

TennisBets

Entrega Final

Artur Abreu Hendler - 231713

Guilherme Zeferino Rodrigues Dobins - 236129

João Vitor Baptista Moreira - 237833

Retomando:

Motivação e Contexto gerador:

Esse projeto tem como objetivo observar a relação entre os *odds* de apostas de uma partida e seu possível resultado. Atrelando fatores como a popularidade de um jogador, país de confronto e terreno do jogo e estatísticas das apostas do passado com o resultado do confronto.



Fontes

Tennis Data UK (xls):
Resultados de jogos, odds e
rankings



Functional Tennis (CSV):
Compilado de Tenistas mais
populares no Instagram em 2019



Instagram (pags. HTML):
Número de seguidores atual



World Cities (CSV):
Lista de países e código do país
referente



Scraping Instagram

Requesting the Web Page:

```
accounts_men = open("../data/raw/menInstagram2019.csv", newline='\n', encoding= 'utf-8')
num_lines_men = sum(1 for _ in accounts_men)
...
for _ in range(num_lines_men - 1):
    item = accounts_men.readline().replace('\x00','').split(',')
    # if the file already exists, it does not request it again
    if(not exists("../data/raw/menInstagramHTMLs/"+item[0]+".html")):
        try:
            urllib.request.urlretrieve(item[-1], "../data/raw/menInstagramHTMLs/"+item[0]+".html")
        except:
            print("request failed")
```

Extraíndo a informação desejada::

```
for _ in range(num_lines_men - 1):
    item = accounts_men.readline().replace('\x00','').split(',')
    men_2021.seek(0,0)
    # if the file exists and is not in the table
    if(exists("../data/raw/menInstagramHTMLs/"+item[0]+".html")):
        if(item[0] not in men_2021.read()):
            men_2021.seek(0, io.SEEK_END)
            men_html_file = open("../data/raw/menInstagramHTMLs/"+item[0]+".html","r", encoding= 'utf-8')
            men_html = men_html_file.read()
            # find in the html file the start of the value of followers
            followers_start = men_html.find("edge_followed_by") + 27
            # finds the end of the value
            followers = men_html[followers_start:men_html.find("}",followers_start)]
            men_2021.write(item[0]+","+followers+"\n")
            men_html_file.close()
```

