

Relatório: Transferências no Futebol

Augusto Dante De Carli Zolet

213356

Resumo:

Este documento especifica os detalhes de desenvolvimento e resultados de um projeto voltado para extração de conhecimento e detecção de padrões em grafos. Mais especificamente, o grafo estudado é composto por clubes de futebol, jogadores e transferências. O banco de dados utilizado é composto por dados do site Transfermarkt, sendo parte do banco originário de um repositório do Kaggle e complementado através de web scraping. Uma vez com o banco que representa o grafo, diversas análises foram feitas e um modelo de previsão simples foi treinado.

Banco de Dados:

O banco de dados é composto por 4 arquivos csv frutos de web scraping e de um repositório no Kaggle (<https://www.kaggle.com/datasets/davidcariboo/player-scores/data>) e organizados da seguinte forma:

1. transfers.csv: Este arquivo contém dados de quase 80 mil transferências, incluindo informações como nome e id do jogador e clubes envolvidos na transferência, data em que ocorreu, os valores envolvidos, entre outras informações. A seguir está a lista de colunas de informações presentes no arquivo:

Lista de colunas:

```
player_id,transfer_date,transfer_season,from_club_id,to_club_id,from_club_name,to_club_name,transfer_fee,market_value_in_eur,player_name
```

Exemplo de linha:

```
149729,2025-07-01,25/26,294,114,Benfica,Besiktas,2000000.000,3800000.000,João Mário
```

Esses dados podem ser interpretados como arestas de um grafo, seja um grafo em que todos os nós são clubes, seja um grafo em que os nós são jogadores e clubes.

2. player_values.csv: Este arquivo contém cerca de 500 mil dados do histórico de valorização de jogadores, incluindo o id do jogador, a data e o valor de mercado no momento.

Lista de colunas:

player_id,date,market_value_in_eur

Exemplo de linha:

10,2004-10-04,7000000

A coleta destes dados ocorreu através de web scraping no site transfermarkt. Os passos para realização desta coleta estão descritos a seguir:

a. Utilização da ferramenta inspecionar:

Ao acessar o endereço:

www.transfermarkt.com.br/<nome-do-jogador>/marktwertverlauf/spieler/<id-do-jogador>

É possível visualizar um gráfico com o histórico de valores do jogador, como ilustrado na figura abaixo:



Figura 1: Interface do site fonte dos dados.

Ao ativar a ferramenta inspecionar do navegador, abrir a aba “Network” e então recarregar a página, é possível ver todos os arquivos baixados através de requests HTTP feitos pela página. Ou seja, os dados utilizados para montagem do gráfico precisam estar presentes nessa lista. Buscando manualmente, arquivo por arquivo dessa lista, foi encontrado um json com os dados de histórico de valorização do jogador, como pode ser visto na imagem a seguir:

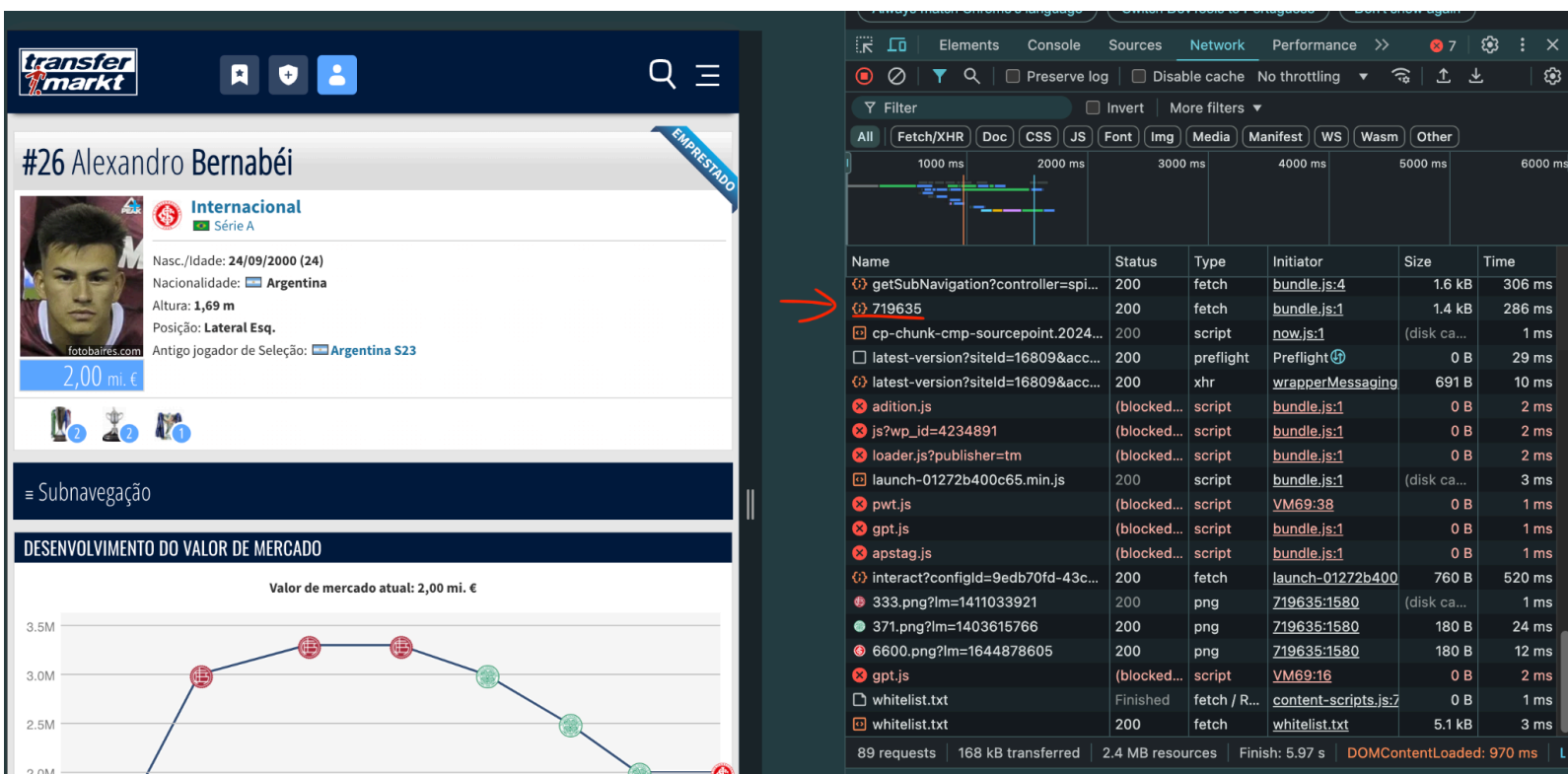


Figura 2: Utilização da ferramenta inspecionar

Clicando neste arquivo, uma nova guia é aberta no endereço utilizado para o request e o conteúdo do objeto retornado é apresentado pelo navegador, neste caso obtemos um json com os dados utilizados para montagem do gráfico.

```

https://www.transfermarkt.com.br/ceapi/marketValueDevelopment/graph/719635

Estilos de formatação ☒

{
  "list": [
    {
      "x": 1584658800000,
      "y": 500000,
      "mw": "500 mil €",
      "datum_mw": "20/03/2020",
      "verein": "CA Lanús",
      "age": "19",
      "wappen": "https://tmssl.akamaized.net/images/wappen/profil/333.png?lm=1411033921"
    },
    {
      "x": 1586296800000,
      "y": 450000,
      "mw": "450 mil €",
      "datum_mw": "08/04/2020",
      "verein": "CA Lanús",
      "age": "19",
      "wappen": ""
    }
  ]
}

```

Figura 3: JSON de dados de valorização.

Com isso, foi possível escrever um script python que faz o request em www.transfermarkt.com.br/ceapi/marketValueDevelopment/graph/<id-do-jogador> para cada player id presente nos outros arquivos csv e armazenar os dados de interesse em uma nova planilha. A seguir, um print das principais operações realizadas pelo script descrito:

```
for player_id in players["player_id"]:
    cont = cont + 1
    data_url = f"{base_url}{player_id}"
    response = requests.get(data_url, headers=headers)

    if response.status_code == 200:
        json_data = response.json() # Converte a resposta em JSON para um dicionário Python

        # Percorre cada item da lista "list" no JSON
        for item in json_data["list"]:
            data_extracted.append({
                "player_id": player_id,
                "value": item.get("mw", ""),
                "date": item.get("datum_mw", ""),
                "club_name": item.get("verein", ""),
                "age": item.get("age", "")
            })
```

Figura 4: Trecho que realiza o request e extrai dados do json.

Ao final, um outro script foi utilizado para formatar as informações visando a coerência com as demais planilhas.

