

Uso de grafos para identificar padrões de jogo de uma equipe de futebol

Arthur Pontes Nader
Luiz Phillippe Pereira
Rita Rezende de Lima

Julho de 2023

1 Introdução

A análise de dados aplicada ao futebol é uma área em crescimento dentro do campo da ciência de dados. Com a disponibilidade crescente de dados e tecnologias para coletá-los, tornou-se possível aplicar técnicas estatísticas e de aprendizado de máquina para avaliar e entender melhor o desempenho dos jogadores e times. Nesse contexto, o uso de grafos, como as Árvore Geradoras Mínimas (AGMs), e Diagramas de Voronoi, emergem como uma abordagem para avaliar os padrões de jogo e o desempenho do time.

O objetivo deste projeto é explorar o uso de Árvore Geradora Mínimas para avaliar a centralidade dos jogadores em uma partida de futebol e Diagramas de Voronoi para analisar a área coberta pelos defensores ao decorrer do jogo.

2 Trabalhos relacionados

Os trabalhos relacionados à análise de futebol com grafos vêm ganhando destaque nos últimos anos, principalmente devido ao aumento da disponibilidade de dados e tecnologias de análise de dados. Esses trabalhos utilizam grafos para modelar e analisar os padrões de jogo de times de futebol, possibilitando a identificação de jogadores-chave, o estudo de estratégias ofensivas e defesa, a avaliação do desempenho individual e coletivo dos jogadores, entre outras possibilidades. Os principais trabalhos relacionados estão disponíveis na seção de referências [2] [1] [3].

3 Metodologia de pesquisa

No caso deste projeto, a metodologia envolveu a coleta de dados de partidas de futebol, incluindo a posição dos jogadores em cada instante do jogo e os eventos que ocorreram, como passes, finalizações, desarmes, entre outros. Esses dados foram processados e utilizados para construir a Árvore Geradora Mínima (AGM) de cada partida. Para avaliar a conectividade de cada jogador usando a AGM, foi utilizado o conceito de centralidade.

Na nossa metodologia de trabalho, utilizamos várias técnicas e *frameworks* para realizar a análise com grafos. Utilizamos a linguagem de programação Python, que oferece uma ampla gama de bibliotecas especializadas para lidar com análise de redes e grafos. Algumas das bibliotecas que utilizamos incluem o *NetworkX*, sendo uma biblioteca poderosa para a criação, manipulação e visualização de grafos. Com o *NetworkX*, conseguimos facilmente construir o grafo a partir dos dados do *DataFrame*, adicionar nós, arestas e atribuir pesos às arestas.

Para realizar cálculos e medidas específicas no grafo, aproveitamos as capacidades da biblioteca NumPy, que oferece suporte a operações numéricas eficientes em *arrays* multidimensionais. Através do NumPy, podemos efetuar cálculos matemáticos nas estruturas de dados do grafo, como somas, médias e transformações nos pesos das arestas. Além disso, utilizamos a biblioteca SciPy para realizar análises estatísticas mais avançadas e calcular métricas de centralidade, como grau, *closeness* e *betweenness*. Essas métricas nos ajudam a entender a importância relativa dos nós no grafo, identificando jogadores-chave na circulação da bola e na posse do time.

Para trabalhar com dados tabulares e manipular o *DataFrame* contendo os eventos da partida, contamos com a biblioteca Pandas. Com o *Pandas*, podemos filtrar e transformar os dados eficientemente, selecionando apenas as linhas relevantes para os passes e manipulando as colunas necessárias para a construção do grafo. Por fim, utilizamos a biblioteca *Shapely* e novamente *SciPy* para trabalhar com a criação e manipulação de Diagramas de Voronoi.

