Contribuição Relativa: Detectando ”Buracos” de Elenco na NBA

Orientador: Pedro Olmo Stancioli Vaz de Melo

Kaio Lucas de Sá
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
desakaio4@gmail.com

***Abstract*—Este documento apresenta a metodologia e a métrica utilizada para encontrar ”buracos” no elenco de um time da National Basketball Association (NBA). A métrica criada pode ser usada tanto para identificar problemas no elenco, que podem ser abordados durante a temporada, ou na temporada seguinte, como também identificar virtudes do elenco. Os dados foram coletados a partir de fontes como a ESPN e o repositório `hoopR-data`, abrangendo as temporadas de 2018 a 2024. A preparação dos dados envolveu a integração das fontes e a derivação de ”shifts” para a análise de formações em quadra, bem como o cálculo da Win Probability Added (WPA) para cada jogador. A análise exploratória inicial revelou padrões na distribuição de vitórias, indicando uma leve vantagem para times da casa, e na frequência de substituições ao longo do jogo e prorrogações. A aplicação da WPA revelou dois fortes vieses, sendo muito sensível tanto a jogadores que passaram pouco tempo em quadra, quanto ao contexto de elenco das equipes, tendendo fortemente a 0. Isso motivou a proposição de uma segunda métrica, chamada de Contribuição relativa, que mostrou-se mais adequada para o contexto, sendo o ponto central de uma metodologia simples e clara para o apontamento de virtudes e defeitos de elencos na liga.**

***Index Terms*—NBA, WPA, substituições, shifts, Contribuição Relativa, scouting, Diagnóstico de elenco**

I. Introdução

O uso de métodos de análise de dados para o contexto do esporte profissional vem se tornando cada vez mais comum. Temos exemplos desse tipo de uso desde o beisebol [3], passando pelo futebol [6] [7] [8] [9], e até no *esport* [2]. Tal como os exemplos anteriores, o **Basquetebol** também oferece um domínio amplo a ser explorado pela análise de dados, de fato, existem trabalhos sobre o uso de métodos estatísticos e computacionais neste contexto [1] [4] [5]. No contexto da *National Basketball Association* (NBA), o primeiro desafio enfrentado por uma equipe é a definição do seu elenco para a temporada. A existência de um teto salarial, bem como o aumento constante da intensidade de jogo na liga nos dois lados da quadra, torna essencial saber quais jogadores serão dispensados, e onde deve-se buscar reforços. Para isso, é preciso realizar análises estatísticas que transcendam o *box-score*. O desafio central reside em traduzir essas estatísticas em um aumento tangível na chance de vitória, garantindo que se evite a penalização de jogadores pelo contexto específico em que atuam. Adicionalmente, deve-se considerar a dificuldade histórica em mensurar o impacto defensivo nas chances de triunfo. Superar essas barreiras é fundamental para diagnosticar em quais pontos o elenco é inferior à média da liga. Nesse sentido, este trabalho investiga como a WPA se comporta no domínio da NBA, buscando validar seu uso como ferramenta de *scouting* e gestão, e, identificando suas limitações, propor uma métrica de contribuição que utilize a informação de WPA de forma simples e eficiente.

Na seção **II** são mostrados trabalhos que atacam problemas similares em contextos diferentes, justificando a ideia de se estudar o comportamento desse tipo de solução no contexto do basquetebol. A descrição das métricas utilizadas pelo trabalho, bem como a explicação do conceito de ***shift***, crucial para o trabalho, podem ser encontrados na seção **III**. Também nessa seção é feita uma descrição da metodologia que pode ser utilizada para identificar ”buracos” nos elencos da liga. A seção **IV** detalha a forma como os dados foram coletados e preparados para montagem, análise, teste e descrição de métrica, além de trazer uma descrição breve dos dados utilizados. Na seção **V** é feita uma análise do comportamento das duas métricas utilizadas pelo trabalho nos dados coletados, bem como uma exemplificação do uso da metodologia explicada para identificar pontos de melhora no elenco de uma equipe. Além disso, esta seção traz uma descrição das outras contribuições relevantes geradas na fase das análises exploratórias feitas inicialmente sobre os eventos de substituição e suas relações com a probabilidade de vitória. A seção **VI** traz um breve resumo de possíveis caminhos para trabalhos futuros que se inspirem neste. Ao final, é feita uma conclusão condensando todos os resultados alcançados pelo trabalho.

