

Sapienza

Data Driven Model of complex sistem

Football transfers network analysis



Table of Contents

- I Data and Cleaning procedures
- II Network and Measures
- III Modularity and Communities
- IV Centrality
- V League comparison
- VI Conclusion



- The data comes from Kaggle, scraped from **Transfer Markt**.
- The datasets regards all the transfer operation in the major 7 European leagues from the **1992/1993** season to the **2021/2022** season.
- The raw data were extremely dirty so a **massive and detailed** cleaning operation was necessary. →

club_name	player_name	age	posit	club_involve	fee	transfer_r	fee_league	season
Cagliari Calcio	Filippo Romagi	20	Cent	Juventus	€7.40m	in	7.4	Serie A 2017/2018
Cagliari Calcio	Dario Del Fabrc	22	Cent	Juventus	€4.50m	out	4.5	Serie A 2017/2018
FC Crotone	Rolando Mandi	20	Defe	Juventus	loan trans	in	0	Serie A 2017/2018
ACF Fiorentina	Federico Berna	23	Righ	Juventus	€40.00m	out	40	Serie A 2017/2018
Hellas Verona	Moise Kean	17	Cent	Juventus	loan trans	in	0	Serie A 2017/2018
Hellas Verona	Simone Ganz	23	Cent	Juventus	End of loa	out	0	Serie A 2017/2018
AC Milan	Leonardo Bonu	30	Cent	Juventus	€42.00m	in	42	Serie A 2017/2018
Juventus FC	Federico Berna	23	Righ	Fiorentina	€40.00m	in	40	Serie A 2017/2018



club_name	player_name	age	posit	club_involve	fee	transfer_r	fee_league	season
Cagliari Calcio	Filippo Romagi	20	Cent	Juventus	€7.40m	in	7.4	Serie A 2017/2018
Crotone	Rolando Mandi	20	Defe	Juventus	loan trans	in	0	Serie A 2017/2018
Hellas Verona	Moise Kean	17	Cent	Juventus	loan trans	in	0	Serie A 2017/2018
AC Milan	Leonardo Bonu	30	Cent	Juventus	€42.00m	in	42	Serie A 2017/2018
Juventus	Federico Berna	23	Righ	Fiorentina	€40.00m	in	40	Serie A 2017/2018

Cleaning procedure:

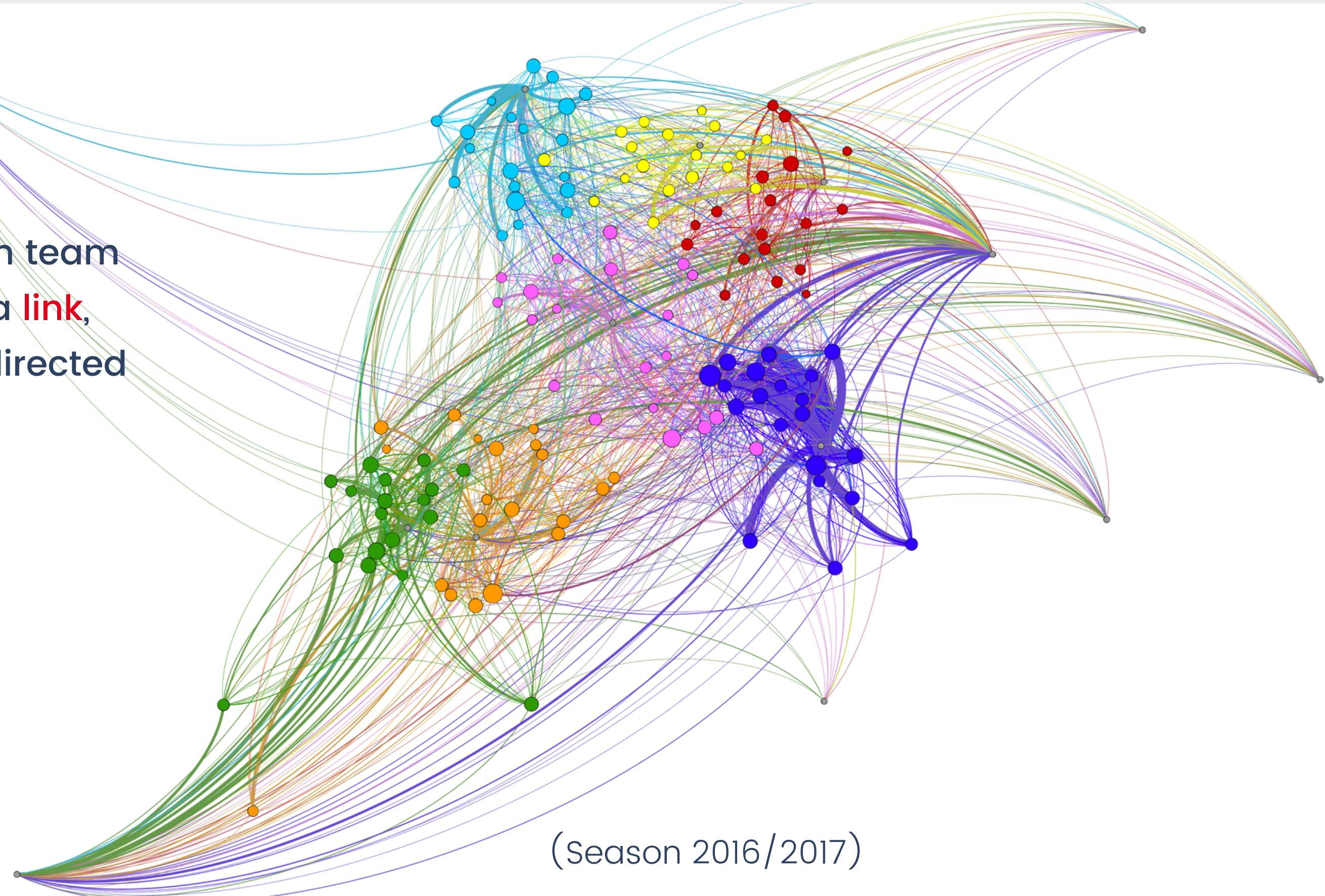
1. **Scrape data from Transfer Markt to build two types of dictionaries:**
 - one that maps every team name for every season its long name to the short one. (ex: "Juventus FC" : "Juventus").
 - one that maps each club not in the 7 top leagues in its continent/country. (ex: "Santos": "South America").
2. **Delete all the "out" and "End of loan" operations to avoid duplicate transactions.**
3. **Check and correct some typos in original data.**