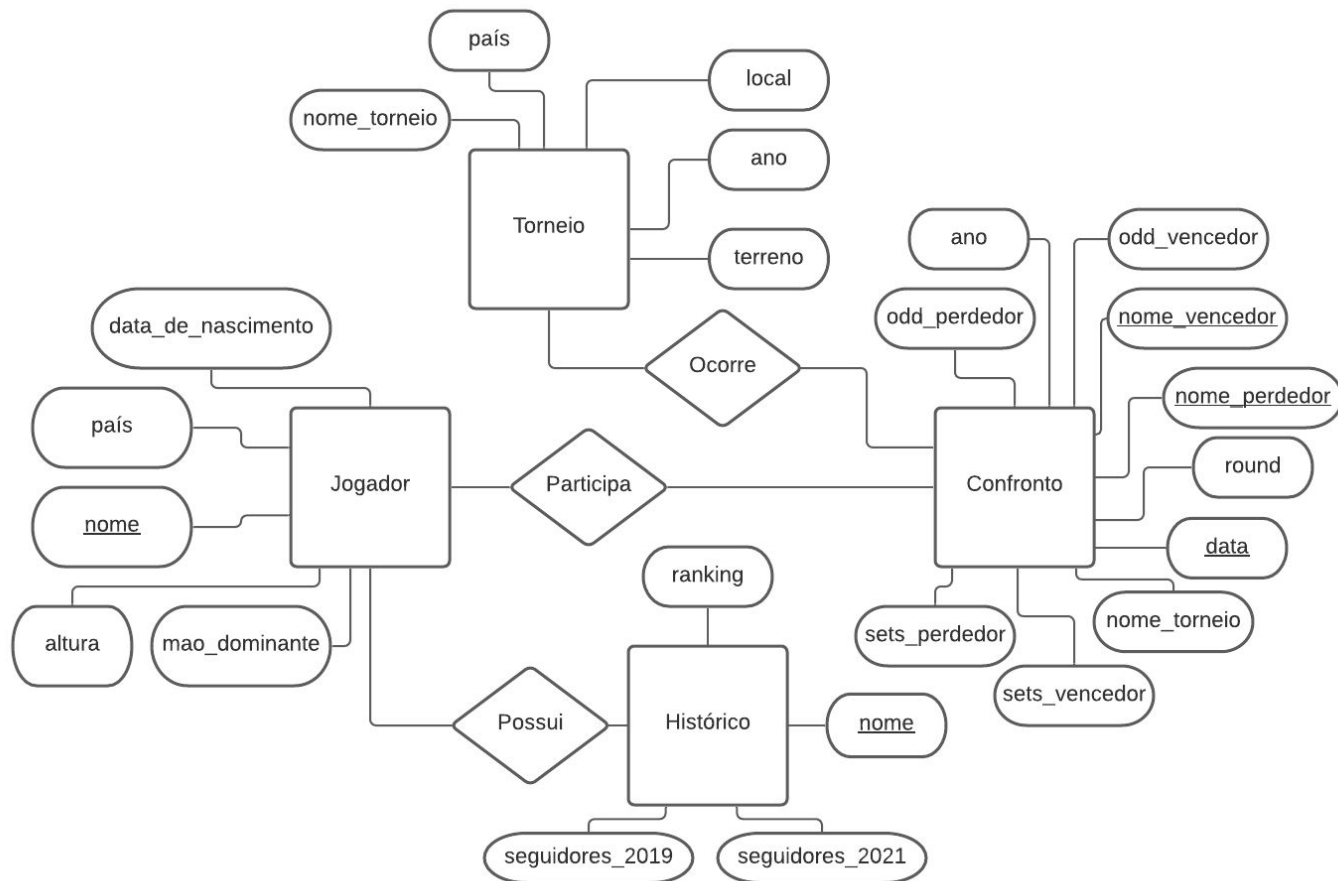


Name Matching



Nome	Nome Transformado
Nadal Rafael	R. Nadal
Herbert P-M.	P.M. Herbert
Herbert P.M..	
Joao P. Pella	J.P. Pella

Modelo Conceitual



Modelo Lógico

Jogador(nome, pais, data_de_nascimento, altura, mao_dominante)

Torneio(nome_torneio,ano,local, pais, terreno)

Historico(nome, seguidores_2019, seguidores_2021, ranking_2010, ranking_2011,
... ,ranking_2020, ranking_2021)

nome chave estrangeira -> Jogador(nome)

Confronto(nome_vencedor, nome_perdedor, data , odd_vencedor, odd_perdedor,
round, nome_torneiro, ano, sets_vencedor, sets_perdedor)

nome_vencedor chave estrangeira -> Jogador(nome)

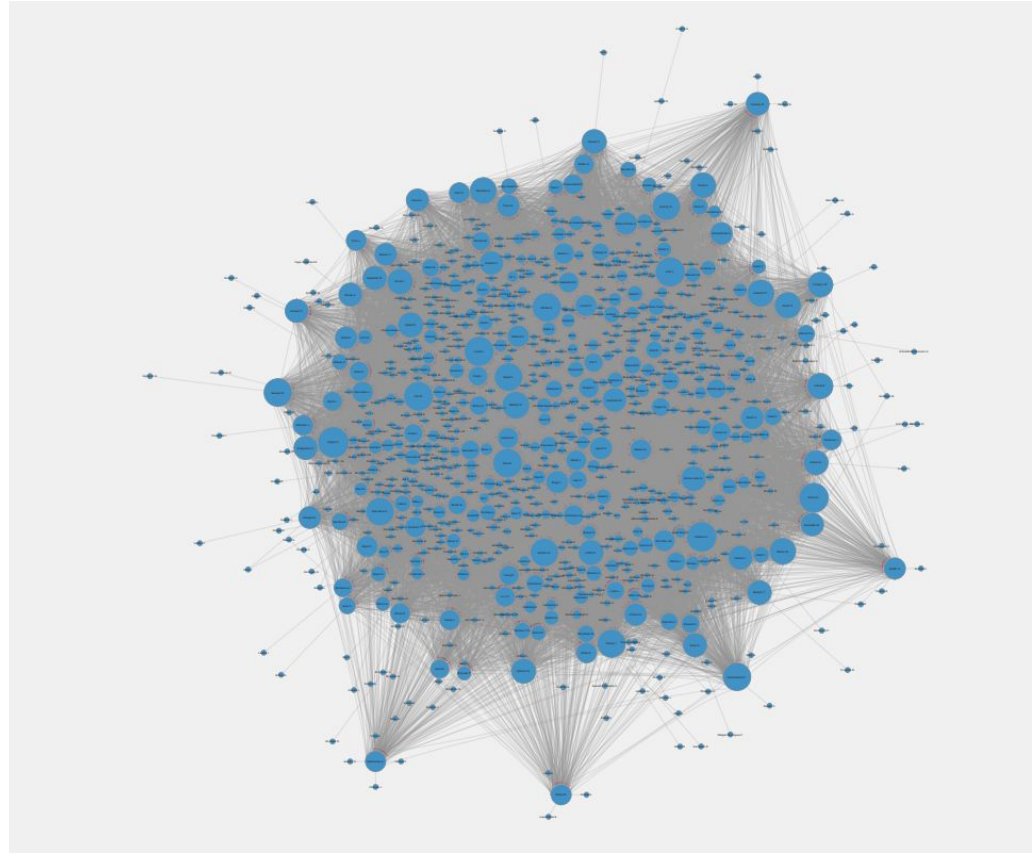
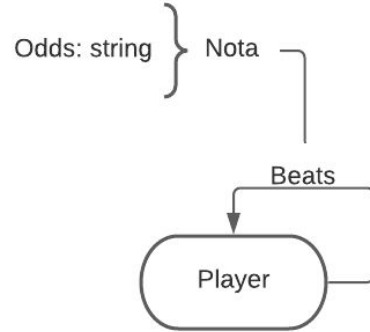
nome_perdedor chave estrangeira -> Jogador(nome)

