

Otimização da formação de equipes em esportes coletivos através de Grafos

Gustavo Henrique D’Anunciação Ferreira

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS

Campus V

Divinópolis, Brasil

ghferreira92@gmail.com

Resumo—Este trabalho apresenta uma abordagem para a otimização da formação de times em esportes coletivos, utilizando técnicas de análise de dados e teoria dos grafos. O foco está na seleção dos melhores jogadores por posição em jogos de basquete, especificamente, utilizando-se de algumas das estatísticas individuais de cada jogador da NBA na temporada regular 22-23, considerando métricas de desempenho calculadas para cada posição baseando-se nesses dados. A metodologia proposta envolve a criação de um grafo representando a conexão entre jogadores, seguida de um algoritmo que busca maximizar o desempenho da equipe ao longo de posições distintas, iterando sob o grafo. O objetivo desse estudo é contribuir para a aplicação prática de técnicas de ciência de dados e análise de redes em contextos esportivos.

Palavras-Chave—Análise esportiva, Ciência de dados, Teoria dos Grafos

I. INTRODUÇÃO

A aplicação de recursos tecnológicos no contexto esportivo tem se tornado cada vez mais comum em diversas modalidades, ferramentas como o VAR (Video Assistant Referee) no Futebol, cada vez mais trazem impacto nos resultados e na dinâmica da prática esportiva [1].

A estratégia central deste estudo consiste em representar a interconexão entre jogadores através de um grafo, no qual os nós correspondem aos jogadores e as arestas refletem relações potenciais entre eles. Ao incorporar métricas de desempenho específicas para cada posição, é desejado construir times que maximizem a eficácia global da equipe.

Esta pesquisa se alinha à tendência de aplicação de ciência de dados em esportes, visando propor-

cionar insights valiosos, além de contribuir para o avanço do conhecimento no campo da análise esportiva.

No restante deste artigo, será explorado em detalhes a contextualização do problema, revisão da literatura sobre trabalhos correlatos, a utilização dos dados e desenvolvimento do algoritmo, apresentação dos resultados obtidos, considerações finais e propostas para trabalhos futuros.

II. CONTEXTUALIZAÇÃO

No contexto do basquete, mais precisamente, neste caso, no contexto da liga profissional de basquete dos Estados Unidos (NBA), a tradicional seleção dos melhores jogadores por posição ao fim de cada temporada sempre gera polêmicas e discussões na mídia, e entre fãs ao redor do mundo. Essa seleção é feita a partir de votos de jornalistas, especialistas e treinadores.

No entanto, a crescente disponibilidade de dados estatísticos detalhados, abre oportunidades para abordagens orientadas por dados, proporcionando insights importantes e relevantes para auxiliar a tomada de decisão de equipes de scout e treinadores no sentido tático e técnico [2].

O desafio dessa abordagem reside na otimização da formação dos times, considerando não apenas as habilidades individuais dos jogadores, mas também relacionar seus desempenhos em quadra. Nesse contexto, a análise dos dados e a representação gráfica das relações entre jogadores são ferramentas

cruciais para entender e aprimorar a dinâmica de equipe.

III. MANIPULAÇÃO DOS DADOS E MODELAGEM GRÁFICA DO PROBLEMA

A. Descrição da base de dados

A base de dados utilizada contém as estatísticas individuais referentes aos 82 jogos da temporada regular da NBA na temporada 2022-2023 de 679 jogadores das 30 equipes que compõem a liga.

Os atributos (Colunas) que compõem a base de dados estão descritos na Tabela I:

| Atributo | Descrição |
|----------|--|
| Rk | Classificação |
| Player | Nome do jogador |
| Pos | Posição |
| Age | Idade do jogador |
| Tm | Time |
| G | Jogos disputados |
| GS | Jogos iniciados |
| MP | Minutos jogados por jogo |
| FG | Arremessos convertidos por jogo |
| FGA | Tentativas de arremesso por jogo |
| FG% | Percentual de acerto de arremessos |
| 3P | Arremessos de 3 pontos convertidos por jogo |
| 3PA | Tentativas de arremesso de 3 pontos por jogo |
| 3P% | Percentual de acerto de arremessos de 3 pontos |
| 2P | Arremessos de 2 pontos convertidos por jogo |
| 2PA | Tentativas de arremesso de 2 pontos por jogo |
| 2P% | Percentual de acerto de arremessos de 2 pontos |
| eFG% | Percentual de acerto efetivo de arremessos |
| FT | Lances livres convertidos por jogo |
| FTA | Tentativas de lances livres por jogo |
| FT% | Percentual de acerto de lances livres |
| ORB | Rebotes ofensivos por jogo |
| DRB | Rebotes defensivos por jogo |
| TRB | Total de rebotes por jogo |
| AST | Assistências por jogo |
| STL | Roubos de bola por jogo |
| BLK | Bloqueios por jogo |
| TOV | Perdas de posse por jogo |
| PF | Faltas pessoais por jogo |
| PTS | Pontos por jogo |

Tabela I

LEGENDA DOS ATRIBUTOS

B. Tratamento dos dados

Inicialmente, a base de dados passa por uma verificação de dados faltantes, inconsistentes ou fora do domínio dos atributos. Neste esse caso, um problema de domínio que relaciona jogadores a mais de uma posição precisou ser tratado, já

que, teria impacto significativo na modelagem do problema.