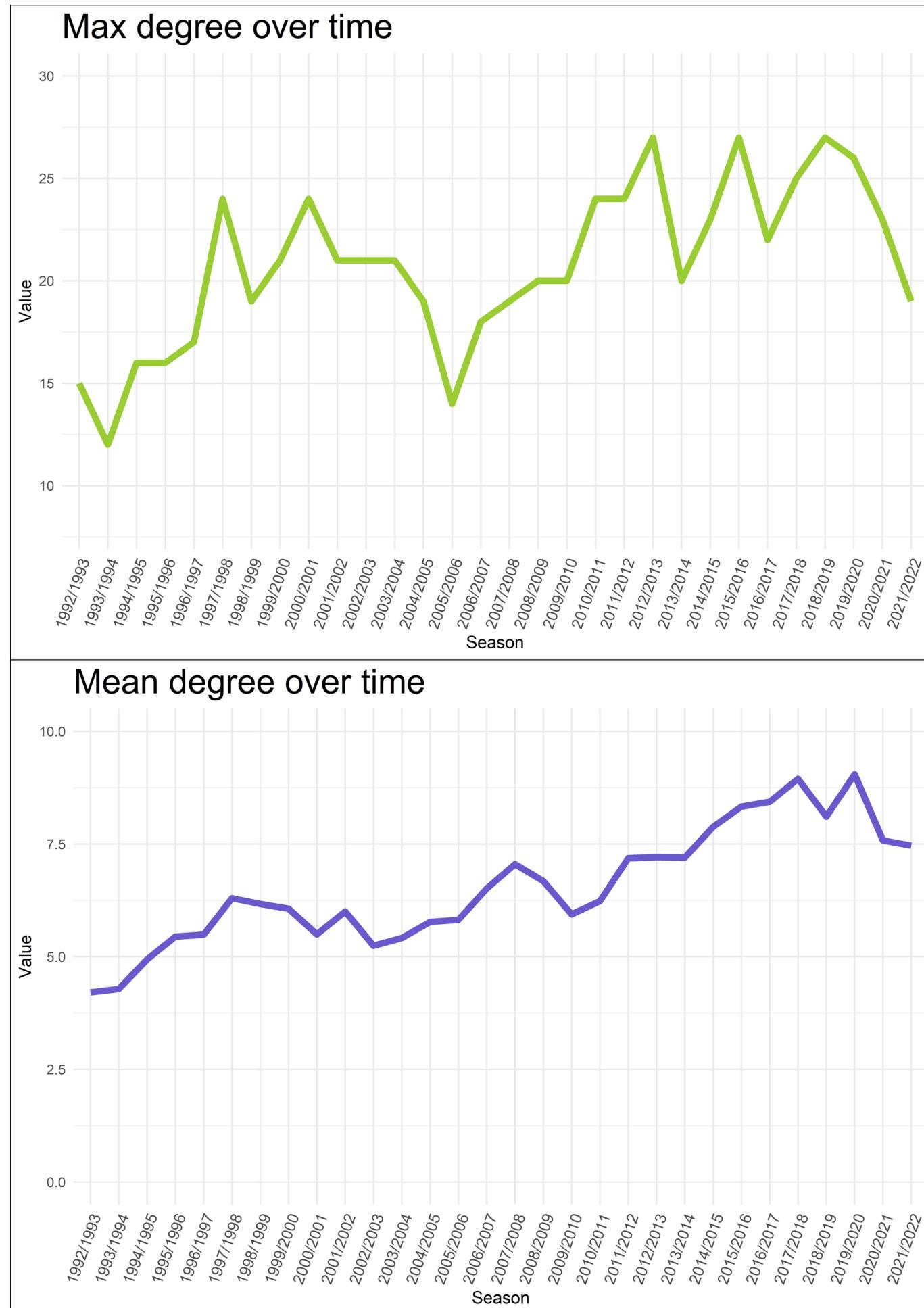
Graph building:

- Every team is a **Node**
- If team A buys a player from team B, they are connected with a **link**, weighted with the fee and directed to B

For every season:

- Nodes: 134
- Links:
 - A) ~900 if one considers all transactions
 - B) ~600 if one considers only transactions from team of Top 7 European Leagues



