



Universiteit
Leiden

Master Computer Science

Identifying successful football player careers within the global transfer market network

Name:	Tiago Cumetti
Student ID:	s3897907
Date:	25/06/2025
Specialisation:	Data Science
1st supervisor:	Frank Takes
2nd supervisor:	Alberto Ceria

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

*"Without data you are just another person with an opinion,
but without an opinion you are just another person with data."*

W. Edwards Deming

LEIDEN UNIVERSITY

Abstract

Faculty of Science
Leiden Institute of Advanced Computer Science (LIACS)

Master of Science in Computer Science

**Identifying successful football player careers within the global transfer market
network**

by Tiago Cumetti

This work identifies career strategies that maximize professional football players' success in the global transfer market. The transfer market has undergone profound changes over the past two decades, driven by globalization, increased data availability and a shift in decision-making power from clubs to players. Existing research has largely focused on club-level strategies, leaving player-centric perspectives underexplored. We present a network science approach that models over 140,000 transfers that occurred between 2005 and 2024 across 50 top-tier leagues as a complex network, analyzed at macro, meso and micro levels. Player success is quantified through a customized metric based on the evolution of their market value, allowing a direct comparison of network patterns between successful and unsuccessful career paths. Our analysis uncovers structural and strategic differences in transfer behaviors, highlighting the distinct roles of elite and development clubs, community structures within leagues, and recurrent successful career motifs. We introduce ClubRank, a novel influence metric for measuring the impact of a club on player development, and classify clubs into archetypes such as elite competitors, development hubs and market intermediaries. Results reveal that successful careers often follow identifiable transfer patterns and exploit specific network positions. These insights provide a data-driven foundation for guiding career decisions in professional football, and open new avenues for integrating temporal dynamics and personalized recommendations based on player-specific characteristics.

Acknowledgements

I would like to express my sincere gratitude to the many people who have helped me throughout this project, making its realization possible.

First and foremost, I would like to thank my first supervisor, Professor Frank Takes, for his expert guidance, insightful feedback and constant support throughout the entire duration of this research. His mentorship has been fundamental to the success of this project and my academic growth.

I am equally grateful to Dr. Alberto Ceria, whose support as second supervisor has played a key role in shaping this work. His detailed advice, continuous suggestions and availability for discussions have made a significant difference in the development of this thesis.

I would also like to thank the Network Science Research Group at Leiden University for providing a very stimulating academic environment. I have been honored to be part of this group and to be able to discuss my project with its many experienced members.

Special thanks go to FACTS Consultancy, and in particular to Gijs Kruikemeier, for their support in the early stages of this project, for helping to identify a research direction that could also be valuable in the context of the real football transfer market, for contributing to my professional growth through an insightful internship journey, and for offering me a position to continue pursuing my ambitions in this field.

On a personal note, I am deeply grateful to my family for always believing in me and allowing me to experience this academic journey abroad, a choice that has definitely changed my life for the better, opening my eyes to a huge and wonderful worldwide community. Their support and sacrifices have been the foundation that sustained me throughout the entire Master's program, and I hope that my ability to successfully conclude this journey and project will make them proud of me.

To my friends and loved ones, the ones who have been there from the beginning, and in particular all the wonderful people I have met over these two years: to Christina, to my housemates, to the true friends I encountered during this academic journey, and to all the other people that the university and the city of Leiden made it possible for me to meet, thank you for everything. For your patience, encouragement, for sharing the same challenges and the same moments of happiness, thank you for inspiring me and making me aware of so many things that are out there and that, since I met you, have also become part of my life. This journey has been both challenging and fulfilling, and I am grateful to have shared it with so many inspiring people. Thank you all for making me a better person, academically, but most importantly, personally.

Finally, I would like to thank myself, that Tiago from two years ago who was brave and mature enough to understand that the choice to take this Master's program and move abroad to face a completely new life experience, giving up so much time with family and friends just to follow his dream, was the right one. Well, after this incredible journey, I can tell you that you were right, Tiago. This experience will be one of the best memories of your life. And I want to thank you for always giving your best in every challenge you faced during this journey. Thank you for never letting yourself down and always believing in yourself. As you have experienced, life always smiles on the brave, those who choose the tough path. Stay true to yourself and enjoy your next journeys just as you did over these two years. Never stop chasing your dreams and continue to follow your ambition.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Motivation	1
1.2 Problem Statement	2
1.3 Contributions	3
1.4 Thesis Structure	3
2 Background and Related Work	5
2.1 Sport Data Science	5
2.2 Network Science and its Applications in Sport	6
2.3 Football Transfer Market Analysis	7
2.3.1 From Clubs to Careers	8
2.3.2 Structural Patterns in Transfers	9
2.3.3 Trade Dynamics and Career Trajectories	9
3 Data	11
3.1 Creation	11
3.1.1 Transfers Data	11
3.1.2 Club Data	12
3.1.3 Player Data	12
3.2 Preprocessing	13
3.3 Description	14
4 Preliminaries	17
4.1 Network Representation	17
4.1.1 Multigraph	17
4.1.2 Weighted Network	19
5 Approach	21
5.1 Success Assessment	21
5.2 Club Level Assessment	23
5.3 Network Analysis	26
5.3.1 Macro-Level	28
Overall Structure Analysis	28
Transfer Direction Analysis	30
5.3.2 Meso-level	31
Community Detection	31
Transfer Flows	32
5.3.3 Micro-level	32
Probabilistic Network Representation	33

2-Node Pattern Analysis	34
3-Node Pattern Analysis	36
5.3.4 ClubRank	37
6 Experiments	39
6.1 Macro-level	39
6.1.1 Overall Structure Analysis	39
6.1.2 Transfer Direction Analysis	44
6.2 Meso-level	47
6.2.1 Successful Player Network Communities	47
6.2.2 Unsuccessful Player Network Communities	49
6.2.3 Transfer Flows	51
6.3 Micro-level	53
6.3.1 2-Node Pattern Analysis	53
6.3.2 3-Node Pattern Analysis	57
6.3.3 ClubRank	59
6.4 Discussion	63
7 Conclusion	65
7.1 Conclusion	65
7.2 Future Work	66
A Football leagues information	67
B Transfer flows analysis with higher thresholds	69
Bibliography	71

List of Abbreviations

ACC	Average Clustering Coefficient
AMV	Average (<i>transfermarkt</i>) Market Value
CR	ClubRank
CSV	Comma Separated Values
FFP	Financial Fair Play
FIFA	Federation Internationale (de) Football Association
LSCC	Largest Strongly Connected Component
NMI	Normalized Mutual Information
SCC	Strongly Connected Component
UEFA	Union (of) European Football Associations
USD	United States Dollar

List of Symbols

$c(s, t)$	number of transfers from node s to node t
k_i	degree of node i
m	number of edges $ E $
n	number of nodes $ V $
r	reciprocity
r_{deg}	weighted degree assortativity (Pearson)
s	source
t	target
$w(e_{s,t})$	weight of the edge from node s to node t
$C_B^{(w)}$	weighted betweenness centrality
D	network density
E	set of all edges in the network
V	set of all nodes in the network
W	weighted edge sum
γ	resolution parameter of the Louvain algorithm
σ_{st}	number of shortest weighted paths from s to t

*To little Tiago, the child who once could only dream of doing this
kind of research for his Master's thesis. . .*

You made it.

Chapter 1

Introduction

1.1 Motivation

Football, also known as soccer, is the most followed sport in the world, currently boasting an estimated 3.5 billion fans across all five continents. Its widespread popularity is largely attributed to its simplicity and accessibility; anyone with a ball can play, even in the poor or underdeveloped regions of the world.

This sport is played professionally in more than 200 countries, creating a global network of people united by a shared passion. The steady growth of this community has led to the continuous creation and expansion of local and international competitions at club and country levels, which have driven the global interest in this sport to increase even more. The passionate football supporters follow club competitions all year long, but country competitions such as the World Cup attract the attention of even more people, that usually are not that interested in football, creating huge national communities. As interest grows, the financial investment in football increases proportionally, resulting in a massive and steadily expanding industry. The football market, which involves organizations, clubs, sponsors, players, supporters, and a large number of different professionals, was valued at USD 3.41 billion in 2024 and is projected to reach USD 4.71 billion in 2033, according to IMARC Group (2025), demonstrating a rapid and constant growth.

This research focuses on a specific segment of the football economy: the football player transfer market. This market is defined by the trading of players between professional football clubs. Each player is contractually tied to a club, and a transfer involves the purchasing club paying a fee to release the player from their current contract and signing a new one with the purchasing club (this is a simplification of the concept, as there are more complex dynamics involved in this network related, for example, to the loan system, which represents a different possibility of trading players, as explained in detail by Bond, Widdop, and Parnell (2020)). The transfer market has existed since the inception of professional football, but has evolved significantly over time.

First of all, the financial amount invested by each member of this market has increased dramatically. As investment in the sport has grown, so has the money spent on acquiring talented players. In addition, globalization and advancements in communication technologies have broadened the scope of scouting and trading, since clubs from all over the world have become more easily connected, thanks also to the increasing amount of international competitions that create occasion for representative of geographically distant clubs to meet. Clubs are no longer restricted to local or regional markets; now they actively scout talent worldwide, including in lower divisions and elite youth academies.

A significant change has also occurred in the distribution of power within the transfer market. Historically, clubs were the protagonists of the football transfer market and had most of the decision-making authority on transfers, while players had limited influence. However, in the modern era, players have emerged as the primary actors in the football ecosystem. The rise of digital media and social networks has elevated their public profiles,

transforming them into global celebrities whose marketability extends beyond the pitch. As highlighted by Bernardo, Ruberti, and Verona (2022), players now generate substantial revenue for themselves, their clubs and associated sponsors through their image and branding rights.

This elevated status has made players more aware of their market value and increased their influence in transfer negotiations. Most of them are now supported by professional entourages, including agents, advisors, branding experts, etc. who strategically manage their careers to maximize earnings, visibility, and sporting success. While previous studies (Bond, Widdop, and Chadwick, 2018; Bond, Widdop, and Parnell, 2020; Xu, 2021; Palazzo et al., 2023; Dieles, Mattsson, and Takes, 2024) have predominantly analyzed the transfer market from a club's perspective, this work shifts the focus to a player-centric viewpoint, in order to capture this ongoing evolution in market dynamics.

Another key motivation for this research is the growing availability and acceptance of data in sports decision-making, with an increasing number of people belonging to the sport field who have started to trust data analysis as a valid means of extrapolating insights from sports. In the past, scouting relied largely on subjective evaluations by so-called "experts." Today, the availability of data allows clubs to use data-driven approaches to objectively assess player qualities and determine whether the observed players would actually fit and have an impact within team strategies and whether they would be worth a certain investment. These analytical strategies have already shown promising results in recent years and are increasingly shaping both transfer strategies and on-field performance, entering a period where an increasing number of football clubs attribute more importance and a more central role to data analysis, as discussed in more detail in Chapter 2.

Moreover, the consistently increasing volume of transfers over the years underscores the relevance of this research. Football clubs are becoming increasingly active on the transfer market with the aim of optimizing team performance. From youth talent exchanges between academies to the trading of the top players among elite clubs, the transfer market has become a critical mechanism, which needs to be analytically studied and understood to achieve sporting success in football (Matesanz et al., 2018; Dieles, Mattsson, and Takes, 2024).

In this work, the football transfer market is studied through the lens of network science, a branch of data science that analyzes systems by modeling entities (nodes) and the relationships between them (edges). In our case, each node represents a football club (or league), and each directed edge represents the transfer of a player from one node to another. This network-based approach enables the detection of structural patterns, recurring paths and influential actors within the market. The analysis is carried out across three complementary scales. At macro-level, we examine the overall topology of the transfer network to capture broad structural trends. At meso-level, we identify and analyze transfer flows and communities, i.e., groups of clubs more densely connected to each other than to the rest of the network, which can reveal interesting clusters. At micro-level, we focus on individual nodes and motifs, studying the most influential clubs and the most common transfer routes (frequent sequences of edges) associated with successful career trajectories. By defining and applying these concepts, this work positions the football transfer market within a rigorous analytical framework that can reveal both high-level structures and fine-grained patterns relevant to career optimization for players.

1.2 Problem Statement

Having established the context and relevance of this study, we now articulate the central research problem. This work adopts a player-centric view of the football transfer market

and aims to generate actionable insights to help players seeking to optimize their future career paths toward success. Specifically, our goal is to assist players in making more data-driven decisions, to identify the most suitable club destinations to achieve their personal and professional goals, maximizing their potential.

To do this, we conducted an analysis of player transfer data obtained from *transfermarkt.com* (Transfermarkt, 2024), spanning from 2005 to 2024. By representing this dataset as a network, we apply techniques from network science (Barabási and Frangos, 2014) to identify frequent patterns in the career trajectories of successful and unsuccessful players. Success, in this context, is assessed using a metric based on the evolution of a player's market value throughout his career.

We then build two distinct network representations, one for successful players and the other for unsuccessful ones, and examine their different structural patterns. These patterns are used to infer strategies that can guide and help football players to make optimal career decisions, particularly in critical situations during their professional journey, when this choice will determine the development of the player's career and life in the upcoming years. In summary, the central research question that guides this thesis can be formulated as follows.

How can insights in the dynamics of the global football transfer market network be leveraged to make informed career decisions for players?

To answer this question, that can obviously be tackled using a wide range of methods, we decided to adopt a multiscale network analysis strategy, examining macro-, meso-, and micro-level dynamics, to extract both high-level trends and fine-grained actionable insights.

1.3 Contributions

This research investigates the field of football analytics and network science by providing the following contributions:

- **Data Creation:** Development of a novel global dataset that contains transactional records of football player transfers between 2005 and 2024, covering 50 top-tier leagues worldwide.
- **Career Success Assessment:** Introduction of a new metric to assess the success of a football player career, based on the evolution of his market value.
- **Network Comparison:** Comparative analysis of the structural and dynamic properties of successful and unsuccessful player networks, identifying distinguishing characteristics.
- **Network Pattern Analysis:** Identification of network patterns that systematically differentiate the career trajectories of successful and unsuccessful players.
- **ClubRank:** Proposal of a new centrality metric to evaluate the role of clubs within successful and unsuccessful player career trajectories.

1.4 Thesis Structure

To conclude this chapter, we provide the reader with an outline of this project structure, as follows. Chapter 2 presents a review of the relevant literature. Chapter 3 describes

the created dataset. Chapter 4 introduces preliminary notion necessary to construct the representations of the football player transfer market network, detailing the approach and the logic behind our network design. Chapter 5 elaborates on our analytical framework, illustrating the methods. Section 5.1 explains how player success is assessed, Section 5.2 focuses on how do we measure a club's appeal for a player, and finally Section 5.3 details the network analysis techniques used at the macro-, meso- and micro-level. Chapter 6 presents the results of our experiments and the figures containing the main takeaways of this study. Finally, Chapter 7 concludes the thesis by summarizing our main findings and proposing directions for future research.

Chapter 2

Background and Related Work

Data science is a discipline with applicability across different domains, as it centers around the analysis of data, that nowadays represent the primary means of storing and understanding information in countless contexts. Among the possible areas of application, one of the fastest-growing is sport science, where data-driven approaches have seen a notable rise in recent years.

Throughout this chapter, our aim is to contextualize the current research by reviewing the findings retrieved by relevant related literature. Section 2.1 explores the use of data science in sports, highlighting its development and areas of impact. Section 2.2 narrows the focus to network science and its multiple applications to sport, particularly in football. Finally, Section 2.3 discusses how network analysis has been used to study the football transfer market, which is the central focus of this research.

2.1 Sport Data Science

Sport science is a broad and multidisciplinary field, and the integration of data science has led to significant advancements across its many branches. Unlike more established areas, sport data science is still evolving, with considerable room for exploration and innovation. The increasing availability of sports-related data has opened new avenues for research to gain deeper insights into individual and team performance.

Sports inherently generate a vast amount of different data. With the advancement of technology, such as sensing and video tracking, the volume and quality of measurable data continue to grow. Sensor data, collected through dedicated devices, measure the physical attributes and movements of athletes during training and competitions. Simultaneously, video analysis, supported by the widespread broadcasting of events and advancements in computer vision, makes it possible to extract a huge amount of data and to identify meaningful patterns from visual recordings. Depending on their origin and purpose, data in sport can be categorized as spatial (generated by the movement and interaction of players in space during a sport event), temporal (describing the evolution of performance over time of a certain athlete, as described in detail by Borrie, Jonsson, and Magnusson (2002)), or annotation-based (manually recorded statistics by analysts, introducing human evaluation into the data individuation process).

Statistics have always played a critical role in sports. The application of modern data science techniques, added to the vast availability of data, enhances these insights, offering analyses that support both public engagement (Kataoka and Gray, 2019) and technical evaluation by athletes themselves, coaches and analysts. These tools allow the validation of expert hypotheses, but also the discovery of new precious insights, that can lead to the generation of new breakthrough advancements related to a wide set of unsolved questions and problems in the sports field, as explained in Balague et al. (2013). Studying data can provide insights applicable to tactics, such as identifying team strengths or exploiting opponent weaknesses (opponent mining), as shown in Miller (2015).

Traditional sport science, which is strongly influenced by medical science, also benefits greatly from data analytics. Performance enhancement (Miller, 2015; Pina, Paulo, and Araújo, 2017; Ribeiro et al., 2017), energy efficiency (Stöggl and Sperlich, 2015), and injury prevention (both during training and competitive activities) (Gabbett, 2016) are central goals to this discipline, and they can all benefit by an approach that involves data science. Data-driven methods help identify the physical or psychological aspects that an athlete should improve to enhance its competitive level and maximize its overall fitness (Van Der Zwaard et al., 2018; Botterill, 1990), and help design optimal training structures and schedules to balance load and recovery (Stöggl and Sperlich, 2015). Given the uniqueness of each athlete's goals, physical structure, working mentality and role, universal programs are ineffective; instead, personalized and continuously adaptive training approaches, where the athlete's performance is constantly evaluated, result particularly more effective (Gabbett, 2016).

Beyond on-field performance, data science has shown its effectiveness in the management of professional sports organizations and individual player careers. This includes studies regarding competitive team formation and administration (Chelladurai, 2014), support staff organization (Eccles and Tenenbaum, 2004), scouting (Ghar, Patil, and Arunachalam, 2021) and career planning for individual players (Meng-Lewis et al., 2022).

The transfer of players between clubs constitutes a global complex market in itself, with high financial and strategic stakes. Understanding the structure and behavior of this market is essential, both for maximizing return on investment and for optimizing athlete career progression. This aspect represents the main area of focus in the present research.

2.2 Network Science and its Applications in Sport

Network science is one of the numerous branches that compose data science, focusing on understanding data by investigating interactions and relationships between entities. These relationships are typically modeled as graphs composed of nodes (entities) and edges (connections). This approach, which is already widely used for studying complex systems in disciplines such as sociology (Wasserman and Faust, 1994), economics (Goyal, 2012) and biology (Barabasi and Oltvai, 2004), in recent years has found growing application in sport research as well (Lusher, Robins, and Kremer, 2010; Passos et al., 2011; Clemente, Martins, Mendes, et al., 2016; Wäsche et al., 2017; Ribeiro et al., 2017; Kim and Yim, 2017). Network-based representations enable researchers to identify structural patterns and interactions that are not immediately evident in raw data. For example, in many real-world systems, researchers have found power-law distributions (Barabási and Albert, 1999) and mechanisms such as preferential attachment (Newman, 2003), which help explain the emergence of hubs and hierarchical organization in complex systems.

In the context of sport, these representations result particularly well-suited to model relational data, which have always been present in a large amount in this field. Early structural analyses such as Gould and Gatrell (1979) demonstrated the value of network-based methodologies in sport. Unlike basic performance statistics, network analysis reveals deeper patterns of connectivity, flow, and influence. It allows for the modeling of complex systems such as the multi-scale movement patterns shaped by specific game contexts and tactical dynamics. As highlighted in Balague et al. (2013), these techniques are particularly effective in improving understanding of the technical, tactical, and physical conditioning aspects of sport performance.

One prominent application of network science in football is tactical analysis. Several works, including Camerino et al. (2012), Pena and Touchette (2012), Cotta et al. (2013), Wäsche et al. (2017), Gürsakal et al. (2018), Buldú et al. (2018), and Bekkers and Dabadghao

(2019), illustrate how motif and pattern analysis within passing networks of both the target and the opponent team, can inform strategic decision-making providing previous insights to the team coaching staff. Such analyses can identify which connections between players should be emphasized during offensive phases and which connections between opponent players should be limited during defensive ones, with the goal of improving the chances of winning a match.

Network science also contributes to player scouting and recruitment. As demonstrated in Peña and Navarro (2015), network features can be used to create a 'digital fingerprint' of a leaving player that needs to be replaced in the team. By comparing this profile with those of players available in the transfer market, clubs can identify candidates whose attributes best align with their tactical systems and style of play, to enhance the probability of finding a right fit for the team.

Another traditional challenge in sport analytics consists in evaluating team and individual performance. In contrast to earlier models that relied mainly on historical or contextual data, Cintia, Rinzivillo, and Pappalardo (2015) introduced a model based on in-game player behavior, analyzed through network metrics. Both basic and advanced network measures have been shown to improve coaching decisions and support training processes by estimating optimal player positioning, quantifying individual contributions to team performance, and offering insights into the broader dynamics of team sports. This perspective is supported by works such as Yamamoto and Yokoyama (2011), Narizuka, Yamamoto, and Yamazaki (2014), Clemente et al. (2015), Gonçalves et al. (2017), Ramos, Lopes, and Araújo (2018), and Arriaza-Ardiles et al. (2018). In particular, networks characterized by high density, high clustering coefficients and low centrality have been consistently associated with superior team performance, as reported in Grund (2012) and Pina, Paulo, and Araújo (2017).

Since football actions frequently involve more than two players simultaneously, some studies have proposed a solution to expand the sport analysis beyond pairwise interactions. In particular, Ramos et al. (2017) introduced the use of hypernetworks to model multi-player interactions, capturing the complexity of simultaneous group dynamics that traditional graphs may overlook.

A fundamental aspect of network modeling involves the analysis of both the topology (the structural configuration of connections) and the dynamics (the flow of information or interactions across the network). Especially relevant are weighted graphs, where edges carry numerical weights that quantify the frequency or strength of connections. This becomes particularly important in domains such as the football transfer market, where each transfer carries a distinct level of significance or value that must be preserved to accurately model the system.

Motivated by these successful applications of network science in sport, the present research concerns itself with the football transfer market. The market is modeled as a weighted, directed graph, allowing us to explore the structural properties and patterns that govern player transfers across clubs and leagues.

2.3 Football Transfer Market Analysis

Within the sport domain, one of the most promising applications of network science is the analysis of the football transfer market. Previous studies have modeled transfers as directed or weighted networks of clubs, analyzing structural features such as degree distributions, centrality measures and small-world properties (Liu et al., 2016; Bond, Widdop, and Chadwick, 2018; Bond, Widdop, and Parnell, 2020). Some have explored community structures or economic inequalities in national and international transfer markets (Clemente

and Cornaro, 2023; Palazzo et al., 2023).

However, despite these contributions, several areas remain underexplored. Most existing studies are limited in temporal or geographic scope, often focusing on specific leagues or a fixed time period. In addition, there is limited research on transfer flow dynamics over time, the role of club strategies in shaping transfer patterns, or the impact of different types of transfer on individual player careers. Network-based methods are rarely used to quantify club influence beyond standard centrality metrics or to distinguish between successful versus unsuccessful player careers.

These gaps leave room for deeper investigation in this field, with this project focusing specifically on understanding the structural patterns of player careers within the football player transfer network, with the aim of distinguishing characteristics of successful versus unsuccessful player trajectories.

2.3.1 From Clubs to Careers

This research builds on work such as Dieles, Mattsson, and Takes (2024), which modeled transaction data related to transfers between clubs from eight major European leagues between 1993 and 2021. Their analysis was focused on individuating successful clubs' features among this network, and it showed that clubs characterized by high degree tend to perform better, revealing a correlation between transfer activity and team success (at least in their domestic league). Similarly, Matesanz et al. (2018) found a correlation between spending power and performance, especially in UEFA competitions (such as Champions League and Europa League), suggesting that the market favors wealthier clubs, creating substantial inequality between clubs and leagues. Liu et al. (2016) further demonstrated that clubs acting as brokers or hubs (central positions within this network) usually achieve better competitive results.

While these studies provide important insights, they primarily focus on club-level analysis. The present study shifts the focus to the player's perspective, aiming to expand previous network representations of the football player transfer market, by analyzing data from 50 leagues worldwide, keeping the focus more on the movement of players across the network and not on the involved financial flows, as considered in other previous works, such as Matesanz et al. (2018). This broader and more individualized approach aims to uncover patterns associated with player career success. In the past, clubs were the part that had the most power in negotiations related to player transfers, while players themselves had less room to decide their favorite destination for the development of their careers. In today's football ecosystem we have witnessed a clear evolution of these roles, with players being the actual main characters in the football transfer market, often supported by a big staff including agents, advisors, and branding experts, they have now more control and influence over career decisions than ever before. As highlighted in Bernardo, Ruberti, and Verona (2022), the *image value* of modern players, understood here as the combined economic and marketing value of a player's personal brand, including their media visibility, social media following, endorsement potential, and overall marketability beyond on-field performance, has grown exponentially. Thanks to increasingly sophisticated marketing strategies and the popularity of social networks, public attention has shifted from clubs to the players themselves. This strengthened position in the market allows modern players to actively navigate transfer opportunities with strategic intent, shaping their own trajectories with the goal of maximizing both their personal revenue and visibility, rather than passively fitting into club strategies.

This evolution explains the necessity for teams to spend more money on the football transfer market and to offer higher wages in an attempt to get hold of the best talents in the world, since their value is no longer strictly related to their playing performance, but also

to their added *image value*. To prevent football from becoming a game in which wealthy clubs could freely spend money to acquire top talents, greatly increasing their chances of winning titles through financial power alone, the International Football Association (FIFA) introduced Financial Fair Play (FFP) in 2010. This act limits the possibility of spending for a football club that wants to be part of the football player transfer market, imposing penalties for the clubs which would not respect the FFP. Some studies such as Ghio, Ruberti, and Verona (2019) focused on the impact generated by the introduction of these limitations (in this specific case on Italian football clubs), and demonstrated how they contributed to reduce the gap between top teams and lower-tier teams.