II. Referencial Teórico

A crescente digitalização do esporte profissional tem impulsionado a adoção de métodos de análise de dados, um fenômeno que se manifesta amplamente em diversas modalidades. Exemplos notáveis incluem o beisebol, o futebol, e até mesmo os *esports* [2]–[4], [10]. A literatura recente demon-

stra que hoje, além de buscar descrever o passado, análises preditivas estão ganhando espaço no campo de pesquisa. Nesse contexto, técnicas de aprendizado profundo vêm sendo utilizadas nos esportes eletrônicos para identificar a decisão que maximize a probabilidade de vitória de um time, ao passo que, no futebol, são utilizadas técnicas estatísticas para estimar o impacto de uma substituição na marcação de gols na partida.

No âmbito da otimização estratégica, Jalovaara (2024) [2] demonstrou que, no contexto de *League of Legends* (LoL), a *Win Probability Added* (WPA) apresenta valor superior a métricas estatísticas básicas, devido à sua sensibilidade ao estado do jogo. O estudo evidencia que quantificar a variação na probabilidade de vitória associada a uma ação específica permite mensurar a contribuição individual considerando o contexto dinâmico da partida.

De forma paralela, no futebol, Amez et al. (2021) [8] investigaram técnicas estatísticas para estimar o impacto das substituições na probabilidade de marcação de gols em partidas de clubes internacionais. Ao analisar dados da UEFA Champions League e UEFA Europa League, os autores identificaram que as equipes tiveram um aumento nas probabilidades de marcar gols após suas primeiras e segundas substituições, o que indica o uso potencial dessas trocas como ferramentas para recuperação dentro de uma partida, especialmente em cenários de desvantagem no placar. Isso sugere que a gestão do elenco durante a partida é um fator determinante para o resultado final.

Embora métricas baseadas em probabilidade de vitória sejam comuns para avaliar lances individuais, existe uma lacuna na aplicação dessa lógica para avaliar a composição de elencos na NBA. Este trabalho une as abordagens supracitadas: utiliza-se a lógica de WPA (Jalovaara) aplicada aos eventos de substituição e formação de *shifts* (análogo a Amez), não apenas para analisar uma partida isolada, mas para diagnosticar deficiências estruturais (”buracos”) e virtudes nas posições de uma equipe durante uma temporada, buscando encontrar possibilidades de melhoria e comparar o desempenho entre temporadas.

III. Metodologia

Essa seção detalha a estrutura analítica utilizada para auxiliar na estruturação de elencos da liga. Diferente de abordagens mais tradicionais, que utilizam estatísticas agregadas como o *box-score*, ou o *plus-minus*, a metodologia aqui apresentada busca valer-se da propriedade da probabilidade de vitória de considerar o contexto e estado do jogo para medir impactos individuais, conforme descrito em Jalovaara [2].

O processo supracitado foi estruturado em quatro etapas essenciais:

- **Cálculo do impacto contextual (WPA):** Atribui-se a cada evento da partida a variação na probabilidade de vitória decorrente desse evento, capturando assim, a importância em termos de probabilidade de vitória do ocorrido.
- **Identificação e Agregação de *Shifts*:** Segmenta-se a partida em blocos contidos entre dois períodos de paradas onde houve substituições, permitindo assim a análise agrupada das métricas de acordo com a formação em quadra.
- **Definição da Contribuição Relativa:** Normaliza-se os resultados acumulados e agregados pelo total da equipe, gerando uma métrica de proporção que representa o quanto um grupo de jogadores contribuiu para a construção do resultado final.
- **Comparação dos Resultados entre as Equipes da Liga e Diagnóstico de Elenco:** Compara-se a contribuição relativa calculada no passo anterior com a média da liga, propondo uma abordagem sistemática para cumprir o objetivo do trabalho.