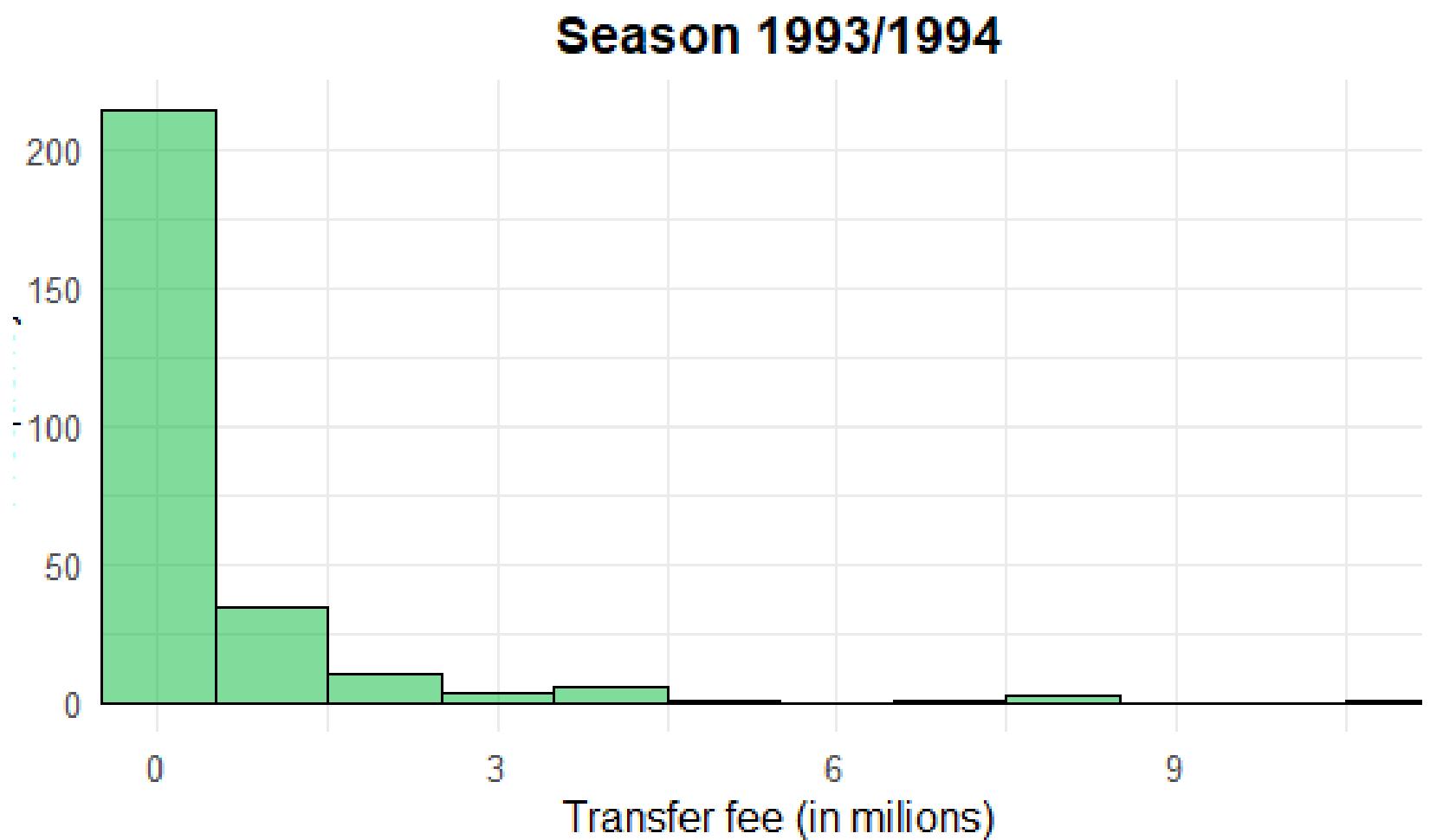


Degree

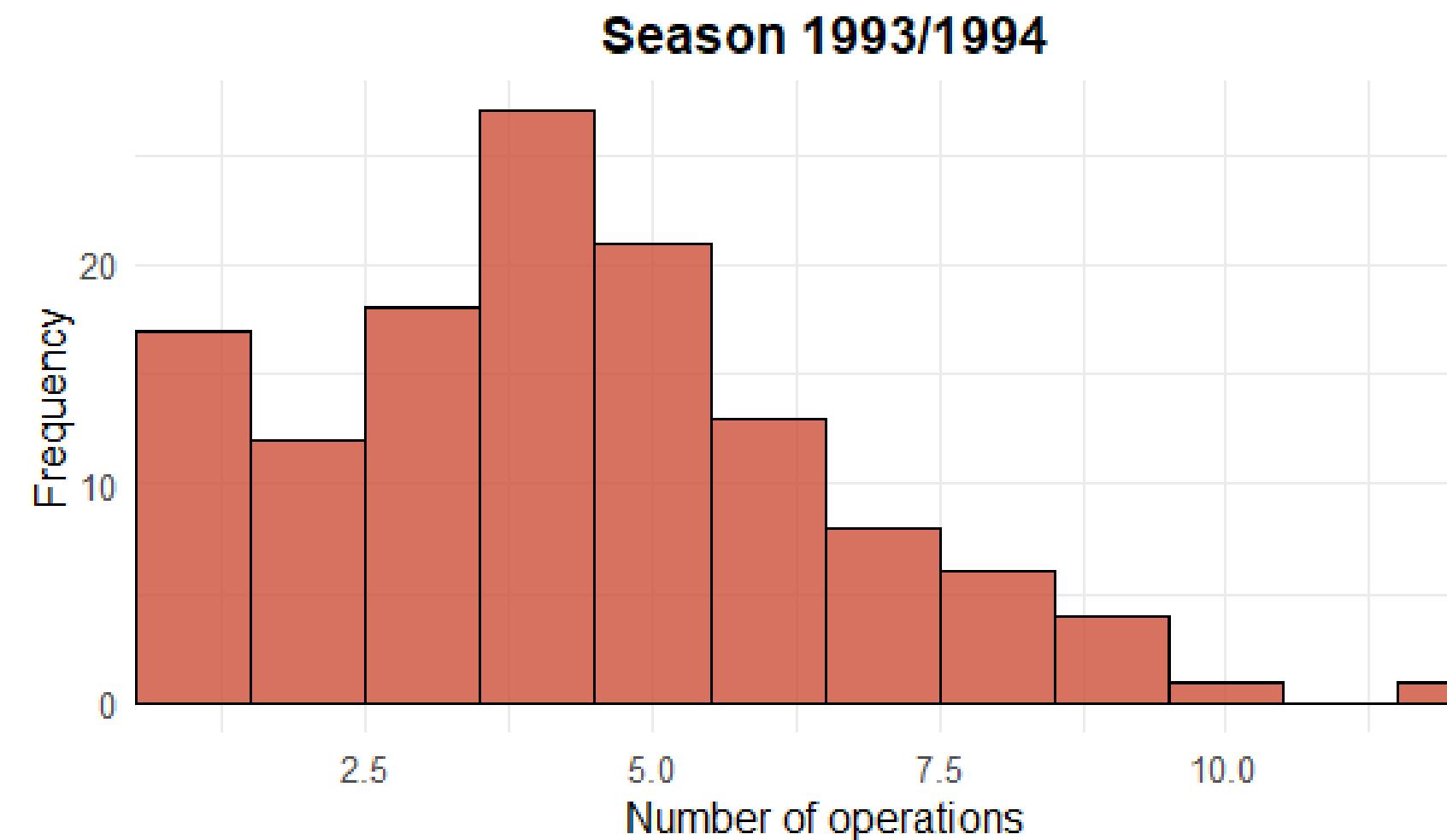
The degree of a team, represents how many different teams it has had negotiations with (both in and out).

- The Average degree of the teams has more or less **doubled** over the seasons (from 4 to 7.5).
- The maximum degree of clubs has **increased** over time, although the effects are less noticeable than the average degree. After all, a team cannot make too many different deals per season, as 11 players always take the field.