2.3.2 Structural Patterns in Transfers

Several studies have explored the structure of the transfer market through network methods with different research interests. Some of them have decided to focus on particular situations concerning the role of a specific league or the characteristics of a single transfer window. This is the case of Palazzo et al. (2023), which examined the community structure related to the 2019 transfer window in Italy's Serie A, revealing different trading strategies between clubs with different goals. Or Nolasco (2019), which studied the role of the Portuguese league in the football player transfer network, as a developer and exporter of football talent, especially to larger European leagues. Identifying the migratory routes of foreign players arriving in Portugal and of Portuguese players who go abroad. Bond, Widdop, and Chadwick (2018) conceptualized the transfer network as a global system, identifying a core of dominant European countries, a semi-periphery of developing football nations and a peripheral group with limited market involvement, where football is less developed.

Velema (2021) retrieved notable results analyzing global transfer flows from 82 countries between 2006 and 2016, showing that while the vast majority of transfers are domestic, international transfers often follow regional paths, with clubs acting as brokers that tend to recruit from countries located in their geographical vicinity. European elite clubs recruit predominantly from top rivals in European leagues, using their superior financial resources to minimize the risk that scouted players do not fulfill their potential. Mid-tier European teams tend to look to Europe, but also to Latin America and emerging markets, scouting widely in the hope of finding talented players to compete with elite teams and to resell them to make some profit and finance their scouting system. Finally, most lower-tier clubs operate primarily on a local scale, occasionally hiring from neighboring countries, lacking financial capabilities to maintain large scouting departments.

Additional context on why clubs tend to scout mainly from the same or similar countries is provided by McGovern (2002), who found that clubs prefer players from culturally, climatically or linguistically similar countries to reduce the risk to invest money on a player who will have to face integration challenges, with the risk of not fitting well into the club's dressing room or playing style. Littlewood, Mullen, and Richardson (2011) emphasized that many leagues, including four of the top-5 European ones, remain heavily dominated by indigenous players, raising caution about the fact that the propensity to adopt global player recruitment strategies may result in becoming dependent development states.

2.3.3 Trade Dynamics and Career Trajectories

Next, we focus on other research that has revealed dynamic structural patterns in player trading. Xu (2021) examined the formation of the transfer network of football player among the top-5 European leagues and showed that clubs often form persistent triadic trading relationships, such that clubs tend to make deals with clubs who shared the same

third trading partners. Studies such as the one by Bond, Widdop, and Parnell (2020) focused on the European football loan system, showing how loan transactions play an important role in the football player transfer network (this type of transfer results particularly prevalent among Italian clubs, whose peculiar situation and systematic need of trading players to maintain themselves is explained in detail in Baroncelli and Lago (2006) and Neri et al. (2023)), indicating that several elite clubs extract value from the loan system, while others play the role of value creators. Similarly, Félix et al. (2019) identified countries that serve primarily as either exporters of talented players (farm countries) or importers (buyer countries). In addition to this, Clemente and Cornaro (2023) used community detection to show that national leagues often cluster based on economic status, influencing market behavior.

Finally, the study by Velema (2018) analyzed player migratory trajectories in the global football transfer market and found that 60% of the football players experience a domestic career, while 40% of them observe frequent cross-border transfers. Moreover, only an elite 10% of players follow predictable upward-trending paths toward top-tier teams; while most experience irregular, uncertain trajectories (often ending up in circulation), highlighting that football transfer market resembles a game of snakes and ladders in which an elite minority of players moves tightly towards the top teams, while the majority of them emphasize the more complex, non-linear nature of career development in professional football.

Building on these foundations, this thesis advances the related literature by taking a player-centric perspective on the football transfer market, whereas most prior studies adopt a club-centric view. Introduces a novel global transactional dataset covering transfers from 2005 to 2024 across 50 football leagues, enabling unprecedented breadth in scope. The analysis applies multiscale network-based methodologies, examining macro-level structural patterns, meso-level community dynamics, and micro-level influential nodes and pathways, to identify patterns in transfer histories that correlate with individual career success. In this context, success is quantified using a metric based on the evolution of a player's market value over time, rather than relying solely on conventional performance statistics. By combining these elements, the research aims to provide actionable data-driven insights into how specific player transfers can shape, accelerate or optimize professional career development.

Chapter 3

Data

In this chapter, we explain the steps followed to create the dataset used in this study. The football transfer dataset has been constructed by collecting data from *www.transfermarkt.com* (Transfermarkt, 2024). This German website is internationally renowned for providing detailed football information, including competition standings, results, club statistics, and, crucially for our research, players' careers and transfers. Its database contains information on over 1.3 million players. For the purpose of our research, we focus on transfers involving the 50 national leagues listed in Table A.1, covering the period from 2005 to 2024. Section 3.1 details the techniques employed to collect the raw data. Section 3.2 outlines the preprocessing steps undertaken to clean the dataset. Finally, Section 3.3 provides a comprehensive explanation of the final version of the dataset.

3.1 Creation

To build the dataset, we collect transfer data from *www.transfermarkt.com*. We target webpages that correspond to the transfer history of specific leagues during specific transfer windows. Our code builds on the work available in Galvin (2023), which we significantly extended by incorporating a larger number of features and including more leagues. To collect our data, we selected a starting and ending year (initially set to 1992 and 2025), and we gathered information about the leagues that we decided to scrape (league name, league ID, and country), corresponding to the contents of Table A.1.

3.1.1 Transfers Data

Using these inputs, we access the relevant webpages for each league and transfer window in the specified time range. For each club participating in a league during a season, the webpage includes two tables: one listing incoming transfers and another listing outgoing transfers for the specified transfer window. Each table contains the following information: player name, age, nationality, position, market value, involved club and transfer fee. All these features are scraped and stored in our dataset.

One challenge encountered was obtaining player nationalities, as these were represented by flags rather than textual names. Moreover, many players hold multiple nationalities, depicted as multiple flags within the same cell. In our final dataset, these were converted into a list of country names, separated by commas.

A second significant challenge was the absence of unique player and club identifiers in the tables. Since names may vary over time within the *transfermarkt* database, it is essential to include consistent identifiers in our dataset. To retrieve these, we accessed the individual webpages of each player and club and extracted the identifiers from the respective URL.

This process was extended to include the identifiers of the clubs involved in each transfer. Exceptions included cases that did not represent real clubs, but denoted player statuses such as "Without Club", "Retired", "Career Break" or "Military Service". These cases were still included in the dataset, as they can represent legitimate steps in a player's career. While "Without Club" and "Military Service" already had assigned IDs on *transfermarkt*, we arbitrarily assigned unused IDs to the remaining two statuses: 'Retired' was mapped to ID 514 and 'Career Break' to ID 513, considering that 'Without Club' corresponds to ID 515.

To improve the clarity of the structure of the dataset, we also added a field representing the direction of each transfer. Transfers appearing in incoming tables were labeled as 'in', while those scraped from outgoing tables were labeled as 'out'.

3.1.2 Club Data

The scraped transfer data only included minimal information about the involved clubs (the names and the introduced identifiers). Moreover, club webpages on *transfermarkt* typically display only current information, which does not reflect the club characteristics at the time of each transfer. As the football landscape and the situations of many clubs have evolved significantly over the past two decades, being able to introduce historical club data related to the moment of each transfer is crucial.

To address this, we scrape club metrics specific to each season, providing as input a new parameter called "season id", which corresponds to the content of the "timestamp" feature in the final dataset used in this work, which is composed of the transfer year followed by a suffix ('s' for summer or 'w' for winter) depending on the transfer window when the transfer occurred. Thanks to this parameter, we were able to construct URLs to access the league standings for the appropriate season. These pages contain information such as: squad size, average age, number of foreign players, average market value and total market value of each club that participated in the competition during the "adjusted season". We define the "adjusted season" of a transfer as follows: if the transfer occurred during the winter transfer window, the adjusted season corresponds to the current season of the transfer; while for summer transfers, it corresponds to the previous season. This distinction is important because it ensures that the club data reflect the conditions closest to the transfer date.

Once scraped, this club data was merged into the main dataset using club and season identifiers. This process was applied to both the source and the target clubs for each transfer, which are the club that is selling the player and the club that is acquiring the player.

3.1.3 Player Data

To further enrich the dataset, we collect additional player-specific information from their individual *transfermarkt* pages. Since we already accessed these pages to retrieve player IDs, we extended the process to incorporate more data which can be used to possibly identify a player without using his *transfermarkt* identifier, with the purpose of possibly merging with different datasets or external sources in future work. We decided to scrape their date of birth (converting it in an appropriate format), current club (for active players) and last club (for retired or inactive players). As before, special cases such as "Without Club", "Career Break", "Retired", and "Military Service" were included.

After scraping this player information, we merged them into the main dataset using player identifiers.

3.2 Preprocessing

After creating the initial datasets, we moved on to dealing with preprocessing tasks aimed at cleaning the data for our analysis. Initially, the dataset contained raw scraped information related to the transfer event, regarding the player himself, involved clubs' details and transfer specifics.

At first, we reorganized the structure of the dataset to enhance the clarity of the transfer directions. The original format included the transfer direction ("in" or "out") and the "involved clubs" features, which often led to confusion, since it was difficult to identify whether the "involved club" corresponded to the selling club or the buying club. We eliminated the direction field and adopted a **source–target model**, where features prefixed with **source** refer to the selling club, while those prefixed with **target** refer to the buying club.

We then introduced several **new features** in the dataset. We categorized players by role, as the original position feature was too fine-grained and inconsistent due to some players' positional flexibility that allows them to play in different roles. In addition, many players change the position they cover on the field over the course of their career. We decided to map positions to four broad roles: Goalkeeper, Defender, Midfielder and Attacker. Then, we introduced delta features, corresponding to the differences between target and source club values, such as squad size, number of foreign players, average age, average market value and total market value. We also added delta features calculated as the differences between the player's age and market value compared to the average age and average market values of the source and target clubs at the moment of the transfer. In all the features of this kind, the values are obtained by subtracting the average value of the club from the value of the player.

Another crucial preprocessing step involved **removing duplicate records**. Our goal is to have a unique record for each transfer in our dataset preserving as much information as possible, but because the transfers were listed in both incoming and outgoing tables on *www.transfermarkt.com*, the presence of duplicates was inevitable during the creation process, and now we need to remove them. To do so, we defined a unique transfer key, comprising the player ID, transfer window ID (also called "timestamp"), source club ID and target club ID. For each set of duplicates, we only retained the record with the fewest missing values, being the one carrying the largest amount of information. After this preprocessing step we obtained the first actual version of the dataset, which consisted of 625,922 transfers involving 147,575 players occurred between 1992 and 2025.

To ensure analytical integrity throughout our experiments, we filtered the dataset to remove records with **missing values** in essential fields. In particular, we excluded transfers missing market value data for the player, as this value is necessary to assess the success of a player, which represents a central part of our study. Looking at Table A.1 we can see how the data coming from the 50 national leagues have different data availability; for most of the leagues the transfer data date back to 1992, while for some of them the data have only become available more recently. Since *www.transfermarkt.com* began recording market values only in 2004, we had to drastically reduce the size of our dataset. Moreover, at the beginning *transfermarkt* introduced the player market value only for players who participated in the top leagues. For this reason, we decided to exclude transfers that occurred in 2004 from our analysis to avoid the introduction of bias. Consequently, we limited the dataset to transfers from 2005 onward.

Furthermore, we excluded transfers involving special cases (“Without Club”, “Retired”, “Career Break”, etc.), as these “clubs” do not belong to any league and consequently lack associated club-level data, leading to the removal of other 39,752 transfers related to these particular occurrences. We also removed transfers from the winter window of 2025, to avoid bias introduced by the temporal misalignment of transfer windows across leagues, since the periods where the clubs are allowed to trade players are different for every national league. We did not want to introduce both transfer data related to a complete national transfer window and some other related to an incomplete transfer window. Thus, the final dataset includes all transfers from 2005 to 2024.

Additional preprocessing experiments were performed to try to include features from external sources from *www.transfermarkt.com*. However, these were not integrated into the final dataset used in this work. For this reason, we decided to not describe them in this work and to consider them as starting points for possible future research directions.

3.3 Description

This section presents a detailed overview of the final dataset, including all the features retained after the preprocessing and cleaning procedure. The dataset captures player movements across **50 national leagues from 2005 to 2024**. After removing duplicate entries and applying the filtering criteria described in the previous sections, the dataset includes a total of **141,538 unique transfers**, involving **41,734 players** and **1,531 clubs**.

All transfers in the final dataset are unique and do not contain missing values across the selected features. The only exception is the set of fee-related variables, which may contain null values. These were retained in the dataset for possible future purposes, although they were not used in the current study. To support further analytical work and the understanding of our methods and experiments, we now provide a comprehensive description of each feature included in the final dataset, organized into four categories: player features, club features, delta features and transfer features.

Player Features

These variables capture the characteristics of the player at the moment of the transfer:

- **Player name** (string): Full name of the player as listed on *www.transfermarkt.com*. Note that inconsistencies may arise when merging with external datasets due to naming conventions, aliases or the inclusion/exclusion of middle names.
- **Player URL name** (string): Version of the player name used in the URL relative to his webpage on *www.transfermarkt.com*.
- **Player ID** (int): Unique numerical identifier assigned by *transfermarkt* to ensure consistent player tracking despite name variations.
- **Date of birth** (date, DD/MM/YYYY): Player’s date of birth in standardized format.
- **Age** (int): Player’s age in years, rounded down to the closest integer number.
- **Nationality** (string): Country represented at national level by the player. For those with multiple nationalities, the names of the countries are separated by commas, with the first nationality being the one used at international level.

- **Role** (categorical): General playing role, categorized into: Goalkeeper, Defender, Midfielder and Attacker.
- **Position** (categorical): More detailed specification of the playing role, which can include values among: Goalkeeper, Defender, Sweeper, Centre-Back, Left-Back, Right-Back, Midfielder, Defensive Midfield, Central Midfield, Attacking Midfield, Left Midfield, Right Midfield, Left Winger, Right Winger, Centre-Forward, Striker and Second Striker.
- **Market value** (float, €M): Player's estimated value (in millions of euros) based on statistical modeling and community evaluation by more than 700,000 registered users on www.transfermarkt.com.
- **Current club** (string): Player's club as of February 2025. For retired players, this field is set to 'Retired'.
- **Last club** (string): Final club affiliation before retirement, or the same as the current club if the player is still active.

Club Features

These variables describe the **source** (selling) and **target** (buying) clubs at the time of the transfer. Features are duplicated with the prefixes "source" and "target" to distinguish between the two involved clubs:

- **Country** (string): Host country of the league in which the club participates. Note that some clubs participate in leagues of countries other than their own (e.g., AS Monaco, Toronto FC).
- **League** (string): Name of the domestic league in which the club is participating.
- **League ID** (int): Unique league identifier assigned by *transfermarkt*.
- **Club** (string): Club name as listed on www.transfermarkt.com.
- **Club ID** (int): Unique club identifier used to disambiguate different representations of the same club, since different abbreviations are used to indicate the same club.
- **Club n players** (int): Number of players in the club's first-team squad.
- **Club foreigners** (int): Number of players in the squad whose primary nationality differs from the country of the league in which the club participates.
- **Club avg age** (float): Average age of the first-team squad, calculated using ages rounded down to integer values for each player.
- **Club avg mv** (float, €M): Average market value of players who are part of the club (in millions of euros).
- **Club tot mv** (float, €M): Total market value of players who are part of the club (in millions of euros).

Delta Features

These variables represent the differences between the source and target clubs (when "source" or "target" are not included as prefixes in the feature name), or between the player values and the involved clubs' average values (when "source" or "target" is specified as prefix in the feature name):

- **N players** (int): Difference in the size of the squad between target and source club.
- **Foreigners** (int): Difference in the number of foreign players between target and source club.
- **Avg age** (float): Difference in average age of the squads between target and source club.
- **Avg mv** (float, €M): Difference in average player market value between target and source club.
- **Tot mv** (float, €M): Difference in total market value between target and source club.
- **Source avg age** (float): Difference between player's age and the average age of the source club's squad.
- **Target avg age** (float): Difference between player's age and the average age of the target club's squad.
- **Source avg mv** (float, €M): Difference between the player's market value and the average market value in the source club.
- **Target avg mv** (float, €M): Difference between the player's market value and the average market value in the target club.

Transfer Features

These variables capture information directly related to the transfer event:

- **Transfer period** (categorical): Indicates whether the transfer occurred in the Summer or Winter transfer window.
- **Year** (int): Year in which the transfer occurred.
- **Timestamp** (string): Unique identifier for the transfer window, constructed by combining the year and a suffix: "s" for summer or "w" for winter (e.g., 2015_s corresponds to Summer 2015, between the season 2014/15 and 2015/2016; while 2015_w corresponds to Winter 2015, during the season 2015/2016).
- **Season** (string, YYYY/YYYY): Season during which the transfer occurred.
- **Adjusted season** (string, YYYY/YYYY): Indicates the season used to retrieve the data related to the clubs. For winter transfers, this is the current season; while for summer transfers, it corresponds to the previous season.
- **Fee** (string/float, €M): Contains different types of information related to the nature of the transfer. In the case of loans or free transfers, this field may reflect non-numeric data, otherwise it may contain the amount paid by the target club to the source club for the loan or the purchase of the player.
- **Fee cleaned** (float, €M): Preprocessed version of the transfer fee, standardized and expressed as a decimal value in millions of euros. Non-numeric values (e.g., "loan transfer" or "free transfer") are left as missing values.
- **Transfer key** (string): Unique identifier for each transfer, constructed concatenating the player ID, timestamp, source club ID and target club ID. This field ensures that duplicate entries are properly resolved and that each transfer is uniquely represented.

This structured dataset forms the foundation for the analysis conducted in the following chapters.

Chapter 4

Preliminaries

In this chapter, we introduce the concepts necessary to understand our research. These notions will allow the reader to understand how our dataset is represented as a network and how the techniques discussed in Chapter 5 relate to this network structure.

4.1 Network Representation

We first model the football transfer market as a multigraph (Subsection 4.1.1) and then as a weighted network (Subsection 4.1.2). The weighted network representation is the most frequently used in our analysis, with weights varying according to the analytical task. Our analysis was performed at two levels of granularity: club level and league level. Although the following explanations focus on club-level networks, they remain valid for league-level networks as well, where the main difference lies in the node representation: corresponding to clubs at club level versus leagues at league level. Although the number of nodes varies a lot depending on the level of aggregation, the total number of edges, which corresponds to the number of transfers in the dataset, remains constant.

4.1.1 Multigraph

We begin by modeling the football transfer market as a **multigraph**. A multigraph is a directed graph in which multiple edges (parallel edges) can exist between the same pair of nodes. Formally, a multigraph is defined as

$$G = (V, E, A_V, A_E),$$

where V is the set of nodes and $E \subseteq V \times V \times A_E$ the set of edges, with A_E denoting a set of edge attributes that distinguish multiple edges. The function $A_V : V \rightarrow \mathcal{X}_V$ assigns time-invariant attributes to nodes (e.g., club name, club ID, league name, league ID, country). Unlike a simple graph, where between two nodes only two edges can exist (one per each direction); in a multigraph the same two nodes can be connected by multiple edges. This property is particularly suitable for systems involving repeated interactions, where each interaction is meaningful and needs to be preserved. For example, in the context of football transfer market, the usage of a multigraph allows the representation of multiple transfers between the same pair of clubs, with each edge representing a distinct transfer event. Note that in a directed graph, each edge is oriented from a source node to a target node. This reflects the directional nature of a football transfer, going from the selling club to the buying club.

The multigraph that we considered in our research is an attribute graph, which allows us to associate attributes with nodes and, in particular, with edges. Nodes in our network represent clubs (or leagues), and thus their attributes must be time-invariant. For this

reason, out of the features present in the data, the only attributes we assign to the nodes are the static ones that allow us to identify them, such as club name, club ID, league name, league ID and country.

In contrast, the edge attributes are dynamic and specific to individual transfers. Consequently, they are suitable for storing all transfer-dependent information, including player-specific information such as player name, date of birth and player ID, which help us identify the transfer movements related to a specific player, and age at the time of transfer, which contributes to understand the temporal order of the transfers related to a certain player, to be able to reconstruct his career trajectory among the network. In addition, club-level metrics regarding the situation of source and target clubs at the time of transfer (including delta features, introduced in Section 3.3) are also stored as edge attributes. As they vary over time, they could not be stored as nodes attribute of a certain club, but by using edge attributes we are able to preserve this precious information in our network representation. Crucially, the possibility of including unique attributes for each edge allows us to introduce a timestamp attribute, enabling the possibility of modeling our multigraph as a **temporal network** (Holme and Saramäki, 2012; Masuda and Lambiotte, 2016).

A temporal network is a graph in which edges (and possibly nodes, but not in our network representation) are associated with timestamps or time intervals, indicating when interactions occur. In many real-world systems, including the football transfer market, relationships evolve over time and make it valuable to introduce temporal dynamics to model this behavior.

In our case, edges are annotated with timestamps that reflect specific transfer windows (e.g., "2011_w" for Winter 2011). These time labels are present as edge attributes in A_E . Temporal dynamics are essential for analyzing patterns over time, such as tracking transfer activity evolution, changes in the trading behavior of specific clubs across seasons and player career trajectories.

Filtering the multigraph by timestamp allows the creation of temporal subgraphs corresponding to specific seasons or periods, containing only the edges related to transfers that occurred during the specified time window. This facilitates our analysis of the network structure and dynamics' evolution over time.

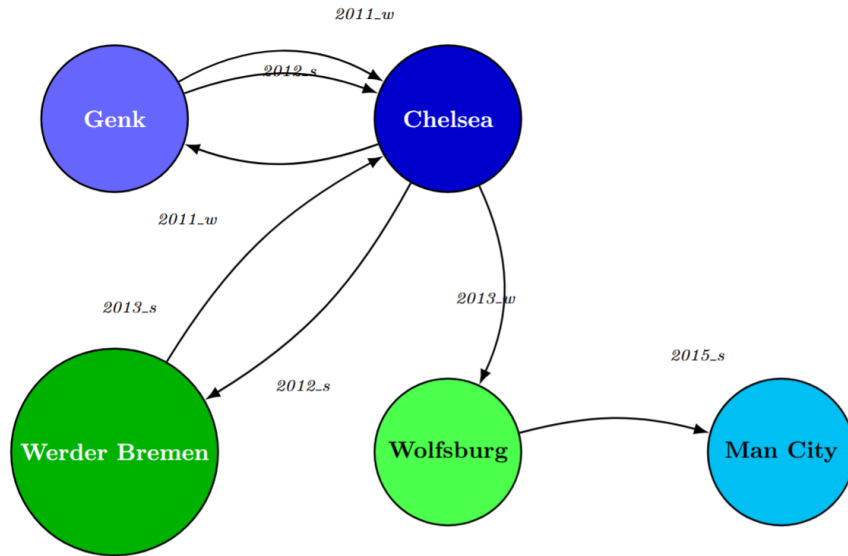


FIGURE 4.1: Career path of Kevin De Bruyne within the multigraph representation of the football transfer network.

Timestamp	Source Club	Target Club	Age	Market Value (€)
2011_w	Genk	Chelsea	20	8.00 mln
2011_w	Chelsea	Genk	20	9.00 mln
2012_s	Genk	Chelsea	21	9.00 mln
2012_s	Chelsea	Werder Bremen	21	9.00 mln
2013_s	Werder Bremen	Chelsea	22	10.00 mln
2013_w	Chelsea	Wolfsburg	22	15.00 mln
2015_s	Wolfsburg	Man City	24	45.00 mln

TABLE 4.1: Kevin De Bruyne’s transfer history with age and market value at each movement of his career trajectory present in our dataset.

To give an idea of what a career path looks like within our network representation, Figure 5.5 visualizes the career trajectory of Kevin De Bruyne across our multigraph, while Table 4.1 details each transaction related to him present in our dataset.

Each transfer generates a directed edge with its corresponding timestamp and attributes. Following a temporal order, we can see how in the winter 2011 he recorded the first transfer of his career, moving from Genk to Chelsea on a permanent deal and getting immediately sent back on loan to Genk as part of the agreement. This situation is represented in our multigraph as a directed edge from Genk’s node to Chelsea’s node, while the simultaneous loan move in the reverse direction is captured as a second edge in the opposite direction between the same two nodes. In summer 2012 the loan finished and De Bruyne went back to Chelsea, here we can see how the multigraph structure allows multiple edges between the same clubs, with a second edge pointing from Genk to Chelsea. After that transfer, he spent an entire season on loan at Werder Bremen, generating two edges going in opposite direction between Chelsea and Werder Bremen. Later, he moved permanently to Wolfsburg and then to Manchester City, where he is still playing in 2025. Each of these movements generated a new edge between two nodes, reflecting a unique transfer event. In summer 2025 his contract is expiring and De Bruyne will be moving to a different club on a free transfer, creating a new edge with a timestamp attribute 2025_s, starting from Manchester City and pointing to another node in the multigraph (probably Napoli).

4.1.2 Weighted Network

The second type of network representation that we use in our study is a **weighted network**, which forms the basis for most of our analytical experiments.

A weighted graph is formally defined as $G = (V, E, w)$, where V is the set of nodes, $E \subseteq V \times V$ is the set of edges and $w : E \rightarrow \mathbb{R}^+$ assigns a positive real-valued weight to each edge $e \in E$. These weights typically represent a property of the relationship between the connected nodes, such as cost, distance or strength. By incorporating weights, the graph captures not only the presence of connections but also their relative significance or intensity, enabling a more detailed analysis of the network.