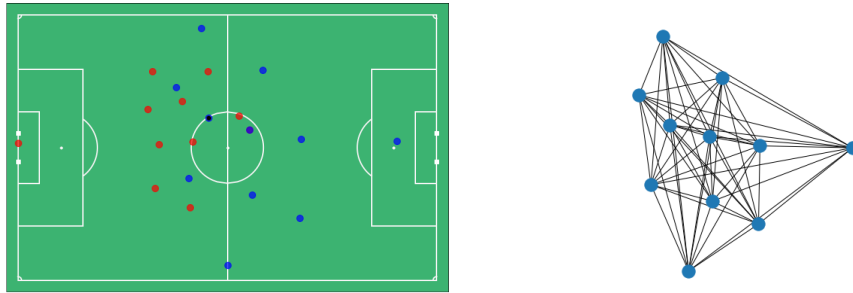
Essas técnicas e *frameworks*, como *NetworkX*, *NumPy*, *SciPy*, *Pandas* e *Shapely*, fornecem uma base para a análise de grafos e nos permitem extrair *insights* valiosos sobre a circulação da bola, posse de time e importância dos jogadores individualmente, contribuindo para uma melhor compreensão do desempenho tático em partidas de futebol.

4 Resultados e Discussões de Análises Feitas com Distâncias

O *DataFrame* utilizado pode ser dividido em *frames*, cada *frame* possui informações sobre um instante da partida, compreendido por alguns segundos.

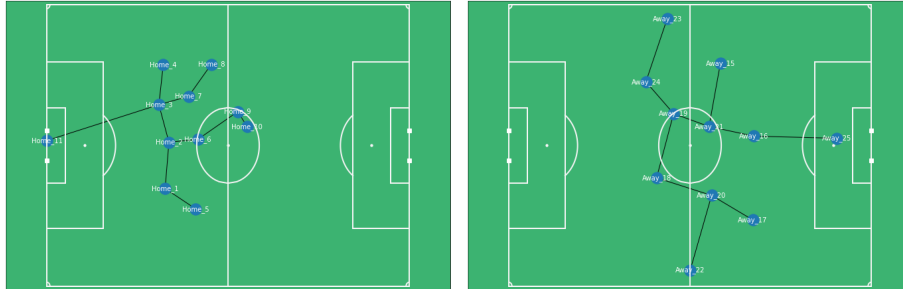
Dentre as informações disponíveis, temos a posição de cada um dos jogadores representadas como pares de coordenadas no plano do campo de jogo. A partir disso, podemos construir um grafo representando as distâncias entre os jogadores em um determinado frame. Esse grafo é completo onde cada vértice representa um jogador e o peso de uma aresta é igual a distância euclidiana dos dois jogadores correspondentes aos vértices.

Figure 1: Posições dos jogadores em um determinado *frame* da partida, em azul o time da casa, em vermelho o time visitante. A esquerda temos o grafo construído a partir das distâncias dos jogadores



Em seguida, podemos construir uma AGM para o grafo obtido. Realizando esse procedimento para cada um dos *frames* da partida, obtemos uma coleção de grafos sobre as quais algumas métricas podem ser analisadas.

Figure 2: Árvore geradora mínima construída a partir da distância dos jogadores durante a partida. A esquerda os jogadores do time da casa, a direita os visitantes.



Por exemplo, ao avaliar a centralidade de intermediação em uma AGM, podemos identificar os jogadores que possuem maior influência na dinâmica coletiva da equipe. Esses jogadores têm a capacidade de receber a bola de um companheiro e distribuí-la para outros jogadores de forma estratégica, influenciando diretamente o ritmo e a eficácia do jogo.

Além disso, a centralidade de intermediação também pode ajudar a identificar

jogadores que desempenham um papel de liderança na equipe, atuando como intermediários-chave na organização tática e na coordenação das jogadas. Eles são capazes de direcionar o jogo, comunicar-se com os companheiros de equipe e facilitar a transição de defesa para ataque. Na tabela a seguir há resultados do primeiro tempo de uma partida:

Valores de centralidade de intermediação		
Jogador	Média	Variância
Home ₁	0.212763	0.216928
Home ₂	0.361789	0.231471
Home ₃	0.442294	0.193520
Home ₄	0.182480	0.177332
Home ₅	0.216881	0.217468
Home ₆	0.463741	0.219104
Home ₇	0.368104	0.232607
Home ₈	0.112763	0.175114
Home ₉	0.186283	0.204877
Home ₁₀	0.198438	0.209548
Home ₁₁	0.008711	0.056687

Jogadores com alta centralidade: Os jogadores com valores de centralidade de intermediação acima da média podem ser considerados jogadores-chave, pois têm uma participação significativa na comunicação e fluxo de informações no grafo. No caso dessa tabela, podemos destacar os jogadores Home3 e Home6, que possuem as maiores médias de centralidade.