Edge plot



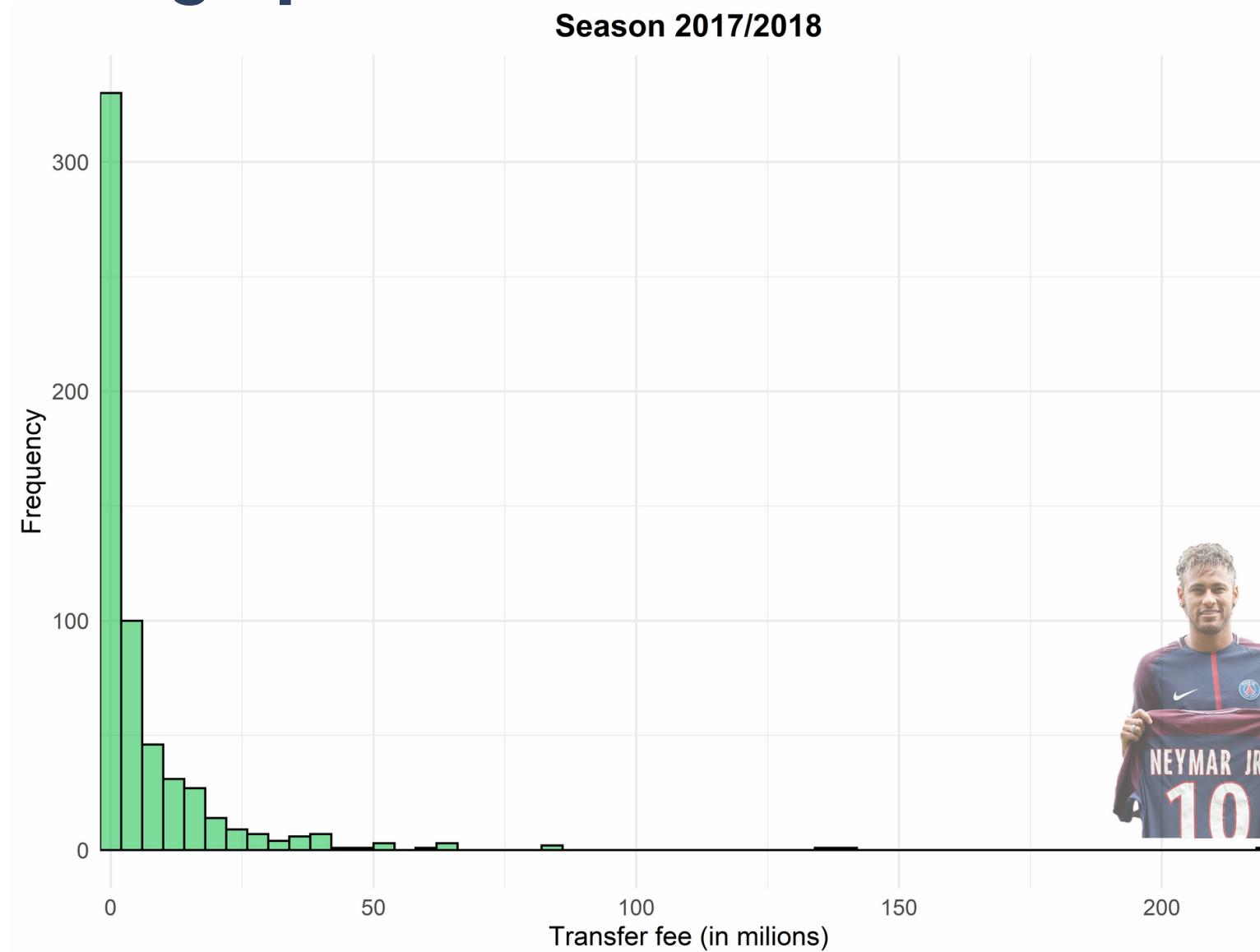
Degree Plot



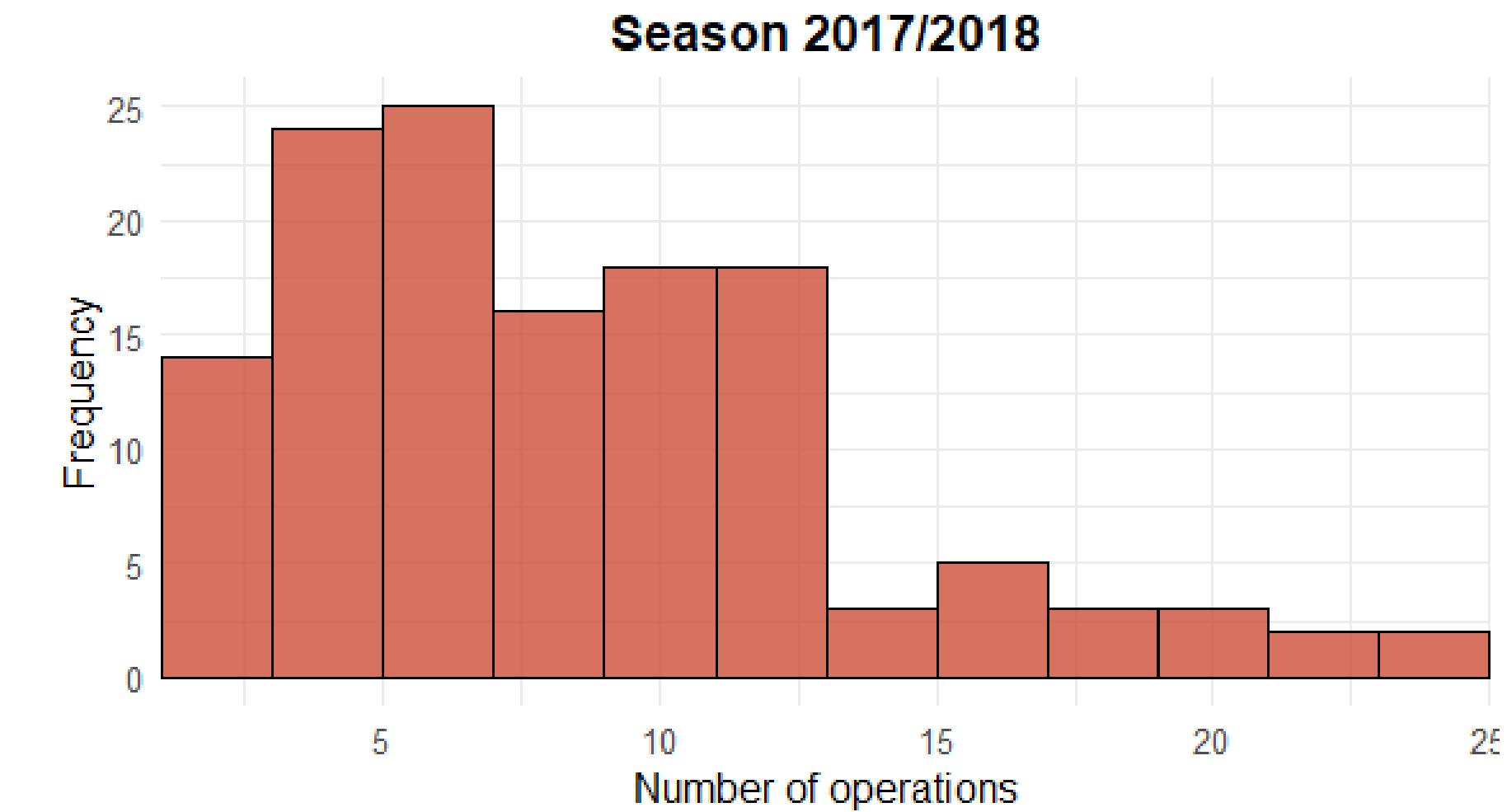
- Most of the operations are **loans**
- There are only few Transfers with a fee > 5 Milion

- Each team on average is involved in deals with **4** other clubs.
- only a few clubs do business with more than **7** teams

Edge plot



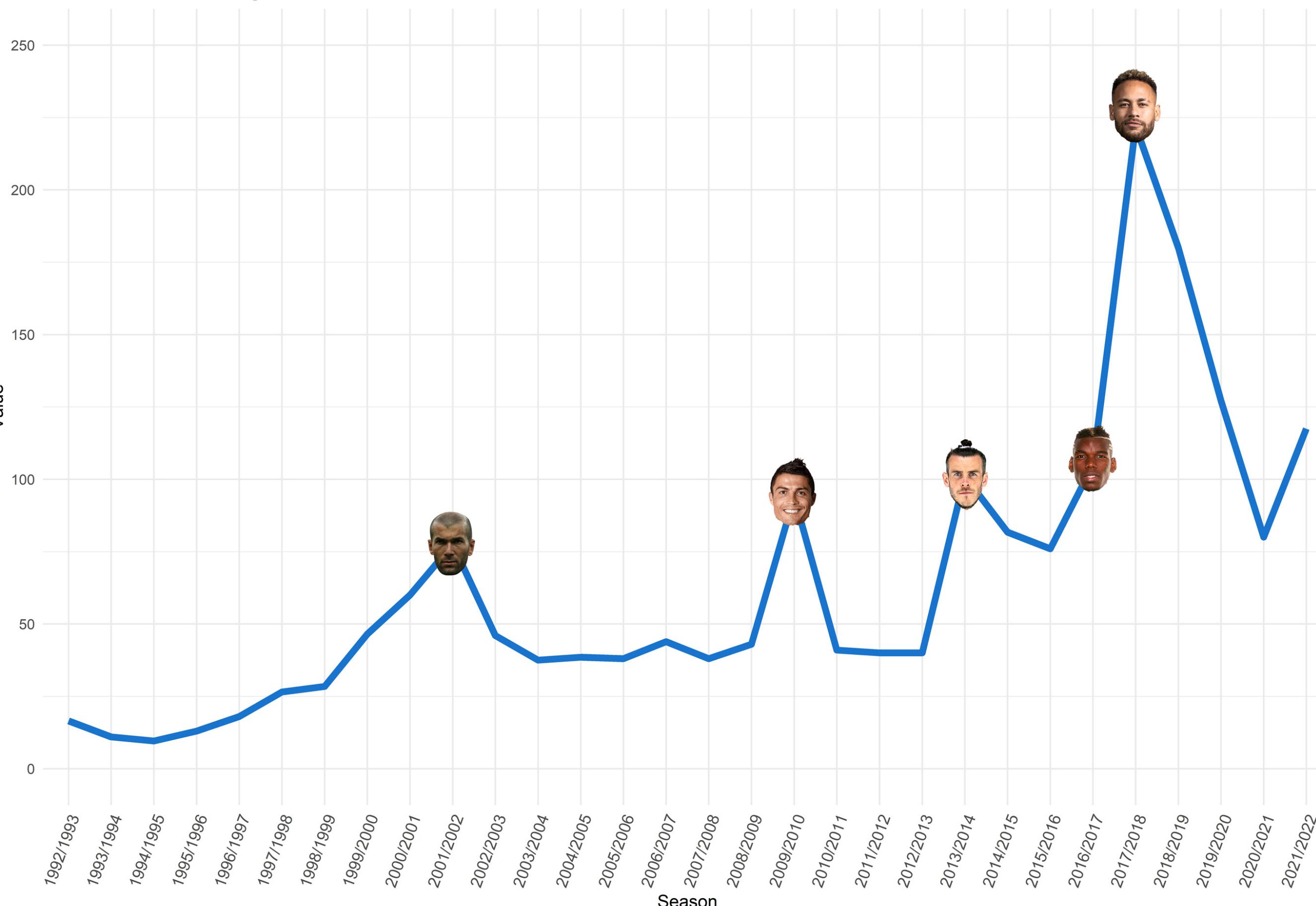
Degree Plot



- Still a lot of transactions are **loans**
- There are some edges with a high charge
- Three purchases exceed **100 million**