No final desta seção também são detalhadas as ferramentas utilizadas para todas as análises supracitadas, bem como uma descrição breve da forma como foram realizadas as análises exploratórias dos dados utilizados pela abordagem aqui descrita.

A. Cálculo do impacto contextual (WPA)

Para a realização deste estudo, optou-se pela utilização de probabilidades de vitória pré-calculadas provenientes da plataforma de analytics da ESPN. A WPA é definida como uma função dessa probabilidade de vitória (WP) e um evento.

Essa métrica mede a variação de WP no período entre dois eventos diferentes, que compõem um intervalo de jogo, e é representada pela equação abaixo:

$$\Delta WP(s) = WP(t_{fim}) - WP(t_{início})$$

Os significados de t_{fim} e $t_{início}$ na equação podem ser explicados, respectivamente, como o momento imediatamente posterior e anterior ao evento ou intervalo analisado.

Percebe-se que essa métrica tem uma direção vetorial clara: um resultado positivo ($\Delta WP(s) > 0$) indica que no fim daquele *shift* a equipe aumentou sua probabilidade de vitória, enquanto um valor negativo indicaria que os eventos ocorridos dentro deste período distanciaram o time de um resultado positivo.

B. Identificação e Agregação de Shifts

Para a análise de desempenho coletivo, torna-se necessária a divisão da partida em unidades temporais de cooperação. Assim, define-se um *Shift* (s) como o intervalo de tempo contido enre duas paradas de jogo onde houve substituições, desta forma garante-se que os 10 jogadores presentes em quadra permaneça a mesma para cada unidade analisada.

Uma grande vantagem de $\Delta WP(s)$ é que podemos medir facilmente o impacto acumulado de um grupo de jogadores nas chances de vitória da equipe. Define-se o WPA Acumulado(WPA_{acc}) como o somatório de $\Delta WP(s)$ de todos os shifts da temporada em que o grupo em questão esteve em quadra.

$$WPA_{acc} = \sum_{s \in \mathcal{S}} \Delta WP(s)$$

De forma análoga, para avaliar o desempenho médio ao longo de uma temporada, calcula-se o WPA Médio por Partida ($\overline{WPA}$), dividindo WPA_{acc} pelo número de jogos (N) em que tal grupo esteve em quadra:

$$\overline{WPA} = \frac{WPA_{acc}}{N}$$

Essas agregações são similares à abordagem adotada por Jalovaara [2] para itens em *League of Legends*, formam a base para o cálculo da contribuição relativa.

C. Definição da Contribuição Relativa:

Dado o objetivo central de identificar lacunas estruturais na construção do elenco, a análise individualizada de cada shift ou jogador torna-se demasiadamente pulverizada para um diagnóstico macroscópico. Para simplificar a modelagem do problema e permitir uma avaliação direta das carências do time, optou-se por agregar os dados segundo as funções táticas dos atletas.

Os jogadores foram agrupados de acordo com as posições presentes nos dados da ESPN. Chamamos de **Grupos Posicionais (**P**)** às posições apontadas pelas fontes dos dados (ex: *Guards, Forwards, Centers*). O desempenho total de uma posição (WPA_P) é definido como a soma de WPA_{acc} de todos os jogadores (j) pertencentes àquela categoria:

$$WPA_P = \sum_{j \in \mathcal{P}} WPA_{acc}(j)$$

Com base nesse agrupamento, calcula-se a **Contribuição Relativa da Posição**(RC_P) para representar a proporção da WPA_{acc} acumulada da equipe que foi resultado das ações daquele setor.

$$RC_P = \frac{WPA_P}{WPA_{acc}(time)}$$

De maneira análoga à $\overline{WPA}$, também é possível calcular a **Contribuição Relativa Média da Posição por Partida** ($\overline{RC_P}$) em uma temporada dividindo RC_P, pelo número total de jogos (N) daquela temporada.

$$\overline{RC_P} = \frac{RC_P}{N}$$

Essa abordagem permite condensar o impacto de milhares de combinações de jogadores em vetores de contribuição média por posição, criando uma nova forma de caracterizar um elenco enquanto facilita identificar os setores que contribuem positiva ou negativamente para os resultados da equipe.