In the context of football transfer networks, edge weights can represent quantities such as the amount or frequency of transfers between clubs, enriching our analysis by incorporating the magnitude of interactions. We used a directed weighted graph, preserving the source-to-target semantics of each transaction, as the reference network representation throughout our experiments. In this representation, only two edges can exist between a pair of nodes (one in each direction). For each ordered pair of source and target clubs, we

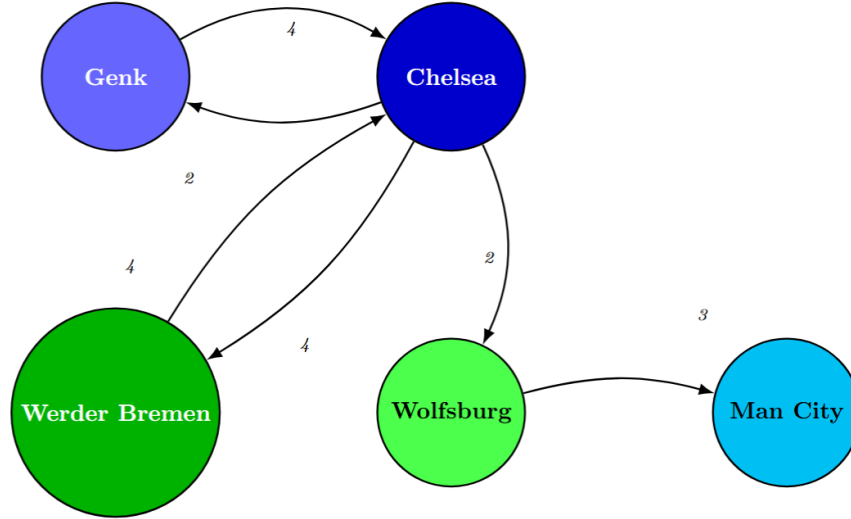


FIGURE 4.2: Example of weighted graph including the clubs that are part of Kevin De Bruyne’s career, where each edge reports the respective weight corresponding to the amount of transfers happened over that specific transfer direction in the 2005-2024 period.

represent their relationship with a single directed edge whose weight corresponds to the total number of transfers from the source to the target over the selected time period.

This contrasts with the multigraph representation, where each transfer is represented by a separate edge. Here, the weight aggregates all such instances, and no additional transfer-related attributes are associated with the edges. The aggregated weighted network simplifies the structure while capturing the magnitude of relationships.

Later in our experiments, we create variations of the weighted network. Once we introduce a definition of player success, we partition the dataset into transfers involving successful versus unsuccessful players. Each partition is then represented as a distinct weighted network containing only nodes and edges corresponding to the clubs and transfers present in the respective dataset, where edge weights reflect the number of transfers in the respective subset.

In Chapter 5, to implement some specific methods, we will further refine our network structure. Specifically, we will maintain the same nodes and edges, but change the weights to reflect relative transfer probabilities instead of raw frequencies. These probability-based weights are calculated based on micro-patterns involving 2- and 3-node motifs that are commonly observed in successful player trajectories. Details of this process are provided in Section 5.3.3.

Chapter 5

Approach

In this chapter, we present the methodology adopted to address our research questions. In Section 5.1, we describe the criteria and metric developed to assess whether a player is considered successful or not. Section 5.2 introduces a method for evaluating the level of a football club, interpreted as its attractiveness to potential players, based on its average *transfermarkt* market value. Section 5.3 focuses on the techniques applied to analyze the football transfer market network.

5.1 Success Assessment

Assessing whether a player has had a successful career is a complex and inherently subjective task. Defining what “success” means in this context and constructing an objective metric for it is challenging, given that different people may prioritize different goals over the course of a player’s career. Our objective is to create a metric to assign a binary label “successful” or “not successful” to each player.

One might argue that the number of trophies a player wins throughout his career defines their success. However, this raises different questions: Does a trophy reflect the value of an individual player or that of the entire team? Should all team members be considered equally successful, or should the contribution of each (e.g., number of appearances or goal involvements), be weighted? Furthermore, should all trophies be valued equally? For instance, is winning a second division title comparable to winning a top-tier league title? And what about players who have never won any trophies? Should they all be uniformly labeled as unsuccessful?

Others may argue that a successful player is one who earns significant income over his career. However, with the emergence of new, highly lucrative leagues in less competitive footballing nations, does this criterion still hold or does it become less meaningful from a sporting perspective?

Another key consideration regards the longevity and consistency of a player’s performance. Is reaching a high level momentarily sufficient or should success require maintaining high-level performance over a longer period or even for the full length of the sportive career?

Acknowledging these complexities, we developed a metric that aims to offer a robust and relatively unbiased approximation of a player’s career success, suitable for use in our research framework.

We base our success metric on a player’s *transfermarkt* market value. The market value, introduced by *transfermarkt.com* (Transfermarkt, 2024) in 2004, represents an estimate of the current value of a football player on the transfer market. It is derived from a combination of different pricing models based on performance metrics, age, positional characteristics, and community assessments from a user base of more than 700,000 members. Although it is not free from biases, such as those related to age, position and market conditions of the player, several studies such as Bryson, Frick, and Simmons, 2013 have

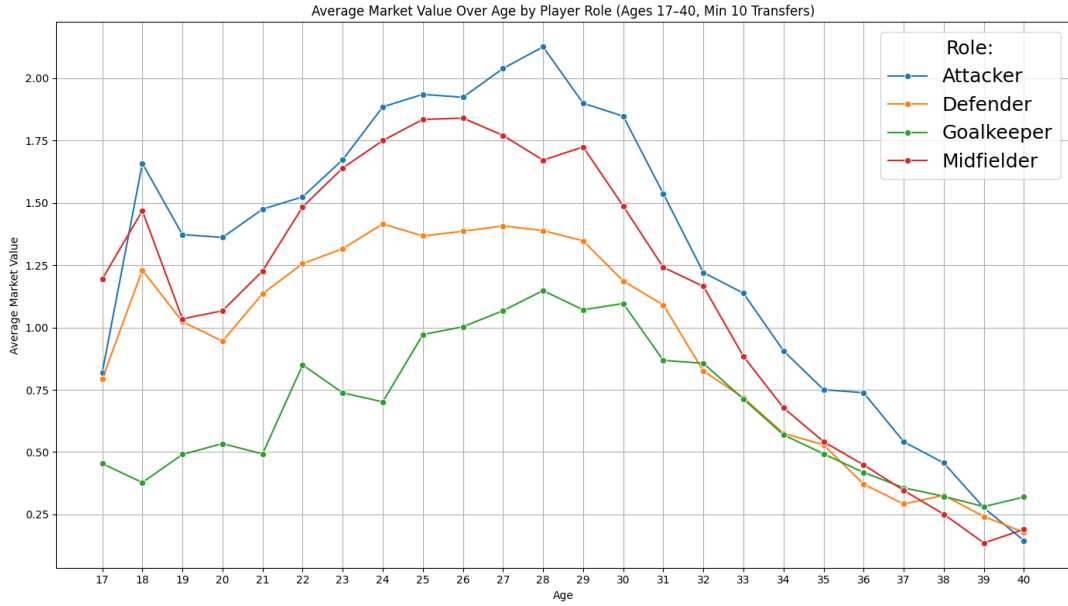


FIGURE 5.1: Average evolution of *transfermarkt* Market Value for football players distinguished by age and role

demonstrated that the estimations are often accurate and its reliability and widespread acceptance within the football community as an established reference point for discussions about player valuations.

To better understand how factors such as age and role influence the market value of a player, Figure 5.1 illustrates how the average market value evolves over the course of a player's career, classified by playing role. Early in their careers, players often experience rapid valuation increases due to potential and hype. Around 19–20 years old, the values tend to stabilize before a more consistent growth phase begins. Players generally reach their peak valuation at different ages depending on their role: defenders around 24, midfielders around 26 and attackers and goalkeepers around 28. Across all roles, market values tend to remain high until approximately 28–29 years old, after which a gradual decline begins, continuing until retirement. Notably, the magnitude of the market value is also influenced by the role, with attackers generally having the highest values, followed by midfielders, defenders, and goalkeepers.

In the construction of our success metric, we account for these variations. For each player in our dataset, we consider all their recorded transfers, which serve as data points representing distinct stages in their careers, each associated with his age, role, and market value at the time of the transfer. For each such data point, we compare the player's market value against the 75th percentile of values for players of the same age and role (the evolution of these thresholds is illustrated in Figure 5.2). If a player's market value exceeds this threshold for the majority of their data points (corresponding to the moments where he moved from a source club to a different target club), they are classified as successful; otherwise, they are considered unsuccessful. In this method, we chose a percentile-based threshold rather than an average-based one to mitigate the influence of outliers, since most players in the dataset have market values below €3 million, but elite players can significantly exceed this, sometimes reaching and surpassing €100 million.

To establish which percentile was the most suitable for our analysis, we experimented with

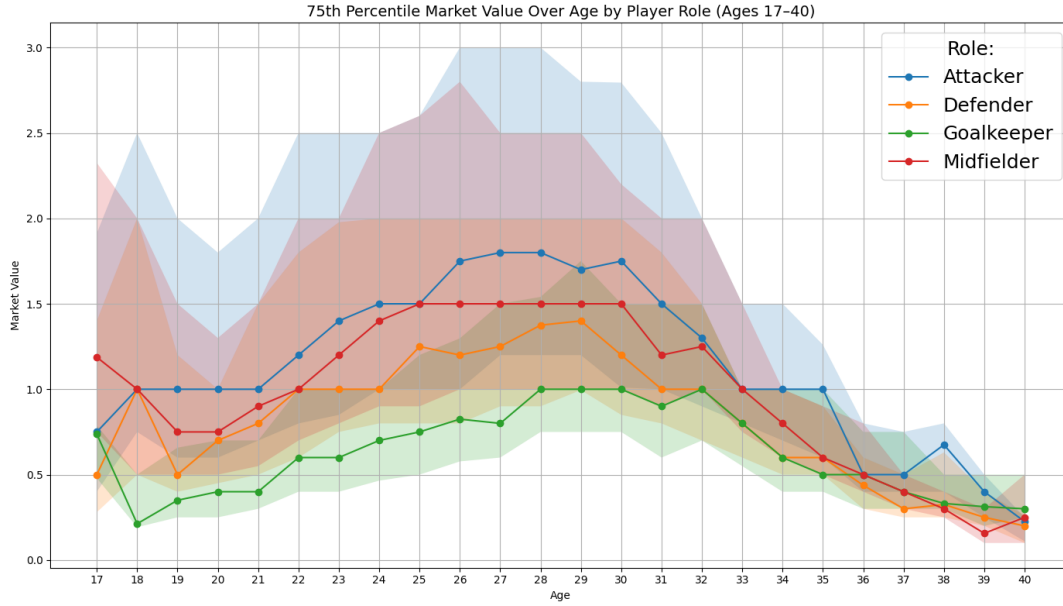


FIGURE 5.2: Evolution of *transfermarkt* Market Value for the 75th percentile of football players distinguished by age and role, with the relative standard deviation shaded areas

various percentile thresholds (70, 75, 80, 85, 90 and 95) and qualitatively evaluated the resulting player groupings based on football knowledge. By choosing a lower threshold, such as 70, we would label as “successful” a large number of players who may not be widely recognized, while a higher threshold, such as 90, would select only elite players consistently playing at top European clubs throughout their whole career. The 75th percentile struck an appropriate balance: it identifies as “successful” players who have reached the top division of a major European league and generally maintained that level or transitioned between top-tier and upper second-tier clubs. We consider this to be in good alignment with a reasonable and inclusive definition of success.

5.2 Club Level Assessment

After having established how to evaluate a player as successful or unsuccessful, we must define a suitable proxy to assess the level of a club at a specific point in time in an objective manner. The *club level* is a metric that represents how attractive a football club is as a destination for a player seeking a transfer. This attractiveness metric is intended to be independent of a player’s specific context, representing instead a general measure of how appealing a club is to the average professional footballer.

Since we use the *transfermarkt* market value as the core component of our metric to assess player success, we define the level of a football club at a given time as the average market value of the players in its squad. The metric is used to compare the individual market value of a player and the average market value of its source/target clubs, enabling us to determine whether a transfer represents a step up, step down, or a step on an even level in their professional career.

TABLE 5.1: Comparison of Club Rankings according to *transfermarkt* Market Value and UEFA Coefficients calculated on the last 5 and 10 years in Summer 2024

Club Name	Market Value Rank	UEFA 5-Year Rank	UEFA 10-Year Rank
Man City	1	2	3
Real Madrid	2	1	1
Barcelona	3	10	4
Arsenal	4	12	12
Paris SG	5	5	5
Liverpool	6	4	6
Tottenham	7	29	19
Bayern Munich	8	3	2
Chelsea	9	7	11
Man Utd	10	11	8
Inter	11	6	14
Leverkusen	12	13	17
Aston Villa	13	46	74
Newcastle	14	83	88
Juventus	15	22	10
AC Milan	16	19	34
West Ham	17	28	54
RB Leipzig	18	20	24
Brighton	19	82	87
Real Sociedad	20	32	49
Atletico Madrid	21	14	7
Napoli	22	34	20
Atalanta	23	16	27
Nottm Forest	24	–	–
Bor. Dortmund	25	8	9

Note: Color-coding indicates performance brackets: green indicates clubs in Top 10, yellow in 11–25, red >25 and gray not ranked. Market value ranks (MV Rank) are based on the estimated current market value of each club's squad. UEFA 5-Year and 10-Year rankings reflect club performance in European competitions across the respective timeframes.

Spearman Correlation	Market Value Rank	UEFA 5-Year Rank	UEFA 10-Year Rank
Market Value Rank	1.000	0.610	0.602
UEFA 5-Year Rank	0.610	1.000	0.881
UEFA 10-Year Rank	0.602	0.881	1.000

TABLE 5.2: Spearman correlation matrix between Market Value Ranking, UEFA 5-Year Ranking and UEFA 10-Year Ranking, based on the top-20 clubs in each ranking.

To validate our metric, we first analyze the ranking of the top 25 clubs worldwide based on their average *transfermarkt* market value in Summer 2024. In Table 5.1, we compare this ranking with the UEFA coefficients of the same clubs, calculated over the previous five and ten year periods (UEFA, 2025). The comparison shows promising alignment: all top-10 clubs according to the ten-year UEFA coefficients, and 9 of the top-10 according to the five-year coefficients, appear in the top-25 of the Market Value ranking. Furthermore, 7 out of 10 clubs from both UEFA coefficient rankings also fall within the Market Value top-10. Remarkably, 24 of the top 25 clubs ranked by market value are found within the top 90 of both UEFA coefficient rankings. Given the large number of clubs considered in the study, the degree of overlap, quantified through the Spearman correlation (Spearman, 1961) reported in Table 5.2, supports the reliability of market value as a meaningful proxy for club level. The correlation coefficients of approximately 0.6 between the market value ranking and the two UEFA coefficient rankings indicate that these measures capture similar dimensions of club quality. This suggests that clubs with the highest average *transfermarkt* market value generally also achieve strong results in European competitions, enhancing their attractiveness to players seeking transfers.

A closer look at Table 5.1 and some football knowledge, also reveal possible explanations behind apparent outliers. For example, Nottingham Forest is not ranked in either UEFA coefficient list, yet its inclusion in the market value top-25 is explained by a significant performance in the 2024/25 season, culminating in qualification for European competitions in the following year. Additionally, several English clubs (e.g., Arsenal, Tottenham, Aston Villa, Newcastle, West Ham, Brighton) appear significantly higher in the market value rankings than in the UEFA coefficient lists. This discrepancy is largely attributable to the English Premier League's huge financial resources, driven by lucrative TV deals and sponsorship agreements, which enable clubs to acquire high-value players and to establish the league as the one with the most valuable players.

Despite having some of the best football players in the world, the coexistence of such a large number of valuable teams in the same championship makes it impossible for all the clubs to consistently qualify for European competitions every year, lowering their UEFA coefficients. Consequently, these teams maintain their high market value and status, continuing to be highly attractive destinations for players, even when they are not regularly competing in UEFA tournaments. This further justifies the use of Market Value as a more player-centric indicator of club appeal.

Recognizing that a club's league affiliation significantly influences the average market value, especially due to the different income levels from broadcasting rights and sponsorship deals, we decided to also assess the validity of our metric at a national level. We used the 2023/24 season of the Italian Serie A as a case study, comparing the final league position of each club with its average *transfermarkt* market value at the end of the season. As shown in Figure 5.3, the market value ranking closely mirrors the final standings of the league. One notable exception is represented by Napoli, which finished tenth despite having the second-highest average market value. However, this discrepancy can be contextualized: Napoli won the league in 2022/23 and again in 2024/25, indicating that their squad value more accurately reflects their overall appeal and competitive strength than their temporary underperformance in a single season. This reinforces the notion that market value can serve as a more stable indicator of club attractiveness over time than short-term competitive results in a specific season.

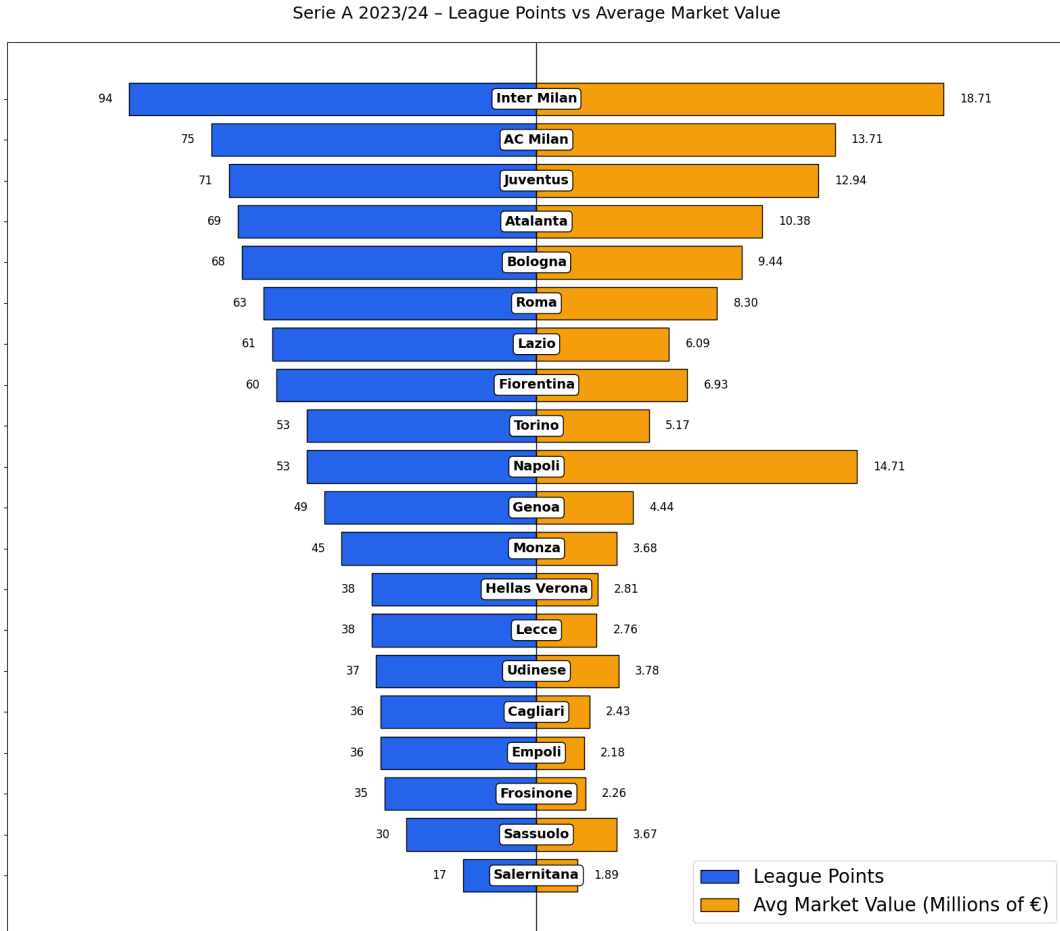


FIGURE 5.3: Comparison between the final positions with the obtained points of Italian football clubs in Serie A 2023/24, and their average *transfermarkt* Market Value in Summer 2024, at the end of the considered season.

5.3 Network Analysis

In this section, we describe the network analysis techniques used to answer our research question. To guide the reader through the different analytical dimensions, Table 5.3 summarizes the set of metrics employed, indicating for each the corresponding graph representation, level of aggregation and analytical scale. This overview provides a systematic framework for the more detailed analyses presented in the following subsections. In Subsection 5.3.1, we adopt a macro-level perspective, analyzing the structure of the overall transfer network, and then focusing in particular on the structural differences between the networks of successful and unsuccessful players. In Subsection 5.3.2, we investigate meso-level dynamics by identifying communities of clubs that appear particularly interconnected and flows of recurrent aggregated transfers between leagues in successful and unsuccessful player networks. Then, in Subsection 5.3.3 we present a micro-level analysis of 2-node and 3-node motifs in football player careers. Finally, Subsection 5.3.4 introduces ClubRank, our method to identify clubs that play a key role in the development of successful football player trajectories within the transfer market.

Metric	Network Representation	Aggregation Level	Scale
Temporal evolution of transfer activity	Multigraph	Club, League	Macro
Number of nodes (n)	Multigraph, Weighted network	Club, League	Macro
Number of edges (m)	Multigraph, Weighted network	Club, League	Macro
Weighted edge sum (W)	Weighted network	Club, League	Macro
Average weighted in-degree ($\bar{k}_w^{in}$)	Weighted network	Club, League	Macro
Average weighted out-degree ($\bar{k}_w^{out}$)	Weighted network	Club, League	Macro
Network density (D)	Weighted network	Club, League	Macro
Average clustering coefficient (ACC)	Undirected projection of weighted network	Club, League	Macro
Largest strongly connected component (LSCC)	Weighted network	Club, League	Macro
Reciprocity (r)	Weighted network	Club, League	Macro
Weighted degree assortativity (r_{deg})	Weighted network	Club, League	Macro
Average weighted betweenness centrality ($C_B^{(w)}$)	Weighted network	Club, League	Macro
Weighted degree distribution	Weighted network	Club, League	Macro
Transfer direction	Plain Dataset	Club	Macro
Community detection (Louvain)	Weighted network	Country	Meso
Transfer flows	Weighted Network	League	Meso
2-node motif analysis	Probabilistic weighted network	Club, League	Micro
3-node motif analysis	Probabilistic weighted network	Club, League	Micro
ClubRank	Probabilistic outgoing weighted network	Club	Micro

TABLE 5.3: Overview of network analysis metrics, indicating the relative graph representation, level of aggregation and analytical scale (macro, meso or micro).

5.3.1 Macro-Level

In this subsection, we concentrate on the structural properties of the complete transfer network, as well as the networks constructed exclusively from successful and unsuccessful player data. By computing and comparing several network metrics across these two networks, our objective is to understand their topological differences and the relative implications.

Overall Structure Analysis

We begin by examining the network that represents the entire dataset, considering both multigraph and weighted representations, and performing the analysis at two levels of aggregation: *league level* and *club level*.

First, we focus on the multigraph representation to validate one of the motivations for this research, the consistent increase in the volume of transfers within the football transfer market over the years. To this end, we construct a series of seasonal subgraphs from 2005 to 2024 and compute the number of edges (transfers) occurring in each transfer window (summer and winter), quantifying the temporal evolution of transfer activity.

Subsequently, we move to the weighted directed network representation. Here, edge weights correspond to the number of transfers between nodes (clubs or leagues). We calculate several standard network metrics to characterize the structure of the complete network and then repeat the process for the networks built from only successful and only unsuccessful players. This allows us to compare the structural properties between the different player categories. The computed metrics include:

- **Number of nodes** ($n = |V|$): total number of unique clubs (or leagues) in the network.
- **Number of edges** ($m = |E|$): total number of unique directed connections representing transfers in the networks.
- **Weighted edge sum** (W): total number of transfers, calculated as the sum of edge weights:

$$W = \sum_{(s,t) \in E} w_{st}$$

Here, w_{st} is the weight of the edge from node s to node t .

Since the network is directed, remind that the edge from node s to node t is different from the edge from node t to node s . Self-edges, going from node s to node s are also included in this network representation and in this analysis.

- **Average weighted in-degree** ($\bar{k}_w^{in}$): the mean number of incoming transfers per node over the full period (2005–2024):

$$\bar{k}_w^{in} = \frac{1}{n} \sum_{s=1}^n \sum_{t \in V} w_{ts}$$

Here, V corresponds to the set of all nodes in the network.

- **Average weighted out-degree** ($\bar{k}_w^{out}$): the mean number of outgoing transfers per node over the same period:

$$\bar{k}_w^{out} = \frac{1}{n} \sum_{s=1}^n \sum_{t \in V} w_{st}$$

- **Network density** (D): overall level of connectedness in the network, defined as the proportion of actual connections to the maximum possible number of directed edges, formally:

$$D = \frac{m}{n(n-1)}$$

Network density values range from 0, indicating a complete disconnected network, to 1 corresponding to a fully connected network (Wasserman and Faust, 1994).

- **Average clustering coefficient (ACC)**: measures the overall tendency of the nodes of a network to form local clusters. It reflects the likelihood that two neighbors of a given node are also connected to each other, forming a triangle:

$$ACC = \frac{1}{n} \sum_{s=1}^n \frac{|\{(t, u) : (s \rightarrow t), (s \rightarrow u), (t \rightarrow u) \in E\}|}{k_s(k_s - 1)}$$

In the case of directed graphs, such as the one studied here, the standard definition of ACC is computed on the corresponding undirected version of the network, obtained by replacing each directed edge with an undirected one (if at least one direction exists between two nodes). Consequently, k_s corresponds to the degree of node s (Watts and Strogatz, 1998).

- **Largest strongly connected component (LSCC)**: the size of the largest subset of nodes such that for any two nodes s and t , there exists a directed path from s to t and from t to s . This is a key indicator of the integrity of the network structure.
- **Reciprocity** (r): quantifies the tendency of pairs of nodes to form mutual connections, defining how often a directed edge from node s to t is reciprocated by and edge from node t to s (Garlaschelli and Loffredo, 2004).

$$r = \frac{|\{(s, t) \in E : (t, s) \in E\}|}{m}$$

Reciprocity values range from 0 (no mutual edges) to 1 (all edges reciprocated).

- **Weighted degree assortativity (Pearson)** (r_{deg}): assess whether the high-degree nodes tend to connect to other high-degree nodes (assortative mixing) or to low-degree nodes (disassortative mixing). In particular, the Pearson coefficient used in this work takes the edge weights into account:

$$r_{deg} = \frac{\sum_{(s,t)} w_{st} (k_s - \bar{k}_w)(k_t - \bar{k}_w)}{\sum_{(s,t)} w_{st} (k_s - \bar{k}_w)^2}$$

Here, $\bar{k}_w$ is the average degree weighted by edge weights.

Its values range from -1 corresponding to complete disassortative mixing, to 1 indicating complete assortative mixing, and values close to 0 indicate the absence of correlation (Newman, 2002).