3. players.csv: Esta planilha armazena informações básicas de cerca de 30 mil jogadores, como nome, id, país onde nasceu, data de nascimento, entre outras informações.

Lista de colunas:

```
player_id,first_name,last_name,name,last_season,current_club_id,player_code,
country_of_birth,city_of_birth,country_of_citizenship,date_of_birth,sub_posi
tion,position,foot,height_in_cm,contract_expiration_date,agent_name,image_ur
l,url,current_club_domestic_competition_id,current_club_name,market_value_in
_eur,highest_market_value_in_eur
```

Exemplo de linha:

```
10,Miroslav,Klose,Miroslav
Klose,2015,398,miroslav-klose,Poland,Opole,Germany,1978-06-09
00:00:00,Centre-Forward,Attack,right,184,,ASBW Sport
Marketing,https://img.a.transfermarkt.technology/portrait/header/10-14484682
```

```
91.jpg?lm=1,https://www.transfermarkt.co.uk/miroslav-klose/profil/spieler/10
,IT1,Società Sportiva Lazio S.p.A.,1000000,30000000
```

4. clubs.csv: Esta planilha armazena informações básicas de cerca de 440 clubes, como nome, id, competição nacional que disputa, entre outras informações.

Lista de colunas:

```
club_id,club_code,name,domestic_competition_id,total_market_value,squad_size
,average_age,foreigners_number,foreigners_percentage,national_team_players,s
tadium_name,stadium_seats,net_transfer_record,coach_name,last_season,filenam
e,url
```

Exemplo de linha:

```
105,sv-darmstadt-98,SV Darmstadt 98,L1,,27,25.6,13,48.1,1,Merck-Stadion am
Böllenfalltor,17810,+€3.05m,,2023,..../data/raw/transfermarkt-scraper/2023/clu
bs.json.gz,https://www.transfermarkt.co.uk/sv-darmstadt-98/startseite/verein
/105
```

É importante destacar que, apesar de serem dados com origem de um dos maiores sites de informações futebolísticas da internet, a própria natureza dos dados implica que haverá informações ausentes, como por exemplo transferências que publicamente sabe-se que ocorreram, mas os detalhes como valores são mantidos em confidencialidade. Além disso, muitos dados relacionados a futebol passaram a ser gerados e armazenados de forma organizada recentemente, sendo assim, há um número menor de transferências antigas registradas, por exemplo.

É possível construir, por fim, alguns tipos de grafos utilizando esses dados. Entre eles, se destacam o grafo composto por nós de clubes e arestas representando transferências de um clube para outro e o segundo composto por nós de clubes e jogadores, sendo as arestas usadas para representar o vínculo entre o jogador e o clube. Em ambos os casos pode-se utilizar o caráter temporal dos dados, entretanto isso torna a representação um pouco mais complexa.

Grafo de Transferências entre Clubes (Amostra)

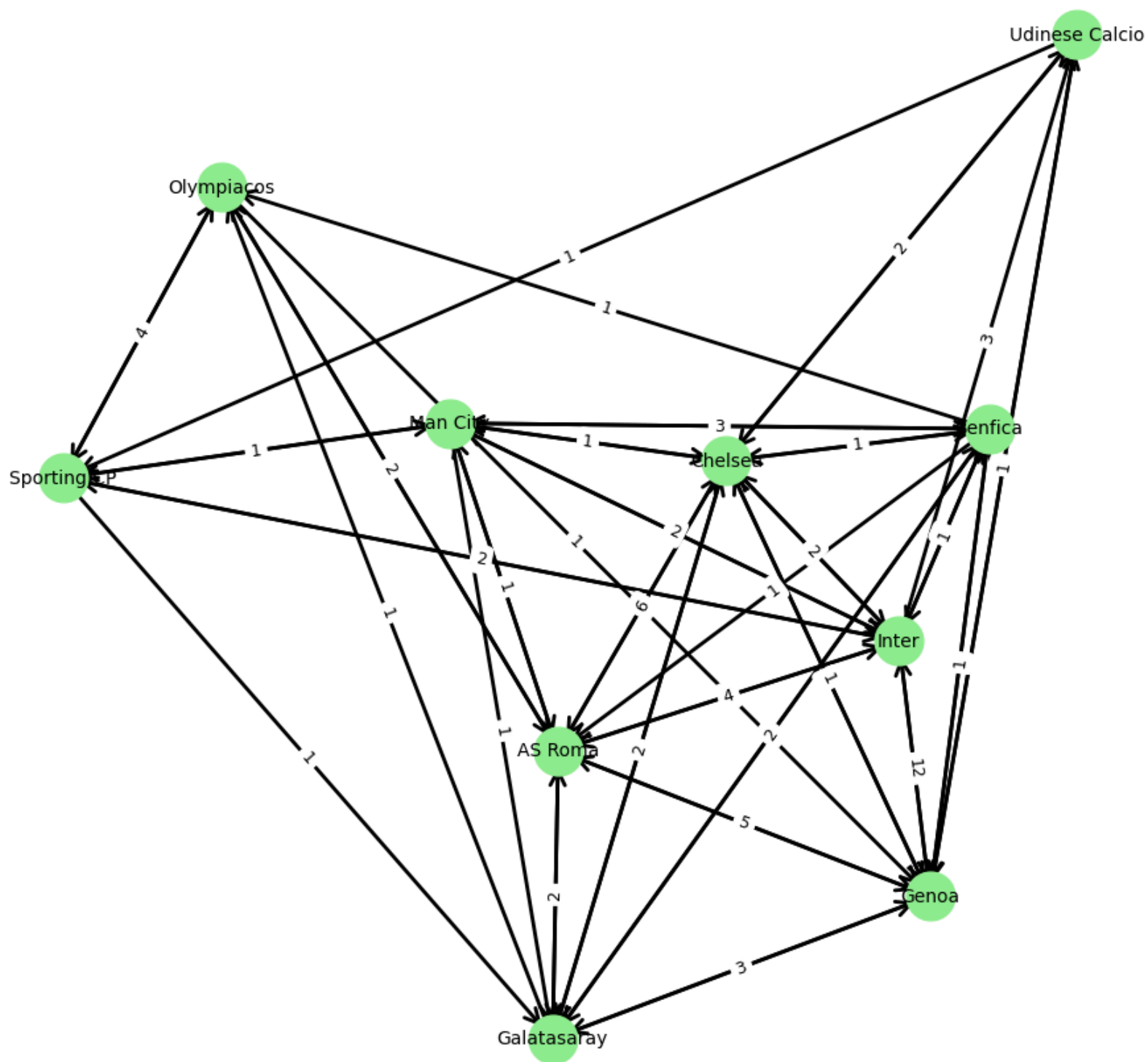


Figura 5: Grafo de transferências entre clubes. Peso das arestas representa o número de transferências.

Grafo de Jogadores para Clubes (Amostra)