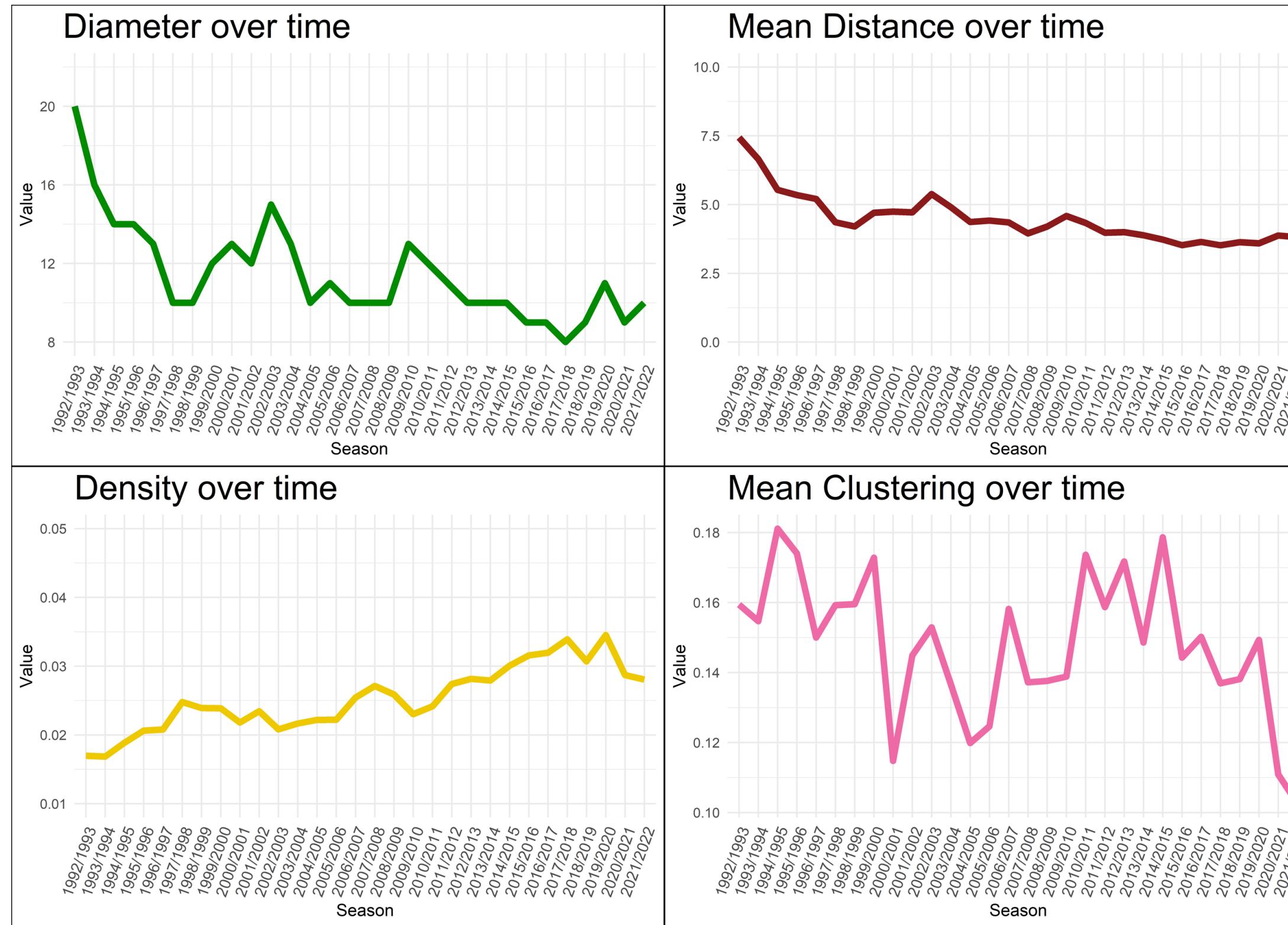
- Each team on average is involved in deals with **9** other clubs.
- 43 clubs over 134 do business with more than 10 teams

Maximum edge fee over time



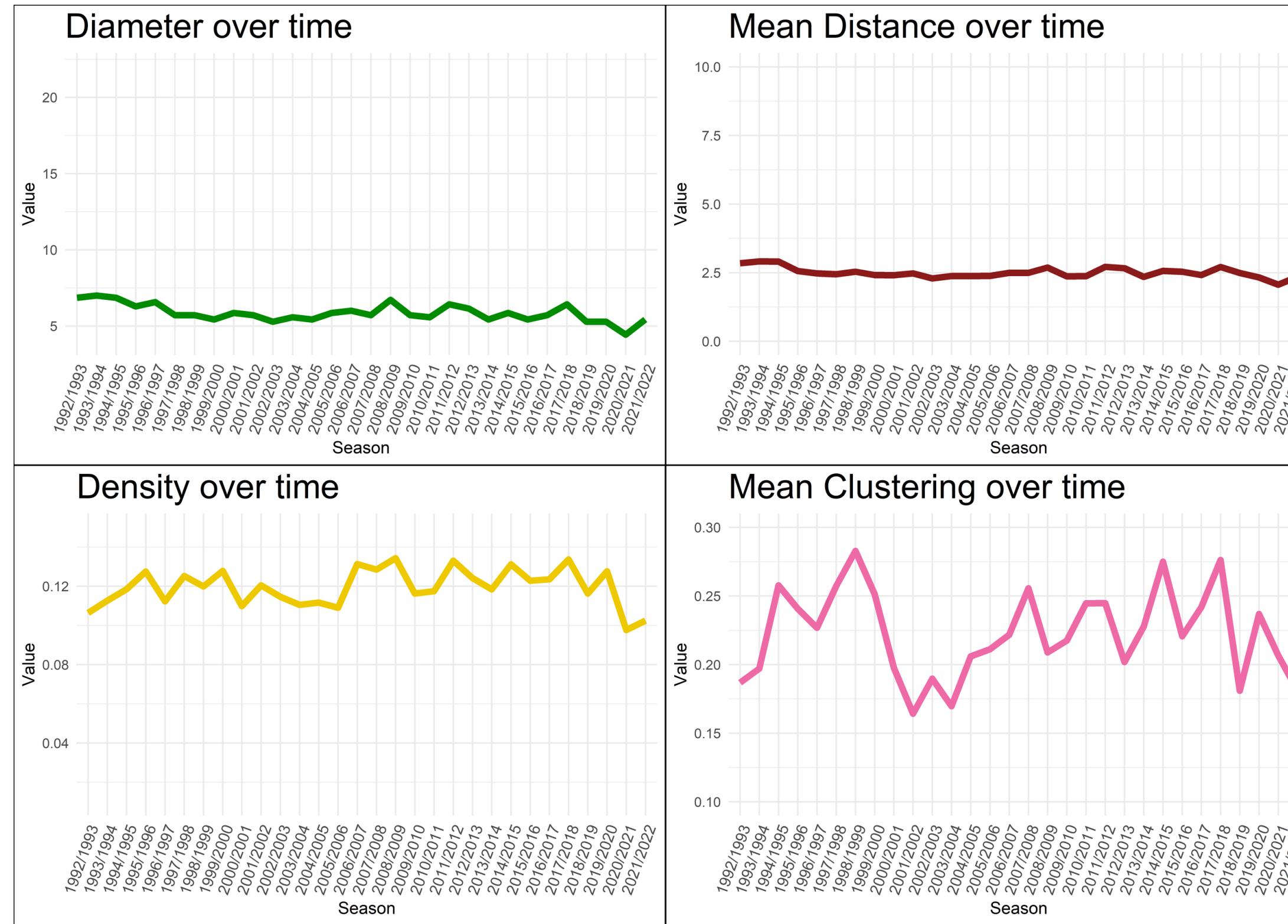
- Until 2000, no purchase had ever exceeded 50 million.
- After the purchase of G. Bale, the Market exploded: at least 75 million was spent each year on a deal.
- Neymar to PSG for 225 million is the most expensive transfer in football history.