- **Average weighted betweenness centrality** ($\overline{C_B^{(w)}}$): measures the extent to which nodes act as bridges along shortest paths between other nodes, accounting for edge weights when computing the shortest paths. It is defined as

$$\overline{C_B^{(w)}} = \frac{1}{n} \sum_{i \in V} \sum_{i \neq s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Here, σ_{st} is the number of shortest weighted paths from source node s to target node t , and $\sigma_{st}(i)$ is the number of those paths that pass through node i .

A higher average betweenness centrality indicates that, on average, nodes tend to play a more central role in connecting other nodes via shortest weighted paths (Freeman, 1977).

Following the calculation of these global metrics, we analyze the degree distributions of the networks, with particular attention to the clubs and leagues with the highest weighted degrees.

Transfer Direction Analysis

Another interesting analysis to assess the differences between the trajectories of successful and unsuccessful players involves evaluating the directionality of their transfers over time. In this context, we define the transfer directions as *Upward*, *Horizontal*, or *Downward*, and employ two different methodologies to label these transitions accordingly.

1. Club Transfer Direction. In the first method, we compare the club levels of the source and target clubs, using their average *transfermarkt* market values (AMV) at the time of transfer. For a given transfer from the source club s to the target club t , we define:

$$\text{Direction} = \begin{cases} \text{Upward} & \text{if } \frac{AMV(t)}{AMV(s)} > 1.2 \\ \text{Downward} & \text{if } \frac{AMV(t)}{AMV(s)} < 0.8 \\ \text{Horizontal} & \text{otherwise} \end{cases}$$

This categorization enables us to visualize and analyze the distribution of transfer directions across player ages and compare how these patterns differ between successful and unsuccessful players over the course of their careers, with the goal of identifying crucial steps within the development of football players.

2. Player Transfer Direction. In the second method, we introduce a novel perspective by considering the ratio between a player's own market value and the average market value of the target club at the time of the transfer. Specifically:

$$\text{Direction} = \begin{cases} \text{Upward} & \text{if } \frac{MV(\text{player})}{AMV(t)} > 1.2 \\ \text{Downward} & \text{if } \frac{MV(\text{player})}{AMV(t)} < 0.8 \\ \text{Horizontal} & \text{otherwise} \end{cases}$$

This allows us to observe how individual player level aligns or conflicts with the level of their target clubs, and how such patterns change across age and between successful and unsuccessful players. For completeness, we also compare player market values with the average market value of the source clubs to assess the typical context from which players are transferred.

5.3.2 Meso-level

We proceeded to examine the networks from a meso-level perspective. At this intermediate scale, the focus shifts to identifying groups of nodes that exhibit similar behaviors and show a higher tendency to cluster internally, forming communities. The objective is to detect these communities within the successful and unsuccessful player networks, assess their stability, and investigate the potential factors underlying their formation. Additionally, we analyze the transfer flows that characterize the networks of successful and unsuccessful players, with the aim of understanding how transfers aggregate and circulate among different groups within professional football leagues.

Community Detection

Community detection is a fundamental technique in network analysis, that aims to identify subsets of nodes, commonly referred to as communities, that are more densely connected internally than they are to the rest of the network. Within the context of football transfer networks, communities can correspond to clusters of clubs that engage more frequently in transfer activities with each other. These clusters may reflect underlying structural, geographic, or economic relationships that influence transfer behavior. By detecting such communities, we can identify patterns in transfer dynamics and gain a deeper understanding of how the global football market is organized.

Among the wide variety of algorithms for detecting communities in networks, this study focuses on the Louvain method (Blondel et al., 2008), a modularity-based approach which has become particularly popular due to its computational efficiency and scalability. Applying the Louvain algorithm to our data revealed that, in the vast majority of cases, clubs from the same country tend to cluster within the same community. This observation led to the decision to change the level of aggregation from individual clubs to country level. Consequently, we assigned each country to the community including the majority of its clubs, effectively treating countries as the fundamental units of analysis in the results related to this community detection approach.

The Louvain algorithm (Blondel et al., 2008) operates as a greedy optimization technique that maximizes a network metric called modularity (Newman and Girvan, 2004), which measures the density of links within communities compared to the expected density of such links in a random network that preserves the degree distribution of the original network. The algorithm operates through two iterative phases: initially, each node is assigned to its own community, and nodes are merged iteratively to maximize the local modularity gain; subsequently, a new network is constructed in which nodes represent the communities identified in the previous phase. This hierarchical process continues until no further modularity improvements can be achieved. The Louvain method's speed, scalability, and ability to uncover meaningful communities without predefining their number make it well-suited for analyzing complex graphs such as the football player transfer network.

To assess the quality and robustness of the detected community structures, we employed two complementary metrics: modularity and Normalized Mutual Information (NMI). Modularity (Newman and Girvan, 2004), as an intrinsic measure, quantifies the strength of the division of a network into communities by comparing the observed density of intra-community links to that expected in a null model (a random network with the same degree distribution). Higher modularity values indicate more pronounced and significant community structures. However, since the Louvain algorithm inherently maximizes modularity, and because in large, complex networks like football transfer markets modularity alone

may not sufficiently characterize community quality, we incorporated NMI as an additional evaluation metric to assess the communities' quality.

NMI (Danon et al., 2005) is a measure used to quantify the similarity between two different community partitions. Usually, it is used to compare an algorithmically derived partition against a reference partition based on external knowledge. In our study, since we do not have such a ground-truth partition, we use NMI to evaluate the quality and consistency of the community partitions obtained across multiple runs of the Louvain algorithm. Since the Louvain method is not deterministic, it can lead to the creation of different community structures on separate executions. By computing the NMI metric between the community partitions obtained from 20 independent runs, we calculate the average NMI score to assess the variability of the results. NMI values range from 0 (indicating that there is no mutual information or agreement) to 1 (perfect agreement), making this metric suitable for evaluating how well the detected communities align over multiple runs of the Louvain algorithm.

Transfer Flows

Having identified country-level communities that demonstrate frequent patterns related to the collaboration within clubs in the transfer market for both successful and unsuccessful players, we further investigated the finer-grained dynamics of player movement between leagues. Specifically, we analyzed how groups of players flow from one league to another and we examined the differences in modularity, leagues involved, and distribution of transfer flows distinguishing between successful and unsuccessful player networks.

To analyze inter-league transfer dynamics, we considered the weighted, directed network aggregated at the league level. The corresponding adjacency matrix directly represents the directional flow of players between leagues, where each entry indicates the number of players transferred from the source league (row) to the target league (column). This matrix captures both the magnitude and directionality of inter-league player movements, providing a structured representation of transfer trajectories.

Using aggregated transfer count data from individual player movements, we filtered the matrix to focus only on transfers occurring more than a defined number of times, emphasizing the most significant and recurrent league-to-league pathways. Initially, we set the thresholds to 36 transfers for successful players and 133 transfers for unsuccessful players, and subsequently we experimented by increasing these thresholds (respectively to 100 and 370) to test the evolution of the results. This approach allows us to visualize and analyze the volume and direction of player flows between leagues, enabling an intuitive understanding of central hubs, feeder leagues, and reciprocal transfer exchanges within the networks of successful and unsuccessful players.

5.3.3 Micro-level

In this section, we describe the network analysis techniques used to identify recurrent patterns and particularly influential clubs within the successful player network. The identification of significant 2- and 3-node patterns at both league and club levels, along with the discovery of the most influential clubs in the development of successful football player careers, represents a core contribution of this study from a micro-level network perspective.

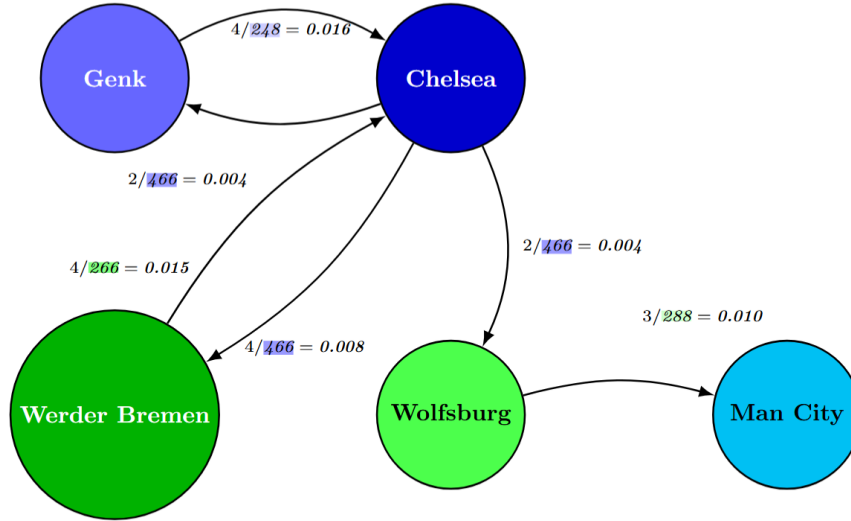


FIGURE 5.4: Example of outgoing probabilistic network including the clubs that are part of Kevin De Bruyne's career path. Each edge reports the respective weight corresponding to the outgoing transfer probability to move from a specific source club to the target club.

Probabilistic Network Representation

To individuate which patterns are most relevant in the football transfer market, we modify its network representation, beginning with a weighted, directed network and reinterpreting the edge weights. Specifically, we focus on the directed weighted network composed of successful players and create two distinct networks where edge weights represent, respectively, the incoming and outgoing probabilities of transitioning to a specific club, given the source or target club. We describe these two methods for assigning edge weights as follows.

In the outgoing probabilistic weighted network, each edge weight is calculated as the probability of moving to a specific target node given the source node. Formally, this is the ratio of transfers from a specific source node to a specific target node, divided by the total number of outgoing transfers from that source node:

$$w(e_{s,t}) = P(t|s) = \frac{c(s,t)}{\sum_{i=1}^n c(s,t_i)} \quad (5.1)$$

Equation 5.1 indicates that the edge weight $w(e_{s,t})$, assigned to the edge e , going from the source node s to the target node t , corresponds to the probability of a transfer to t given that the player starts at s . The numerator $c(s,t)$ denotes the count of such transfers, corresponding to the weight of the edge $e_{s,t}$ in the previous weighted network representation, while the denominator sums all outgoing transfers from s , effectively representing its weighted out-degree in the original network representation.

In the incoming probabilistic weighted network, each edge weight is computed as the probability of originating from a specific source node given the target node. This is the ratio between the number of transfers from that source node to the target node, and the total number of incoming transfers to the target node:

$$w(e_{s,t}) = P(s|t) = \frac{c(s,t)}{\sum_{i=1}^n c(s_i,t)} \quad (5.2)$$

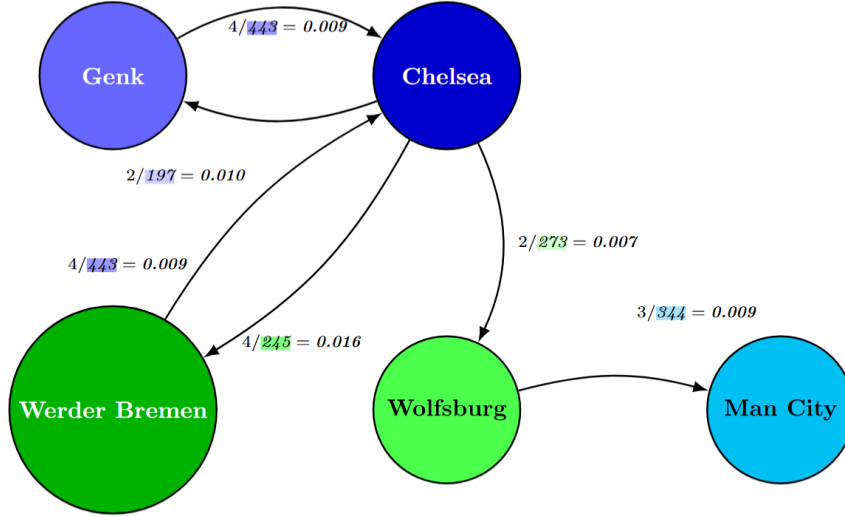


FIGURE 5.5: Example of incoming probabilistic network including the clubs that are part of Kevin De Bruyne’s career path. Each edge reports the respective weight corresponding to the incoming transfer probability to move from the source club to a specific target club.

As shown in Equation 5.2, the edge weight $w(e_{s,t})$ reflects the likelihood of arriving from the source node s , conditioned on a transfer arriving at the target node t . Here, $c(s,t)$ is again the number of transfers from node s to node t , while the denominator represents the count of all transfers going into the target club t , corresponding to the weighted in-degree of t in the previous weighted network representation.

Although explained with reference to the club-level aggregation, these same probabilistic representations and formulas were analogously applied to league-level aggregated networks.

2-Node Pattern Analysis

The goal of the 2-node pattern analysis is to identify transfers that are consistently present in the careers of successful players and are more frequent compared to those of unsuccessful players. Practically, this means detecting 2-node paths (a source and a target nodes representing either clubs or leagues) that result overrepresented in the successful player network compared to the overall network. These should also be statistically significant, meaning that we should aim to avoid including outlier paths that appear only a few times in the dataset, but focus on transfers happening consistently among successful player careers.

To quantify overrepresentation, we initially calculated the log-ratio of the transition probabilities (edge weights) in the successful network compared to the total network, according to the formula stated in Equation 5.3 (Dunning, 1993).

$$\text{log-ratio}(e_{s,t}) = \log \left(\frac{w_{suc}(e_{s,t})}{w_{tot}(e_{s,t})} \right) \quad (5.3)$$

The log-ratio compares the observed frequency of a specific 2-node path in the successful player network to its expected frequency according to the overall network. This allows for quantifying deviations from randomness and capturing how much more likely a specific transfer path is among successful careers relative to its baseline frequency. A high log-ratio indicates that a given trajectory appears more frequently in successful careers

than expected under the baseline distribution.

However, raw log-ratios can be unstable, especially for rare paths, leading to the risk of identifying outlier 2-node paths as significant. To address this, we used a smoothed log-odds approach. This incorporates Laplace smoothing (Jurafsky and Martin, 2009) into probability estimates, adding a small constant to mitigate the impact of low counts and provide more robust estimates. Specifically, for each direct edge, the conditional transition probability is computed according to Equation 5.4.

$$w(e_{s,t}) = P_{\text{smoothed}}(t|s) = \frac{c(s,t) + 1}{\sum_{i=1}^n c(s,t_i) + n} \quad (5.4)$$

In Equation 5.4, n denotes the number of distinct target nodes connected by a directed edge starting from the source node s . This formulation ensures that even rarely observed edges receive a non-zero probability estimate, improving robustness. We computed these probabilities separately for the successful player network and the overall transfer network.

Next, we transformed these smoothed probabilities into log-odds (Monroe, Colaresi, and Quinn, 2008) according to Equation 5.5.

$$\text{log-odds}(e_{s,t}) = \log \left(\frac{P_{\text{smoothed}}(t|s)}{1 - P_{\text{smoothed}}(t|s)} \right) \quad (5.5)$$

Using log-odds instead of raw probabilities symmetrizes the scale, making overrepresentation and underrepresentation of a path equally interpretable. Moreover, it linearizes multiplicative differences in probabilities, facilitating direct subtraction between the successful and overall player networks.

Finally, as reported in Equation 5.6, we compute the difference in log-odds between the successful and total player networks, which represents a more stable indicator of overrepresentation.

$$\Delta \text{log-odds}(e_{s,t}) = \text{log-odds}_{\text{suc}}(e_{s,t}) - \text{log-odds}_{\text{tot}}(e_{s,t}) \quad (5.6)$$

Positive values of $\Delta \text{log-odds}$ imply that a transition path is more frequent among successful players' transfers than expected. This method emphasizes paths that are not just common, but disproportionally linked to successful players' career paths, enabling us to retrieve interpretable insights.

Although Laplace smoothing helps mitigate the effect of zero- or low-frequency transitions by adjusting probability estimates, it does not fully eliminate the influence of rare events. To reduce the impact of such statistical noise, we introduced a minimum support threshold based on the number of occurrences in the successful players' network. Specifically, we required that a transfer must occur at least 10 times (league-level) or 5 times (club-level) in the successful network to be included in the analysis. This threshold balances robustness with the granularity of the network, ensuring to retrieve surprising, but yet robust patterns.

In particular, our analysis focused on finding the most overrepresented source and destination for a transfer involving each considered club or league. To do so, we selected all leagues and a subset corresponding to the top 40 clubs (based on the current UEFA coefficients (`uefa_coeff`)), computing the top source and target nodes for each based on

both log-ratios and smoothed log-odds. This allows us to recommend the best potential next destination (club or league) for a player given his current club (or league), or the most promising recruitment source for a club, based on historical success of specific 2-node patterns within football players' careers.

3-Node Pattern Analysis

To explore higher-order patterns beyond individual transfers, we analyzed 3-node motifs, defined as sequences of three chronologically ordered transfers made by a certain player. These triplets capture more complex behavior than 2-node paths and were extracted from both the complete dataset and the successful player subset, at both club and league levels. To better underline this complex behavior, particularly interesting in the league-aggregated network, we classified the motifs into four different categories that characterize the players' career trajectories.

- **Chains:** Sequences of three distinct nodes (A-B-C).
- **Loops:** Sequences where the first and last nodes are the same (A-B-A).
- **Partial repetitions:** Sequences in which a node appears twice consecutively (A-A-B or A-B-B).
- **Triple repetitions:** Sequences where all nodes are the same (A-A-A).

For club-level aggregated networks, only chain and loop motifs were considered relevant, as self-edges are less meaningful for our research focus. After having classified the different motifs in these four clusters, we move the focus back to our main interest. As with 2-node patterns, our objective is to determine which 3-node motifs are overrepresented in successful player careers, compared to their presence in the overall network. To do this, we calculated motif frequencies, applied Laplace smoothing, and computed both raw log-ratios and smoothed log-odds for each motif, to assess how much more (or less) likely a motif is to occur among successful players' careers compared to the general population. To filter out noise, we require a motif to occur at least 10 times in the full dataset and 5 times in the successful set for league-level aggregation; and at least 5 and 3 times, respectively, at club level.

We then visualized the most over- and underrepresented 3-node motifs in terms of their raw log-ratios and smoothed log-odds, providing insight into whether certain sequences of transitions across leagues (or clubs) represent an indication of success, being consistently associated with successful players' career trajectories.

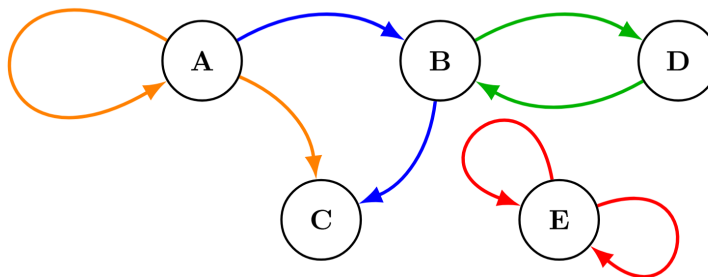


FIGURE 5.6: Illustration of the four categories of 3-node motifs: Chains (blue), Loops (green), Partial repetitions (orange) and Triple repetitions (red).

By incrementing the number of nodes considered in the motifs, we always retrieve a smaller number of occurrences in the network. In particular, consistent motifs including more than 3 nodes are extremely rare in club-level aggregated networks. Therefore, this is the reason why our micro-level analysis focused on 2- and 3-node patterns, where the results are still significant and interpretable.

5.3.4 ClubRank

The final part of our network analysis, tackling our network representations from a micro-level perspective, focuses on identifying the most influential clubs in the development of successful players' careers. Inspired by the original application of the PageRank algorithm (Page et al., 1999) in web networks, where it measures the importance of a webpage based on the structure of hyperlinks, we applied a similar approach, which we call ClubRank, to our transfer network.

PageRank is an iterative algorithm that operates on the principle that a node is important if it is connected to other important nodes. Formally, the algorithm simulates a random walk through the network, where at each step the walker follows an outgoing edge with a probability proportional to its weight, or teleports to a random node with a small fixed probability. The resulting steady-state probabilities indicate the relative importance of each node, representing their long-term visitation likelihood.

The structure of the probabilistic outgoing club-to-club transfer network among successful players forms a probabilistic, directed graph that closely resembles the flow-based nature of the web, with clubs acting as nodes and weighted edges representing the likelihood of a transfer between them. This structural similarity motivated the use of the PageRank algorithm to assess the influence of a club, not in terms of visibility, but in terms of their centrality within successful player development paths.

By computing PageRank scores over the successful outgoing transfer network, we constructed a "Club Rank": a data-driven measure of how structurally influential each club is in forwarding players who go on to have successful careers. This metric reflects both the volume and the quality of a club's connections within the successful player network, capturing the idea that clubs positioned as key stepping stones in successful trajectories gain greater importance. High-ranking clubs are not necessarily those with the most activity, but those best embedded within the successful transfer structure, serving as hubs or key stepping stones. In our context, clubs with high Club-rank are those that are frequently reached through the flow of successful transfers and are well-connected to other influential clubs, suggesting their central role in successful player career development.

In Chapter 6, where we discuss the results of our experiments, we investigate in detail the resulting rankings based on the different centrality measures and interpret the structural roles that different clubs play within these networks.

Chapter 6

Experiments

In this chapter, we describe and discuss all the experimental results retrieved applying the methodologies explained in Chapter 5.

In Section 6.1, we focus on the overall structure of our network representations, analyzing the graph related to the complete dataset and highlighting the differences between the successful and unsuccessful player networks. Section 6.2 presents the obtained communities, assesses the partition quality, and displays the transfer flows characterizing the successful and unsuccessful player networks. Section 6.3 contains the results related to the 2-node and 3-node pattern analysis, as well as the application of ClubRank to the player transfer network, identifying which patterns are particularly associated with success and which clubs play a key role in the development of successful players. Finally, in Section 6.4, we discuss the obtained results and contextualize them with findings reported in the related literature.

6.1 Macro-level

In this section, we present the results related to the macro-level perspective of the football transfer network. We begin by describing the overall network created from the entire dataset, and then shift to the main contribution of this section: identifying the global structural differences between the networks of successful and unsuccessful players.

6.1.1 Overall Structure Analysis

As a first analysis, we focused on verifying one of the fundamental motivations behind this work, the consistent increase in the number of transfers per year and per transfer window over time.

As shown in Figure 6.1, the number of transfers per transfer window has increased steadily over the past 20 years. In 2005, the total number of player movements across the 50 considered football leagues was approximately 4,000, while in 2024 this number exceeded 8,000, demonstrating that it has more than doubled in two decades. The trend is clearly visible, exhibiting an almost monotonic increase, which was temporarily disrupted in 2020 by the COVID-19 crisis. This global event significantly affected the football market, leading to a reduction in the number of transfers during that year. Following the return to normal conditions, the clubs registered the highest number of transfers in a single year in 2022. Since then, the trend has slightly decreased, appearing to stabilize around 8,000 transfers per year.

Another important observation from Figure 6.1 concerns the consistent difference in the number of transfers occurring in the winter versus the summer windows. Summer transfers consistently account for approximately twice as many movements as winter ones.

This phenomenon can be attributed to the shorter duration of the winter transfer window. Furthermore, clubs are generally more inclined to finalize transfers during the summer,

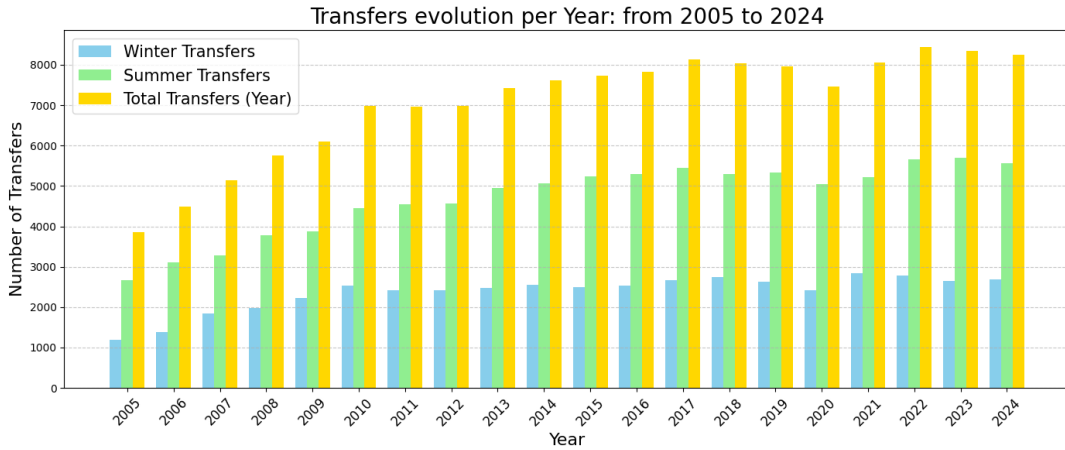


FIGURE 6.1: Distribution of the number of transfers over a 20-year time span, aggregated by year and by transfer window.

when competitions have concluded and there is time to strategically plan the upcoming season. During this period, teams typically reshape their squads, replacing departing players and satisfying the demands of newly appointed coaching staff. In contrast, the winter window is mainly used to make minor adjustments or replace injured players, often without disrupting the established team chemistry in the middle of the season.

After verifying the hypothesis regarding the increasing number of transfers per year, we proceeded to analyze the global structure of our network representations using the network metrics introduced in Section 5.3. We begin by examining the entire dataset and comparing the results at different levels of aggregation.

Table 6.1 presents the values of key network metrics, providing insight into the structure of our networks. Naturally, both the league-level and club-level networks contain the same total number of transfers (141,538), but these occurred across 1,531 unique clubs participating in 50 different national competitions over the last 20 years. Consequently, the number of nodes and edges is significantly larger for the club-level network, which also exhibits a lower average weighted degree.

Metric	League-Level Network	Club-Level Network
Number of Nodes	50	1531
Number of Edges	2099	70070
Total Transfer Count (Edge Weights Sum)	141,538	141,538
Average Weighted In-Degree	2830.76	92.45
Average Weighted Out-Degree	2830.76	92.45
Network Density	0.8567	0.0299
Average Clustering Coefficient	0.9360	0.3806
Largest SCC Size	50	1494
Reciprocity	0.9271	0.6261
Weighted Degree Assortativity (Pearson)	0.0188	0.2060
Average Weighted Betweenness Centrality	0.027631	0.001108

TABLE 6.1: Comparison of network metrics between football transfer networks: League-Level aggregation vs. Club-Level aggregation.