O domínio correto do atributo "Posição" é: ['PG', 'SG', 'SF', 'PF', 'C'], e está descrito na Tabela II.

| Classificação | Descrição |
|---------------|-------------|
| PG | Armador |
| SG | Ala-Armador |
| SF | Ala |
| PF | Ala-Pivô |
| C | Pivô |

Tabela II

LEGENDA DAS POSIÇÕES

Dados com o atributo "Posição" classificados como 'PF-SF', foram reclassificados como 'SF', já dados que tinham as classificações 'SG-PG' e 'SF-SG' foram reclassificados como 'SG'.

C. Seleção de atributos

A seleção de atributos é uma parte importante para o cálculo das métricas de desempenho. Levando em conta os atributos listados na Tabela I, foi feita uma análise para selecionar os atributos que mais influenciam no desempenho geral do jogador durante a temporada. Essa análise considerou, principalmente, o quanto o atributo em questão influencia no resultado final de uma partida.

Os atributos selecionados para compor o cálculo das métricas de desempenho estão listados na Tabela III.

| Atributo | Descrição |
|----------|--|
| G | Jogos disputados |
| eFG% | Percentual de acerto efetivo de arremessos |
| TRB | Total de rebotes por jogo |
| AST | Assistências por jogo |
| STL | Roubos de bola por jogo |
| BLK | Bloqueios por jogo |
| TOV | Perdas de posse por jogo |
| PF | Faltas pessoais por jogo |
| PTS | Pontos por jogo |

Tabela III

ATRIBUTOS SELECIONADOS

D. Cálculo das métricas de desempenho

Após a seleção de atributos, a métrica de desempenho, nesse modelo, chamada de coeficiente de desempenho (CD), pode ser calculada.

O cálculo do 'Coeficiente de desempenho' utiliza dos valores dos atributos listados na Tabela III, e é dado por :

$$CD = PTS \times (2 \times eFG\%) + G + P1 \times AST + P2 \times STL + P3 \times BLK + P4 \times TRB - (P5 \times (TOV + PF)) \quad (1)$$

Na equação (1), P1, P2, P3, P4 e P5 representam os pesos, que são ajustados no cálculo do 'Coeficiente de desempenho' por posição. A tabela IV lista os pesos utilizados nesta abordagem.

| Posição | P1, P2, P3, P4, P5 |
|---------|-----------------------|
| PG | 2, 1, 1, 1, 1.5 |
| SG | 1.75, 1, 1, 1, 1 |
| SF | 1.5, 1, 1, 1, 1 |
| PF | 1.25, 2, 1.75, 1.5, 1 |
| C | 1.25, 1.25, 2, 2, 1 |

Tabela IV
PESOS POR POSIÇÃO

Com o 'Coeficiente de desempenho' calculado para cada posição, o atributo 'COEF.Desempenho' é adicionado à base de dados, contendo o valor da métrica para cada jogador.

E. Modelagem do grafo

A modelagem gráfica do problema foi feita utilizando-se dos conceitos da Teoria dos grafos [3]. O grafo foi construído utilizando uma idéia de "regiões", onde cada região contém vértices representando os jogadores de uma posição. Esses vértices possuem atributos, que são eles: 'Nome do jogador', 'Posição', 'Coeficiente de desempenho'.

Todos os vértices de uma região estão ligados por arestas a todos os vértices da região vizinha, formando, visualmente, uma "ponte" entre as regiões. Uma aresta que liga dois vértices, está ponderada com a soma dos coeficientes de desempenho dos dois vértices que ela conecta.

A Figura 1 ilustra o grafo gerado nos termos dessa abordagem. Cada aglomerado de vértices em verde, representa uma região de jogadores de uma determinada posição. Da Direita para a esquerda, estão as regiões na seguinte ordem: PG, SG, SF, PF, C.

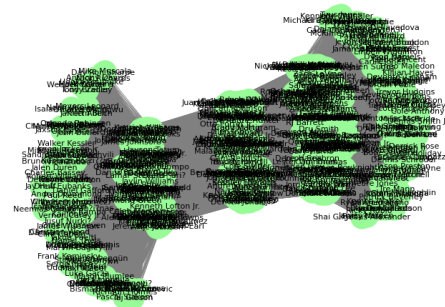


Figura 1. Grafo modelado

IV. ITERAÇÃO SOB O GRAFO

A modelagem e o algoritmo de iteração sob o grafo foram feitos utilizando a linguagem Python, e os arquivos estão disponíveis no repositório do GitHub [4].