Jogadores com baixa centralidade: Os jogadores com valores de centralidade de intermediação abaixo da média podem ser considerados jogadores com baixa influência na rede. Eles podem ter menos participação na comunicação e no fluxo de informações entre os jogadores. Nessa tabela, podemos observar que os jogadores Home8 e Home4 possuem valores de centralidade de intermediação muito baixos, indicando uma menor importância na rede. Esses jogadores são da lateral esquerda do campo, o que é um indício de que o time não prioriza muito esse lado do campo para evolução do jogo.

5 Resultados e Discussões Análises Feitas com Passes

Recebemos um *DataFrame* contendo informações de eventos de uma partida, em que cada linha representava um evento específico. Para nosso objetivo, filtramos apenas as linhas onde o evento era um passe. Esse *DataFrame* possuía colunas de início e fim, indicando os identificadores dos jogadores envolvidos nos passes. Com essas informações, conseguimos criar arestas no nosso grafo.

Utilizamos todas as linhas de passe para construir um grafo, no qual os 11 jo-

Figure 3: A esquerda o grafo cujo peso das arestas é igual ao número de passes obtido entre os jogadores, a direita o grafo onde o peso é igual a 1 sobre esse número.

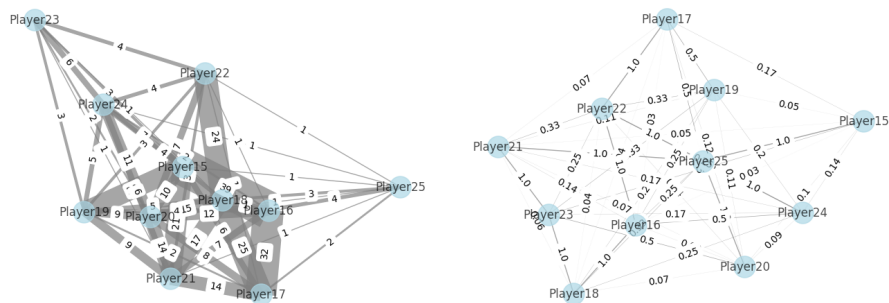
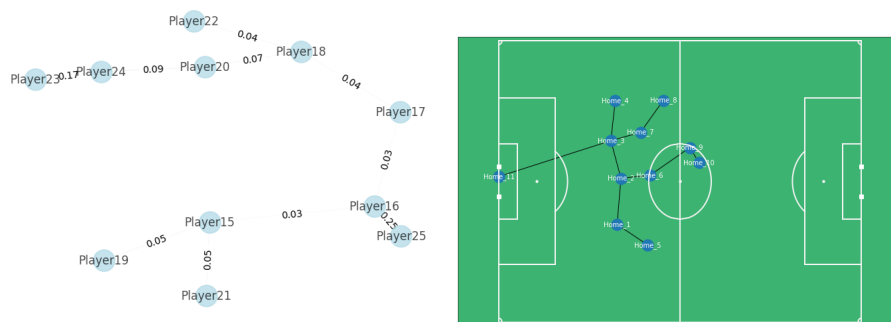


Figure 4: AGM construída a partir do grafo do inverso do número de passes.



5

da equipe. Esses índices de centralidade nos fornecem informações sobre o papel de cada jogador no jogo, destacando aqueles que têm maior influência na circulação da bola e no desempenho ofensivo da equipe.

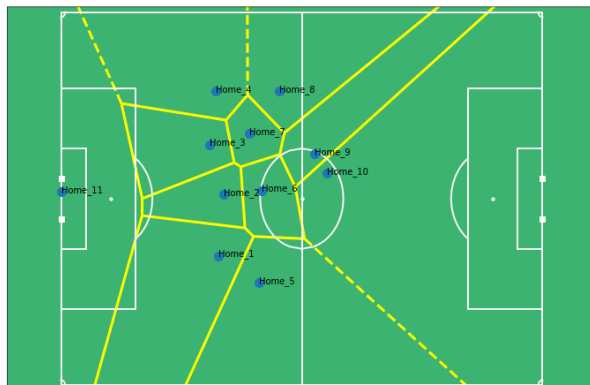
Valores de centralidade calculados			
Jogador	Degree	Betweenness	Closeness
Player 1	0.1	0.000000	0.333333
Player 11	0.1	0.000000	0.270270
Player 6	0.2	0.555556	0.500000
Player 3	0.1	0.000000	0.333333
Player 8	0.2	0.200000	0.357143

6 Resultados e Discussões Diagramas de Voronoi

As áreas dos diagramas de Voronoi podem ser utilizadas como uma métrica para verificar o desempenho e a cobertura dos zagueiros em uma partida de futebol. O diagrama de Voronoi divide o campo de jogo em regiões, onde cada região representa a área mais próxima a um determinado jogador.