It is not coincidental that the average weighted in-degree and out-degree are equal across both networks. This is a direct result of how weighted edges contribute simultaneously to the out-degree of the source node and the in-degree of the target node, keeping the total sums balanced.

Examining the other metrics, the league-level network appears particularly well connected, with high values for both density and average clustering coefficient. In contrast, the club-level network shows a lower density and a relatively lower (but still significant) average clustering coefficient. This difference is expected since, in the aggregated league-level view, it is highly probable that at least one club from each league has been involved in transfers with a club from every other league. However, in the finer granularity at club-level, it is unrealistic to expect each of the 1,531 clubs to trade with all the other clubs present in the network.

With each league involved in an average of 2,830 transfers during the considered period, it is likely that they form a highly connected graph. Clubs, however, average only about 92 transfers each, reducing the probability of being universally connected. In this light, the clustering coefficient at club level remains impressively high, indicating the frequent formation of transfer triangles.

The sizes of the largest strongly connected components further support the interpretation of a well-connected network. All 50 leagues and 1,494 out of 1,531 clubs are part of their respective largest strongly connected components. Furthermore, the reciprocity metric suggests a high level of bidirectional exchange between leagues and, to a slightly smaller extent, among clubs. This implies the presence of long-term trading relationships, such as elite clubs loaning young talent to smaller clubs that offer them valuable playing time, creating a mutually beneficial exchange.

Degree assortativity results close to zero at league level, showing almost no correlation between the degree of the nodes that tend to form edges. At club level, there is a slight positive correlation, suggesting that active clubs tend to interact more with other similarly active clubs. Finally, as expected, betweenness centrality is higher for the smaller and more densely connected league-level network, where nodes are more likely to fall on the shortest paths between other nodes, compared to the club-level network.

Metric	League-Level		Club-Level	
	Successful	Unsuccessful	Successful	Unsuccessful
Number of Nodes	48	50	1028	1531
Number of Edges	1199	2077	18414	58612
Total Transfer Count (Weight Sum)	29,934	110,833	29,934	110,833
Average Weighted In-Degree	623.62	2216.66	29.12	72.39
Average Weighted Out-Degree	623.62	2216.66	29.12	72.39
Network Density	0.5315	0.8478	0.0174	0.0250
Average Clustering Coefficient	0.8178	0.9291	0.2670	0.3668
Largest SCC Size	47	50	833	1492
Reciprocity	0.8157	0.9263	0.6306	0.6093
Weighted Degree Assortativity	-0.0803	0.0181	0.0653	0.0984
Average Weighted Betweenness Centrality	0.024854	0.026584	0.001642	0.001149

TABLE 6.2: Comparison of network metrics across successful and unsuccessful groups of players in both league-level and club-level aggregated football transfer networks.

After analyzing the structure of the overall network, we now turn our attention to a comparison of the same network metrics, calculated for the subgraphs representing the career trajectories of successful and unsuccessful players in the football transfer market. Table 6.2 presents all the metric values, from which we can extract several meaningful insights about the structural differences between the two subsets.

First, it is important to note that, by setting the threshold for success at the 75th percentile, the number of successful players in the dataset is naturally lower than that of unsuccessful players. Specifically, there are 6,593 unique successful players, compared to 35,000 unsuccessful ones. As a consequence, the number of transfers included in each respective network differs: the successful player network contains 29,934 transfers, while the unsuccessful player network comprises 110,833.

Calculating the average number of transfers per player reveals that successful players tend to have more dynamic careers in terms of mobility, with an average of 4.540 transfers per player, compared to 3.167 for unsuccessful players.

Looking at the number of nodes and edges, we find that unsuccessful players have registered transfers involving all 50 leagues and all 1,531 clubs considered in the analysis. In contrast, successful players have registered transfers that involve only 1,028 clubs and 48 leagues, where the two leagues not included in the successful player network are the Estonian Premium Liiga and the Dutch Tweede Divisie. Furthermore, the number of unique transfer directions is higher in the unsuccessful player network for both league- and club-level aggregations. However, neither of the two networks reaches the total number of directions seen in the overall network. This suggests that some transfer trajectories, likely those between elite leagues and top-tier clubs, are strictly related to successful players, reinforcing our hypothesis that transfer behavior differs between player quality levels.

The average weighted in- and out-degrees show that, on average, clubs and leagues engage in a higher number of transfers involving unsuccessful players, a result that is expected given that they are five times more represented in the full dataset.

Network density and average clustering coefficient further confirm that the unsuccessful player network is more densely connected, with a higher tendency to form closed triads at both league and club levels, with metric values that are close to those observed in the overall network. In contrast, the successful player network has significantly lower density, indicating that successful players tend to follow more specific and selective transfer routes. Examining the largest strongly connected components (LSCCs), we see that all leagues are part of the LSCC in the unsuccessful player network. While, in the successful player network, the Irish Premier League is excluded due to only a single outgoing transfer to the English Premier League and no incoming transfers. At club level, the LSCC in the unsuccessful network includes most of the nodes, whereas the LSCC in the successful network omits a more substantial portion of clubs, further emphasizing the more constrained movement of successful players.

Reciprocity values reveal a high level of mutual transfer flows between leagues in both networks. Interestingly, the value is slightly lower for successful players, suggesting that some directional flows, likely from lower-tier to top-tier leagues, are not reciprocated. At club level, we observe the opposite: reciprocity is slightly higher in the successful player network, supporting the idea that clubs tend to establish stronger, more reciprocal relationships when it comes to trading successful players, possibly between clubs of similar stature aiming to optimize squad quality and between clubs of different status involved in the loan system.

Weighted degree assortativity is close to zero across all networks, indicating no substantial correlation in the degree of nodes that tend to trade with each other, either at league or club level.

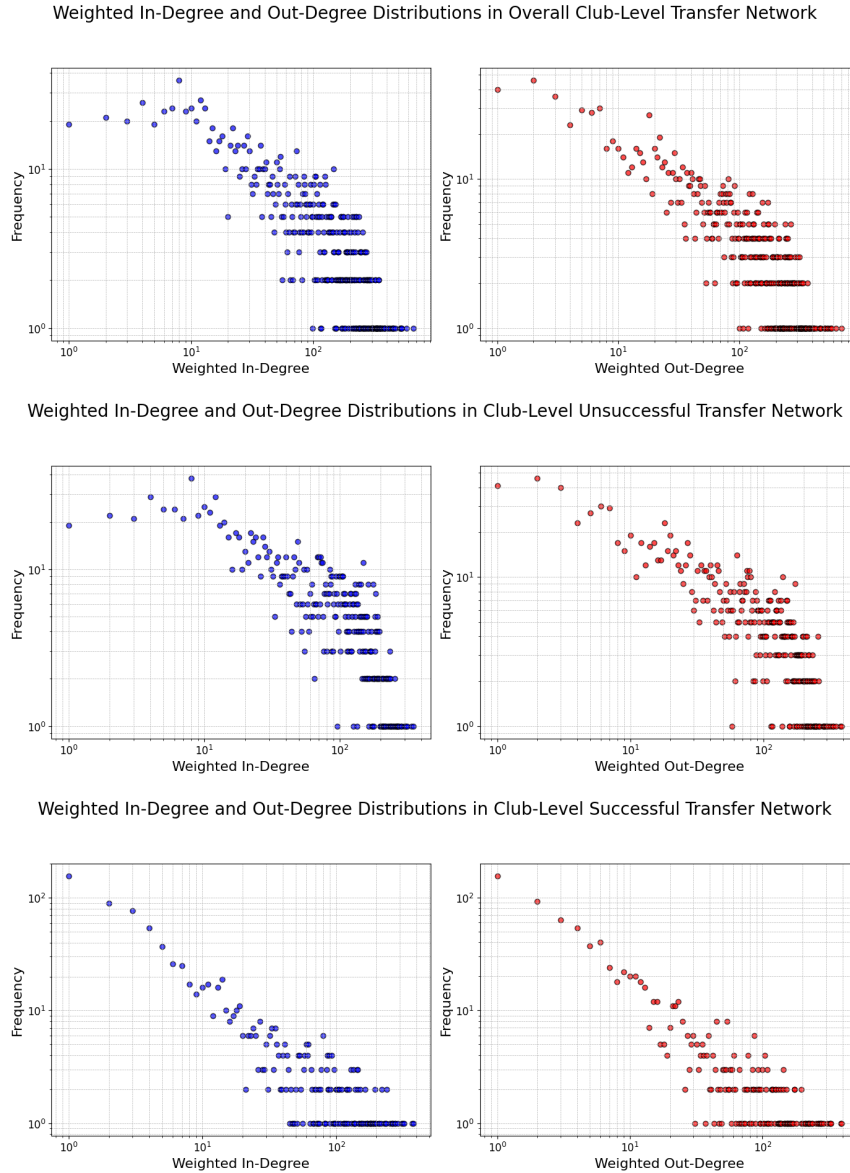


FIGURE 6.2: Comparison of In-Degree and Out-Degree distributions for the club-level transfer networks corresponding to the overall dataset, and the successful and unsuccessful player subsets.

Finally, average betweenness centrality does not provide strong differentiating insight: values remain relatively low and similar across both subsets, and are also close to those found in the overall network.

Figure 6.2 compares the In-Degree and Out-Degree distributions of club-level networks for the overall dataset and the two subsets of successful and unsuccessful players. From the scatterplots, we observe a clear similarity between the overall and unsuccessful player networks, while the distribution patterns for successful players differ more noticeably. In the successful player network, a relatively high number of nodes exhibit very low in-degree and out-degree values, compared to the degree distributions in the other networks. These nodes are likely lower-tier clubs that appear only rarely in the careers of successful players. Some may have served as launching pads from which players moved directly to higher-level clubs, explaining the high Out-Degree but low In-Degree occurrences, while others may have been final destinations for players at the end of their careers, reflected in

high In-Degree and low Out-Degree values.

Across all networks, the in-degree and out-degree distributions show similar shapes, indicating that there is no strong asymmetry in the transfer directionality. All three networks exhibit heavy-tailed, right-skewed distributions, with a few clubs acting as major hubs in the transfer market, while most clubs have mid-to-low degree values. However, in the successful player network, the prevalence of extreme degree values is slightly higher, suggesting a more polarized structure where a few elite clubs dominate activity while many others participate only occasionally.

6.1.2 Transfer Direction Analysis

After exploring the results obtained by the application of standard network metrics, we now focus on the analysis of transfer directions across the career trajectories of successful and unsuccessful players.

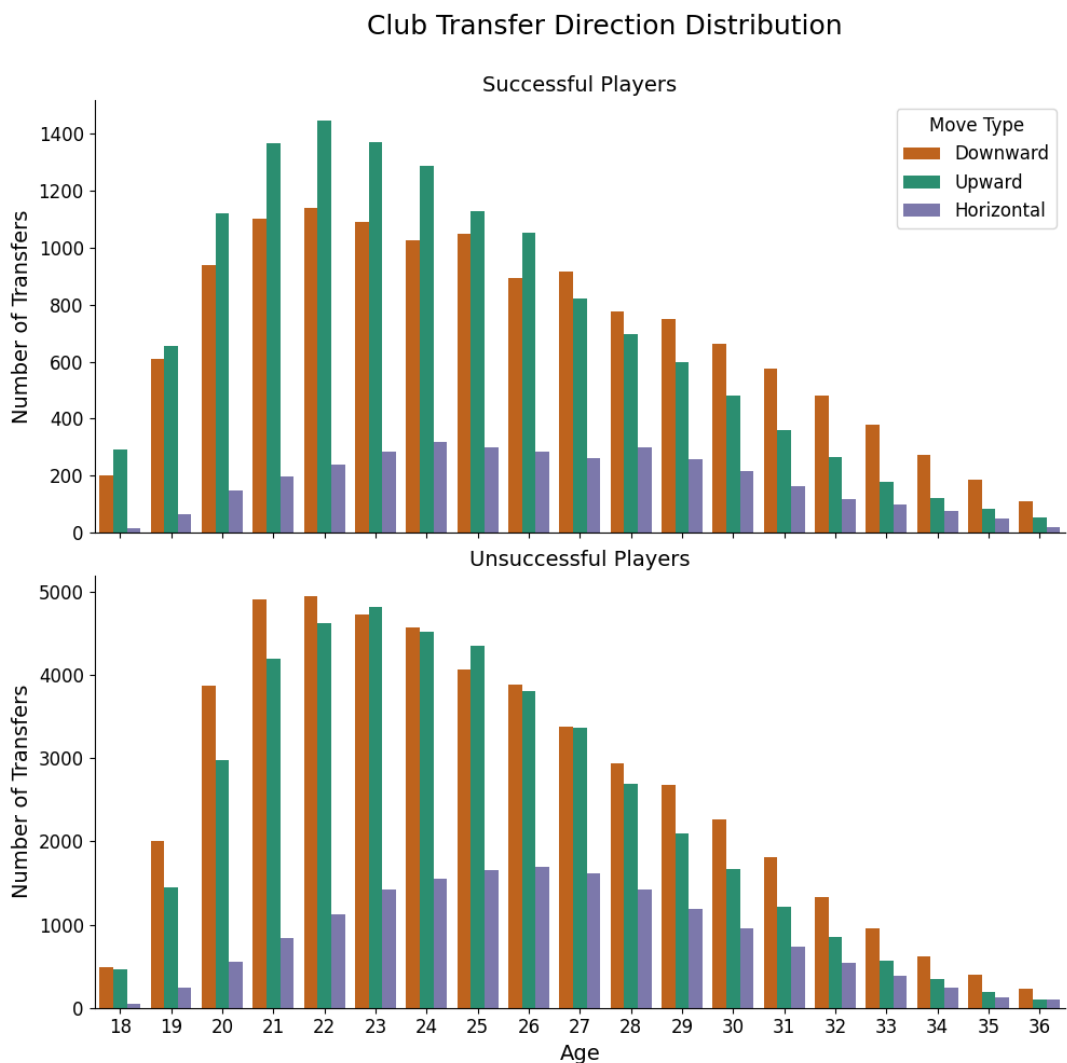


FIGURE 6.3: Distribution of club transfer directions over the careers of successful and unsuccessful players. Transfer classes are determined by comparing the average *transfermarkt* market value of the source and target clubs.

Figure 6.3 displays the distribution of 'Upward', 'Horizontal' and 'Downward' transfers based on the comparison between the average market values of the source and target clubs involved in each transfer.

For successful players, we observe a strong tendency to move to clubs with higher average market value at the beginning of their professional careers, indicating a progression toward more prestigious clubs during the developmental phase, particularly from the teenage years through their early to mid 20s. Around the age of 26, typically considered the peak of a player's career, the pattern reverses, with players increasingly transferring to clubs of lower average market value. This trend is likely due to natural aging and declining physical performance, leading to moves away from elite clubs to clubs of slightly lower status.

In contrast, unsuccessful players begin their careers with a marked tendency to transfer to lower-tier clubs. This may reflect their need to find an environment that matches their level of performance and offers sufficient playing time. This phase appears to stabilize around the ages of 23–24, at which point their career trajectories show a balance between upward and downward moves. After a few years, approximately around age 27–28, a decline begins, with players increasingly moving to lower level clubs, reflecting the aging-related decline seen in successful players, although from a different baseline.

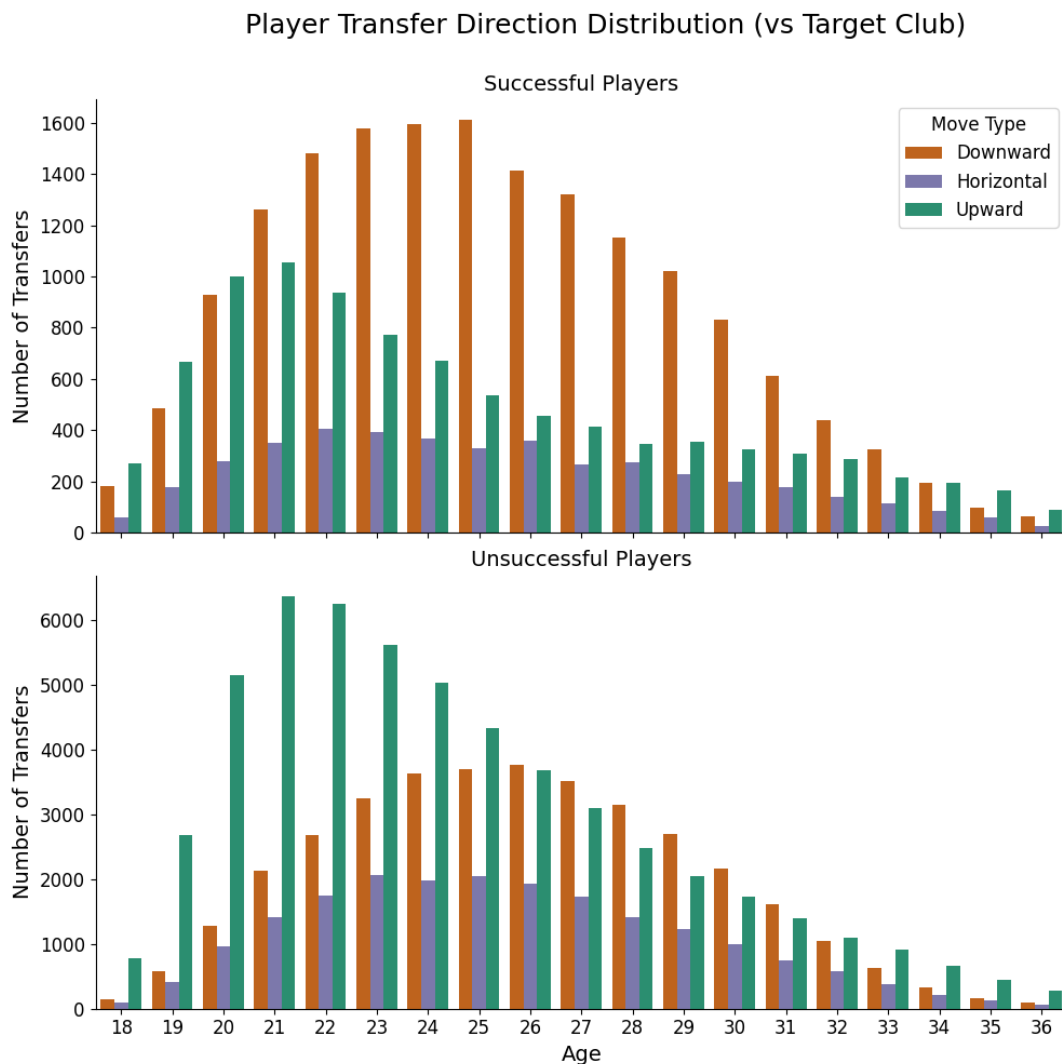


FIGURE 6.4: Distribution of player transfer directions based on the comparison between player market value and the average market value of the target club.

Horizontal transfers, characterized by moves between clubs of comparable average value, follow an approximately Gaussian distribution, especially among unsuccessful players, with a clear peak around age 26. This suggests that mid-career players tend to maintain their level, without significant upward or downward movement. However, among successful players, this peak is less pronounced, indicating a tendency to pursue more polarizing career moves, either upward or downward, rather than remaining at a consistent level during their prime.

Figure 6.4 shows the distribution of transfer directions through a different method of assessment: the comparison between a player's market value and the average market value of the target club.

Here, the career trajectories of successful and unsuccessful players diverge significantly. Successful players tend to move to clubs with higher average market value during the early stages of their careers, reflecting their ambition to join top-tier teams where they can further develop. However, around age 21, the frequency of 'upward' transfers for successful players declines significantly. A plausible interpretation is that, by this age, many of these players have already substantially increased their own market value. From that point onward, even transfers to a club with greater prestige may be classified as 'downward' by this metric, since the player's individual market value may already exceed the average market value of the destination club. This trend continues until about age 34, at which point their market value drops significantly due to aging and declining performance. Consequently, transfers beyond this point tend to be to clubs where the player falls below the average market value.

Unsuccessful players, on the other hand, display the opposite trend. For most of their development years, up until around age 27, they tend to move to clubs where their market value is lower than the average. Around their late 20s, when they typically reach peak performance, some players begin transferring to clubs where their market values result above the average, to challenge themselves in such contexts, though this phase has usually a limited duration. After age 32, their declining market value results in predominantly 'downward' moves.

Lastly, Figure 6.5 presents the transfer direction distribution based on the relationship between the player's market value and the average market value of the source club. In this context, an 'upward' transfer indicates that the current market value of the involved player is higher than the average market value of the source club (the club that the player is leaving).

This perspective further highlights the contrast between the two groups.

Successful players tend to begin their careers with market values lower than those of their source clubs, a natural consequence of their youth and inexperience. However, around age 21, they typically exceed the average market value of their current club. At this point, their transfers often reflect a search for more challenging environments: they leave clubs where they are already outperforming teammates looking for opportunities in higher-level teams. This behavior persists until around age 32, when their declining market value leads to more frequent moves away from clubs where they are below average.

Unsuccessful players follow a different path. They remain below the source club's average market value for a much longer period, probably due to slower development or limited impact on the pitch. Only around age 26 some of them begin to transfer away from clubs where they are above the average, indicating a brief window of relative success. However, by age 30, their value decreases again, and they return to a pattern of 'downward' transfers from clubs where they are below average.

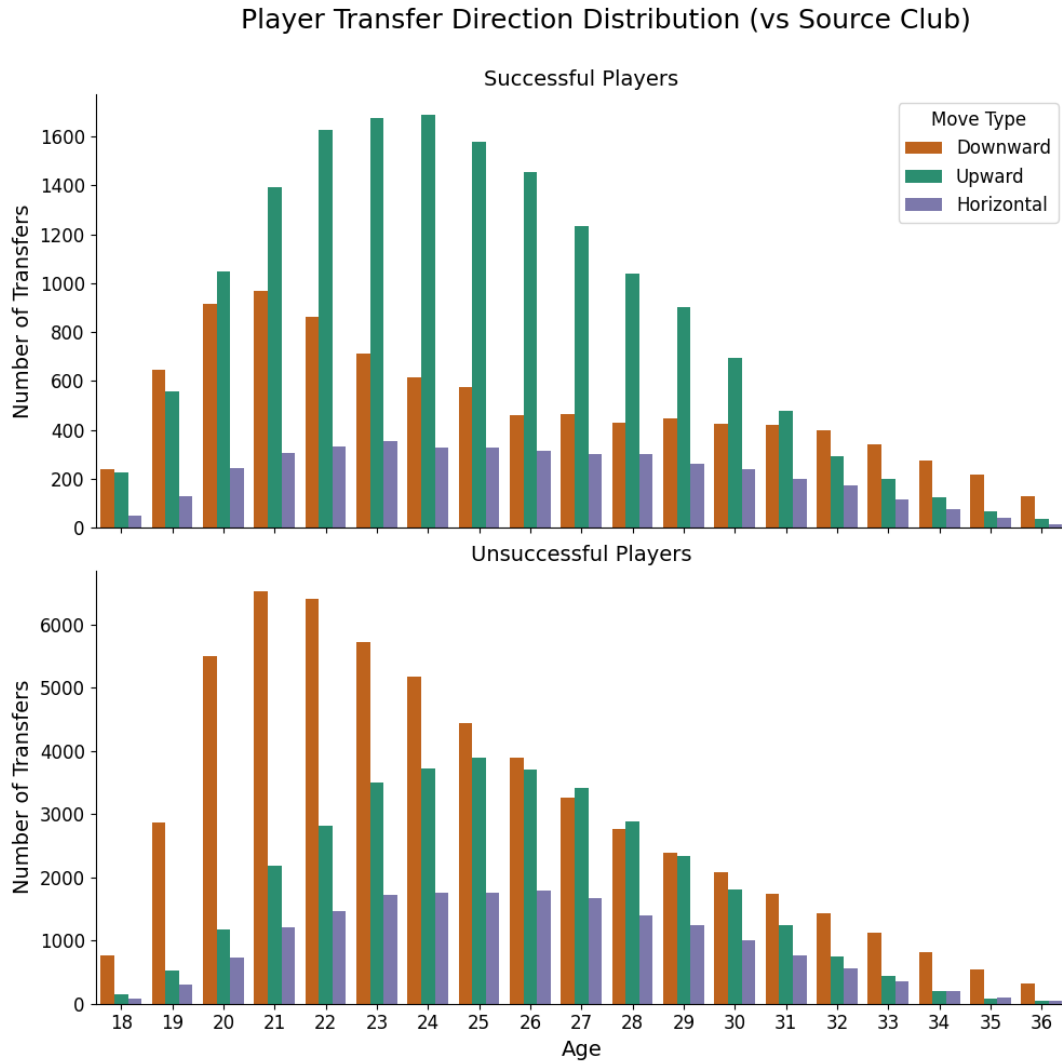


FIGURE 6.5: Distribution of player transfer directions based on the comparison between player market value and the average market value of the source club.

6.2 Meso-level

In this section, we present the results of our meso-level network analysis.

We first discuss the results related to the community detection, which was performed using the Louvain algorithm, applied to an undirected weighted version of the successful and unsuccessful player networks, to ensure compatibility with the method, where edge weights correspond to transfer counts and the resolution parameter was set to $\gamma = 1.0$ by default. To compare the partitions obtained for the successful and unsuccessful player networks, we performed a qualitative inspection of them, focusing on which major football leagues appeared consistently in both networks.

Subsequently, we analyze the transfer flows between football leagues at varying levels of aggregation, comparing how these differ across the two player career types.