A partir do grafo ilustrado na Figura 1, é possível chegar à seleção dos melhores jogadores por posição.

O algoritmo itera sob as regiões, partindo da direita para a esquerda, mais precisamente, iniciando na região contendo os armadores (PG) e finalizando na região contendo os pivôs (C).

Iniciando na primeira região, todos os vértices são visitados e é feita a verificação nas arestas, a fim de encontrar a aresta com o maior valor ponderado que faz a ligação a um vértice da região vizinha. Quando essa aresta é encontrada, as informações a respeito dos vértices que ela liga, além do seu valor ponderado, são armazenados.

Partindo do vértice da segunda região, que está conectado através da aresta de maior valor ponderado com o vértice da primeira região, a verificação das arestas se repete sucessivamente até atingir a região que contém os pivôs, somando os valores ponderados nas arestas a cada transição de região.

Ao fim, a lista com os jogadores selecionados é criada e mostrada no terminal, juntamente com o valor total do coeficiente de desempenho da equipe, resultando na seleção dos melhores jogadores por posição de acordo com as métricas calculadas.

V. RESULTADOS E TRABALHOS FUTUROS

A. Resultados

Tradicionalmente, a seleção dos melhores jogadores por posição ao fim de cada temporada da NBA é feita a partir do voto de treinadores, especialistas e jornalistas envolvidos no contexto da liga.

Na temporada 2022-2023, que é a temporada relacionada aos dados desta abordagem, a seleção desses jogadores, chamada de All-NBA teams, já foi realizada e divulgada.

As listas abaixo contém, respectivamente, os jogadores selecionados para a primeira e segunda seleção da temporada.

All-NBA First Team

Luka Dončić
Shai Gilgeous-Alexander
Jayson Tatum
Giannis Antetokounmpo
Joel Embiid

All-NBA Second Team

Donovan Mitchell
Stephen Curry
Jimmy Butler
Jaylen Brown
Nikola Jokić

Ao fim da iteração do algoritmo sob o grafo, a seleção dos melhores jogadores por posição nos termos desta abordagem são mostrados no terminal. Os resultados gerados estão descritos na lista abaixo:

Resultado da análise do grafo

Luka Dončić
Anthony Edwards
DeMar DeRozan
Jayson Tatum
Nikola Jokić

O valor do Coeficiente de desempenho da equipe gerada nos resultados foi de 949,49.

Comparando as três listas, há 3 jogadores em comum entre os resultados da análise do grafo e os All-NBA Teams montados na realidade: Luka Dončić e Jayson Tatum estão no primeiro time e Nikola Jokić está no segundo time, apesar de ter liderado seu time rumo ao título da temporada.

B. Trabalhos futuros

No que se refere a trabalhos futuros, algumas adaptações podem ser estudadas a fim de melhorar o cálculo da métrica de desempenho.

Outros atributos, além dos selecionados, podem passar a fazer parte da modelagem matemática do desempenho a fim de considerar outros pontos que podem impactar no resultado final de uma partida, ou dos resultados finais da temporada.

Além disso, técnicas de validação cruzada podem ser utilizadas para fazer um ajuste dos 5 pesos diferentes para as posições, com objetivo de se aproximar ainda mais dos resultados reais.

REFERÊNCIAS E TRABALHOS CORRELATOS

- [1] A. C. Oliveira, "A nova tecnologia no Futebol: diálogos sobre a influência do VAR," Aug. 26, 2020. <http://www.rbff.com.br/index.php/rbff/article/view/898>
- [2] L. Chen and W. Wang, "Analysis of technical features in basketball video based on deep learning algorithm," Signal Processing: Image Communication, vol. 83, p. 115786, Apr. 2020, doi: 10.1016/j.image.2020.115786.
- [3] P. P. Da [Unesp Costa, "Teoria dos grafos e suas aplicações," Dec. 01, 2011. <https://repositorio.unesp.br/items/29bde62e-cf36-46d9-ade6-98835427297a>
- [4] Gustavo Ferreira, "All-NBA-Team-with-graphs-and-data-analysis". GitHub Repository. URL: <https://github.com/Lokinha92/All-NBA-Team-with-graphs-and-data-analysis>
- [5] M. Ahmadalinezhad and M. Makrehchi, "Basketball lineup performance prediction using edge-centric multi-view network analysis," Social Network Analysis and Mining, vol. 10, no. 1, Aug. 2020, doi: 10.1007/s13278-020-00677-0.