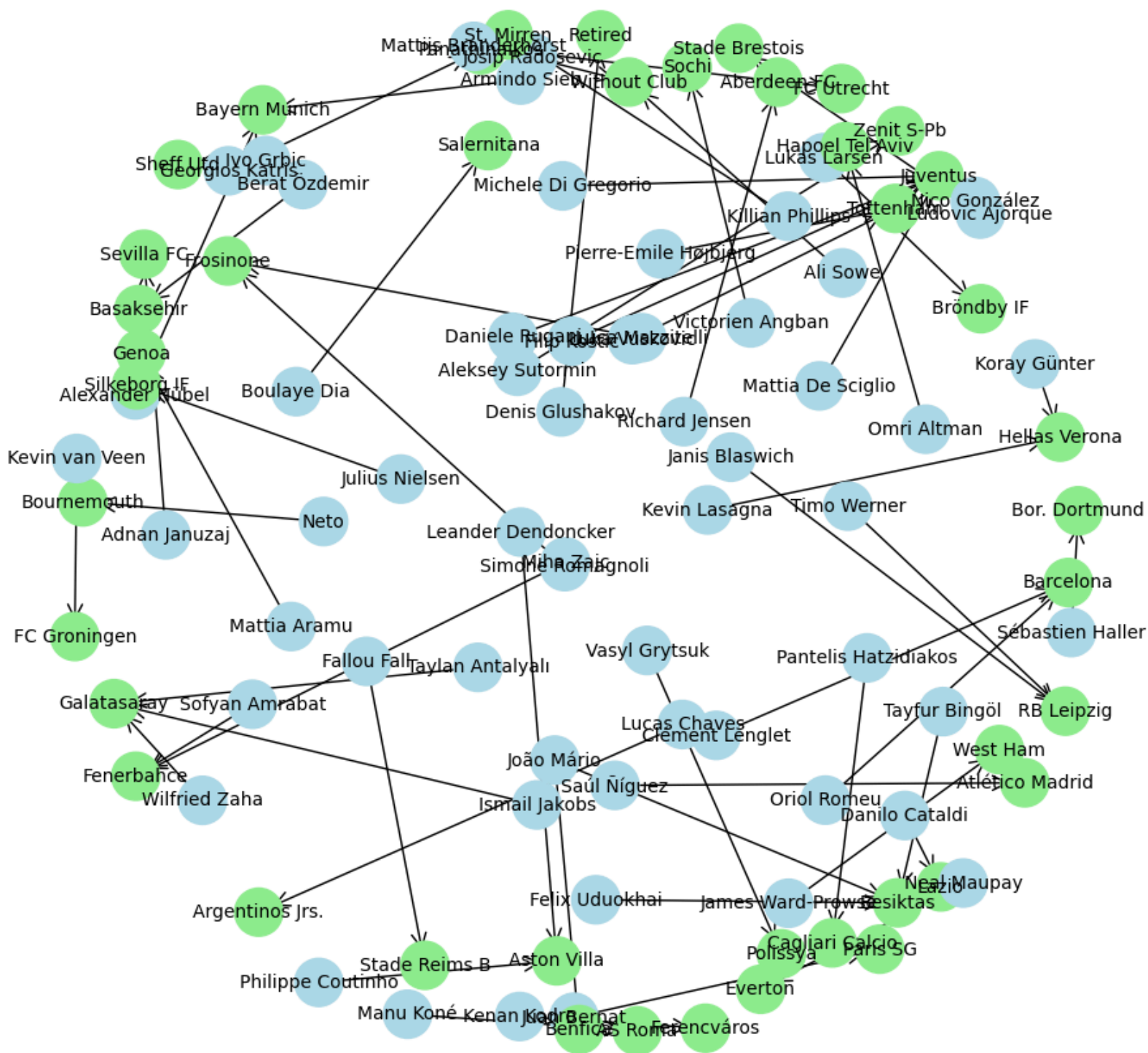


Figura 6: Grafo de vínculos entre clubes e jogadores.

Análises:

Uma vez com os dados em mãos e organizados de forma a representar um ou mais grafos, alguns exercícios de análise foram realizados visando detectar padrões e extrair conhecimentos de diversos tipos. Esses exercícios aconteceram na plataforma Google Collab, dado que nela é possível obter uma maior organização do trabalho e também manter os progressos obtidos armazenados de forma segura na nuvem.

Além disso, muitas bibliotecas Python foram utilizadas, como Pandas para carregar e manipular os dados em data frames e matplotlib para geração de gráficos.

Análise 1: Distribuição do Valor de Mercado dos Jogadores

Visando entender como ocorrem os movimentos de valorização e desvalorização de jogadores, bem como a distribuição desses números nos dados, inicialmente foram aplicados alguns filtros para obter o maior valor de mercado por jogador e, usando ferramentas do matplotlib, foram gerados gráficos que apresentam o “Número de jogadores” x “Maior valor de mercado”:

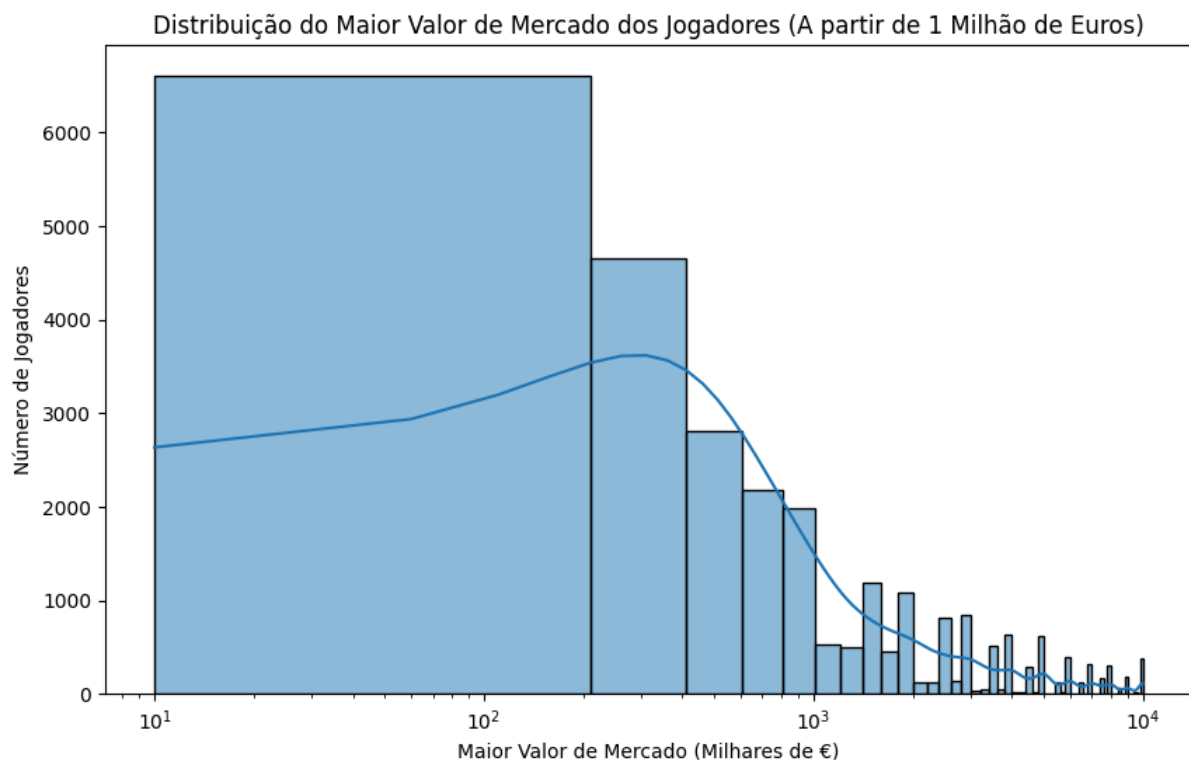


Figura 7: Distribuição do Maior Valor de Mercado dos Jogadores (A partir de 1 Milhão de Euros).

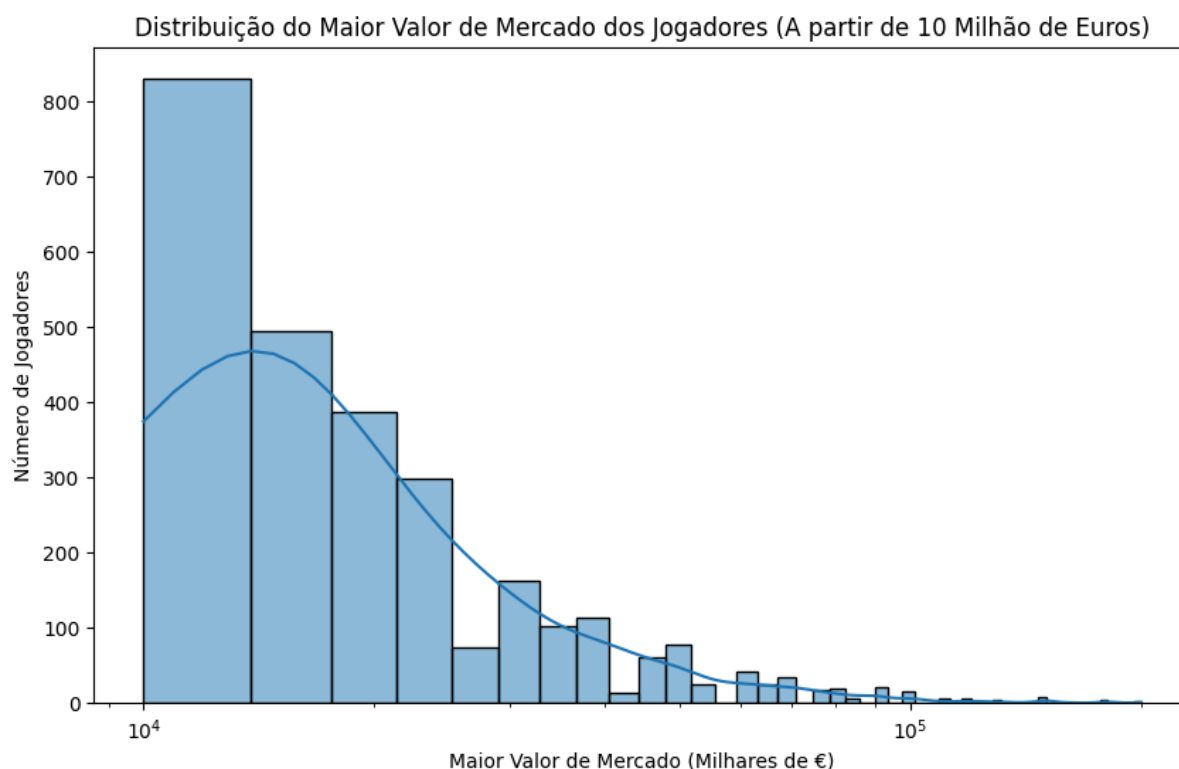


Figura 8: Distribuição do Maior Valor de Mercado dos Jogadores (A partir de 10 Milhões de Euros).

