, we see from the visualisation in Fig. 1 that teams appear to form eight clear communities, where there is a greater density of transfers within the identified groups than between them. We confirm, using the teams in the top domestic leagues, that teams fall into a community together with other teams that appear in the same league. In fact, we found that league as a categorical variable and community as a categorical variable match one-on-one. Teams that do not appear in one of the eight leagues fall into the community that they trade the most with. This means that teams tend to

Table 4 Whole-graph metrics of the European football transfer network

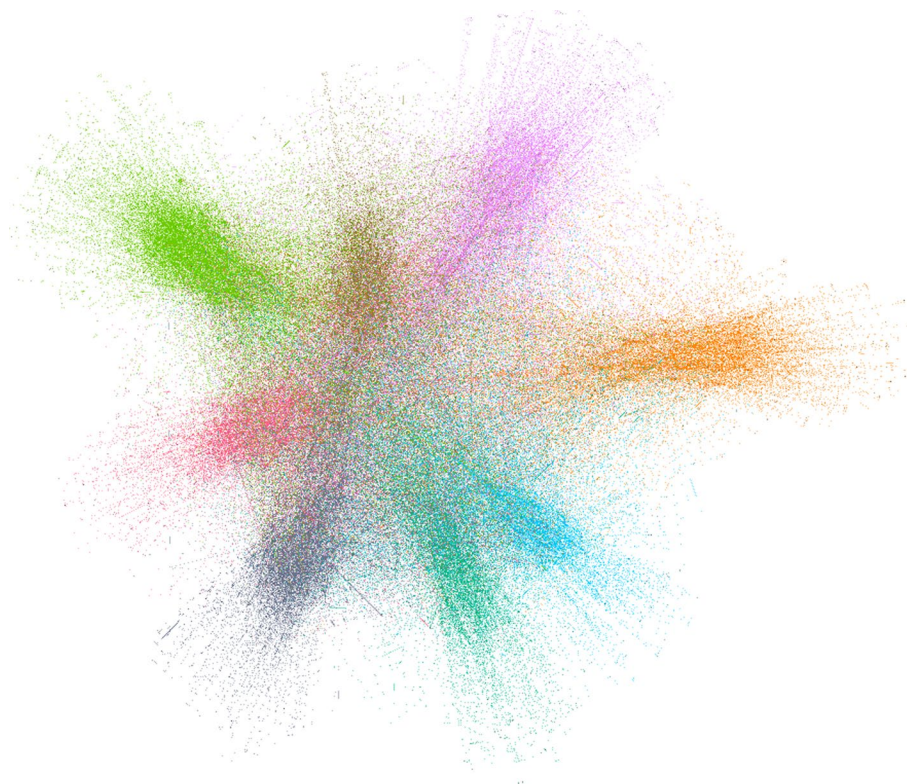| Symbol	Metric	Value
n	Nodes	4,876
m	Edges	46,079
$\overline{D}$	Average degree	15.6
$\overline{W}$	Average weighted degree	37.5
$\overline{d}$	Average path length	3.18
d_{max}	Diameter	6
ρ	Graph density	0.002
Q	Modularity (r=1.0)	0.596
$\overline{CC}$	Average clustering coefficient	0.314

Fig. 1 Visualisation of the European football transfer network, coloured by community. Legend: Purple: Portugese Liga NOS, 18.91%, Lime: Italian Serie A, 15.91%, Orange: Russian Premier Liga, 14.95% Blue: Dutch Eredivisie, 11.77%, Teal: German 1.Bundesliga, 10.6% Red: French Ligue 1, 10.15%, Grey : English Premier League, 9.89% Gold: Spanish La Liga, 7.81%

trade more with those in the same league whereas transfers between leagues are relatively scarce. In a similar vein, we can see that the average clustering coefficient $\overline{CC}$ is 0.314 which is about 160 times higher than expected from the graph density ρ. High clustering is a common feature of real-world social networks, specifically, and indicates that teams have a propensity to trade also with trading partners of their trading partners. This underscores the tendency of teams to trade with a familiar set of teams — often, linked to their respective league, and that the trading behaviour of teams differs on a domestic or international scale. Similar findings were presented in Liu et al. (2016); Wand (2022); Bond (2020). On this basis, teams seem to form communities with teams of similar geography or agent-involvement (Palazzo et al. 2023).