6.2.1 Successful Player Network Communities

Figure 6.6 shows the community structure detected by applying the Louvain algorithm to the successful player network. The method identified 8 communities, where clubs tend to

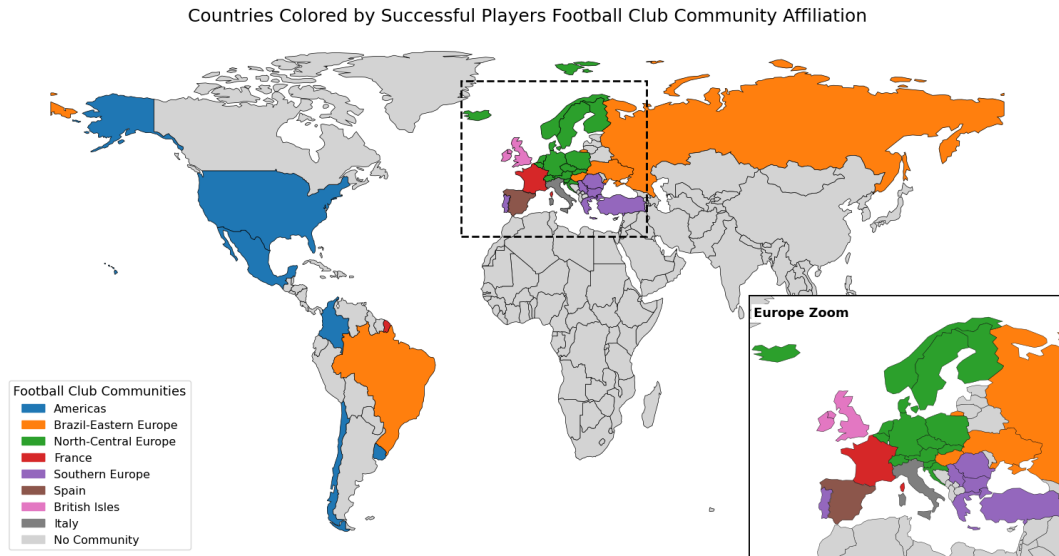


FIGURE 6.6: World countries assigned to their corresponding community based on the Louvain algorithm applied to the successful football player network. Each country is assigned to the community that includes the majority of its clubs.

exchange players more frequently within their group than with clubs outside of it. Since most clubs within the same league and all leagues within the same country were grouped together, we visualized the results at country level, coloring each country according to the community to which it belongs.

Several interesting patterns emerge from this community partitioning.

First, three of the top-5 European leagues, namely France, Spain and Italy, form individual communities that also include their second divisions. This indicates that a substantial portion of successful player transfers involving these leagues occur within national boundaries, particularly between the top and second divisions.

Another clearly defined cluster is formed by the British Isles, including England, Scotland and Ireland. Beyond geographic proximity, these countries share the same language and several cultural traits, likely facilitating both club-to-club trading interactions and player adaptation.

The largest identified community covers northern and central Europe. This cluster includes Germany, the Netherlands, Belgium, Denmark, Finland, Sweden, Norway, Iceland, Austria, Switzerland, Slovenia, Croatia, Czech Republic, Slovakia and Poland. These countries exhibit strong mutual transfer activity, with Germany functioning as a central hub. Talents from smaller leagues often move to intermediate leagues (like the Netherlands, Belgium or Denmark) before making a final step to top-tier German clubs.

A particularly notable community includes Portugal alongside several Southeastern European countries: Serbia, Romania, Bulgaria, Greece and Turkey. Although Portugal may not seem a natural fit in this group at first, the data reveal a strong exchange of successful players between Portuguese clubs and teams in Greece, Turkey and Romania. Successful players from these countries often use Portuguese clubs as intermediate steps before moving on to Europe's top-5 leagues. Moreover, both Turkey and Portugal are established as global hubs in the football transfer network, acting as bridges between elite European leagues and the broader international periphery, and since they also have a high trading activity between each other, the final result of having Portugal as part of this community acquires even more value.

Another surprising yet meaningful community combines Brazil with eastern European countries, especially Russia and Ukraine. Despite the geographical distance, this connection reflects long-standing football transfer patterns, Brazilian players frequently join Russian or Ukrainian clubs (such as in the well-known cases of Zenit and Shakhtar Donetsk) before advancing to top-tier European teams.

Finally, a distinct community emerges among American countries; Mexico, the United States, Colombia, Chile and Uruguay. In addition to geographical proximity, most of these countries (except the USA) share a common language and cultural background, which facilitates player integration. Furthermore, Mexico and the US function as gateways to greater visibility for South American talents aiming to gather the attention and interest of European clubs.

The modularity score for this partition is 0.3922, which is relatively low. However, modularity tends to underestimate the quality of communities in large, complex networks such as this. Notably, the average Normalized Mutual Information (NMI) over 20 Louvain runs results 0.8472, indicating the high robustness and consistency of the detected communities even at club level.

6.2.2 Unsuccessful Player Network Communities

Figure 6.7 shows the partition of the communities for the unsuccessful player network. As in the previous case, the Louvain algorithm was applied at club level and then aggregated at country level. The vast majority of clubs within the same country are grouped into the same community, supporting this aggregation approach.

Here, the algorithm identified 13 communities, suggesting a more fragmented network structure for unsuccessful player careers, with transfers occurring predominantly within

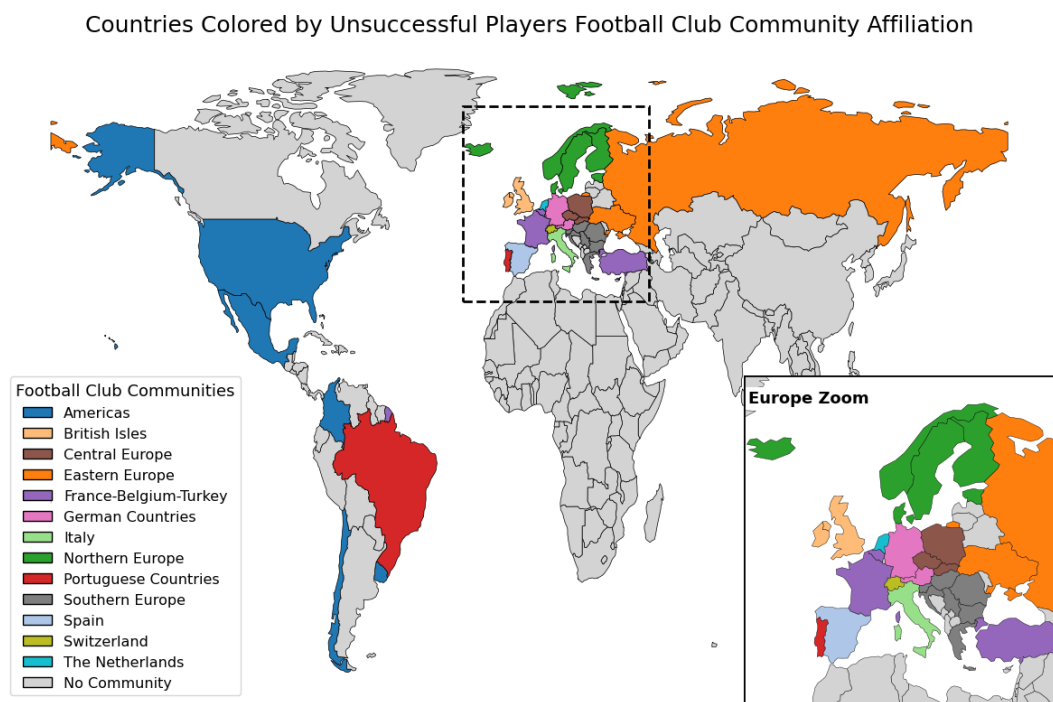


FIGURE 6.7: World countries assigned to their corresponding community based on the Louvain algorithm applied to the unsuccessful football player network. Each country is assigned to the community that includes the majority of its clubs.

smaller local clusters. In several cases, a single country forms an entire community, as seen with Italy, Spain, the Netherlands and Switzerland. This implies that unsuccessful players often spend most of their careers within domestic leagues.

As with the successful player network, we find the British Isles and American countries forming separate, cohesive communities, further reinforcing the role of geography, language and culture in shaping transfer behaviors.

Interestingly, the Eastern Europe community now only includes Russia and Ukraine, unlike the successful network where it was linked to Brazil. This suggests that while successful Brazilian players often use Russia and Ukraine as a gateway to access the main European leagues, such transfers are less common or less appealing, for players whose careers are labeled as unsuccessful.

Several previously unified communities in the successful network now appear fragmented. For example:

- Germany and Austria form a separate "German-speaking" cluster.
- A Northern European cluster includes Iceland, Denmark, Norway, Sweden, Finland and Estonia.
- A Central Europe community is represented by Poland, Czech Republic and Slovakia.
- Southeastern Europe (Slovenia, Croatia, Serbia, Hungary, Romania, Bulgaria and Greece) also forms its own localized group.

These smaller, culturally cohesive clusters indicate that unsuccessful players are less likely to take ambitious steps abroad, and more likely to stay within culturally familiar environments. This suggests that career success may require not only talent, but also the willingness to leave comfort zones and pursue challenging opportunities in unfamiliar leagues.

Two additional communities offer further insights into the trajectories of unsuccessful players. The "Portuguese countries" cluster links Brazil and Portugal, revealing a strong connection between these linguistically connected nations. Interestingly, this connection is more consistent among unsuccessful players compared to successful players, with Brazilian players who show to have more success, on average, by accessing Europe from Russian and Ukrainian clubs, rather than Portuguese ones.

The final community includes France, Belgium and Turkey. While France and Belgium have clear cultural and linguistic ties, the presence of Turkey may seem less intuitive. However, Turkey maintains strong football connections with both countries, especially France. Transfers are facilitated by the large Turkish presence in France (Turkish people represent the fifth most present population in the country) and by Turkey's strategic role as a gateway to the main European leagues. In the opposite direction, for unsuccessful players from France, Turkey and Belgium often represent possibilities to move to clubs where they can increase their playing time at a lower but still competitive level.

This partition retrieves a modularity score of 0.6330, significantly higher than in the successful player network. This suggests a stronger internal cohesion within communities and fewer transfers between clubs belonging to different communities. Furthermore, the average NMI over 20 runs of the Louvain algorithm is exceptionally high: 0.9922. This indicates extremely robust and stable community structures at both club and country levels.

6.2.3 Transfer Flows

In this section, we discuss the results of the analysis of transfer flows between national leagues for both successful and unsuccessful players.

Figure 6.8 presents the aggregated movements of successful players (6.8a) and unsuccessful (6.8b) players. To make a fair comparison, we set different thresholds for inclusion in the visualization: a minimum of 36 transfers for the successful network and 133 transfers for the unsuccessful one. These thresholds were selected to maintain a consistent ratio between the number of transfers in the two networks, with reference to the original unfiltered dataset, as the sizes and distributions of the two subsets differ significantly. This choice allows for a comparable number of leagues and connections in the two graphs.

In Figure 6.8a, we observe that only 29 of the 48 leagues in the successful player network participate in at least 36 successful player transfers with another league. Among these, Croatia and the Czech Republic mainly exchange successful players within their own domestic leagues, with limited interaction with foreign leagues. Other lower-tier leagues are involved in transfers mainly within their local ecosystem or along well-established historical pathways; for example, Austria and Switzerland with Germany, Russia and Ukraine with Brazil, and Scotland with England and France. Outside of Europe, Mexico and the USA connect strongly with Spain and England. These patterns reflect strong geographical, linguistic and cultural connections that facilitate international player movement.

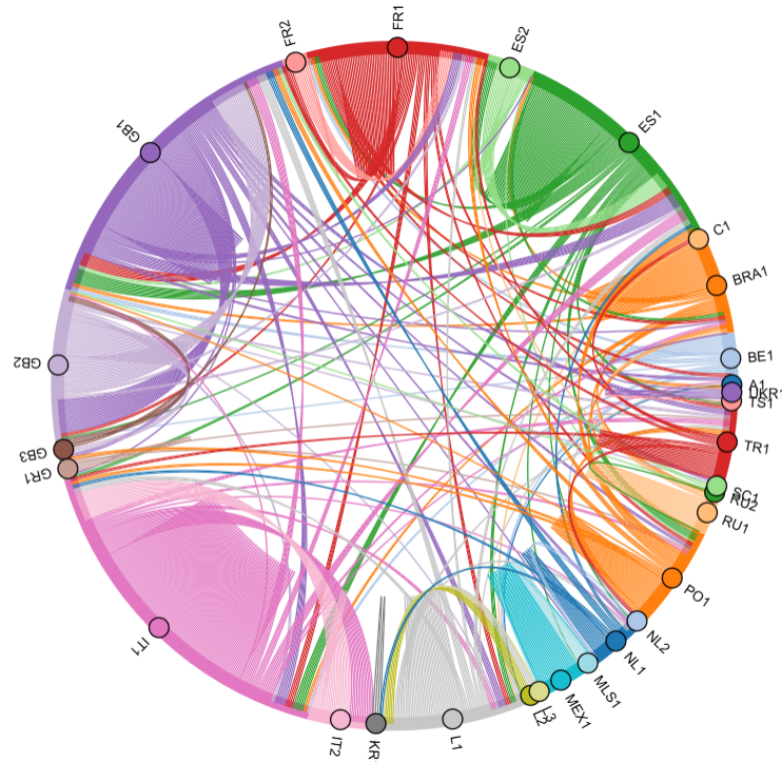
Several mid-tier leagues such as Brazil, Belgium, Greece, Turkey, Portugal and the Netherlands appear as key intermediaries in the development of successful players' careers, exporting talent to a variety of higher-level destinations. At the top of the hierarchy, the top five European leagues dominate the successful player trading network, with a high volume of internal transfers, robust connections to each other, and strong connections to intermediate leagues that play a key role as successful players' suppliers.

An important pattern involves national second and third divisions, which engage in consistent trading with their own top-tier divisions. This suggests a strong local internal circulation of successful players. The Italian Serie A has the highest number of local transfers involving successful players, while the English Premier League registers the most successful international exchanges.

Turning to Figure 6.8b, we find that almost all leagues in the unsuccessful player network participate in at least 133 transfers with another league, with the exception of the Dutch Tweede Divisie. The vast majority of these transfers occur either within the same league or between the top and lower domestic divisions of the same country. This includes not only lower-tier leagues from various continents, but also three of the top five European leagues: Italy, Spain and Germany. These same leagues, which are also associated with standalone communities according to the results of the Louvain algorithm, exhibit little international activity for the trading of unsuccessful players, emphasizing the strong domestic focus of their player circulation at this performance level.

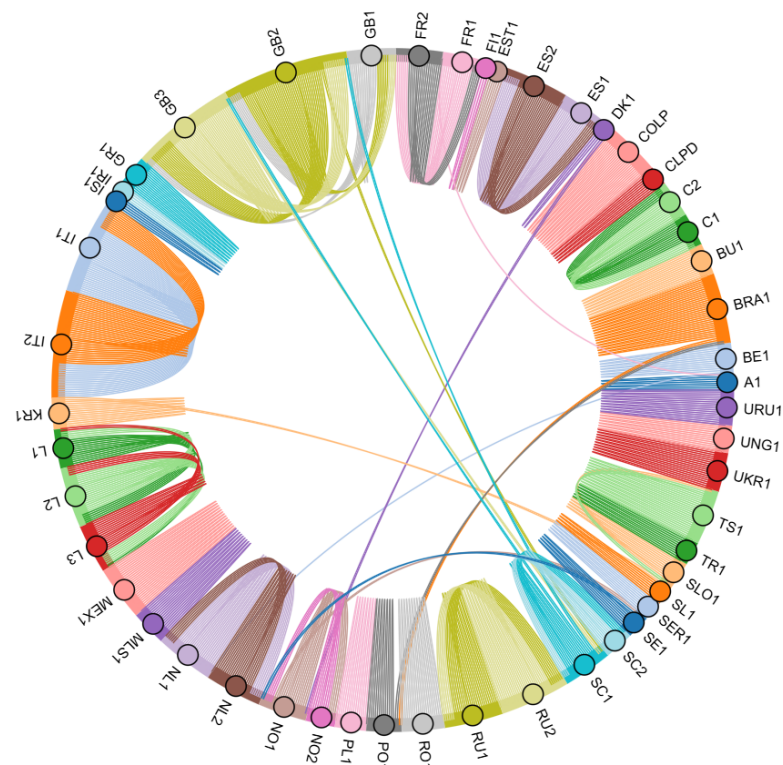
Cross-border flows of unsuccessful players are limited and mostly follow predictable patterns based on geography or shared history. Notable connections include England–Scotland, Czech Republic–Slovakia and Slovenia–Croatia, all of which involve countries with shared pasts or ongoing cultural ties. Other connections, such as France–Belgium–Netherlands and Denmark–Norway–Sweden, rely on regional proximity and cultural similarity. Finally, Brazil and Portugal show a language-based transfer relationship, especially regarding the trading of unsuccessful players.

Successful Player Transfers between National Leagues (over 36 transfers)



(A) Successful Player Network (minimum 36 transfers)

Unsuccessful Player Transfers between National Leagues (over 133 transfers)



(B) Unsuccessful Player Network (minimum 133 transfers)

FIGURE 6.8: Comparison between the transfer flows of successful and unsuccessful players among national leagues, using a low threshold.

Comparing the two graphs in Figure 6.8, we observe distinct structural patterns. Transfers involving successful players are more centralized and international, often circulating among the same relatively small subset of leagues. These leagues, especially the European top five, engage in dense interconnections and have access to wider talent pools through the connections with mid-tier leagues that tend to export successful players. In contrast, transfers involving unsuccessful players are widespread and localized, spanning almost all leagues evenly but rarely crossing borders unless strong cultural ties exist.

In Appendix B we further analyze the transfer flows that characterize successful and unsuccessful player networks, when using higher thresholds. The results support the highly localized nature of unsuccessful player trajectories, where national-level circulation is predominant and almost no significant foreign flows beyond the minimum threshold are present.

6.3 Micro-level

This section presents the results of the micro-level analysis of the network, focusing on uncovering recurrent and successful patterns involving sequences of two and three nodes. We begin by analyzing the most overrepresented transfers among successful career paths using 2-node and 3-node patterns. Then, we identify the most influential clubs based on the ClubRank algorithm and compare them with rankings derived from various centrality measures.

6.3.1 2-Node Pattern Analysis

By applying the methodologies described previously to the probabilistic network representations, we identify the most overrepresented transfers within the successful player network, compared to the overall network. These analyses were conducted in both directions of transfers, allowing us to determine, for each node, both the **optimal source** (the origin league or club from which transfers are most strongly associated with successful football player careers) and the **optimal target** (the destination league or club to which transfers are most strongly associated with successful football player careers). This methodology was applied at both league and club levels, and in this section, we present results related to a subset of those leagues and clubs.

Table 6.3 displays the optimal source and target leagues for the top 40 leagues in the network, selected based on the volume of successful player transfers. The rankings are derived from Laplace-smoothed log-odds differences, shown in parentheses. High values of these differences indicate a clear overrepresentation of the corresponding transfers among successful player careers. As such, they can serve as useful indicators for football players when making decisions about the next steps in their careers.

In Table 6.3, only the most overrepresented transfer for each league is reported, based on successful career paths. However, the complete ranking for each league was computed, which may serve as a valuable tool to analyze potential destinations interested in acquiring specific players.

Unsurprisingly, the most overrepresented leagues among successful players, both as sources and targets, are also the most prestigious in the world. Among the top 40 leagues listed, 33 of the top destinations and 25 of the top source leagues belong to the top five European leagues. In particular, the English Premier League emerges as the most prestigious league, where most successful players spend part of their careers. It appears as the top destination for 18 leagues and the top origin for 15 leagues. Furthermore, it is considered the best

trading partner in both directions by all other top five leagues, reinforcing its status as the best football league in the world.

Interestingly, the most common source of successful players arriving in the English Premier League is represented by the Italian Serie A, while the most common destination for successful players leaving the English Premier League is the German Bundesliga. Some lower-tier leagues exhibit lower or even negative log-odds difference values (reported in italic and with a star in the table), indicating that although the displayed leagues are still the top transfer partners for their respective league, the transfers are not particularly overrepresented among successful paths, they are more present in successful player career trajectories, but they also figure in unsuccessful player careers. This outcome is influenced by the robustness filter applied in the analysis, which considers only transfers that have occurred at least seven times among successful players. Consequently, for leagues with a small number of recorded successful transfers, the top transfers may simply result as those that passed the threshold, meaning these results should be interpreted with caution due to limited data.

Beyond the top five European leagues, several leagues act as important intermediate steps in a successful career trajectory. Countries such as Chile, Czech Republic, Denmark, the Netherlands, Sweden, and Switzerland often serve as scouting grounds for elite European clubs, playing a crucial transitional role in a player's development.

Top Source (Log)	League	Top Target (Log)	Top Source (Log)	League	Top Target (Log)
ITA1 (1.432)	ENG1	GER1 (1.430)	<i>POL1 (-2.267)*</i>	POL1	GER1 (0.801)
RUS1 (1.155)	ENG2	FRA1 (1.183)	BRA1 (1.161)	DEN1	ENG1 (0.836)
<i>ENG1 (-0.928)*</i>	ENG3	FRA1 (1.288)	ENG1 (1.094)	AUT1	ENG1 (1.193)
ENG1 (1.415)	ITA1	ENG1 (1.432)	TUR1 (1.018)	SWI1	ESP1 (1.173)
FRA1 (1.067)	ITA2	ESP1 (1.152)	<i>SWI1 (-3.058)*</i>	SWI2	<i>SWI1 (-2.400)*</i>
ENG1 (1.417)	ESP1	ENG1 (1.410)	FRA1 (0.864)	SCO1	GER1 (1.193)
ENG1 (1.044)	ESP2	ENG1 (1.264)	RUS1 (0.887)	SWE1	ENG1 (1.067)
ENG1 (1.430)	GER1	ENG1 (1.427)	ENG1 (1.219)	HRV1	ENG1 (1.257)
POR1 (0.941)	GER2	ESP1 (1.124)	ITA1 (0.171)	SER1	ENG1 (1.124)
<i>GER1 (-1.966)*</i>	GER3	<i>GER1 (-0.566)*</i>	<i>POR1 (-0.678)*</i>	ROM1	TUR1 (0.118)
ENG1 (1.408)	FRA1	ENG1 (1.421)	ENG1 (1.150)	UKR1	ENG1 (1.267)
ENG1 (0.949)	FRA2	FRA1 (1.193)	ENG1 (1.259)	RUS1	ENG1 (1.301)
MEX1 (1.434)	NTL1	ENG1 (1.180)	<i>RUS1 (-2.940)*</i>	RUS2	<i>RUS1 (-1.483)*</i>
<i>NTL1 (-2.679)*</i>	NTL2	<i>NTL1 (-1.083)*</i>	<i>BUL1 (-3.458)*</i>	BUL1	<i>BUL1 (-3.458)*</i>
ENG1 (1.213)	POR1	ENG1 (1.283)	ENG1 (1.292)	BRA1	ENG1 (1.379)
ENG1 (1.219)	BEL1	ENG1 (1.243)	BRA1 (0.398)	CHL1	ESP1 (0.635)
ENG1 (1.295)	TUR1	ENG1 (1.339)	<i>USA1 (-0.351)*</i>	COL1	<i>USA1 (-0.206)*</i>
BEL1 (0.359)	CZE1	GER1 (0.511)	ITA1 (1.261)	MEX1	NTL1 (1.434)
ENG1 (1.117)	GRE1	ENG1 (1.206)	ESP1 (0.507)	URU1	ESP1 (0.810)
<i>DEN1 (-1.283)*</i>	NOR1	ITA1 (0.787)	ESP1 (1.005)	USA1	GER1 (1.242)

TABLE 6.3: Top Source and Target leagues for the top-40 leagues, based on the highest amount of successful player transfers, according to Laplace-smoothed log-odds differences. The European leagues are ordered according to UEFA coefficients for countries in 2025, while the international leagues are ordered alphabetically. Negative values (not strongly overrepresented) are shown in italic, and they present a star. The top 5 European leagues are indicated in bold.

Another important observation concerns the second and third divisions. Contrary to the intuitive expectation that the recommended path to success would be represented by staying within the same country and advance to the top local division, the results show that successful transfers frequently involve international moves. That is more often correlated with success, for players in lower divisions which move abroad, rather than upward within the domestic leagues.

We now move to club-level aggregation results. Table 6.4 shows the top source and target clubs for each of the top 40 clubs, based on the 2025 UEFA coefficients. The smoothed log-odds differences, reported in brackets, motivate these rankings.

As expected, these values are generally lower than those observed at the league level due to the lower frequency of direct transfers between individual clubs. Nonetheless, all values, except for the top source for AZ Alkmaar, are positive, confirming that the identified transfers are overrepresented among successful players. Additionally, each transfer pair has occurred at least four times, meeting the robustness threshold.

These findings can assist players in identifying the most suitable destination clubs, based on the club where they are currently playing. As in the league-level analysis, while Table 6.4 shows only the top source and target for each club, full rankings are available.

From Table 6.4, several meaningful insights emerge. For example, some clubs, such as AC Milan, appear frequently as top destinations but never as top sources. This suggests that they act as final destinations or "arrival points" in many successful player careers.

For most Italian clubs, excluding AC Milan and Juventus, there is a noticeable preference for domestic transfers among successful players, with both their top sources and destinations being other Italian clubs.

Several clubs show strong bilateral transfer relationships, in which the same club appears as both the top source and the top target. These partnerships may be between clubs of similar level, involving the exchange of elite players, or hierarchical, where a higher-level club loans players to a lower-level club to benefit of their development before retrieving them. Prominent examples include: Inter Milan and AS Roma, Inter Milan and Napoli, AC Milan and Chelsea, PSV Eindhoven and Chelsea, Liverpool and West Ham, PSG and Eintracht Frankfurt, Feyenoord and Utrecht, and Club Brugge and Charleroi.

Another point of interest regards the ratio of top clubs which prefer local over international trading. Of the 40 top source clubs, 22 are domestic (local) to their associated clubs. Similarly, 21 of the top target clubs are local. This suggests a relatively balanced preference between local and international markets among elite clubs.

However, it is important to note that this table focuses exclusively on top-tier clubs. The trading behavior of lower-tier clubs may differ significantly, as their players are more likely to pursue moves to higher-level teams abroad to advance their careers. In contrast, successful players who already play at elite clubs may have less concern regarding geographic movement, since they are likely mainly interested in competing in international top competitions regardless of location.