D. Comparação dos Resultados entre as Equipes da Liga e Diagnóstico de Elenco:

As métricas acima definidas servem para descrever o perfil de uma equipe na liga, porém, para utilizá-las de forma diagnóstica é necessário definir um padrão ao qual comparar afim de identificar se é necessário um salto de qualidade para competir por títulos, ou se o elenco atual já atende as expectativas competitivas.

Para isso, foi utilizado como *benchmark* para as comparações define-se a **Contribuição Média da Liga**(μ_P). Esta métrica é calculada através da média dos valores de $\overline{RC_P}$ de todas as K equipes da liga, onde $K = 30$:

$$\mu_P = \frac{1}{K} \sum_{k=1}^{K} \overline{RC_{Pk}}$$

Onde $\overline{RC_{Pk}}$ representa a contribuição média da posição P na equipe k.

Através disso podemos calcular o **Devio Posicional**(δ_P), que mensura a distância da média da equipe para a média da liga.

$$\delta_P = \overline{RC_P}_{Time} - \mu_P$$

A interpretação de δ_P permite identificar as características construtivas do elenco:

- Valores positivos ($\delta_P > 0$) indicam que a posição possui uma contribuição para o resultado da equipe superior à média da liga (Virtude).
- Valores negativos ($\delta_P < 0$) sinalizam que a posição contribui menos do que o padrão esperado pelos competidores, configurando um ”buraco” ou deficiência estrutural.

E. Análise Estatística

Durante esse projeto, foi feita uma análise estatística dos dados, que pode ser separada nas seguintes etapas:

- **Análise Exploratória dos Dados (EDA):** Nesta etapa iremos entender o comportamento dos eventos de substituições e sua relação com a probabilidade de vitória. Isso inclui a visualização de distribuições, como a proporção de vitórias por time (casa vs. visitante) e a distribuição temporal das substituições ao longo das partidas e prorrogações. Serão exploradas questões como a influência do mando de quadra e os padrões de substituição em diferentes fases do jogo.
- **Análise de $\overline{WPA}$:** Foi realizado uma análise do comportamento da WPA dos jogadores da liga, buscando enxergar como esta métrica se distribui nos dados utilizados, bem como entender suas limitações e virtudes.
- **Análise de $\overline{RC_P}$:** Também foi feita uma observação do comportamento dessa segunda métrica nos dados apresentados, também buscando descrever sua distribuição.

A intenção é que, a partir dessas análises, consigamos identificar ”buracos” de elenco em uma equipe, buscando apontar oportunidades de melhora. A EDA nos fornecerá uma base sólida para a compreensão dos dados e para a formulação de hipóteses.

F. Ferramentas e Tecnologias

Para a implementação das etapas de coleta, preparação e análise dos dados, serão utilizadas as seguintes ferramentas e tecnologias:

- **Python:** Linguagem de programação principal para manipulação de dados, automação de processos e desenvolvimento de scripts de análise.

- **Bibliotecas Python (Pandas, NumPy, Matplotlib, Seaborn):**
  - **Pandas:** Essencial para a manipulação e análise de dados tabulares.
  - **NumPy:** Utilizada para operações numéricas de alta performance.
  - **Matplotlib e Seaborn:** Para a criação de visualizações e gráficos que auxiliam na Análise Exploratória dos Dados.
- **R e hoopR:** Fontes programáticas para acesso aos dados da NBA.
- **Modelos de Linguagem (LLMs):** O modelo Gemini (Google) foi utilizado como ferramenta de apoio para revisão gramatical, refinamento de estilo acadêmico, formatação de equações em LaTeX e estruturação de roteiros de escrita. A concepção intelectual, a análise dos dados e a validação final de todo o conteúdo textual e matemático são de autoria integral do aluno.