Nota-se, a partir desses gráficos, que ambas as distribuições mostram uma diminuição acentuada no número de jogadores conforme o valor de mercado aumenta. Esse comportamento sugere uma distribuição que lembra uma gaussiana, dado que a curva cresce até um pico e depois volta a diminuir. Entretanto, isso provavelmente ocorre devido ao fato de que uma porcentagem menor de jogadores menos relevantes possuem seus dados catalogados, sendo assim é mais interessante imaginar que a realidade seja mais semelhante com uma curva que decresce exponencialmente. Nesse contexto aparenta ser mais válido direcionar o foco para as áreas de maior valor em euros do gráfico.

De forma geral percebe-se um contexto onde poucos jogadores têm valores de mercado extremamente altos, enquanto a maioria está concentrada em valores médios e baixos.

A curva suavizada nas distribuições mostra uma tendência decrescente, indicando que quanto maior o valor de mercado, menor é a quantidade de jogadores que atingem esse valor, algo completamente plausível que mostra que, mesmo que não perfeitamente, os dados refletem a realidade de certa forma.

Uma forma adicional de avaliar essa questão é usando porcentagem da representação. Essa abordagem alternativa também indica para as mesmas conclusões.

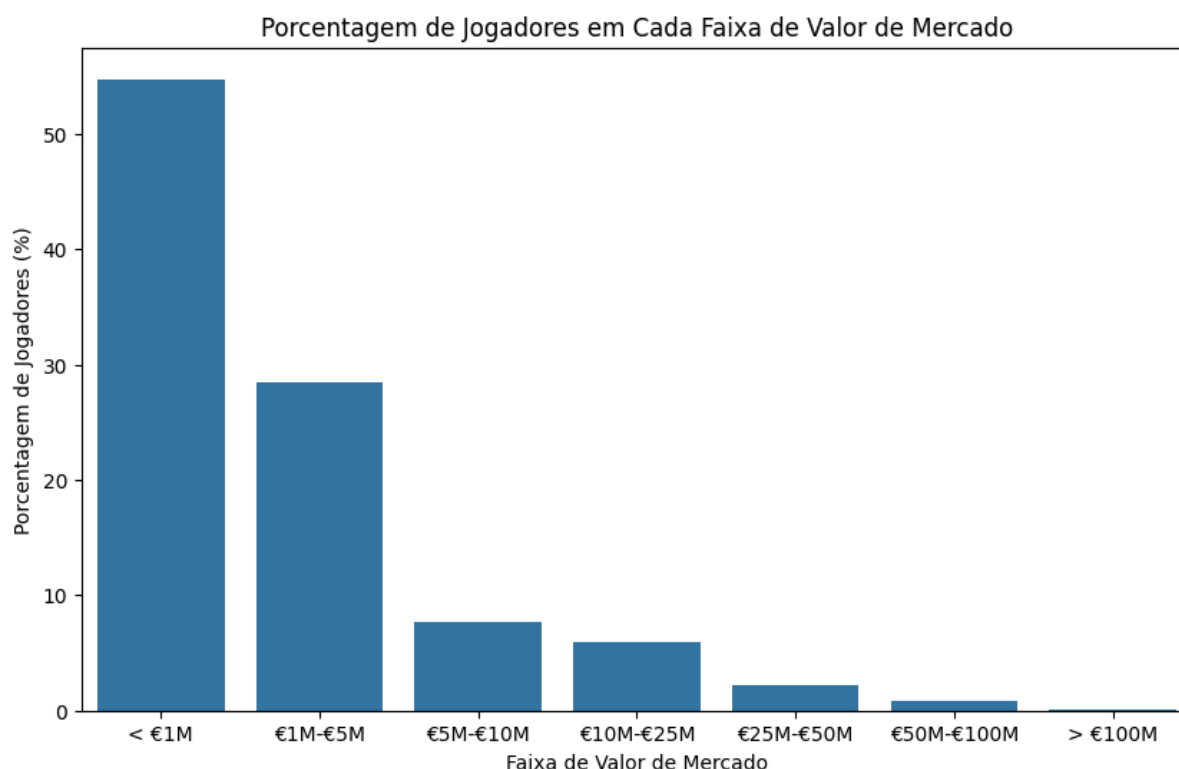


Figura 9: Porcentagem de jogadores em cada faixa de valor de mercado.

Porcentagem de jogadores com valor de mercado \geq €1,000,000: 45.28%

Porcentagem de jogadores com valor de mercado \geq €5,000,000: 16.89%

Porcentagem de jogadores com valor de mercado \geq €10,000,000: 9.22%

Porcentagem de jogadores com valor de mercado \geq €25,000,000: 3.24%

Porcentagem de jogadores com valor de mercado \geq €50,000,000: 1.00%

Porcentagem de jogadores com valor de mercado \geq €100,000,000: 0.15%

Análise 2: Evolução do Valor de Mercado de um Jogador ao Longo do Tempo

Neste conjunto, o tempo é um elemento importante, uma vez que temos vários dados de valorização e transferências envolvendo mesmos clubes e jogadores, sendo a diferença entre eles a data em que ocorreram. Com isso em mente, o primeiro passo para entender um pouco melhor a progressão de alguns destes dados ao longo do tempo é avaliar como ocorre a evolução no tempo do valor de mercado de alguns jogadores.

Sendo assim, gráficos envolvendo o histórico de valorização de alguns jogadores foram criados visando evidenciar padrões iniciais dos dados.

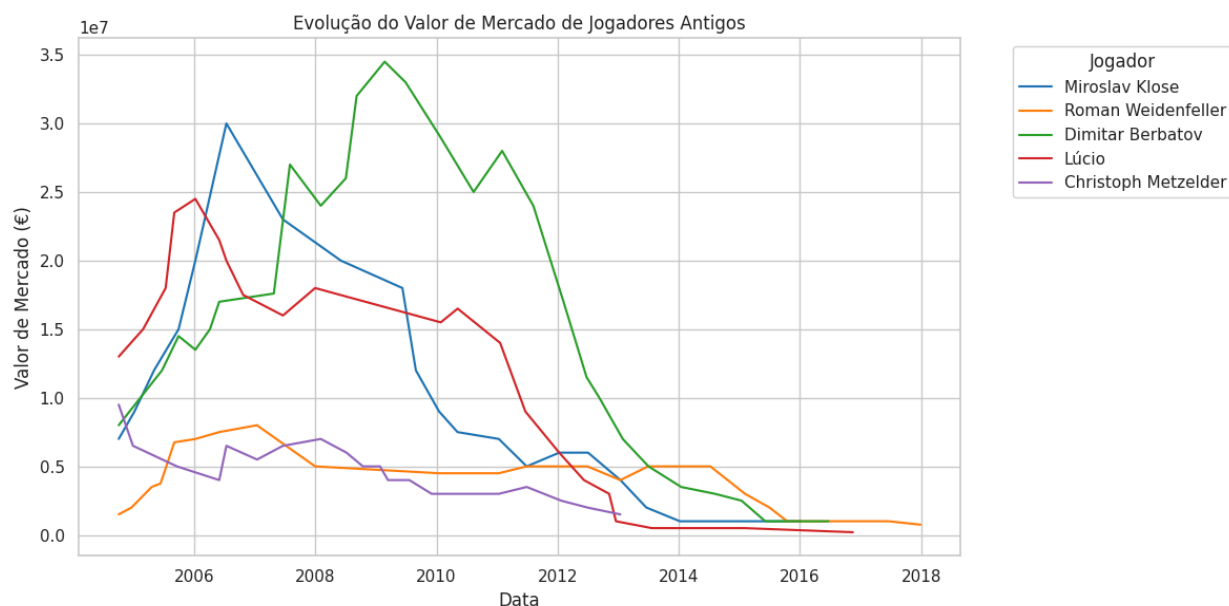


Figura 10: Evolução de valor de mercado de jogadores antigos.