Ao analisar as áreas dos diagramas de Voronoi, é possível identificar como os zagueiros estão distribuídos pelo campo e quais áreas eles estão cobrindo com maior frequência. Dessa forma, o uso das áreas dos diagramas de Voronoi permite uma análise visual do desempenho e da cobertura dos zagueiros, fornecendo insights sobre a organização defensiva da equipe e identificando possíveis áreas de melhoria na marcação e no posicionamento dos zagueiros. Essa análise pode ser valiosa para os treinadores e analistas na tomada de decisões táticas e no aprimoramento da defesa da equipe.

Figure 5: Diagrama de Voronoi representando as áreas do campo cobertas pelos jogadores



Área cobertas pelas zagueiros ao longo da partida		
Jogador	Área média	Variância
Home2	329.96	275.22
Home3	300.89	236.87
Away15	335.19	326.34
Away16	216.70	149.62

Sabendo-se que o placar final foi 3 a 0 para a equipe da casa, pode-se inferir que os jogadores Home2 e Home3 tiveram um desempenho mais efetivo em ocupar e cobrir áreas significativas do campo, enquanto os jogadores Away15 e Away16 tiveram um desempenho inferior nesse aspecto. Apesar disso não poder ser validado devido a pouca quantidade de jogos que se têm os dados de localização ao longo da partida, de certa forma isso é refletido pela maior área média somada dos jogadores da casa em comparação com os visitantes.

7 Conclusão

Em suma, nossa metodologia permitiu a proposição e implementação de métricas para a análise da circulação da bola e do comportamento dos jogadores em uma partida de futebol. Também conseguimos extrair informações tanto em termos de visualização quanto de resultados numéricos.

Através das métricas de centralidade, como grau, closeness e betweenness, pudemos avaliar a importância de cada jogador individualmente na posse do time. Essas métricas nos forneceram uma compreensão clara sobre o papel de cada jogador no jogo, identificando aqueles que desempenharam um papel crucial na circulação da bola e na progressão do ataque.

Além disso, a visualização dos grafos gerados a partir dos dados dos passes proporcionou uma compreensão visualmente impactante dos trajetos comuns da bola e da posse do time. Esses gráficos nos permitiram identificar os padrões de movimentação da bola e as conexões entre os jogadores, revelando visões importantes sobre a dinâmica coletiva e a interação entre os jogadores durante a partida.

Complementando os gráficos, também foram gerados resultados numéricos tabelados, fornecendo uma análise individual detalhada do desempenho dos jogadores na posse da bola. Esses resultados numéricos permitiram uma avaliação mais precisa das contribuições de cada jogador, destacando seus pontos fortes e áreas que precisam ser aprimoradas.

Em suma, por meio da aplicação de métricas concretas e da geração de gráficos, nossa metodologia proporcionou uma análise abrangente dos times de futebol. Isso nos permitiu compreender tanto os aspectos coletivos da circulação da bola quanto o desempenho individual dos jogadores durante a posse, fornecendo per-

cepções para aprimorar a estratégia e tática da equipe em torno dos jogadores que possuem maior centralidade.

References

- [1] Markus Brandt and Ulf Brefeld. Graph-based approaches for analyzing team interaction on the example of soccer. 01 2015.
- [2] Javier M. Buldú, Javier Busquets, Johann H. Martínez, José L. Herrera-Diestra, Ignacio Echegoyen, Javier Galeano, and Jordi Luque. Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game. *Frontiers in Psychology*, 9, 2018.
- [3] Vinicius Machado, Roger Leite, Felipe Moura, Sergio Cunha, Filip Sadlo, and João Comba. Visual soccer match analysis using spatiotemporal positions of players. *Computers Graphics*, 68, 08 2017.