A escolha dessas ferramentas se justifica pela sua robustez, flexibilidade e ampla utilização na comunidade de ciência de dados e pesquisa acadêmica, permitindo um desenvolvimento eficiente e replicável do projeto.

IV. Dataset

Os dados que serão utilizados nesse trabalho provém de duas fontes: **os dados de evento da ESPN** e do projeto **hoopR-data repository's NBA Data**. A ESPN fornece informações detalhadas sobre eventos de jogo, como substituições e probabilidades de vitória ao longo da partida, enquanto o hoopR-data repository's NBA Data complementa com informações de formações dos times.Para possibilitar a execução de nossas análises, diversas etapas de preparação foram cruciais para integrar e padronizar esses conjuntos de dados.

A primeira etapa consistiu em realizar um casamento dos jogos das duas fontes, buscando correspondências entre os registros. Este processo envolveu a identificação de jogos únicos com base em critérios como data, equipes participantes e identificadores de jogo. Foi possível encontrar uma correspondência para **7555 dos 8890** jogos coletados, abrangendo as temporadas da NBA de 2018 a 2024. Os jogos para os quais não foi possível encontrar um casamento em ambas as fontes foram descartados da base de dados para garantir a integridade e a completude das informações utilizadas.

A segunda preparação realizada foi a seleção dos eventos de substituição, início e fim de partida, visto que a partir deles conseguimos derivar todas as estatísticas relevantes para nossa análise. Esta fase envolveu a filtragem de grandes volumes de dados de eventos para isolar apenas aqueles que são diretamente pertinentes ao nosso estudo, garantindo a eficiência das etapas subsequentes. A justificativa para essa seleção é que as substituições são os eventos primários que alteram as formações em quadra, assim, quando uma substituição ocorre estamos trocando o shift da partida.

A partir dessa seleção, executamos a terceira preparação nos dados: o cálculo do número do "shift" da partida. Para isso, agrupamos os eventos de substituição pelo tempo de jogo em que foram executados. Cada agrupamento de substituições marca o início de um novo shift. O primeiro shift da partida, s_1 é definido como o período entre o início da partida e o primeiro agrupamento de substituições. Os demais shifts consistem no período entre o agrupamento anterior de substituições e o seguinte. Este processo de segmentação temporal é fundamental para contextualizar a WPA, permitindo-nos analisar o impacto das formações de equipe em períodos específicos do jogo.

A partir da categorização desses períodos, montamos a formação de cada equipe em cada um desses intervalos de tempo. Isso envolve a identificação dos cinco jogadores em quadra para cada equipe em cada shift. A acurácia na determinação das formações é crucial para a aplicação da métrica WPA, pois ela se baseia diretamente nas mudanças de elenco.

Depois foi realizado o cálculo das duas métricas descritas anteriormente: $WPA_{Acumulada\ média}$ para cada jogador, e em seguida a contribuição relativa de cada par ordenado posição, time existente na base de dados.

V. Resultados e Discussão

Nesta seção, iremos discutir e exemplificar os resultados que encontramos durante o projeto. Os dados utilizados serão os dados de substituições e início de partida do conjunto de dados, bem como os dados de probabilidade de vitória da ESPN. O período a ser analisado está compreendido entre os anos de 2018 e 2024. Antes de entrarmos em detalhes na relação das substituições com a variação de probabilidade de vitória, vamos analisar a distribuição das vitórias e das substituições.

A. Distribuição das Vitórias

Podemos perceber que durante as temporadas de 2018 e 2024, houve uma leve vantagem dos times da casa em relação aos times visitantes em termos de vitórias. O curioso nesse caso é a temporada de 2020. É importante lembrar que essa temporada, por conta da pandemia de COVID-19, foi chamada de "temporada da bolha", onde a liga isolou os times em Orlando, e fechou os jogos para o público, com o objetivo de diminuir a exposição dos jogadores ao vírus. Dessa forma os times atuaram nessa temporada em "território neutro". Apesar disso, ainda conseguimos perceber uma leve vantagem para os times de casa. Isso nos indica que em um futuro modelo de previsão de probabilidade de vitória, o fator time da casa/visitante pode ser uma feature a ser utilizada.