Prominent teams

The presence of highly connected nodes is a common feature of real-world networks, and an important feature of the European football transfer network. While the average degree is 15.6, there is a small group of so-called 'hubs' with many more connections (Barabási 2016). Highly skewed degree distributions with prominent 'hubs' have been identified in football transfer networks analyzed per season (Wand 2022) and cumulatively (Liu et al. 2016). For our network, Table 5 presents the ten nodes with the highest

Table 5 The left column displays the hubs in the European transfer network and the right column showcases the most dominant teams in their respective leagues

Rank	Team u	$D(u)$	League	Team u	$P(u)$	League
1	Udinese Calcio	440	L_{IT}	Bayern Munich	15.46	L_{DU}
2	Parma Calcio 1913	409	L_{IT}	FC Barcelona	15.37	L_{ES}
3	SL Benfica	401	L_{PT}	Ajax Amsterdam	15.11	L_{NL}
4	Sporting CP	366	L_{PT}	Real Madrid	15.06	L_{ES}
5	Genoa CFC	345	L_{IT}	PSV Eindhoven	15.05	L_{NL}
6	Inter Milan	336	L_{IT}	Manchester United	15.03	L_{GB}
7	Vitória Setúbal FC	333	L_{PT}	FC Porto	14.98	L_{PT}
8	CS Marítimo	320	L_{PT}	SL Benfica	14.29	L_{PT}
9	AS Roma	320	L_{IT}	Juventus FC	13.94	L_{IT}
10	SC Braga	319	L_{PT}	Sporting CP	13.89	L_{PT}

degree, that is, the ten teams with the most number of trading partners. The large disparity between the average degree and the degree of the hubs means that the distribution of degree is right-skewed, or heavy-tailed, and the majority of nodes have relatively few trading partners. The high degree hubs are instrumental in efficiently connecting the transfer network. This is underlined by the strong correlation between degree and the tendency of being located on shortest paths between nodes (degree D and betweenness centrality B share $r^2 = 0.83$ for the top 15 teams in each league). Hubs help minimize the distance between nodes and thus serve to connect the teams in the transfer network across different leagues.

Teams with a higher degree also tend to perform better, according to prior research by Liu and colleagues (Liu et al. 2016). However, this does not tell the whole story. In Table 5, we compare the top-10 teams by league performance and by degree in the European football transfer market. While there is minor overlap, the differences convey that a team's success is not trivially related to this one measure of position in the network. Therefore, further exploration of the relation between network position and performance is needed which Bond's loan network also suggests (Bond 2020).

Transfer network and team performance

This section presents the results of various linear regression models which highlight the relations between the node measures and the league performance measures. We first consider team performance per season and then the overall league performance measure. In both cases we include as explanatory variables the level of engagement with the transfer market and a set of network position metrics. The models that are displayed in this table have standardized coefficients so that comparison of coefficients between attributes is simplified.

In Table 6, the columns denote different linear regression models. The first three models correspond to regressions for per-season league performance while the second three refer to overall league performance. The rows represent the variables. When a variable is used in a model, the (rounded) coefficient of said variable is displayed in the corresponding cell. Heteroskedasticity-consistent standard errors are used in computing *p*-values.

Model 1 includes only our control variables, capturing autocorrelation in league performance and the distinct competitive profiles of the different leagues — as observed by

Table 6 Results of multiple linear regression models per season and overall

Variable Model	(1)$P(y)$	(2)$P(y)$	(3)$P(y)$	(4)P	(5)P	(6)P
$P(y-1)$	2.845**	2.357**	2.265**	–	–	–
M_i		0.710**	0.668**	1.212*	1.181*	1.126
M_o		−0.010	-0.071	1.300*	1.228*	1.094
W_i		−0.985**	−1.192**	−2.270**	-2.429**	-2.333**
W_o		0.721**	0.552*	2.165**	0.864	1.079
D			x		1.694**	x
B			0.317**		x	0.649
C			0.264*		x	0.891*
CC			0.063		x	-0.144
L_{DE}	0.047	0.448	0.370	3.349	2.464	1.659
L_{ES}	0.107	0.363	0.286	2.290	1.974	1.231
L_{FR}	0.040	0.491	0.390	3.306*	2.264*	1.481
L_{GB}	0.054	−0.442	−0.506	−0.502	−0.906	−1.431
L_{NL}	0.146	0.972**	0.950**	6.039**	4.941**	3.933**
L_{PT}	0.059	1.065**	0.913**	5.399**	3.860**	2.907**
L_{RU}	0.324	0.794	0.683	3.570**	3.616**	2.550*
Obs.	3480	3480	3480	120	120	120
AIC	18680	18420	18390	529.3	499.9	510.6
BIC	18730	18500	18490	562.8	536.1	552.4
r^2	0.394	0.439	0.443	0.681	0.754	0.740
max VIF	1.75	2.40	2.77	9.34	11.20	11.15

Note: *$p < 0.05$ **$p < 0.01$ x excl. VIF

Vales et al. (2018). The Dutch and the Russian leagues have the largest coefficients relative to our reference level, the Italian Serie A. The top 15 teams place consistently higher in these leagues, implying that they are the least competitive; this corresponds nicely to the league performance statistics presented in Table 2.