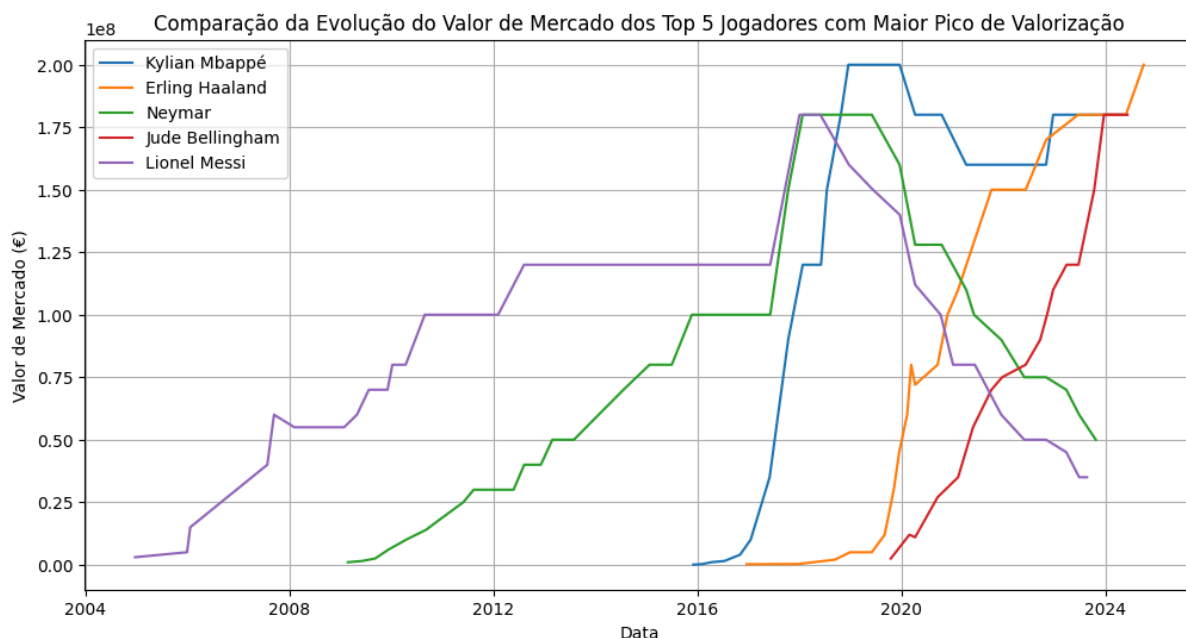


Figura 11: Evolução de valor de mercado de jogadores mais valorizados.

Estas curvas de valorização de jogadores apresentam diferentes comportamentos, entretanto há uma indicação forte para confirmar a hipótese de que de fato, conforme se aproximam do fim da carreira, a valorização dos jogadores diminui. Além disso, alguns jogadores encontram um crescimento acentuado no início da carreira até encontrarem seu ápice. É importante destacar, porém, que isso não é uma análise suficiente para cravar com certeza que todos os jogadores obedecem essa tendência de comportamento, haja vista que no próprio gráfico da figura 10, a curva pertencente ao jogador Christoph Metzelder, começa no auge e se mantém relativamente estável ao longo da carreira.

Essa breve análise é suficiente para concluirmos que, de uma forma ou de outra, ocorrem momentos máximos de valorização e esses momentos são breves. Esse raciocínio motivou a análise a seguir que observa o tempo de duração dos momentos de maior valorização de cada jogador.

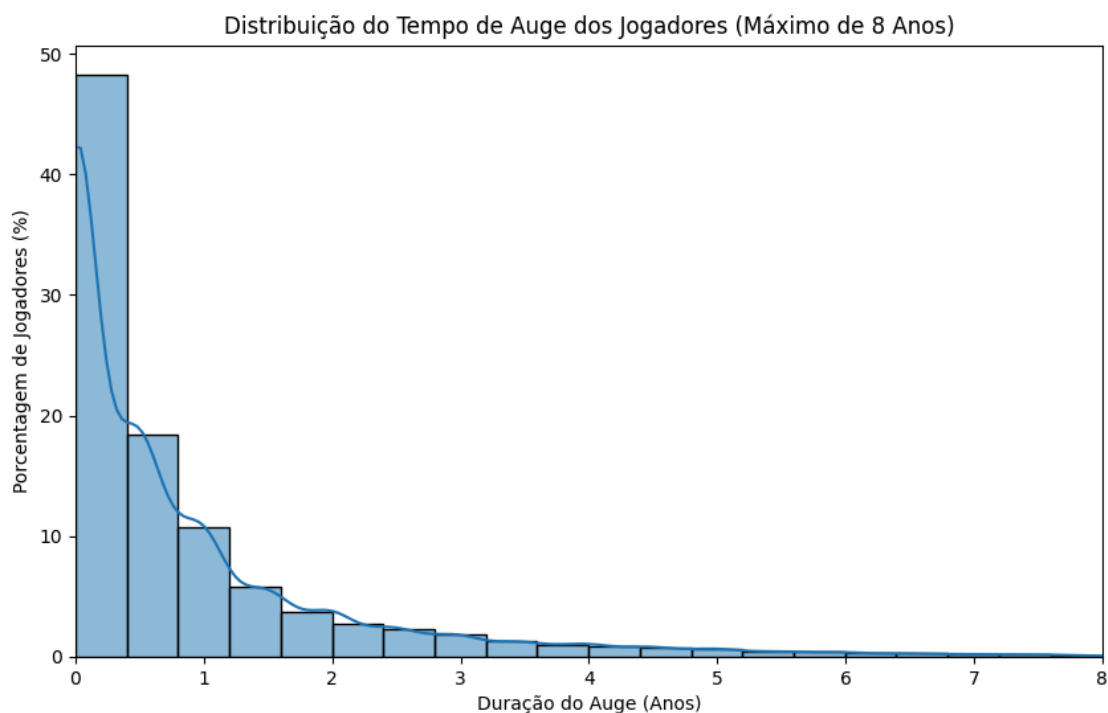


Figura 12: Distribuição do tempo de auge dos jogadores.

Estatísticas sobre a duração do auge dos jogadores (em anos, até 8 anos):

```
count:    30738
mean:     0.843653
std:      1.265307
min:      0.000000
25%:      0.000000
50%:      0.427105
75%:      1.043121
max:      7.994524
```

De forma semelhante aos exercícios de análise de distribuição do valor de mercado, a distribuição de tempo de auge também indica uma semelhança com curvas de decaimento exponencial.

Análise 3: Tempo Decorrido até a Melhor Transferência

Seguindo a linha da análise anterior que aponta para possibilidade de informações importantes serem extraídas quando se leva em consideração o aspecto temporal, nessa sessão será explorado a questão do tempo que se leva para um jogador alcançar seu ápice.

Uma vez que a idade em que o jogador estréia no futebol profissional não está diretamente relacionada com seu potencial como jogador, mas sim relacionada com as necessidades de jogadores que a equipe enfrenta, foi escolhido substituir a idade pelo tempo decorrido até a melhor transferência a partir do início da carreira profissional.

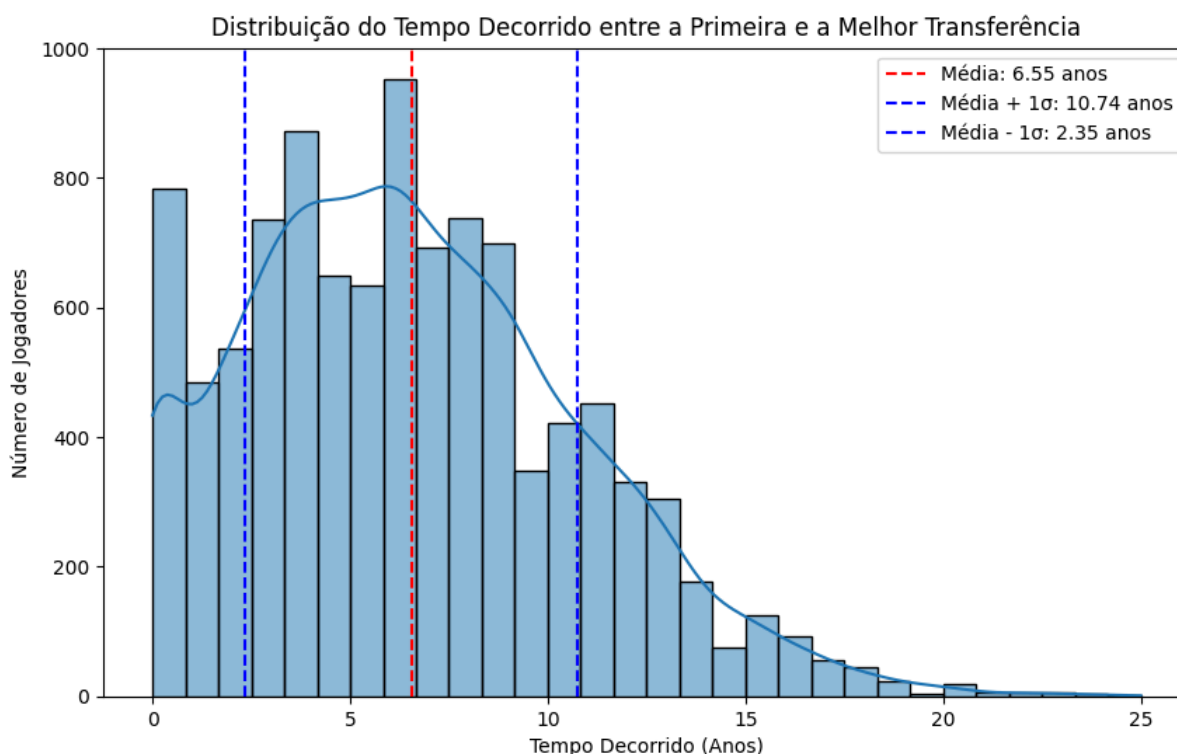


Figura 13: Distribuição do tempo decorrido entre a Primeira e a Melhor Transferência.