B. Distribuições de Substituições

Em relação à frequência e distribuição das substituições, notamos uma maior concentração desses eventos no início de cada quarto, com exceção do primeiro. Esses picos podem ser atribuídos a momentos estratégicos predefinidos pelos treinadores para ajustar formações e gerenciar o desgaste físico dos jogadores. Com exceção desses picos iniciais de quarto,

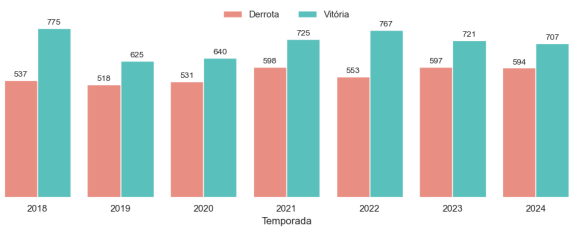

Fig. 1. Distribuição de vitórias do time que tem mando de quadra nas temporadas 2018-2024.

as substituições são relativamente bem distribuídas ao longo dos períodos, aproximando-se de uma distribuição normal.

Observamos também que, nas prorrogações (**overtimes**), as substituições ocorrem predominantemente no primeiro período de prorrogação e tendem a diminuir drasticamente nos períodos subsequentes. Isso pode ser reflexo da intensidade e da curta duração das prorrogações, onde os treinadores buscam manter os melhores jogadores em quadra o máximo possível, realizando apenas substituições estritamente necessárias ou em caso de emergência. A Figura 2 apresenta a distribuição das substituições ao longo do tempo de jogo.

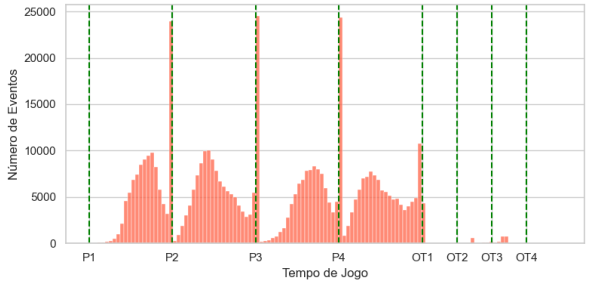

Fig. 2. Frequência de eventos de substituição ao longo do tempo de jogo e prorrogações.

C. Correlação Shift x WP

Um resultado interessante e contraintuitivo que conseguimos alcançar nesta fase do projeto é a ausência de uma correlação clara entre o número de ”shifts” (períodos de jogo com formações de jogadores consistentes entre substituições) de uma partida e a probabilidade de vitória de um time. Conforme ilustrado na Figura 3, a variação da probabilidade de vitória não parece ser diretamente influenciada pela quantidade de trocas de formação realizadas. Esse resultado inicial reforça a necessidade de investigarmos mais a fundo o impacto da escolha dos jogadores que entram em quadra em determinado momento, em vez de apenas o evento de substituição em si. Ou seja, o ato de substituir, por si só, pode não ter um impacto significativo na chance de um time ganhar ou perder uma partida, mas sim quem é substituído e por quem.

Apesar dessa falta de correlação direta, outros fatores podem justificar essa observação, como, por exemplo, a diferença de pontos entre as duas equipes naquele shift. É plausível que a situação do placar e o contexto do jogo no momento da substituição (e do shift) sejam variáveis mais influentes do que a simples frequência das trocas de formação. A relação entre essas variáveis adicionais (como a diferença de pontos e o número de shifts) será investigada mais aprofundadamente na segunda fase do projeto, onde buscaremos modelos mais complexos que considerem a interação desses fatores.

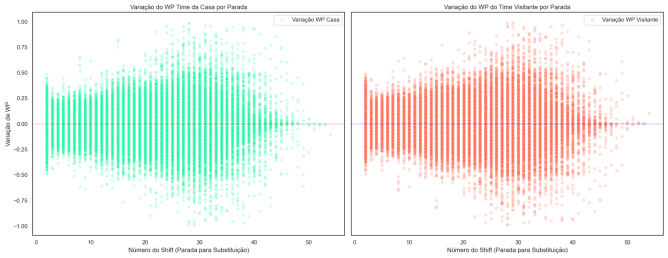

Fig. 3. Correlação entre o número de shifts (trocas de formação) e a variação da Probabilidade de Vitória.