Model 2 goes on to confirm established results on teams' engagement in the European football transfer market. The ability to pay for expensive incoming players in a season is strongly related to league performance that season; football is indeed a "money game" (Liu et al. 2016). Incorporating also the number of transferred players, and we see that having many incoming players is inversely related to league performance that season. This finding remains consistent and significant across all models and furthermore coincides with research by Xu (2021), whose analysis does not include transfers with lower tier teams. In this model, the coefficients on the league controls absorb also average differences in the engagement of teams across leagues. Less engaged are the Dutch, Portuguese, and Russian leagues; note that the latter has fewer observations and thus less statistical power. The English Premier League is consistently the most engaged in the European transfer market, on the whole.

Network position metrics are incorporated first in Model 3. Both closeness and betweenness centrality in the season's transfer network are strongly related to performance. Playing a central role has been linked to better match performance by previous literature (Liu et al. 2016; Bond 2020). We avoid degree because it introduces substantial multicollinearity alongside total transfers (VIF = 15.43). This is because, in a single season, it is not so common for a team to engage in transfers of multiple players with the

same trading partner. Next we consider the cumulative network, where degree becomes more relevant (Bond 2020).

For overall league performance, we first consider a regression with just our engagement variables. Model 4 shows that the relation between high-value transfers and league performance is now symmetrical, presumably because well-resourced teams will tend to find themselves on both sides of high-value transfers over multiple seasons. The negative relation between league performance and receiving many incoming players persists as we consider the European transfer market over multiple seasons. At the same time, having a prominent position in the unweighted, undirected, cumulative transfer network is indicative of overall league performance. Models 5 and 6 introduce either degree or a set of three network position metrics: betweenness, closeness, and clustering. Introducing these variables together produces multicollinearity with degree (VIF = 28.50). Degree becomes a useful network position metric over multiple seasons; other network metrics can be used together but they are not as precise. However, some caution is warranted in that the network position measures are less reliable for lower ranked teams. This limitation is discussed in the following section and regression tables with versions of Models 5 including more and fewer teams are shown in Appendix A: extended regression table. The fitted coefficients are of the same sign and with overlapping confidence intervals with the top 10 or the top 20 teams in each league; note that degree has a *p*-value of 0.12 with the top 10 teams.

Limitations

Network analysis of football transfer markets, including this one, generally focus on top teams. In particular, data collection on player transfers has centered on teams appearing in top national leagues (Liu et al. 2016; Matesanz et al. 2018; Bond 2020; Wand 2022; Xu
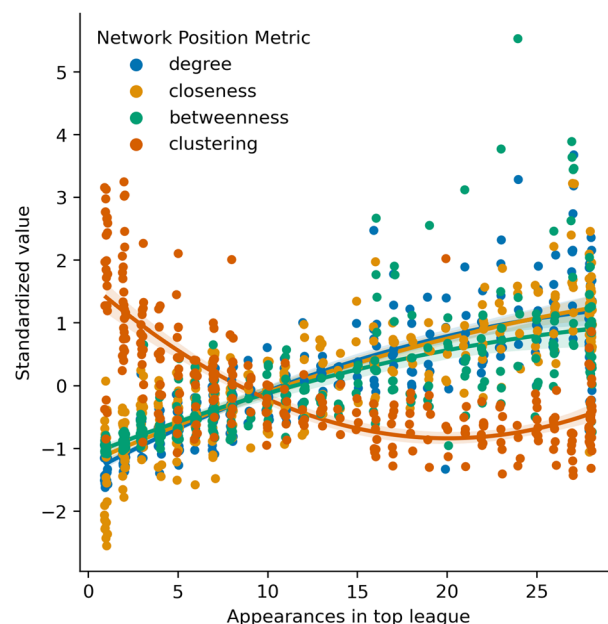

Fig. 2 Network position metrics of teams as a function of their number of appearances in the top leagues. Polynomial fit of order two is shown. The number of appearances for Portuguese teams may be off by one due to missing standings from the 1993/94 season. Teams from the Russian league are excluded