A partir deste gráfico, que mostra a distribuição do tempo decorrido entre a primeira e a melhor transferência dos jogadores, algumas conclusões e análises interessantes podem ser feitas.

A média está em torno de 6,55 anos após a primeira transferência, indicando que a maioria dos jogadores leva um tempo considerável para atingir o ponto alto de valor de mercado ou transferência. Combinado a isso, a curva está levemente inclinada para a direita, isso significa que há um número significativo de jogadores que leva mais tempo do que a média para alcançar a melhor transferência, embora a maioria esteja próxima da média.

Além disso, as linhas de desvio padrão indicam que grande parte dos jogadores atinge a melhor transferência entre 2,35 e 10,74 anos após a primeira. Essa faixa cobre aproximadamente 68% dos jogadores se aproximarmos essa curva a uma distribuição normal. Isso pode estar relacionado ao desenvolvimento de habilidades e à experiência que aumenta o valor dos jogadores no mercado.

É interessante notar que para as principais transferências das últimas temporadas que geralmente envolvem jogadores muito jovens que representam grandes promessas, como pode ser visto nas curvas dos jogadores jovens de valorização repentina e precoce Erling Haaland, Jude Bellingham e Kylian Mbappe no gráfico de valorização da figura 11, podem deixar a impressão que esse é um processo comum: por estar no início da carreira, o jogador se valoriza pelo potencial. Entretanto, o gráfico da figura 13 indica que, para maioria dos jogadores, o processo é mais lento e mais relacionado com o desempenho apresentado do que com a estimativa de potencial.

Como próximo passo, inspirado pela análise do gráfico anterior, foi feita uma tentativa de criação de um modelo de aprendizado de máquina simples para prever o tempo decorrido (em dias) entre a primeira transferência de um jogador de futebol e sua transferência de maior valor. Para isso, utilizamos dados detalhados dos jogadores, clubes e transferências.

As variáveis selecionadas para treinar o modelo foram dados como o valor da primeira transferência do jogador, a altura em centímetros, o clube origem e o clube destino da primeira transferência, a posição e subposição em que atua, variáveis sobre o local de nascimento e cidadania e o pé dominante (destro, canhoto ou ambidestro). Nossa variável alvo foi o tempo entre a primeira e a transferência mais valiosa de cada jogador.

Para o pré-processamento dos dados, utilizamos StandardScaler para padronizar variáveis numéricas e OneHotEncoder para transformar as variáveis categóricas em dummies.

O modelo escolhido foi o Random Forest Regressor, que demonstrou boa capacidade de lidar com a variabilidade dos dados. Dividimos o conjunto de dados em amostras de treinamento, validação e teste usando train test split, e configurando o random state para garantir a reprodutibilidade dos resultados. Avaliamos o modelo com as métricas MAE (erro médio absoluto) e RMSE (raiz do erro quadrático médio), para entender a precisão das previsões em termos absolutos e com penalização para erros maiores.

Para realizar esse procedimento, foram utilizadas as bibliotecas scikit-learn e torch.

O resultados obtidos estão apresentados a seguir:

Desempenho no conjunto de validação:

MAE: 854.18 dias

RMSE: 1106.54 dias

Desempenho no conjunto de teste:

MAE: 875.95 dias

RMSE: 1173.91 dias

Exemplos de previsões no conjunto de teste:

id	Real	Previsto	Erro Absoluto
8046	697	1480.65	783.65
7808	2922	1725.93	1196.07
3658	2495	1233.61	1261.39
318	6040	2476.90	3563.10
3014	1868	2529.90	661.90
1078	2753	2347.37	405.63
8946	515	851.65	336.65
3880	2203	2262.51	59.51
5651	2924	3661.34	737.34

7847 1826 1709.30 116.70

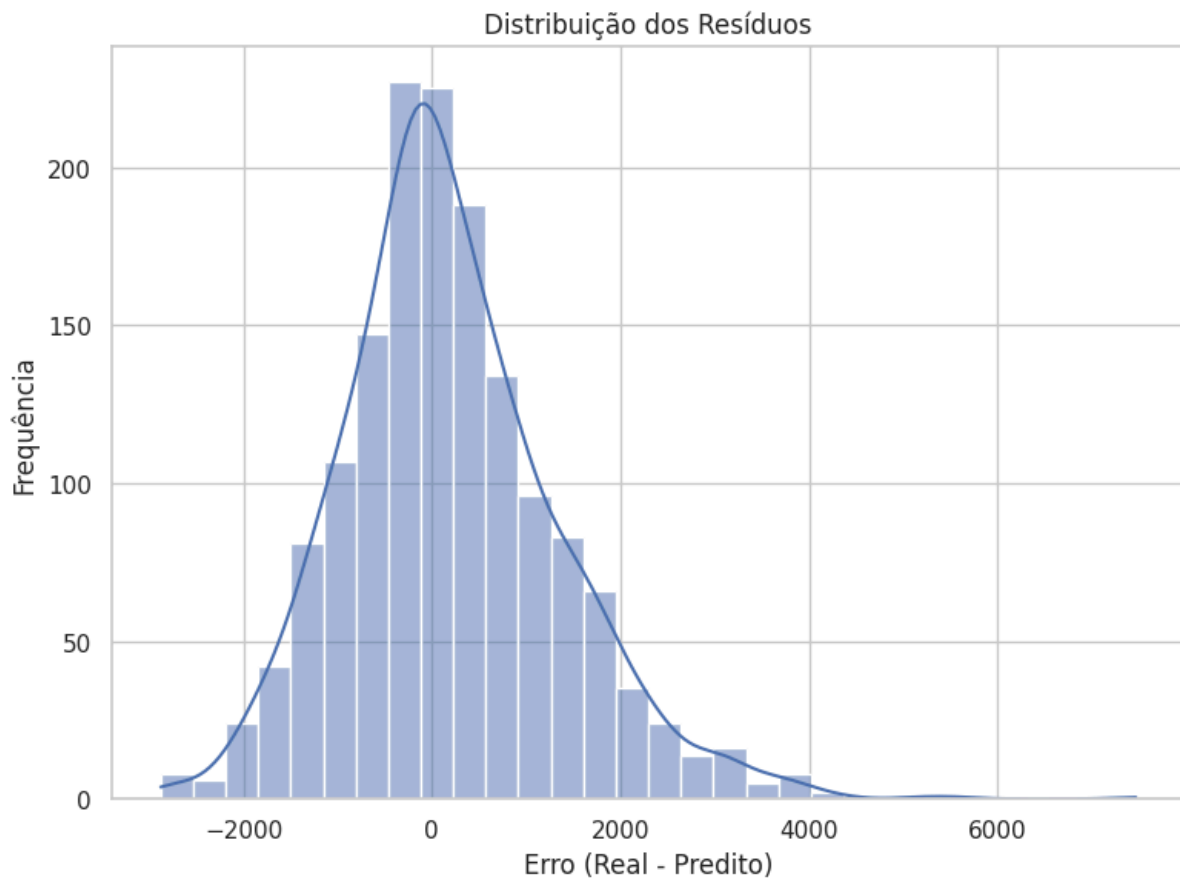


Figura 14: Distribuição de Resíduos (Erro x Frequência).

Analisando os exemplos de previsões no conjunto de teste e as estatísticas de erro, observamos que o modelo é capaz de fornecer estimativas próximas aos valores reais em alguns casos, como o jogador com ID 3880, que teve um erro absoluto de apenas 60 dias. No entanto, em outros casos, os erros são mais significativos, como no jogador com ID 318, onde o erro absoluto ultrapassou 3500 dias. Essa variação sugere que, embora o modelo capture certos padrões nos dados, talvez seja preciso uma abordagem alternativa e um conjunto de dados mais robusto.