Measures on the overall network



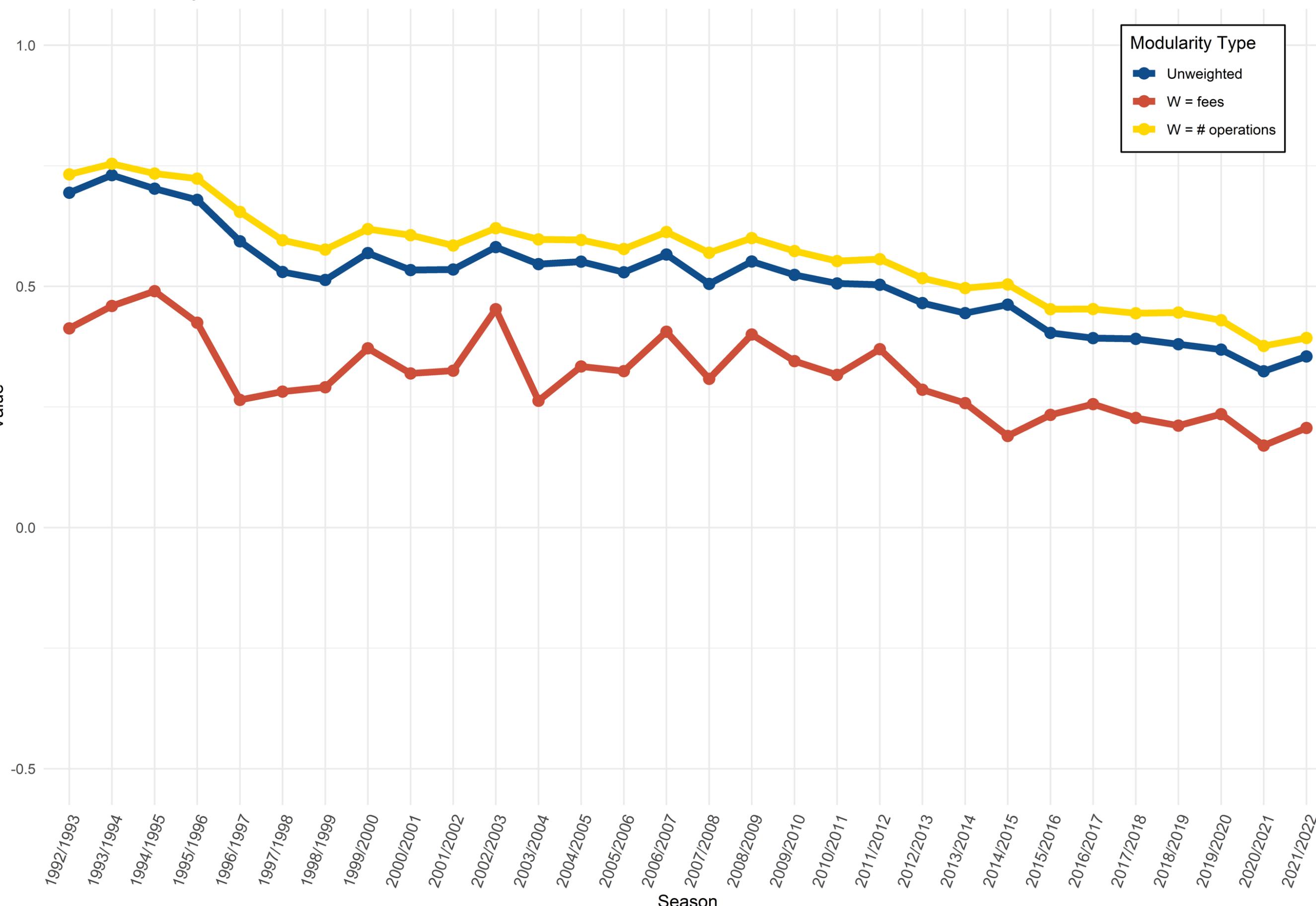
- The **Diameter** has decreased quite a bit over the seasons (from 20 to 10 and in 2017/2018 was equal to 8).
- Also the **Mean distance** decreased a lot.
- The **Density** has increased, but it's still very low.
- The **Mean clustering coefficient** has varied over the seasons, but on average, apart from the last 3 years, it has not changed much from its initial value.

Measures averaged by each league sub-network



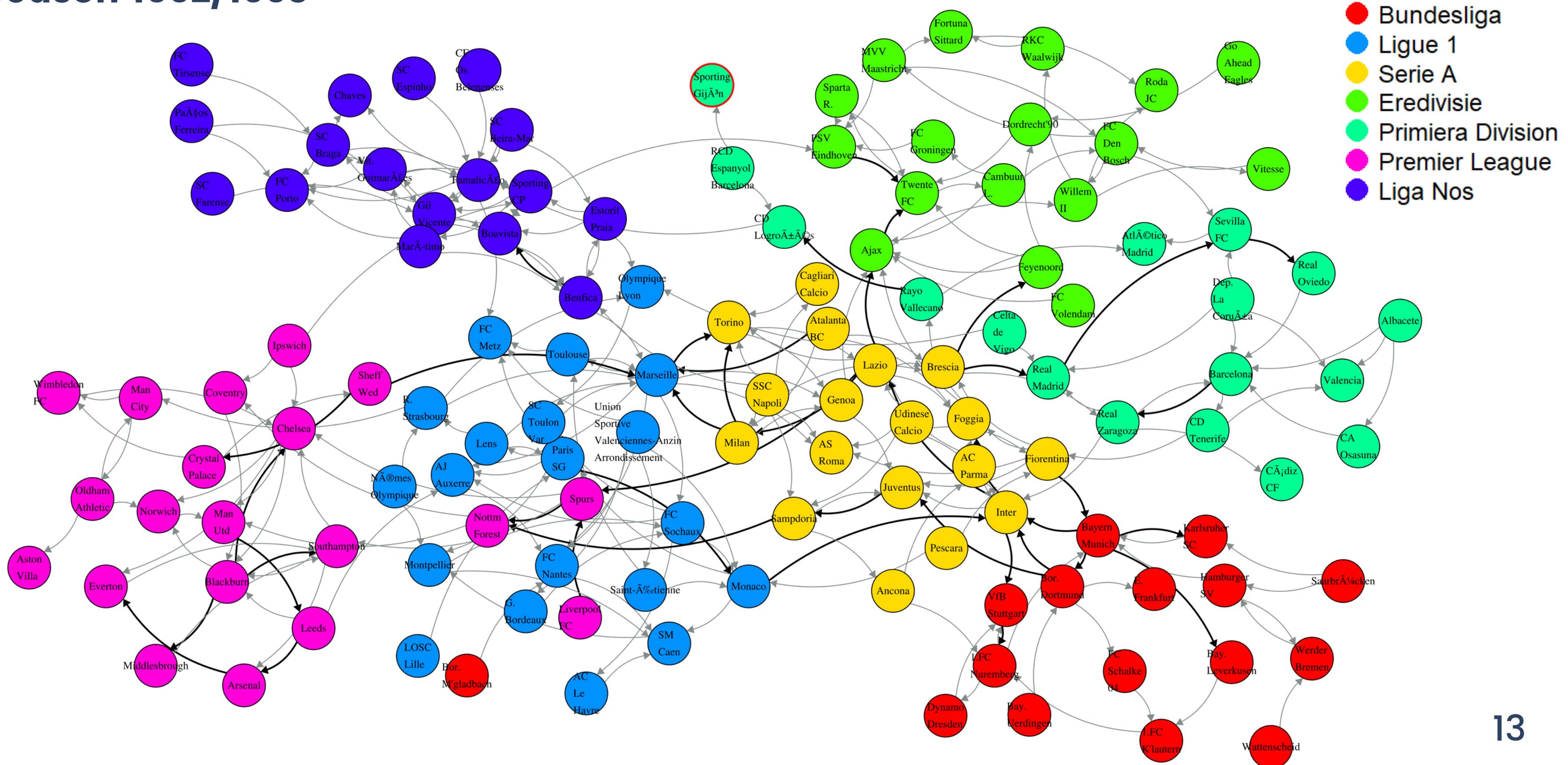
- All four measures, calculated inside the leagues remain very constant over the years.
- We can see that the **Diameter** and the **Mean Distance** are lower than those calculated in the entire network.
- The **Density** and **The Mean clustering coefficient** are higher than those in the whole network.