2021; Clemente 2023). As noted in (Wand 2022), this is largely due to reasons of data availability and further insights might be gained by considering also lower-level leagues. Here we extend this observation to make concrete one specific way that standard data collection practices limit the insights that can be gained from network analysis. Figure 2 shows several network position metrics as computed on the European football transfer network for seasons 1993/94-2020/21 and their dependence on the number of appearances of that team in the top league. Not all clubs appear in the top league for all seasons and the cumulative network is missing transfers to and from teams during periods when they were relegated to lower leagues. We see that this has a dominant effect on the network metrics for teams with few appearances in the top leagues. In our case, this complicates the interpretation of network position metrics for lower performing teams and we limit our regression analysis to top teams in each league. Extending the analysis to teams with fewer appearances in the top leagues would require additional data collection and is a promising avenue for future work.

We include this extended note on this limitation in that network data collection centered on top teams is common and likely to affect the observed network structure of football transfer markets in other ways of relevance well beyond this work. It is difficult to say, for instance, if the average clustering coefficient $\overline{CC}$ presented in Table 4 would be higher or lower were the data collection to include all transfers between teams also in lower leagues. To the best of our knowledge, the effects of imposing network boundary criteria as would be applicable to football transfer markets has not been systematically studied.

Conclusion

This paper has approached the European football transfer market from a network science perspective. We found that the football transfer network shares characteristics that are also present in other real-world networks. Notably, a low density of links and some prominent 'hubs' with exceptionally many trading partners. We see that teams cluster into communities with teams from the same league and tend to be 'embedded' in their direct neighbourhood. There are some teams that transcend this phenomenon and that connect the different communities within the network, leading to the emergence of a small-world effect. Overall, this makes for a connected network on a local level that is bridged through the hubs.

In relating a team's position in the European football transfer market to league performance we considered networks both per season and overall. Some aspects of a team's trading behaviour coincides with higher performance both in the same season and over the longer term. Teams that receive many players tend to do consistently worse while teams that can spend substantial amounts on incoming players tend to do consistently better. This speaks to recent work applying sophisticated network statistics to transfers *among* teams in top European leagues (Xu 2021) and serves to extend the finding to include transfers with lower tier teams. This may be especially interesting with respect to differences in the loaning behaviour of teams depending on their revenue as documented by Pantuso and Hvattum (2021); Bond (2020). When accounting for league and

these aspects of engagement in the transfer market, teams that *also* have a more central position in the network perform better, on average, in their domestic league. These findings are in well line with previous literature finding that a more central network position overall suggests a better performance (Liu et al. 2016; Bond 2020). However, regression analysis allows for more nuance, showing, for instance, that network centrality is not so easy to disentangle from engagement in the market over multiple seasons.

This paper indicates multiple directions of further research. An obvious improvement would be to extend data collection beyond top European leagues. Lower leagues play an important role in developing talent but are much less often considered, as noted also in Wand (2022). The football transfer market is also connected globally, with talent flowing from and to different parts of the world. Extending this research to other regions will explore the robustness of our findings in countries which fundamentally play a different role in the global transfer market (Clemente 2023). On the other hand, taking a closer look at specific leagues, such as the research by Bond (2020); Palazzo et al. (2023), can expose more detailed trading patterns. Overall, rapidly advancing network science approaches may aid coaches and managers in the overall strategical decision-making and positioning of their team in the football transfer market.

Appendix A: extended regression table

Reported in the main text are the regression results including the top 15 teams in each league. Table 7 reports values with more and fewer teams per league. Note that the regression coefficients are of the same sign and with overlapping confidence intervals with the inclusion of five more or five fewer teams per league; we take this to indicate

Table 7 Results of Model 5 for different sets of top teams

Variable Model	(5)*P, Top*20	(5)*P, Top*15	(5)*P, Top*10	(5)*P, Top*5
M_i	1.143±0.52	1.181±0.57	0.777±0.64	0.345±0.55
M_o	1.509±0.50	1.228±0.53	1.777±0.64	0.621±0.42
W_i	−2.011±0.38	−2.429±0.39	−1.673±0.65	−2.232±0.84
W_o	0.050±0.60	0.864±0.67	0.352±0.98	1.465±0.66
D	2.007±0.29	1.694±0.40	1.030±0.66	−0.915±0.75
L_{DE}	1.692±1.05	2.464±1.28	2.530±1.78	−1.171±1.38
L_{ES}	1.559±0.96	1.974±1.13	1.135±1.88	−2.479±1.54
L_{FR}	1.678±0.95	2.264±1.10	1.885±1.57	−0.396±1.37
L_{GB}	−1.095±0.76	−0.906±0.94	−0.100±1.01	−2.095±1.21
L_{NL}	3.960±1.05	4.941±1.18	4.639±1.72	1.659±1.08
L_{PT}	3.055±0.92	3.860±1.01	3.449±1.31	3.939±1.30
L_{RU}	3.027±0.93	3.615±1.08	3.143±1.52	-1.838±1.38
Obs.	160	120	80	40
AIC	676.0	499.9	334.0	128.6
BIC	716.0	536.1	365.0	150.5
r^2	0.778	0.754	0.666	0.757
max VIF	12.20	11.20	19.25	18.78