In general, club-trading strategies vary widely. Some clubs aim to buy already established, successful players from elite clubs, either through high investments or opportunistic purchases of talents that are not satisfied at their current club; others focus on scouting rising stars from top clubs in foreign leagues, and others mainly focus on acquiring the talents from local lower-tier teams. Similarly, selling strategies differ: some clubs sell players at their peak to global elite clubs, while others maintain steady local relationships that lead them to sell their talents to the local bigger clubs or other clubs aim to only sell aging players to less competitive leagues, once they have enjoyed their performance in their prime. These strategies depend on a combination of factors such as financial resources, scouting capabilities and ambition. Although these variables are not directly captured in the data

Top Source (Log)	Club	Top Target (Log)
Chelsea (0.082)	Real Madrid	AC Milan (0.122)
Valencia (0.282)	Manchester City	Middlesbrough (0.290)
Shalke 04 (0.224)	Bayern Munchen	Borussia Dortmund (0.312)
AS Roma (0.257)	Liverpool	Cardiff City (0.287)
Benfica (0.271)	PSG	AS Roma (0.387)
AS Roma (0.362)	Inter Milan	AS Roma (0.394)
Reading FC (0.182)	Chelsea	AC Milan (0.183)
Bayern Munich (0.454)	Borussia Dortmund	Real Madrid (0.529)
Genoa (0.410)	AS Roma	Inter Milan (0.460)
Valencia (0.110)	Barcelona	Sevilla (0.117)
Crystal Palace (0.229)	Manchester United	West Ham (0.284)
Chelsea (0.267)	Arsenal	Barcelona (0.318)
AS Roma (0.312)	Bayer Leverkusen	Eintracht Frankfurt (0.187)
Chelsea (0.152)	Atletico Madrid	Deportivo de La Coruña (0.180)
Atlético Madrid (0.382)	Benfica	Valencia (0.401)
Monza (0.770)	Atalanta	Udinese (0.818)
Tottenham (0.314)	Villareal	AC Milan (0.397)
São Paulo (0.382)	Porto	Inter Milan (0.406)
Chelsea (0.463)	AC Milan	Chelsea (0.499)
PSG (0.447)	RB Leipzig	AS Roma (0.636)
Inter Milan (0.683)	Lazio	Fiorentina (0.806)
Fiorentina (0.488)	Juventus	Bayern Munich (0.524)
PSG (0.494)	Eintracht Frankfurt	PSG (0.666)
Charleroi (0.440)	Club Brugge	Charleroi (0.598)
Liverpool (0.944)	Glasgow Rangers	Birmingham City (0.988)
Utrecht (0.444)	Feyenoord	Utrecht (0.624)
Swansea City (0.312)	Tottenham	Everton (0.374)
Chelsea (0.533)	PSV Eindhoven	Chelsea (0.587)
Liverpool (0.487)	West Ham	Liverpool (0.574)
Barcelona (0.472)	Ajax	Sevilla (0.490)
Auxerre (0.450)	Lille	Olympique Lyon (0.527)
Real Madrid (0.565)	Real Sociedad	Real Betis (0.639)
AS Roma (0.477)	Fiorentina	Juventus (0.599)
Inter Milan (0.506)	Napoli	Inter Milan (0.562)
Barcelona (0.371)	Sporting CP	Legia Warszawa (0.355)
Barcelona (0.271)	Sevilla	AC Milan (0.357)
AS Monaco (0.769)	Olympiakos	Porto (0.770)
Sporting CP (1.384)	Dinamo Zagreb	Inter Milan (1.371)
<i>Sparta Rotterdam (-0.041)*</i>	AZ Alkmaar	Ajax (0.936)
Real Sociedad (0.506)	Real Betis	Atlético Madrid (0.618)

TABLE 6.4: Top Source and Target clubs for each club in the top-40 based on UEFA coefficients in 2025, according to Laplace-smoothed log-odds differences (values in brackets). Negative values are shown in italic and indicated with a star.

reported in the table, they remain crucial for a more in-depth understanding of football transfer market dynamics.

6.3.2 3-Node Pattern Analysis

After discussing the results related to 2-node patterns, we now turn to the analysis of 3-node patterns with the objective of identifying the most successful motifs at both league and club aggregation levels.

Figure 6.9 illustrates the league-level aggregated results, showing the distribution of 3-node motifs, defined as two consecutive transfers between leagues, based on the smoothed log-odds differences between successful player motifs and average motifs, as well as the number of motifs related to successful players.

The distribution reveals a large number of motifs that occur only a few times among successful players, alongside a smaller subset of motifs that occur frequently. Some of these are significantly overrepresented among successful players, while a vast number are underrepresented, implying a stronger presence in the career paths of less successful players. One of the most insightful observations from this visualization concerns the distribution of motif types. As detailed in Section 5.3, to better analyze the complex behavior of football players in the football transfer market, we classified the 3-node motifs into four different categories (chains, loops, partial repetitions and triple repetitions) based on the sequence of nodes included in the motif. For example, the red dots, representing the *triple repeat* class, correspond to many of the extreme values in this distribution. These motifs are among the most underrepresented when they involve lower-tier leagues, such as the second divisions in Germany and Russia, or national leagues from Bulgaria, Czech Republic, Hungary, Romania, Ukraine, Uruguay, Chile and Colombia. In contrast, they are the most frequently occurring and overrepresented motifs among successful players when involving two consecutive transfers within one of the top five European leagues. In particular, two consecutive transfers within the Italian Serie A represent the most frequent

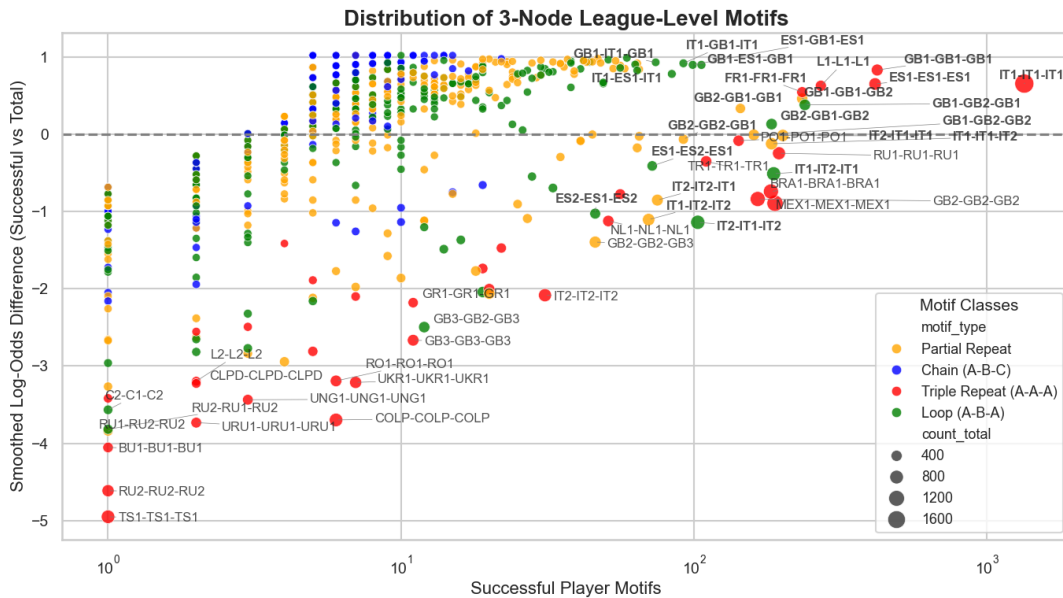


FIGURE 6.9: Distribution of 3-node motifs aggregated by leagues over smoothed log-odds differences and count of motifs related to successful players. Colors indicate motif types, while sizes reflect the total number of such motifs present in the dataset. Motifs including the top 5 European leagues are reported in bold.

motif among successful player careers, while two consecutive transfers within the English Premier League are the most overrepresented motif in the successful subset out of the *triple repeat* cluster.

Another interesting motif type is the *loop* cluster. The motifs of this class are significantly underrepresented among successful players when involving lower-tier leagues, such as oscillations between the first and second divisions in Russia or Switzerland. Somewhat higher counts (though still underrepresented) are seen in Spanish and Italian leagues. However, the same pattern involving English clubs is associated with successful careers. Among motifs with more than 70 occurrences in successful player careers, the most overrepresented patterns are all loops, particularly regarding oscillations between English, Italian and Spanish clubs, which frequently feature in successful trajectories.

The *partial repeat* class represents another category with notable trends. It is generally underrepresented when involving multiple steps in lower-level leagues but becomes overrepresented when it includes repeated steps in elite European leagues.

Chain movements are the least frequent among the 3-node motif classes. However, they display similar patterns to other motif types: they are underrepresented among successful players when involving low-tier leagues and overrepresented when composed of transfers between different top European leagues.

Figure 6.10 presents a similar visualization for club-level aggregation, where each 3-node motif represents two consecutive transfers between clubs.

Since self-edges are not considered in this context, only *loop* and *chain* motifs appear in this analysis. The number of 3-node motifs that meet the thresholds of at least five occurrences in total or at least three among successful players is significantly lower than in the league-level scenario. In fact, only six motifs occur more than six times among successful players.

Loop transfers are substantially more common than *chain* transfers. This is largely due to the consistent presence of loan arrangements in which a player spends a short period of time at another club to gain playing time before returning to the original club.

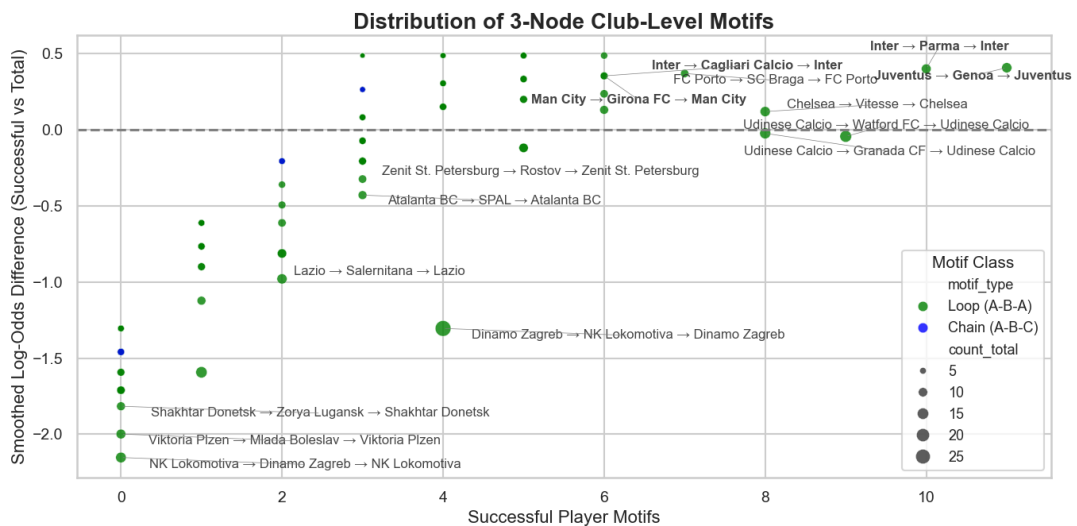


FIGURE 6.10: Distribution of 3-node motifs aggregated by clubs over smoothed log-odds differences and the count of motifs related to successful players. Colors indicate motif types, while sizes reflect the total number of such motifs present in the dataset. Motifs including only clubs that are currently participating in the top 5 European leagues are reported in bold.

The most underrepresented motifs among successful players are loop transfers involving clubs from lower-tier leagues, such as Croatian, Czech and Ukrainian clubs. While loop transfers involving clubs in the Italian second division and Russia show higher smoothed log-odds scores, they remain underrepresented in the successful player subset.

A particularly interesting case is represented by the player exchanges between Udinese, Watford, and Granada; three football clubs located in different countries that share the same ownership. Despite a high number of interactions, these motifs show smoothed log-odds values close to zero, suggesting no clear correlation with either successful or unsuccessful career paths.

However, the motifs overrepresented in successful player careers are mainly composed of structured trading relationships between high-level and mid-level clubs. These often involve subordinate loan arrangements in which the elite club loans young promising players to the lower-tier club for development, before reintegrating them into the team. At national level, this is exemplified by relationships such as Juventus - Genoa, Inter - Parma, Inter - Cagliari and Porto - Braga. At international level, such partnerships are less common, but they do exist; most notably between Chelsea and Vitesse, and between Manchester City and Girona, with the latter pair of clubs which also share the ownership.

6.3.3 ClubRank

The final results we present in this section concern the application of the ClubRank algorithm to the outgoing probabilistic transfer network. The method inspired by PageRank leads to a ClubRank score for each club, calculated separately for the networks of successful and unsuccessful players.

Table 6.5 reports the rankings of the 25 most influential clubs, according to ClubRank, in the career trajectories of successful (left) and unsuccessful (right) players.

A first notable observation is the disparity in the magnitude of the ClubRank scores. The clubs in the successful network exhibit substantially higher values than those in the unsuccessful network. This difference arises from the structural characteristics of the two networks: the successful player network is smaller and more centralized. Consequently, the clubs listed on the left side of Table 6.5 are more frequently involved in the careers of successful players. These clubs are not only structurally central, but also appear to form a tightly connected core of influential clubs that actively participate in the development and transfer of high-potential talent. In contrast, the clubs on the right result central within the unsuccessful player network, while having less influence on the career development of successful players.

Something interesting to underline, is that several clubs appear in both rankings. These include Genoa, Udinese, Atalanta, Sampdoria and Olympiacos. These clubs, in particular the Italian ones, are known for their high activity in the transfer market, engaging in established relationships with a wide range of trading partners with which they exchange a large amount of players of varying levels (both successful and unsuccessful). At international level, Olympiacos serves as a notable example: the club exports local Greek talent to a wide range of different national leagues, while also attracting international players, maintaining a strong presence in both successful and unsuccessful transfer paths.

Another key insight is the correlation between ClubRank and traditional network centrality metrics. To illustrate this, in Table 6.5, we highlight in green the clubs that appear in the top 10 of at least one centrality metric in the successful network, and in red those in the top 10 of at least one centrality metric in the unsuccessful network. The correlation is particularly strong in the successful network, where 15 of the top 25 clubs by ClubRank also appear in the top 10 of some centrality measures. In the unsuccessful network, this is true for only 8 out of 25 clubs. This result supports the validity of ClubRank as a

Rank	Club	Score	Rank	Club	Score
1	Chelsea	0.009615	1	Steaua Bucharest	0.002709
2	Inter	0.009140	2	Olympiacos	0.002428
3	Genoa	0.007941	3	Sparta Prague	0.002405
4	FC Porto	0.007921	4	Genoa	0.002320
5	SL Benfica	0.007795	5	CFR Cluj	0.002311
6	Juventus FC	0.007759	6	Parma	0.002298
7	AS Roma	0.007580	7	SK Slavia Prague	0.002209
8	Fiorentina	0.007574	8	FC Dinamo	0.002163
9	Udinese Calcio	0.007222	9	Rostov	0.002153
10	Sporting CP	0.006789	10	Atalanta BC	0.002149
11	Manchester City	0.006731	11	Udinese Calcio	0.002122
12	AC Milan	0.006679	12	Videoton FC	0.002105
13	Atlético Madrid	0.006077	13	Chievo Verona	0.002096
14	Liverpool FC	0.006016	14	Slavia Sofia	0.002064
15	Atalanta BC	0.006014	15	Baltika Kaliningrad	0.001992
16	Tottenham Hotspur	0.006007	16	SKA Khabarovsk	0.001953
17	Sevilla FC	0.005928	17	Shakhtar Donetsk	0.001948
18	Sampdoria	0.005873	18	FC Khimki	0.001947
19	Olympiacos	0.005653	19	Ural Yekaterinburg	0.001926
20	Aston Villa	0.005644	20	Rubin Kazan	0.001909
21	AS Monaco	0.005582	21	GNK Dinamo Zagreb	0.001886
22	Manchester United	0.005182	22	Levski Sofia	0.001843
23	Arsenal	0.005172	23	Sampdoria	0.001837
24	Valencia	0.005118	24	Debreceni VSC	0.001833
25	Fulham	0.005112	25	CSKA Sofia	0.001796

TABLE 6.5: Top 25 Clubs by ClubRank scores on successful (left) and unsuccessful (right) outgoing probabilistic transfer networks. The clubs that figure in the top 25 for both successful and unsuccessful player networks are reported in bold. The clubs that appear in the top 10 for at least one network metrics applied on the relative network (successful or unsuccessful) are colored. See Tables 6.6 & 6.7 for the comparison.

meaningful indicator of structural influence in these networks.

The top clubs according to the application of ClubRank on the successful network span a variety of roles, including:

- **Elite clubs:** Established top teams that compete at the highest levels and frequently exchange successful players with the goal of building the best team to aim at winning international competitions (e.g., Chelsea, Inter, Juventus, Manchester City, AC Milan, Atlético Madrid, Liverpool, Tottenham, Manchester United, Arsenal).
- **Development-focused clubs:** Teams that scout and acquire local talented players, later selling them to elite clubs (e.g., Porto, Benfica, Roma, Fiorentina, Sporting CP, Atalanta, Sevilla, Olympiacos, Aston Villa, AS Monaco, Fulham).
- **Market intermediaries:** Clubs that specialize in acquiring under-the-radar players from worldwide and selling them at a profit once their value increases (e.g., Genoa, Udinese, Sampdoria).

A closer look at the centrality rankings in Tables 6.6 and 6.7 highlights more structural patterns. Unsurprisingly, clubs with high degree centrality also rank highly in in-degree and out-degree, but we decided to include all three metrics to better understand the transfer activity profiles of the reported clubs. Some clubs are prominent in both incoming and outgoing transfers of successful players, such as Benfica, Chelsea, Sporting CP, Porto, Manchester City, Inter and Roma, functioning as key hubs in the successful transfer

Rank	Degree Centrality		Betweenness Centrality		In-Degree (Arrivals)		Out-Degree (Departures)	
1	SL Benfica	0.3106	SL Benfica	0.0277	SL Benfica	157	SL Benfica	162
2	Chelsea	0.2921	Sporting CP	0.0222	Chelsea	151	Sporting CP	157
3	Sporting CP	0.2892	FC Porto	0.0208	Sporting CP	140	Chelsea	149
4	FC Porto	0.2648	Fiorentina	0.0201	FC Porto	130	FC Porto	142
5	Inter	0.2473	Olympiacos	0.0183	Manchester City	123	Inter	133
6	Manchester City	0.2454	Udinese Calcio	0.0171	Sevilla FC	122	Manchester City	129
7	AS Roma	0.2366	Fulham	0.0157	Fiorentina	122	AS Roma	124
8	Sevilla FC	0.2308	CSKA Moscow	0.0155	Inter	121	Liverpool	120
9	Fiorentina	0.2278	Chelsea	0.0154	AS Roma	119	AS Monaco	117
10	Liverpool	0.2269	Dynamo Kyiv	0.0150	Udinese Calcio	117	Arsenal	117

TABLE 6.6: Top 10 Clubs by different centrality metrics applied on the original successful player network. The clubs that figure in the top 10 for more than one metric are reported in bold.

Rank	Degree Centrality		Betweenness Centrality		In-Degree (Arrivals)		Out-Degree (Departures)	
1	Udinese Calcio	0.2046	CSKA Sofia	0.0126	Udinese Calcio	154	Udinese Calcio	159
2	Genoa	0.1771	CFR Cluj	0.0104	Genoa	127	Genoa	144
3	Olympiacos	0.1719	Red Star	0.0102	Olympiacos	126	Olympiacos	137
4	Watford	0.1627	Udinese Calcio	0.0096	Watford	125	Standard Liège	126
5	CFR Cluj	0.1614	HJK Helsinki	0.0087	CFR Cluj	122	CFR Cluj	125
6	Birmingham	0.1595	Levski Sofia	0.0086	Birmingham	120	Watford	124
7	Celtic	0.1556	FC Dinamo	0.0078	Wolverhampton	118	Birmingham	124
8	Parma	0.1542	HNK Rijeka	0.0077	Celtic	115	RSC Anderlecht	124
9	CSKA Sofia	0.1529	Partizan	0.0072	CSKA Sofia	115	Celtic	123
10	HNK Rijeka	0.1529	Lech Poznan	0.0068	Barnsley	115	Parma	123

TABLE 6.7: Top 10 Clubs by different centrality metrics applied on the original unsuccessful player network. The clubs that figure in the top 10 for more than one metric are reported in bold.

ecosystem. While others appear predominantly as exporters or importers.

A similar distribution is present in the unsuccessful network. Clubs like Udinese, Genoa, Olympiacos, Watford, CFR Cluj, Birmingham City and Celtic feature in both In-Degree and Out-Degree top-10 lists. Meanwhile, clubs such as Wolverhampton primarily import unsuccessful players, while Standard Liège, Anderlecht and Parma tend to export them. Betweenness centrality offers a slightly different perspective, capturing clubs that act as bridges in career trajectories rather than hubs. This highlights clubs that facilitate progression from lower-tier to top-tier leagues. Notable examples include Olympiacos, CSKA Moscow and Dynamo Kyiv in the successful network, and CSKA Sofia, CFR Cluj, Red Star Belgrade and HJK Helsinki in the unsuccessful network. These clubs often serve as stepping stones to move to better clubs from local low-tier leagues, particularly in international transfers.

To further validate the ClubRank results, we compare them with a ranking based on total income from outgoing transfers over the last 10 years. Transfer income provides a useful proxy for clubs that successfully develop local talents or acquire undervalued players, giving them a nice environment to grow and then resell them at a higher price, making profit. Table 6.8 shows the top 20 clubs by transfer income over the past decade, highlighting the presence of substantial alignment with our ClubRank-based rankings. 11 of the top 20 income-generating clubs, including all of the top four, also appear in the top 25 by ClubRank. Of the nine remaining clubs, four fall within the top 50, and the other five within the top 105, which still represent a valuable result given the amount of clubs present in the dataset.

Income Rank	Club	Income (millions)	CR Rank	CR Value
1	Chelsea	1.300	1	0.009615
2	SL Benfica	1.270	5	0.007795
3	AS Monaco	1.140	21	0.005582
4	Juventus FC	1.090	6	0.007759
5	Borussia Dortmund	1.030	82	0.003201
6	Ajax	1.000	69	0.003599
7	Barcelona	973	41	0.004316
8	Atlético Madrid	937	13	0.006077
9	Manchester City	877	11	0.006731
10	FC Porto	846	4	0.007921
11	Lyon	816	58	0.003838
12	AS Roma	815	7	0.007580
13	PSG	789	40	0.004356
14	Inter	788	2	0.009140
15	Atalanta BC	782	15	0.006014
16	Lille	749	76	0.003384
17	RB Leipzig	738	105	0.002794
18	Real Madrid	731	48	0.004158
19	Sporting CP	722	10	0.006789
20	Stade Rennais	676	78	0.003374

TABLE 6.8: Top 20 Clubs by Transfer Income (2015–2025) compared to ClubRank Rankings. The color scheme indicates on green clubs in the top 25 according to ClubRank, in yellow clubs in the top 50, and in red clubs outside the top 50.

It should be noted that elite clubs such as Barcelona, PSG and Real Madrid appear higher on the income list due to the large value of the players they sell, but these clubs typically represent final destinations rather than stepping stones in a player's development. For this reason, they do not rank highly according to ClubRank. In contrast, clubs such as Borussia Dortmund, Ajax, Lyon, Lille, RB Leipzig, and Stade Rennais excel in the development of talented players in the first stages of their careers and serve as springboards for future stars, rather than central steps of their careers. Consequently, they are not structurally central in the transfer pathways of successful players to appear in the top 25 according to ClubRank.

Ideally, we would compare ClubRank to transfer income over the same 20-year span as our dataset. Since many clubs were more active or effective in the past than in the recent decade, a full-period analysis might lead to even more significant results.

6.4 Discussion

After presenting the results of our main experiments, analyzing the football transfer network from macro-, meso- and micro-level perspectives, this section contextualizes our results with the findings obtained by other studies conducted in the field, that we introduced in Chapter 2.

Several of our findings highlight specific structural dynamics within national leagues. For instance, the 3-node pattern analysis reveals a consistent tendency among Italian elite clubs to engage in recurrent transfer activity with mid- and lower-tier local clubs. These relationships, often involving the loaning of young promising players, appear to be strategic and are frequently associated with successful career trajectories. In contrast, similar trading patterns among lower-tier clubs, such as between Serie A lower-tier clubs and Serie B teams, are more often linked to unsuccessful career paths. These dynamics in the Italian transfer market, particularly around the developing role played by the loaning system, are consistent with observations made in Palazzo et al., 2023; Bond, Widdop, and Parnell, 2020; Baroncelli and Lago, 2006; Neri et al., 2023.

Another significant insight concerns the Portuguese football league, which stands out as a major developer and exporter of football talent, especially to Europe's top five leagues. Our ClubRank metric, along with traditional centrality measures, places clubs such as SL Benfica, FC Porto and Sporting CP among the most influential teams in the network of successful players. These clubs not only develop local talent through a structured loan system similar to the one in Italy, but also consistently generate significant income from player transfers, reinforcing the findings of Nolasco, 2019. The Portuguese league also serves as a key point of access to the main European football leagues, attracting players from all around the world, particularly from countries like Turkey, Greece and Romania. Additionally, it remains the main destination for Brazilian players due to linguistic and cultural similarities. However, many of these transfers are associated with unsuccessful career outcomes, likely due to oversaturation and intense competition among players of similar profile, who try to follow the same path from Brazilian to European football.

As described in Bond, Widdop, and Chadwick, 2018, the global transfer network follows a core-periphery structure. At the core, we find the top five European leagues, which are also central to most successful career trajectories, as shown by our 2- and 3-node pattern analyses. A semiperiphery is composed of leagues from countries such as Portugal, Belgium, the Netherlands, Turkey and Russia, which serve as important points of access to the core for successful players. Finally, the periphery includes a wide range of national leagues, especially from outside Europe, that play limited roles in the development of successful players.

Our findings are also in line with Velema, 2021, demonstrating that most transfers still follow local or regional routes. Clubs tend to transfer players within their own league systems or to local second divisions, and also our community detection methods cluster clubs within the same national league into the same communities. Furthermore, the transfer strategies we observe at club level reflect those reported in Velema, 2021: elite clubs primarily trade top-tier players among themselves, mid-level clubs focus on identifying and reselling talent for profit, and finally smaller clubs result largely engaged in domestic trading, constrained by limited financial resources.

The community partition of the successful player network further illustrates these regional patterns. Four of the top five European leagues (treating the British Isles as a single cultural unit) form self-contained communities, indicative of dominant internal transfer flows. The exception is Germany, which tends to trade more with clubs from central and northern Europe. These observations align with the findings of Littlewood, Mullen, and Richardson, 2011, where it was observed that most top leagues remain dominated by indigenous players, with Germany as a key outlier.

Economic stratification is another factor that contributes to shape the community partition, consistent with the findings in Clemente and Cornaro, 2023. Our results also align with McGovern, 2002, which highlights the role of geographic, linguistic and cultural proximity in guiding transfer activity. This phenomenon affects both clubs (in scouting) and players (in destination preferences). However, the current study also finds that opting for less popular, more challenging career paths can sometimes correlate with greater long-term success.