Pode-se dizer que o modelo desenvolvido demonstra potencial na tarefa proposta, de forma que aponta, com uma boa margem de erro, o momento da carreira que o jogador conquistará sua melhor transferência. Considerando que a questão abordada é muito complexa e alguns elementos muito importantes não estão presentes nos dados, como relacionamentos entre empresários e jogadores, talvez uma abordagem mais interessante para esses dados, possivelmente, seria um modelo de link prediction em que, dado o grafo temporal de jogadores, clubes e arestas que representam as transferências ou vínculos entre jogadores e clubes, o algoritmo aprenderia com os padrões de evolução do grafo e tentaria prever novas arestas a serem formadas no futuro.

De fato houve uma tentativa, fracassada, de aplicar o método de link prediction, porém no processo de tentativa também foi constatado que o número de “snapshots”

relevantes (representações do grafo para determinados momentos no tempo) possivelmente seria insuficiente para o treinamento.

Análise 4: Centralidade e Comunidades

A figura 5 “ *Grafo de transferências entre clubes*” apresenta uma possibilidade de interpretação dos dados em forma de um grafo em que os nós representam clubes e as arestas representam as transferências entre eles.

Nesse contexto, foram aplicadas algumas análises para observar as propriedades do grafo. Primeiramente, o foco foi direcionado para atributos mais básicos, como graus de entrada e saída de cada nó. Entradas (In-Degree) representam transferência em que o clube destino é o nó avaliado e saída (Out-Degree) são as transferências de jogadores que estão deixando o nó.

Top 10 Clubes Compradores (In-Degree) :

id	Club	In-Degree
43	Olympiacos	140
136	Benfica	128
427	Inter	128
400	AS Roma	123
148	Genoa	121
87	Udinese Calcio	120
413	Chelsea	117
218	Galatasaray	115
70	Hellas Verona	112
165	Sporting CP	110

Top 10 Clubes Vendedores (Out-Degree) :

id	Club	In-Degree
43	Olympiacos	126
136	Benfica	124
427	Inter	114
400	AS Roma	108
165	Sporting CP	105
1078	Man City	98
413	Chelsea	97
83	Standard Liège	96
148	Genoa	96
464	Wolves	92

Analisando estes aspectos dos dados, cria-se a expectativa de que clubes como Olympiacos, Benfica, Inter, AS Roma apareçam entre os nós com maior medida centralidade de intermediação. E, de forma complementar a essa análise, foi calculado o lucro total para os clubes envolvidos nas transferências registradas, na expectativa de encontrar alguma relação com os dados da análise anterior.

Top 10 Clubes que Mais Lucraram com Transferências:

id	Clube	Lucro (euros)
556	Benfica	555700000.0
180	Ajax	514110000.0
1340	FC Porto	451920000.0
2947	RB Salzburg	397640000.0
3438	Sporting CP	372917000.0
2137	LOSC Lille	319930000.0
2687	Olympique Lyon	289380000.0
1062	Dinamo Zagreb	215255000.0
1126	E. Frankfurt	214900000.0
3074	River Plate	210169000.0

Refletindo sobre estes resultados, percebe-se que para os 10 clubes que mais lucraram a ordem de grandeza do lucro é a mesma, na casa das centenas de milhões de euros. Entretanto, o único clube presente nos três levantamentos é o Benfica, fato que indica para possibilidade de que apesar de contratar muito, o clube vende os jogadores contratados de forma significativamente mais lucrativa do que os demais clubes com grande número de transferências. Percebe-se também que, dado que o Benfica é uma exceção, a lucratividade não está necessariamente associada a um grande número de operações.

Utilizando a função *betweenness centrality* oferecida na biblioteca NetworkX, calculou-se o grau de centralidade de intermediação dos nós do grafo. Essa medida se refere à importância de um nó em termos de sua posição como intermediário nos caminhos mais curtos entre outros pares de nós no grafo. Ela indica quais nós (neste caso, clubes) são cruciais para conectar diferentes partes da rede, funcionando como “pontes” entre outros clubes.

Top 10 Clubes Intermediários (Betweenness Centrality):

id	Club	Betweenness Centrality
573	Go Ahead Eagles	0.018378
1341	Sporting Yth.	0.016833
1950	FC Twente Youth	0.016444
2632	Juventude	0.016001
5323	Esportivo	0.015579
936	Benfica U17	0.015421
5507	Juventude U20	0.015421
3282	Benfica U19	0.015045
960	Waasland-Beveren	0.014440
517	Ajax Youth	0.013211
1528	Barça Youth	0.012919
2469	Arsenal Youth	0.012711
1141	Chelsea Youth	0.012233
4144	Granollers	0.010847
2155	KRC Genk Youth	0.010242
115	Vejle BK	0.010139
43	Olympiacos	0.009502

796	CSKA Moscow	0.009487
2390	Vitória Setúbal	0.009439
1397	Shakhtar D.	0.009150
131	PAOK Salonika	0.008427
83	Standard Liège	0.008092
127	Fenerbahce	0.007862
1116	PEC Youth	0.007795
4520	Man City Youth	0.007702
770	PSV Youth	0.007519
698	Dynamo Kyiv	0.007169
4769	Créteil Youth	0.006988
1204	Santa Clara	0.006922
70	Hellas Verona	0.006893

Analisando esses dados de centralidade de intermediação, nota-se, primeiramente, que os graus de entrada/saída não têm tanta importância na centralidade e, além disso, que muitos clubes participam das conexões das diferentes partes da rede. Isso se dá porque a biblioteca detectou que são muitos os nós que aparecem frequentemente nos caminhos mais curtos entre outros clubes. Esses clubes, muitos dos quais são equipes de categorias de base ou “youth” (como “Sporting Yth.”, “Benfica U19”, “Ajax Youth”), funcionam como importantes pontos de passagem para jogadores que transitam entre clubes, especialmente para talentos jovens ou em desenvolvimento.

As equipes de base e os clubes de menor expressão, como “Go Ahead Eagles” e “Juventude”, atuam como intermediários que facilitam o movimento de jogadores entre clubes de maior e menor renome. Isso sugere que esses clubes são frequentemente escolhidos para desenvolver jovens talentos antes de sua transição para equipes maiores ou mais competitivas assim então posicionados como “conectores” dentro do mercado.

Um próximo passo interessante para entender melhor a estrutura do grafo que estamos trabalhando seria tentar aplicar um algoritmo de detecção de comunidades. Comunidades em grafos são grupos de nós que possuem uma maior densidade de conexões entre si, formando assim subestruturas mais conectadas internamente. Essas comunidades são úteis para identificar padrões e agrupamentos naturais dentro de uma rede e, nesse sentido, espera-se que aplicando essa análise aos dados que trabalhamos haja formação de comunidades voltadas para questões regionais, como por exemplo uma comunidade formada por clubes da Europa.

Para isso, foi utilizado o algoritmo de Louvain, implementado na biblioteca python `community_louvain`. Esse algoritmo é um método de detecção de comunidades em grafos, que visa maximizar a modularidade, ou seja, a qualidade das divisões em comunidades. Ele começa atribuindo cada nó a sua própria comunidade e, em seguida, avalia se mover um nó para a comunidade de um vizinho aumenta a modularidade; se sim, o nó é realocado. Essa fase local é repetida até que a modularidade não possa ser mais aumentada. Em uma segunda etapa, o algoritmo “colapsa” cada comunidade identificada em um único nó, formando um novo grafo simplificado que mantém as conexões entre comunidades. O processo é então repetido iterativamente em várias escalas até que a modularidade atinja um valor máximo global, revelando uma estrutura hierárquica de comunidades no grafo.

Aplicando essa técnica no grafo, obteve-se 64 comunidades diferentes. Apesar do grande número de clubes por comunidades, fica claro que o algoritmo apontou para formação de comunidades de clubes geograficamente próximos. de modo que, por exemplo, as 7 primeiras comunidades se definem por:

Comunidade 1: Predominantemente Rússia, com algumas equipes do Cazaquistão e Moldávia.

Comunidade 2: Predominantemente Ucrânia, com clubes da Eslováquia e Azerbaijão.

Comunidade 3: Predominantemente países escandinavos (Dinamarca, Suécia, Noruega) e Islândia.

Comunidade 4: Principalmente Alemanha e Áustria, com presença da Suíça.

Comunidade 5: França e algumas equipes da Bélgica e Suíça.

Comunidade 6: Grécia e Israel, com alguns clubes de Chipre.

Comunidade 7: Predominantemente Portugal e Espanha, com presença de clubes do Brasil.

Tentativas de representar graficamente essas comunidades por completo falharam em função da necessidade de alto poder computacional. Entretanto, foi possível aplicar a mesma técnica a um subgrafo dos 50 clubes mais conectados e gerar uma representação visual apresentada na figura 15 da próxima página.