that the qualitative findings are robust. Including only the top 5 teams in each league leads to different coefficients for many of the variables (even controls). We take this to indicate that this subset is too small or too special to draw reliable conclusions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-024-00675-7.

> **Additional file 1.**

Author contributions

TD, CM, and FT conceptualized the study and methodology. TD and CM wrote the manuscript. CM and FT supervised and administered the project. TD prepared the dataset and visualized the results. TD and CM performed the experiments. All authors reviewed and approved the manuscript.

Declarations

Competing interests

The authors declare that they have no Conflict of interest.

References

Matheson V.A (2003). European football: A survey of the literature. Technical report, Williams College, Department of Economics . https://books.google.nl/books?id=1DLzMgEACAAJ

Frick B (2007) The football players' labor market: Empirical evidence from the major European leagues. Scottish J Polit Economy 54(3):422–446

Goes FR, Meerhoff LA, Bueno MJO, Rodrigues DM, Moura FA, Brink MS, Elferink-Gemser MT, Knobbe AJ, Cunha SA, Torres RS, Lemmink KAPM (2020) Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. Europ J Sport Sci 21(4):481–496

Li C, Zhao Y (2020) Comparison of goal scoring patterns in the big five european football leagues. Front Psychol 11:619304

Fabíola Zambom-Ferraresi VR, Lera-López F (2018) Determinants of sport performance in European football: what can we learn from the data?'. Decis Supp Syst 114:18–28

Gelade GA (2018) The influence of team composition on attacking and defending in football. J Sports Econom 19(8):1174–1190

Langville A.N, Meyer C.D (2012). The Science of Rating and Ranking. Princeton University Press

Vales A, López CC, Gómez P, Pita HB, Olivares JS (2018) Competitive profile differences between the best-ranked european football championships. Human Movem 18(5):97–105

Pantuso G, Hvattum LM (2021) Maximizing performance with an eye on the finances: a chance-constrained model for football transfer market decisions. TOP 29:583–611

Mourao PR (2016) Soccer transfers, team efficiency and the sports cycle in the most valued european soccer leagues-have european soccer teams been efficient in trading players? Appl Econom 48(56):5513–5524

Liu XF, Liu Y-L, Lu X-H, Wang Q-X, Wang T-X (2016) The anatomy of the global football player transfer network: Club functionalities versus network properties. PLOS ONE 11(6):1–14. https://doi.org/10.1371/journal.pone.0156504

Dobson S, Goddard J (2012). The Economics of Football. Cambridge University Press

Matesanz D, Holzmayer F, Torgler B, Schmidt SL, Ortega GJ (2018) Transfer market activities and sportive performance in European first football leagues: a dynamic network approach. PLOS ONE 13(12):1–16. https://doi.org/10.1371/journal.pone.0209362

Barabási A.-L (2016). Network Science. Cambridge University Press

Newman M (2018). Networks. Oxford University Press

John Bond Alexander PWDP (2020) Topological network properties of the European football loan system. Europ Sports Manag Quart 20(5):655

Palazzo L, Rondinelli R, Clemente FM, Ievoli R, Ragozini G (2023) Community structure of the football transfer market network: the case of Italian series. J Sports Analyt 9(3):221–243

Wand T (2022) Analysis of the football transfer market network. J Statist Phys 187(3):27

Clemente GP, A.C, (2023) Community detection in attributed networks for global transfer market. Ann Operat Res 325:57–83

Xu Y (2021) The formation mechanism of the player transfer network among football clubs. Soccer Soc 22(7):704–715

Beck N, Meyer M (2011) Modeling team performance: theoretical and empirical annotations on the analysis of football data. Emp Econom 43:335–356

Easley D, Kleinberg J (2010). Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Statist Mech Theory Exper 2008(10):10008–10012

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.