nome_torneio chave estrangeira -> Torneio(nome_torneio)



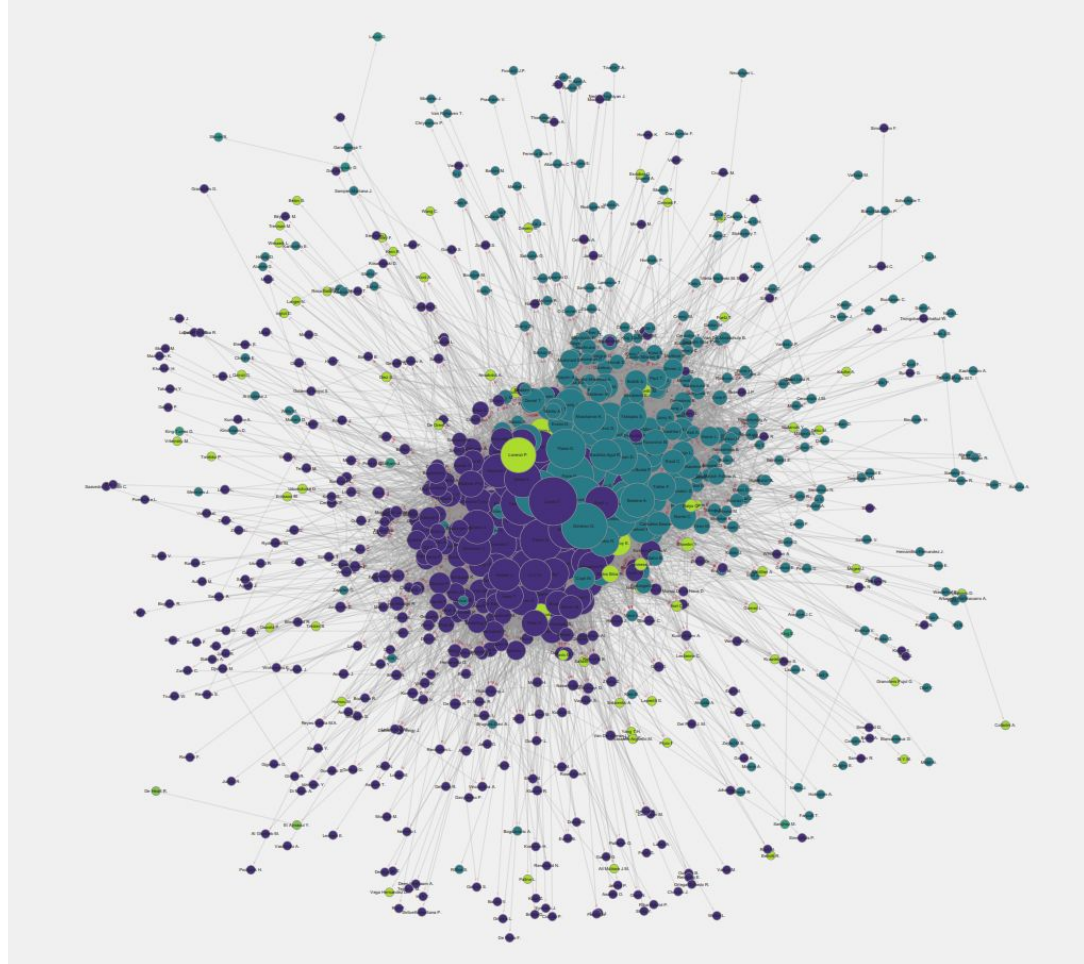
Grafo

Modelo de Grafo e
instância:

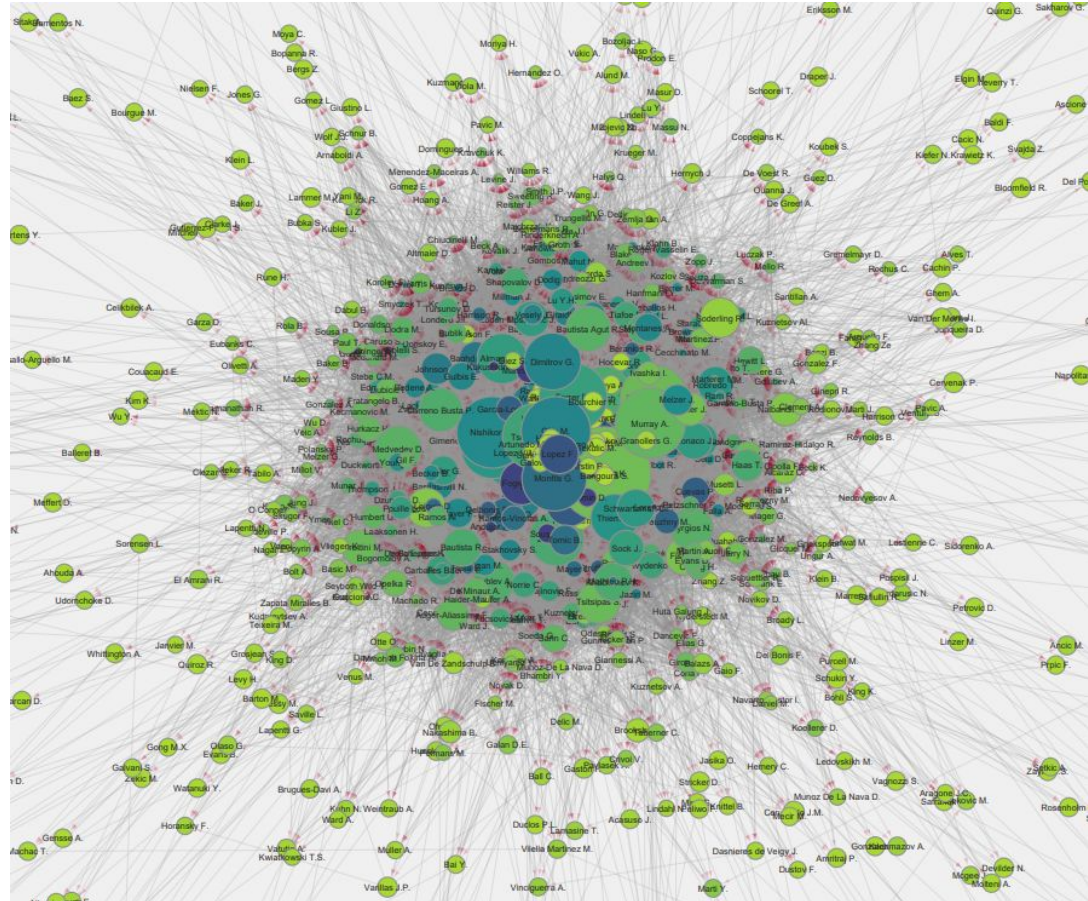


Análise do Grafo

Comunidades da Rede:



Análise de centralidade:

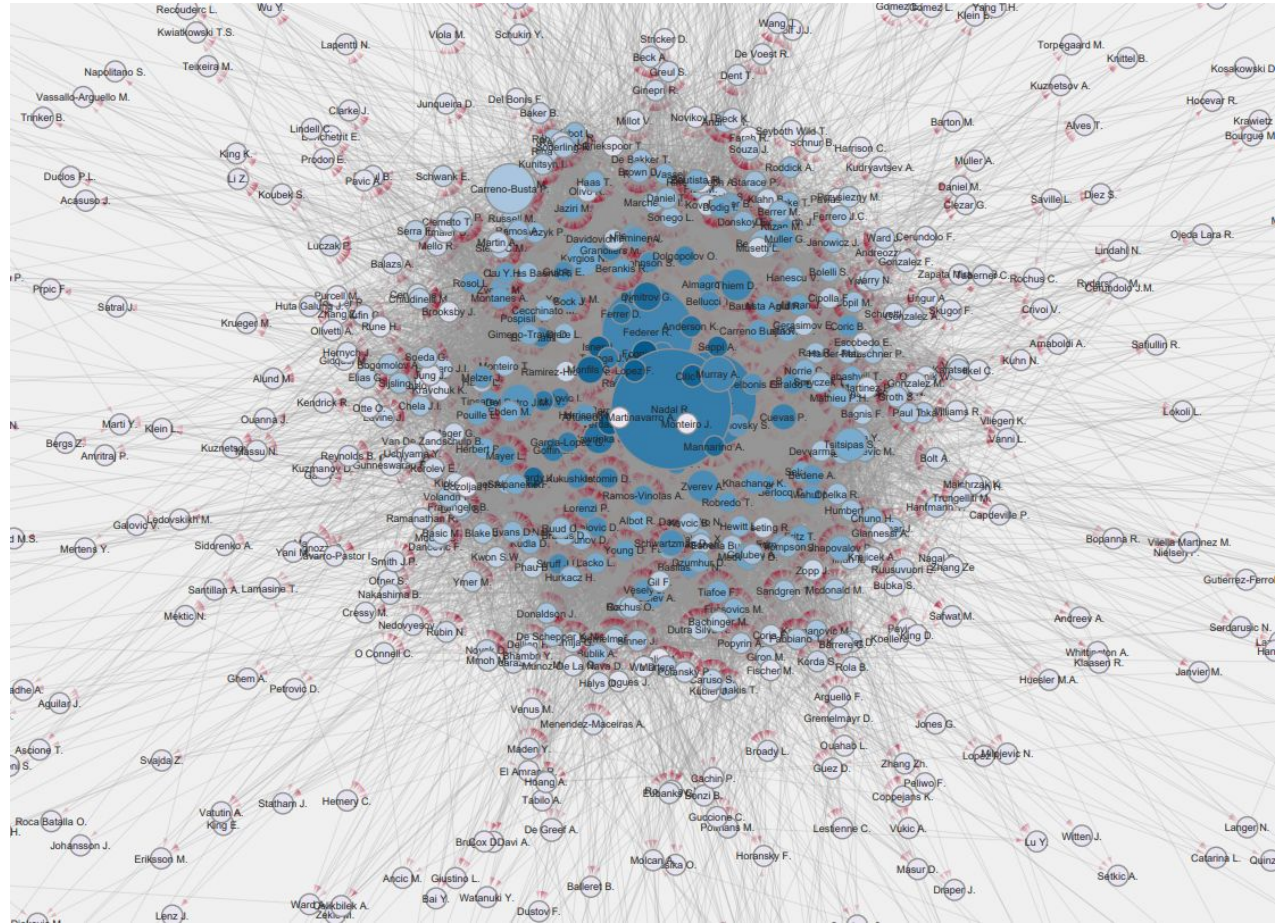


Análise do Grafo



Análise de popularidade dos jogadores:

Análise do Grafo



Tennis Bets



CONSULTAS SQL

Tennis Bets



- Quais países possuem mais pontos por jogador no ranking em 2019?

Tennis Bets



index	LOCAL	PONTOS_POR_JOGADOR
0	GRE	2651
1	GEO	1450
2	URU	1097
3	SRB	874
4	BOL	743
5	SUI	634
6	MDA	533
7	CAN	486

Tennis Bets



- Quais jogadores tiveram as melhores performances individuais em uma temporada nos últimos 10 anos?

Tennis Bets



index	NAME	FOLLOWERS_2019	FOLLOWERS_2021	ANO_2010	ANO_2011	ANO_2012	ANO_2013
0	Djokovic N.	5100000	9465474	6240	13630	12920	12260
1	Murray A.	1700000	1802054	5760	7380	8000	5790
2	Nadal R.	6434996	12349083	12450	9595	6690	13030

Tennis Bets



- Quais foram os maiores enganadores nas partidas?

Tennis Bets



Jogadores Subestimados

index	WINNER	ENGANAÇÃO
0	Wang C.	10.39
1	Kopriva V.	8.94
2	Erler A.	8.85
3	Gomez A.	8.72
4	Wessels L.	8.52
5	Cacic N.	8.27

Tennis Bets



PERGUNTAS
FUTURAS

Tennis Bets



- Observando as primeiras relações criadas por um jogador, é possível determinar se este é um profissional promissor?
- Como os análises feitas para a categoria masculina também vale para a feminina?
- Como a idade de um jogador afeta sua chance de avançar no torneio?

Dificuldades e evolução do projeto

- Scraping do instagram - bloqueio após algumas tentativas
- Nomes escritos de diferentes formas - precisamos transformar em uma pessoa só. Exemplo R. Nadal e Rafael Nadal
- Pessoas com nomes iguais - geram conflitos nos grafos. Li Z.
- Dados faltantes na data de nascimento - chave primária do jogador
- Ranking masculino e feminino



Obrigado!