Modularity evolution over time

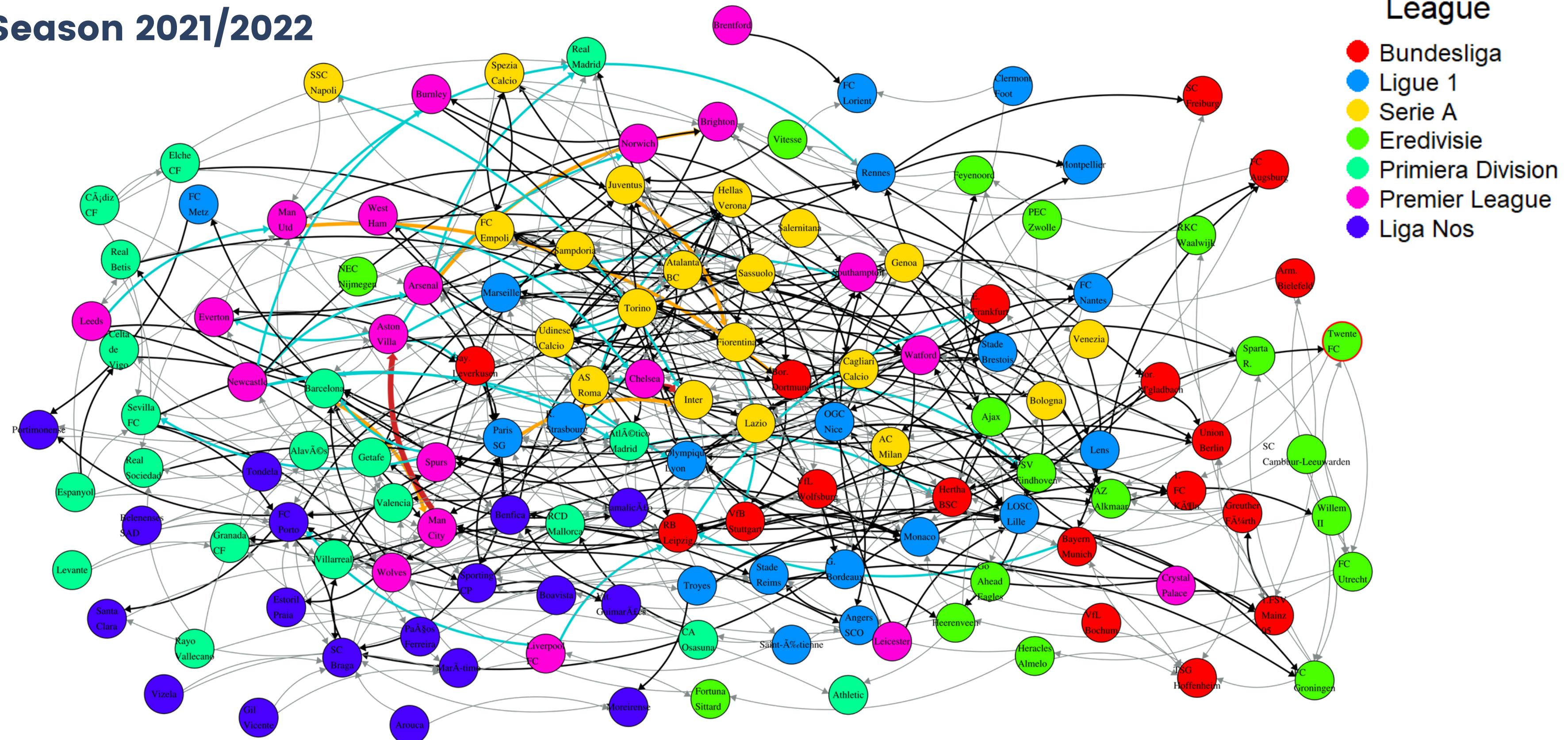


- in the initial four seasons, the score was very high.
- The results of all three cases calculated are in agreement with each other, and show a continuous decrease in this value.
- The market has become more and more global.

Season 1992/1993

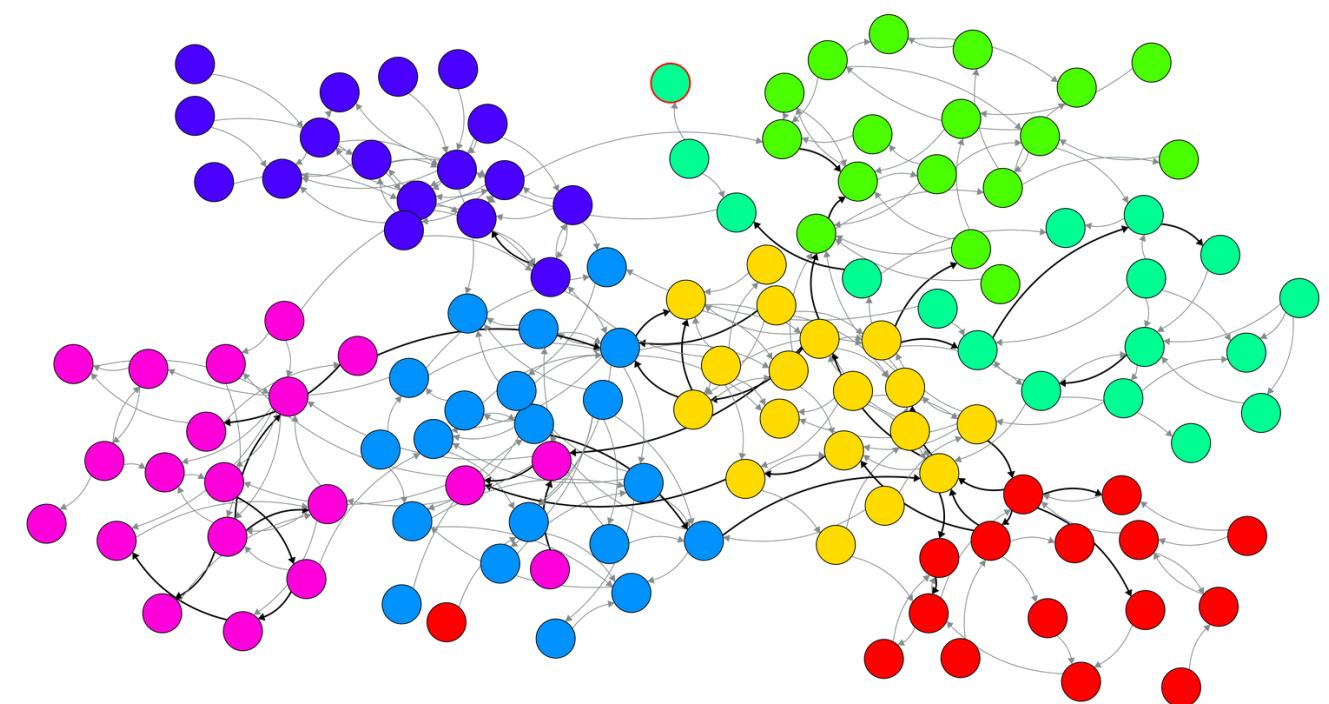


Season 2021/2022

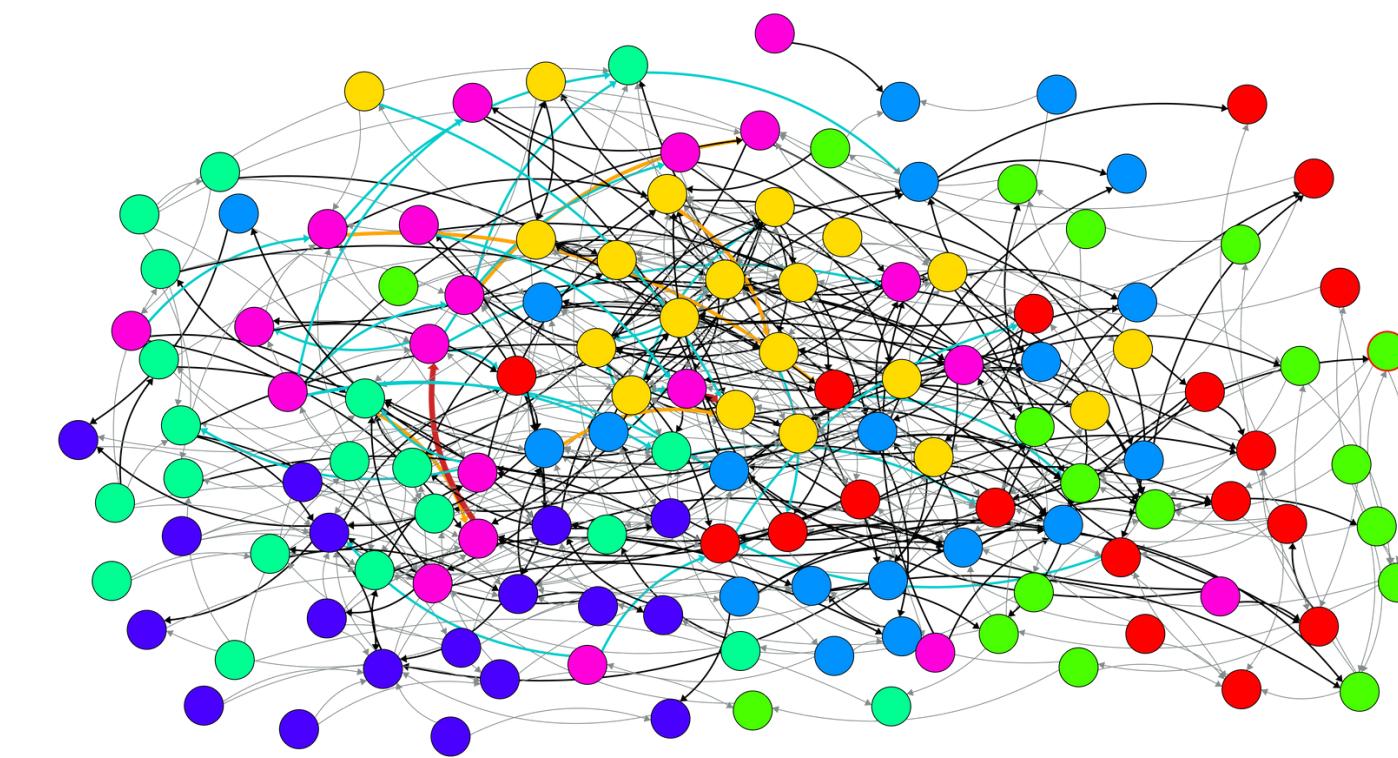


- League**
- Bundesliga
 - Ligue 1
 - Serie A
 - Eredivisie
 - Primera Division
 - Premier League
 - Liga Nos