As quatro comunidades apontadas foram:

Comunidade 1: Brentford, Sevilla FC, Tottenham, Galatasaray, OGC Nice, Fenerbahce

Comunidade 2: E. Frankfurt, Olympiacos, Paris SG, Besiktas, Standard Liège, Club Brugge, Everton, SC Braga, Sporting CP, Liverpool, FC Porto, Man Utd, Nottm Forest

Comunidade 3: LOSC Lille, PAOK Salonika, Wolves, Arsenal, Monaco, Brighton, Atlético Madrid, Trabzonspor, Chelsea, Celtic, Benfica, Shakhtar D., KRC Genk, RSC Anderlecht, Newcastle, Man City, VfL Wolfsburg

Comunidade 4: Fiorentina, Lazio, Udinese Calcio, AS Roma, Torino, SSC Napoli, Genoa, Hellas Verona, Inter, Atalanta BC, AC Milan, Juventus

Percebe-se que, mesmo para essa amostra extremamente limitada do grafo, o algoritmo continuou apontando para formação de comunidades regionais, sendo uma delas composta totalmente por times italianos (comunidade 4). Nesse cenário, é natural concluir que transferências de jogadores ocorrem de forma mais intensa entre clubes próximos geograficamente e, em alguns casos, de forma ainda mais intensa entre clubes dentro de um mesmo país.

Comunidades Detectadas no Grafo de Transferências entre Clubes

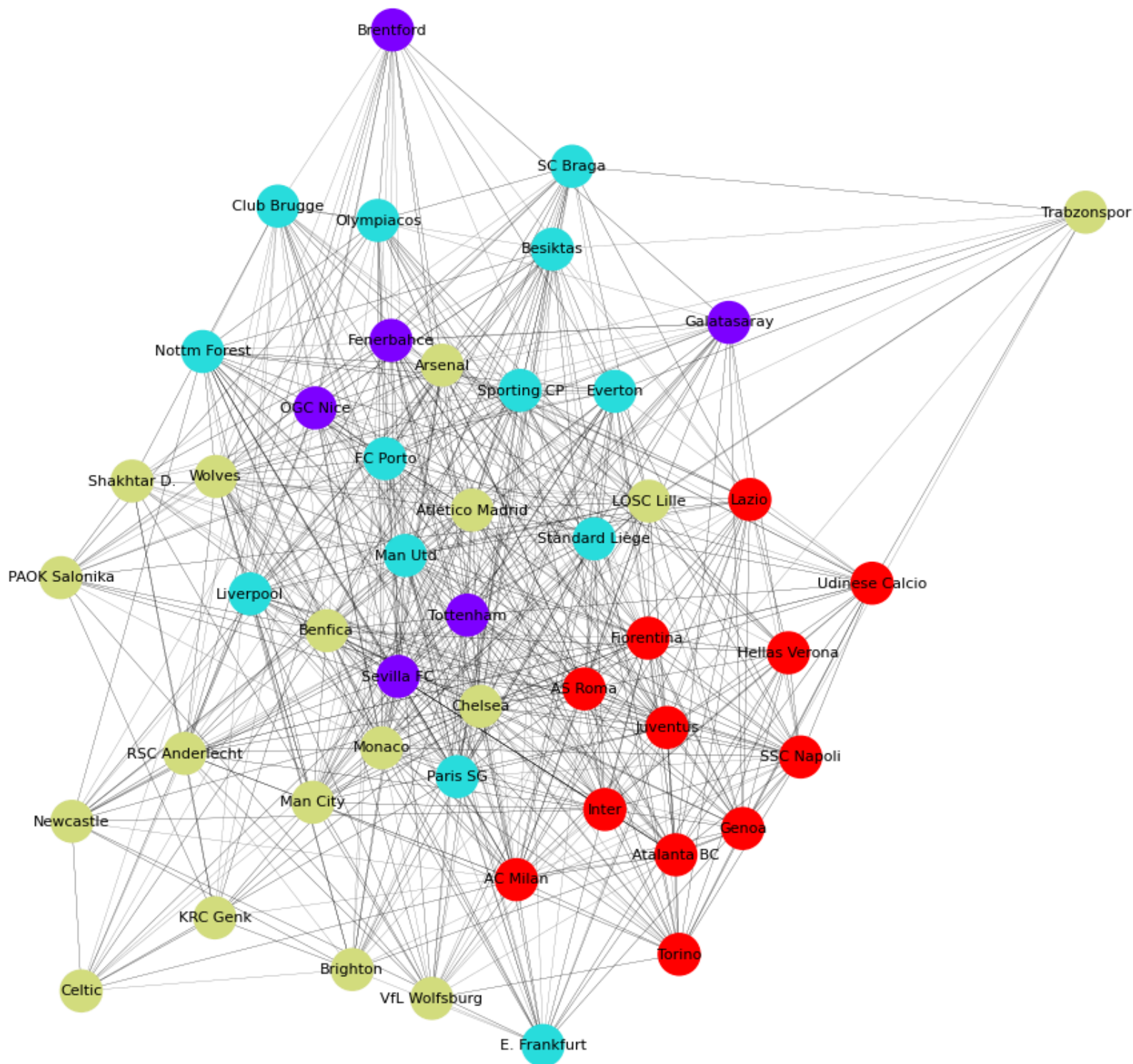


Figura 15: Comunidades detectadas no subgrafo de transferências entre clubes.

Conclusão:

Este projeto realizou uma exploração detalhada dos dados de transferências no futebol, com o objetivo de extrair padrões e insights sobre a valorização de jogadores e as dinâmicas entre clubes. A análise da distribuição do valor de mercado dos jogadores revelou que uma pequena parcela atinge valores extremamente altos, enquanto a maioria possui valores mais modestos. Essa distribuição assimétrica, com tendência exponencial decrescente, reflete a realidade de um mercado onde poucos jogadores concentram grande parte do valor total.

Ao examinar a evolução do valor de mercado ao longo do tempo, observamos que muitos jogadores atingem seu pico de valorização em momentos específicos de suas carreiras, geralmente declinando à medida que se aproximam do final da atividade profissional. No entanto, nem todos seguem esse padrão, indicando a diversidade de trajetórias no futebol. A análise do tempo até a melhor transferência mostrou que, em média, os jogadores alcançam sua transferência mais valiosa cerca de 6,55 anos após a primeira, com uma variação significativa. Isso sugere que o desenvolvimento e a valorização ocorrem ao longo de vários anos, contrariando a percepção criada por talentos emergentes que rapidamente alcançam seu ápice no mercado.

A tentativa de prever o tempo até a melhor transferência utilizando um modelo de aprendizado de máquina (Random Forest Regressor) apresentou resultados limitados, com erros médios absolutos em torno de 2 a 3 anos. Embora o modelo tenha capturado certos padrões, a variabilidade nos dados e a ausência de fatores críticos, como desempenho individual, histórico de lesões e influências externas (por exemplo, agentes e condições contratuais), limitaram sua precisão. Isso indica que a progressão da carreira de um jogador é influenciada por uma combinação complexa de fatores que não foram completamente representados nos dados disponíveis.

A análise de centralidade e a detecção de comunidades no grafo de transferências proporcionaram insights valiosos sobre a estrutura do mercado. Identificamos clubes com alta centralidade de intermediação, muitos dos quais atuam como pontes entre diferentes partes da rede, frequentemente facilitando a transição de jogadores em desenvolvimento para clubes maiores. A aplicação do algoritmo de Louvain revelou que as comunidades formadas tendem a refletir proximidades geográficas, com transferências ocorrendo mais intensamente entre clubes do mesmo país ou região. Isso reforça a influência de fatores regionais e culturais nas movimentações do mercado.

Para aprimorar futuras análises, sugere-se a incorporação de dados adicionais que capturem aspectos qualitativos e quantitativos mais abrangentes, como estatísticas de desempenho em campo, métricas de condicionamento físico, influência de agentes e políticas dos clubes. Além disso, a aplicação de modelos mais sofisticados, como algoritmos de previsão de links em grafos temporais e técnicas de aprendizado profundo, pode oferecer previsões mais precisas e capturar padrões mais sutis nas dinâmicas de transferências. Considerar o contexto temporal de forma mais granular também pode melhorar a compreensão das tendências e flutuações no valor de mercado dos jogadores.

Em resumo, o projeto demonstrou que a análise de grafos e técnicas de aprendizado de máquina podem fornecer insights significativos sobre o mercado de transferências no futebol. No entanto, para capturar plenamente a complexidade deste fenômeno, é essencial expandir o escopo dos dados e adotar abordagens analíticas que considerem a multiplicidade de fatores que influenciam a carreira dos jogadores e as estratégias dos clubes.