The high Average Clustering Coefficient values, obtained by our analysis, also support the hypothesis proposed by Xu, 2021, which suggests that clubs are more likely to make deals with clubs that share the same third trading partners. However, our deeper network analysis indicates that this clustering tendency is more present in the network of unsuccessful players, compared to the successful player one.

The 2-node pattern analysis at league level identifies leagues that primarily act as exporters of successful players, especially lower-tier leagues, and leagues that function mainly as importers, in particular including the top five European leagues. These findings confirm earlier work by Félix et al., 2019.

Finally, our 3-node pattern analysis reinforces the conclusions of Velema, 2018, which describes the career paths of many players as highly localized. According to our results, players often transfer between clubs within the same leagues or return to former clubs, forming loops. In contrast, linear upward chain progressions, characterized by successive transfers to clubs and leagues of an increasingly higher level, are less common. Our results show that players who remain within the top five leagues during their career paths are much more likely to be classified as successful. Furthermore, cross-league circulation among these top-tier competitions is more strongly correlated with career success than fully domestic career paths.

Chapter 7

Conclusion

This chapter concludes the study by summarizing the key findings and addressing the research question introduced in Chapter 1. In addition, it outlines several promising directions for future work that could further enhance understanding of the football transfer network and its dynamics.

7.1 Conclusion

In this work, we investigated the dynamics of the global football transfer market from a network science perspective, aiming to provide insights that can inform player career decisions. We constructed a comprehensive dataset spanning 2005–2024, including 141,358 transfers involving 41,734 players across 50 top-tier leagues worldwide. Each transfer record includes 57 features capturing player attributes, club characteristics, and transfer details, providing a robust basis for subsequent network analysis.

To assess player success, we introduced a novel metric based on the evolution of their *transfermarkt* market value, allowing the classification of careers into successful and unsuccessful trajectories. This classification aligns with observable trends in player development and supports the use of the average market value as a proxy for club attractiveness in the transfer market.

At the macro-level, league- and club-level network analyses revealed that successful and unsuccessful player transfers exhibit distinct structural patterns. Successful players generally move consistently upward in the competitive hierarchy, interacting over a higher number of transfers with a smaller set of clubs, while unsuccessful players experience broader but less strategic transfer patterns. Community detection and flow analyses at the meso-level highlighted that successful players circulate within multi-league communities, often involving top European leagues, whereas unsuccessful players tend to move across culturally or geographically proximate leagues. Micro-level analysis further identified recurrent motifs associated with career success, including repeated transfers among top European leagues, reinforcing their role as central points in the careers of successful players; while unsuccessful trajectories are characterized by oscillations between lower-tier or culturally similar leagues.

Finally, we proposed *ClubRank*, a tailored centrality metric that captures the influence of clubs on player development. ClubRank distinguishes three categories of high-impact clubs: elite clubs, development clubs, and market intermediaries, providing actionable insights for career planning and scouting strategies.

Overall, our findings demonstrate that network-based analyses of transfer dynamics can uncover structural and temporal patterns that differentiate successful and unsuccessful careers. By combining career success assessment, network structure characterization, motif analysis and ClubRank, this study offers a comprehensive framework to inform strategic decisions for players, clubs, and stakeholders in the global football market.

7.2 Future Work

Although the results of this research are promising, some limitations emerged throughout the study, presenting opportunities for significant future advancements. In the following, we outline some directions that could further expand the scope of the analysis in this domain.

First, even if our dataset is large and rich, it is constrained to a specific time window and it includes only 50 leagues. Incorporating data from additional leagues, particularly from regions that are not considered in this study, such as Africa, Asia and Oceania, would provide a more complete and globally representative picture of the transfer market. Furthermore, many features within the dataset created by us remain unexplored. Future work could incorporate variables such as transfer fees and types (e.g., loan, free transfer, permanent deal) to analyze the economic and financial dimensions of the player transfer network more thoroughly.

The binary classification approach used to distinguish successful players from unsuccessful players could also be refined. Future studies might explore how different threshold values affect the results of the implemented experiments; or might implement a multi-class approach to capture varying degrees of success. A more objective and continuous method could involve tracking market value changes over fixed intervals (e.g., each season or transfer window) rather than only during transfer events. Moreover, replacing *transfermarkt* market value with a different index that would not suffer from the influence of factors such as the age and the contractual situation of a player, would represent an even better assessment approach. A valid alternative that could be considered is the Player Level metric provided by *Hypercube* (Hypercube Business Innovation, 2025), which represents an objective means of assessing a player's impact on their team's performance; determined by analyzing the changes in a team's winning odds when a player is on the field.

Another interesting direction could be related to the deeper integration of temporal patterns into the analysis. Studying how long players remain at clubs of different levels, and how these durations correlate with career success, may reveal data-driven transfer strategies that could lead a player to a successful career development.

Finally, future work could enhance the personalization of the current analysis. By incorporating player-specific attributes such as age, position, and physical characteristics, we could provide customized career recommendations. This would enable the identification of historically successful career paths for players with similar profiles, helping both athletes and scouts make more informed decisions.

Appendix A

Football leagues information

Table A.1 displays the fundamental information regarding the 50 football leagues considered in our analysis. Reporting their names, IDs, affiliated countries and data availability on *www.transfermarkt.com*.

TABLE A.1: Football leagues considered in this study and their transfermarkt IDs

League Name	League ID	Country	Data Availability
Bundesliga	A1	Austria	1992 - present
Jupiler Pro League	BE1	Belgium	1992 - present
Campeonato Brasileiro Serie A	BRA1	Brazil	1996 - present
Efbet Liga	BU1	Bulgaria	1992 - present
Primera Division De Chile	CLPD	Chile	1992 - present
Liga Dimayor I	COLP	Colombia	2010 - present
1 Hnl	KR1	Croatia	2005 - present
Fortuna Liga	TS1	Czech Republic	1992 - present
Superligaen	DK1	Denmark	1992 - present
Premier League	GB1	England	1992 - present
Championship	GB2	England	2004 - present
League One	GB3	England	2004 - present
Premium Liiga	EST1	Estonia	1997 - present
Veikkausliiga	FI1	Finland	1992 - present
Ligue 1	FR1	France	1992 - present
Ligue 2	FR2	France	1992 - present
1 Bundesliga	L1	Germany	1992 - present
2 Bundesliga	L2	Germany	1992 - present
3 Liga	L3	Germany	2008 - present
Super League 1	GR1	Greece	1992 - present
Nemzeti Bajnoksag	UNG1	Hungary	1992 - present
Pepsi Max Deild	IS1	Iceland	1992 - present
Premier League	IR1	Ireland	1996 - present
Serie A	IT1	Italy	1992 - present
Serie B	IT2	Italy	1992 - present
Liga MX Clausura	MEX1	Mexico	1996 - present
Eredivisie	NL1	Netherlands	1992 - present
Keuken Kampioen Divisie	NL2	Netherlands	1992 - present
Tweede Divisie	NTD	Netherlands	2016 - present
Eliteserien	NO1	Norway	1992 - present
Obos Ligaen	NO2	Norway	2002 - present
Pko Ekstraklasa	PL1	Poland	1992 - present
Liga Nos	PO1	Portugal	1992 - present
Liga 1	RO1	Romania	2002 - present
Premier Liga	RU1	Russia	1992 - present
1 Division	RU2	Russia	1992 - present
Scottish Premiership	SC1	Scotland	1997 - present
Scottish Championship	SC2	Scotland	1998 - present
Super Liga Srbije	SER1	Serbia	2004 - present
Fortuna Liga	SLO1	Slovakia	2003 - present
Prva Liga	SL1	Slovenia	1999 - present
Primera Division	ES1	Spain	1992 - present
La Liga 2	ES2	Spain	1992 - present
Allsvenskan	SE1	Sweden	2000 - present
Super League	C1	Switzerland	1992 - present
Challenge League	C2	Switzerland	1992 - present
Super Lig	TR1	Turkey	1992 - present
Premier Liga	UKR1	Ukraine	1992 - present
Primera Division Apertura	URU1	Uruguay	2006 - present
Major League Soccer	MLS1	USA	1996 - present

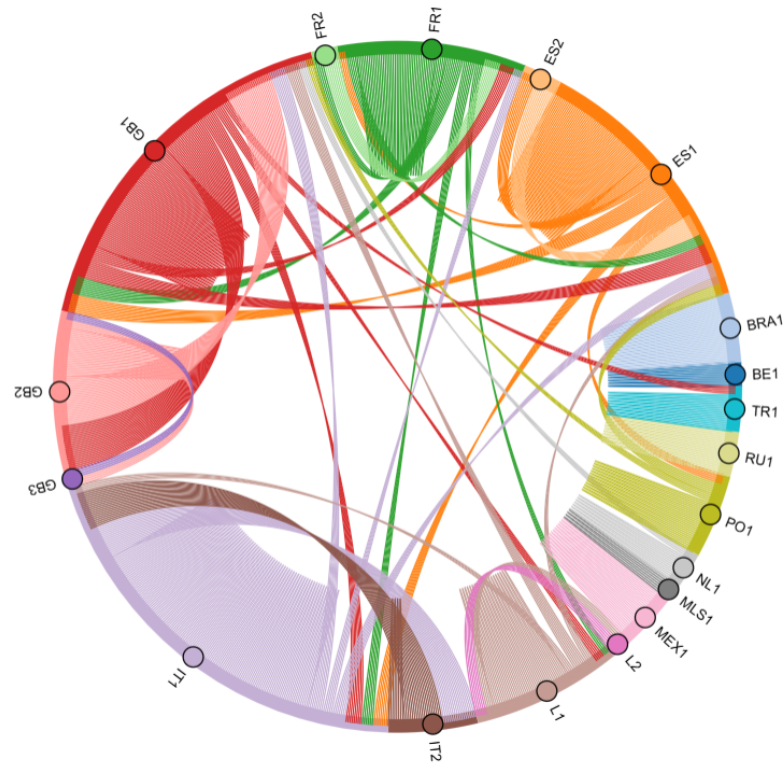
Appendix B

Transfer flows analysis with higher thresholds

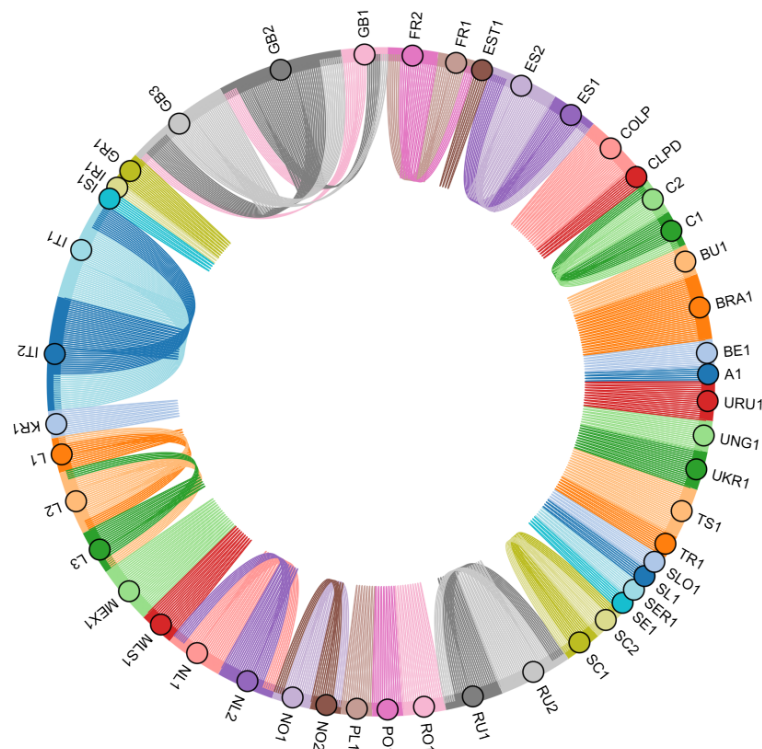
Figure B.1 highlights the structure of transfer flows when using a higher threshold for the minimum number of transfers to consider, 100 for successful players and 370 for unsuccessful ones. These stricter criteria reveal only the most significant and persistent transfer channels.

In the successful player network (Figure B.1a), only 19 of the 48 leagues exceed 100 transfers with another league. Most of the remaining flows involve domestic exchanges or well-established international pathways. For example, Turkey predominantly exports successful players to France, the Netherlands to England, and Portugal to both England and Spain. These directional flows indicate not only cultural or geographic proximity, but also long-term transfer relationships shaped by strategic career moves toward top-tier European leagues. At this level of aggregation, more than 75% of successful player transfers involve the top five European leagues and their second divisions.

In contrast, Figure B.1b shows that even with such a high threshold, 47 out of 50 leagues in the unsuccessful player network remain active and present in the final visualization. Most of the interactions occur within national borders, involving second and third divisions. Only three leagues fall below the threshold: the Dutch Tweede Divisie, the Finnish Veikkausliiga and the Danish Superligaen. This further confirms the highly localized nature of unsuccessful player trajectories, dominated by national-level circulation and almost no significant foreign flows beyond the minimum threshold.

Successful Player Transfers between National Leagues (over 100 transfers)

(A) Successful Player Network (minimum 100 transfers)

Unsuccessful Player Transfers between National Leagues (over 370 transfers)

(B) Unsuccessful Player Network (minimum 370 transfers)

FIGURE B.1: Comparison between the transfer flows of successful and unsuccessful players among national leagues, using a high threshold.

Bibliography

- Arriaza-Ardiles, Enrique et al. (2018). "Applying graphs and complex networks to football metric interpretation". In: *Human movement science* 57, pp. 236–243.
- Balague, Natàlia et al. (2013). "Overview of complex systems in sport". In: *Journal of Systems Science and Complexity* 26, pp. 4–13.
- Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *science* 286.5439, pp. 509–512.
- Barabási, Albert-László and Jennifer Frangos (2014). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. Basic books.
- Barabasi, Albert-Laszlo and Zoltan N Oltvai (2004). "Network biology: understanding the cell's functional organization". In: *Nature reviews genetics* 5.2, pp. 101–113.
- Baroncelli, Alessandro and Umberto Lago (2006). "Italian football". In: *Journal of sports economics* 7.1, pp. 13–28.
- Bekkers, Joris and Shaunak Dabadghao (2019). "Flow motifs in soccer: What can passing behavior tell us?" In: *Journal of Sports Analytics* 5.4, pp. 299–311.
- Bernardo, Giovanni, Massimo Ruberti, and Roberto Verona (2022). "Image is everything! Professional football players' visibility and wages: evidence from the Italian Serie A". In: *Applied Economics* 54.5, pp. 595–614.
- Blondel, Vincent D et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10, P10008.
- Bond, Alexander John, Paul Widdop, and Simon Chadwick (2018). "Football's emerging market trade network: Ego network approach to world systems theory". In: *Managing Sport and Leisure* 23.1-2, pp. 70–91.
- Bond, Alexander John, Paul Widdop, and Daniel Parnell (2020). "Topological network properties of the European football loan system". In: *European Sport Management Quarterly* 20.5, pp. 655–678.
- Borrie, Andrew, Gudberg K Jonsson, and Magnus S Magnusson (2002). "Temporal pattern analysis and its applicability in sport: an explanation and exemplar data". In: *Journal of sports sciences* 20.10, pp. 845–852.
- Botterill, Cal (1990). "Sport psychology and professional hockey". In: *The sport psychologist* 4.4, pp. 358–368.
- Bryson, Alex, Bernd Frick, and Rob Simmons (2013). "The returns to scarce talent: Footedness and player remuneration in European soccer". In: *Journal of Sports Economics* 14.6, pp. 606–628.
- Buldú, Javier M et al. (2018). "Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game". In: *Frontiers in psychology* 9, p. 1900.
- Camerino, Oleguer Foguet et al. (2012). "Dynamics of the game in soccer: Detection of T-patterns". In: *European journal of sport science* 12.3, pp. 216–224.
- Chelladurai, Packianathan (2014). *Managing organizations for sport and physical activity: A systems perspective*. Routledge.

- Cintia, Paolo, Salvatore Rinzivillo, and Luca Pappalardo (2015). "A network-based approach to evaluate the performance of football teams". In: *Machine learning and data mining for sports analytics workshop, Porto, Portugal*.
- Clemente, Filipe Manuel, Fernando Manuel Lourenço Martins, Rui Sousa Mendes, et al. (2016). "Social network analysis applied to team sports analysis". In.
- Clemente, Filipe Manuel et al. (2015). "Using network metrics in soccer: a macro-analysis". In: *Journal of human kinetics* 45, p. 123.
- Clemente, Gian Paolo and Alessandra Cornaro (2023). "Community detection in attributed networks for global transfer market". In: *Annals of Operations Research* 325.1, pp. 57–83.
- Cotta, Carlos et al. (2013). "A network analysis of the 2010 FIFA world cup champion team play". In: *Journal of Systems Science and Complexity* 26.1, pp. 21–42.
- Danon, Leon et al. (2005). "Comparing community structure identification". In: *Journal of statistical mechanics: Theory and experiment* 2005.09, P09008.
- Dieles, Tristan J, Carolina ES Mattsson, and Frank W Takes (2024). "Identifying successful football teams in the European player transfer network". In: *Applied Network Science* 9.1, p. 65.
- Dunning, Ted (1993). "Accurate methods for the statistics of surprise and coincidence". In: *Computational Linguistics* 19.1, pp. 61–74.
- Eccles, David W and Gershon Tenenbaum (2004). "Why an expert team is more than a team of experts: A social-cognitive conceptualization of team coordination and communication in sport". In: *Journal of Sport and Exercise Psychology* 26.4, pp. 542–560.
- Félix, Lucas GS et al. (2019). "A social network analysis of football with complex networks". In: *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*. SBC, pp. 47–50.
- Freeman, Linton C. (1977). "A set of measures of centrality based on betweenness". In: *Sociometry* 40.1, pp. 35–41. DOI: [10.2307/3033543](https://doi.org/10.2307/3033543).
- Gabbett, Tim J (2016). "The training—injury prevention paradox: should athletes be training smarter and harder?" In: *British journal of sports medicine* 50.5, pp. 273–280.
- Galvin, Ewen (2023). *Transfers: An R package to access and analyze football transfer data*. URL: <https://github.com/ewenme/transfers>.
- Garlaschelli, Diego and Maria I Loffredo (2004). "Patterns of link reciprocity in directed networks". In: *Physical Review Letters* 93.26, p. 268701.
- Ghar, Shantanu, Sayali Patil, and Venkhatesh Arunachalam (2021). "Data Driven football scouting assistance with simulated player performance extrapolation". In: *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, pp. 1160–1167.
- Ghio, Alessandro, Massimo Ruberti, and Roberto Verona (2019). "Financial constraints on sport organizations' cost efficiency: The impact of financial fair play on Italian soccer clubs". In: *Applied economics* 51.24, pp. 2623–2638.
- Gonçalves, Bruno et al. (2017). "Exploring team passing networks and player movement dynamics in youth association football". In: *PloS one* 12.1, e0171156.
- Gould, Peter and Anthony Gatrell (1979). "A structural analysis of a game: the Liverpool v Manchester United Cup Final of 1977". In: *Social Networks* 2.3, pp. 253–273.
- Goyal, Sanjeev (2012). "Connections: an introduction to the economics of networks". In: *Connections*. Princeton University Press.
- Grund, Thomas U (2012). "Network structure and team performance: The case of English Premier League soccer teams". In: *Social Networks* 34.4, pp. 682–690.

- Gürsakal, Necmi et al. (2018). "Network motifs in football". In: *Turkish Journal of Sport and Exercise* 20.3, pp. 263–272.
- Holme, Petter and Jari Saramäki (2012). "Temporal networks". In: *Physics reports* 519.3, pp. 97–125.
- Hypercube Business Innovation (2025). *Sports Division — Hypercube: Charting the effects of innovation in sports*. <https://www.hypercube.nl/en/sports/>.
- IMARC Group (2025). *Football Market Size, Share, Trends and Forecast by Type, Size, Distribution Channel, and Region, 2025–2033*. <https://www.imarcgroup.com/football-market>. Report ID: SR112025A5678.
- Jurafsky, Daniel and James H Martin (2009). *Speech and Language Processing*. 2nd ed. Pearson Education.
- Kataoka, Yasuyuki and Peter Gray (2019). "Real-time power performance prediction in tour de France". In: *Machine Learning and Data Mining for Sports Analytics: 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD 2018, Dublin, Ireland, September 10, 2018, Proceedings* 5. Springer, pp. 121–130.
- Kim, Kyung-Sik and Brian H Yim (2017). "Utilizing social network analysis in social sciences in sport". In: *Asia Pacific Journal of Sport and Social Science* 6.2, pp. 177–196.
- Littlewood, Martin, Chris Mullen, and David Richardson (2011). "Football labour migration: an examination of the player recruitment strategies of the 'big five' European football leagues 2004–5 to 2008–9". In: *Soccer & Society* 12.6, pp. 788–805.
- Liu, Xiao Fan et al. (2016). "The anatomy of the global football player transfer network: Club functionalities versus network properties". In: *PloS one* 11.6, e0156504.
- Lusher, Dean, Garry Robins, and Peter Kremer (2010). "The application of social network analysis to team sports". In: *Measurement in physical education and exercise science* 14.4, pp. 211–224.
- Masuda, Naoki and Renaud Lambiotte (2016). *A guide to temporal networks*. World Scientific.
- Matesanz, David et al. (2018). "Transfer market activities and sportive performance in European first football leagues: A dynamic network approach". In: *PloS one* 13.12, e0209362.
- McGovern, Patrick (2002). "Globalization or internationalization? Foreign footballers in the English league, 1946–95". In: *Sociology* 36.1, pp. 23–42.
- Meng-Lewis, Yue et al. (2022). "Understanding complexity and dynamics in the career development of eSports athletes". In: *Sport Management Review* 25.1, pp. 106–133.
- Miller, Thomas W (2015). *Sports analytics and data science: winning the game with methods and models*. FT press.
- Monroe, Burt L, Michael P Colaresi, and Kevin M Quinn (2008). "Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict". In: *Political Analysis* 16.4, pp. 372–403.
- Narizuka, Takuma, Ken Yamamoto, and Yoshihiro Yamazaki (2014). "Statistical properties of position-dependent ball-passing networks in football games". In: *Physica A: Statistical Mechanics and its Applications* 412, pp. 157–168.
- Neri, Lorenzo et al. (2023). "Football players and asset manipulation: the management of football transfers in Italian Serie A". In: *European Sport Management Quarterly* 23.4, pp. 942–962.
- Newman, M. E. J. (2002). "Assortative mixing in networks". In: *Physical Review Letters* 89.20, p. 208701.
- Newman, Mark EJ (2003). "The structure and function of complex networks". In: *SIAM review* 45.2, pp. 167–256.

- Newman, Mark EJ and Michelle Girvan (2004). "Finding and evaluating community structure in networks". In: *Physical review E* 69.2, p. 026113.
- Nolasco, Carlos (2019). "Player migration in Portuguese football: a game of exits and entrances". In: *Soccer & Society* 20.6, pp. 795–809.
- Page, Lawrence et al. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Tech. rep. 1999-66. Stanford InfoLab. URL: <http://ilpubs.stanford.edu:8090/422/>.
- Palazzo, Lucio et al. (2023). "Community structure of the football transfer market network: the case of Italian Serie A". In: *Journal of Sports Analytics* 9.3, pp. 221–243.
- Passos, Pedro et al. (2011). "Networks as a novel tool for studying team ball sports as complex social systems". In: *Journal of science and medicine in sport* 14.2, pp. 170–176.
- Peña, Javier López and Raúl Sánchez Navarro (2015). "Who can replace Xavi? A passing motif analysis of football players". In: *arXiv preprint arXiv:1506.07768*.
- Pena, Javier López and Hugo Touchette (2012). "A network theory analysis of football strategies". In: *arXiv preprint arXiv:1206.6904*.
- Pina, Tiago J, Ana Paulo, and Duarte Araújo (2017). *Network characteristics of successful performance in association football. A study on the UEFA champions league*.
- Ramos, João, Rui J Lopes, and Duarte Araújo (2018). "What's next in complex networks? Capturing the concept of attacking play in invasive team sports". In: *Sports medicine* 48, pp. 17–28.
- Ramos, João et al. (2017). "Hypernetworks reveal compound variables that capture cooperative and competitive interactions in a soccer match". In: *Frontiers in Psychology* 8, p. 1379.
- Ribeiro, João et al. (2017). "Team sports performance analysed through the lens of social network theory: implications for research and practice". In: *Sports medicine* 47, pp. 1689–1696.
- Spearman, Charles (1961). "The proof and measurement of association between two things." In.
- Stöggl, Thomas L and Billy Sperlich (2015). "The training intensity distribution among well-trained and elite endurance athletes". In: *Frontiers in physiology* 6, p. 295.
- Transfermarkt (2024). *Transfermarkt: Football Transfers, Market Values, Rumours and Stats*. URL: <https://www.transfermarkt.com>.
- UEFA (2025). *UEFA Club Coefficients*. <https://www.uefa.com/nationalassociations/uefarankings/club/>.
- Van Der Zwaard, Stephan et al. (2018). "Critical determinants of combined sprint and endurance performance: an integrative analysis from muscle fiber to the human body". In: *Faseb Journal* 32.4, pp. 2110–2123.
- Velema, Thijs A (2018). "A game of snakes and ladders: Player migratory trajectories in the global football labor market". In: *International Review for the Sociology of Sport* 53.6, pp. 706–725.
- (2021). "Globalization and player recruitment: How teams from European top leagues broker migration flows of footballers in the global transfer network". In: *International Review for the Sociology of Sport* 56.4, pp. 493–513.
- Wäsche, Hagen et al. (2017). "Social network analysis in sport research: an emerging paradigm". In: *European Journal for Sport and Society* 14.2, pp. 138–165.
- Wasserman, Stanley and Katherine Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Watts, Duncan J and Steven H Strogatz (1998). "Collective dynamics of 'small-world' networks". In: *Nature* 393.6684, pp. 440–442.

- Xu, Yu (2021). "The formation mechanism of the player transfer network among football clubs". In: *Soccer & Society* 22.7, pp. 704–715.
- Yamamoto, Yuji and Keiko Yokoyama (2011). "Common and unique network dynamics in football games". In: *PloS one* 6.12, e29638.