Season 1992/1993



Season 2021/2022



- Very **closed** structure.
- Few clubs buy outside their league
- Transfers are **fewer** and the amounts are **very low**.

- The structure is very **open**, it almost looks like one league.
- There are **many** transfers and prices are often **high**.

Top 3 nodes by Betweness Centrality



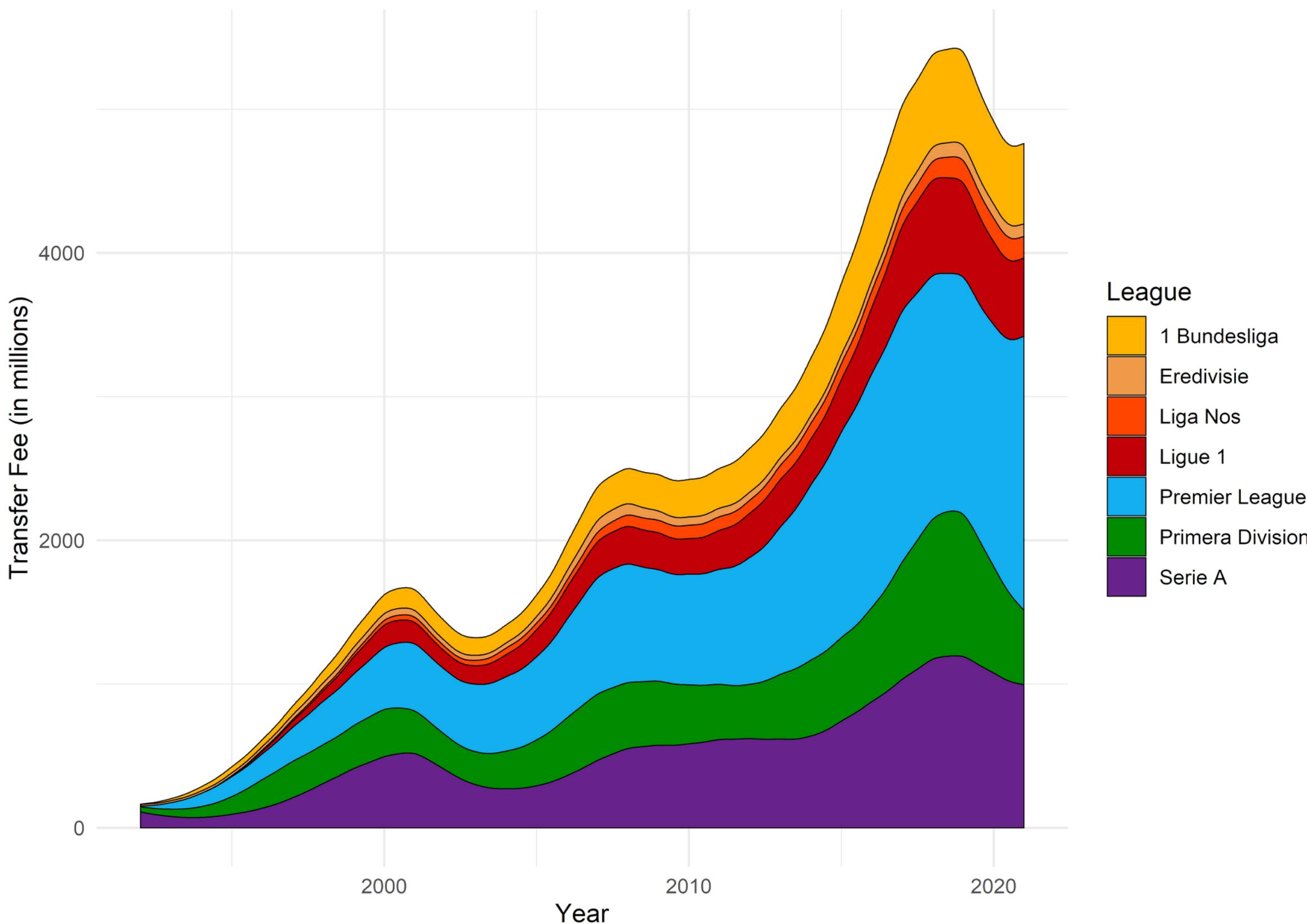
- It is interesting to note that, although each season has different negotiations, the nodes with the **highest betweness** centrality are very often the **same**.

- Almost all of them are **mid-to-high-end clubs**: this makes his players attractive to all the big teams in Europe but also to the lower-end ones.

League comparison

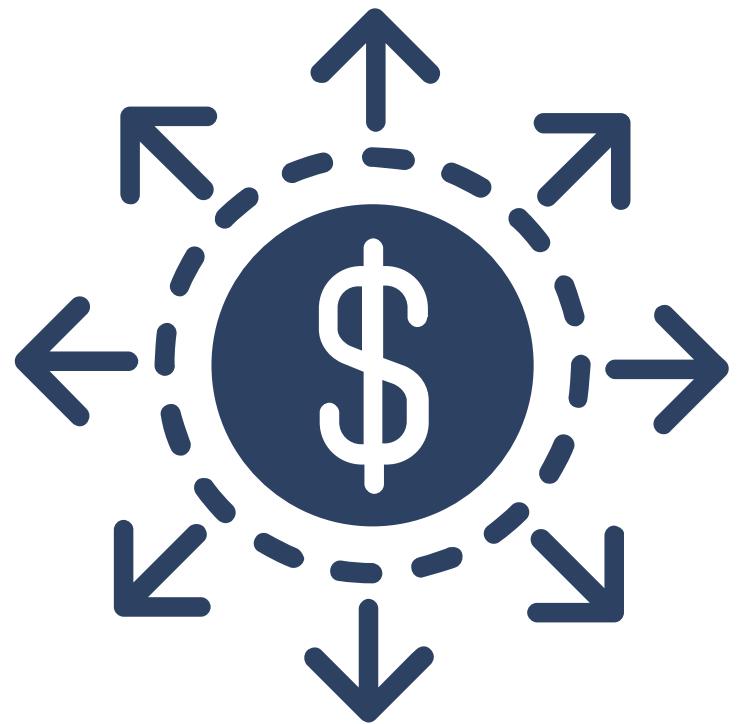
- For the first three seasons more than 50% of the money spent in European football **came from Italian teams**, from 2005 onwards this amount dropped below 25 per cent. Since then its place has been taken by the **Premier league**.
- In recent years, **more than half** of the money spent in Ligue 1 comes from PSG.
- Covid-19 **stopped** the insane growth of spending in all leagues **except the Premier league**.

Stream Plot of Transfer Fees by League and Year





The market has become increasingly **global** over the years. This is evidenced by the decreasing diameter, average distance and modularity.



Over the course of the seasons, the **changing balance** of power between the leagues can also be seen in the money spent by the various clubs in them. Furthermore, since 2013 there has been a strong **price increase**.



Some clubs like Roma, Porto and Chelsea have, in many seasons, always extremely **high betweenness centrality values**.