D. Distribuição da $\overline{WPA}$

Além das análises exploratórias, investigou-se o comportamento da métrica $\overline{WPA}$ por jogador. A distribuição apresentada na Figura 4 demonstra uma concentração acentuada de valores em torno de zero. Este comportamento é teoricamente esperado, uma vez que, a maioria dos jogadores são aqueles que não tem um impacto grande nas chances de sua equipe vencer, os chamados *roleplayers*.

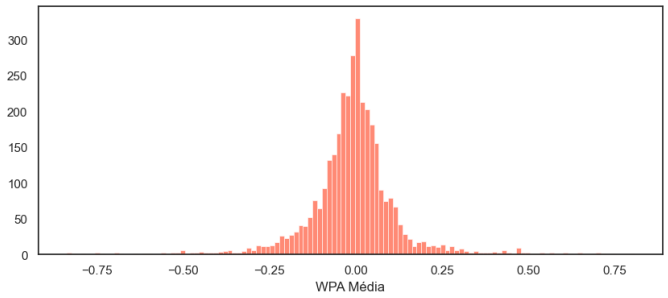

Fig. 4. Histograma da distribuição da WPA Acumulada Média por jogador.

No entanto, observa-se a presença de *outliers* significativos em ambas as extremidades. Ao filtrar esses dados, percebeu-se que tais atletas tiveram uma amostragem muito baixa de jogos na temporada (a maioria com menos de 15 jogos na temporada regular), conforme evidenciado na Figura 5.

Em contrapartida, ao analisar os candidatos ao prêmio de MVP (Figura 6), jogadores de alto volume e impacto reconhecido, nota-se que seus valores de WPA acumulado médio tendem a convergir para zero. Isso evidencia um **viés de volume e contexto**, jogadores com poucos minutos estão sujeitos à alta variância, podendo acumular WPA artificialmente alto ou baixo, enquanto estrelas que atuam por longos períodos têm seu impacto médio diluído. Essa constatação inviabiliza o uso isolado do WPA absoluto para diagnósticos de elenco robustos.

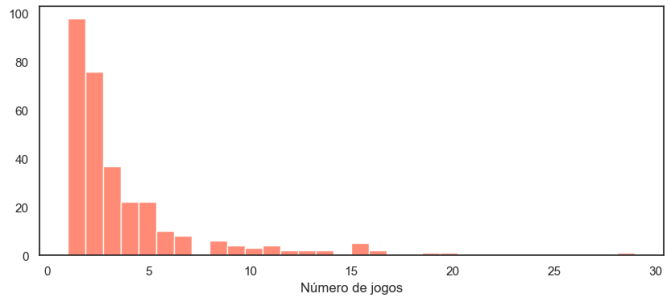

Fig. 5. Relação entre WPA extrema e número de jogos disputados na temporada.

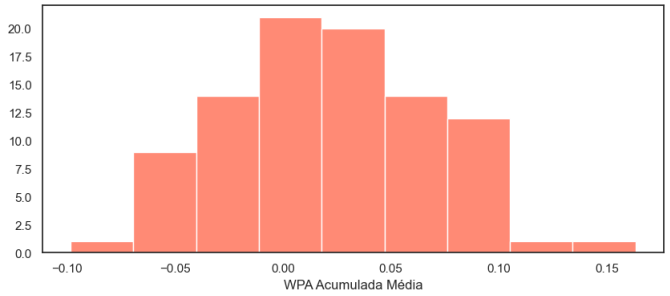

Fig. 6. Distribuição da WPA Acumulada Média dos candidatos ao prêmio de MVP (2018-2024).

E. Contribuição Relativa (RC_P)

Com o objetivo de mitigar os vieses de volume identificados na subseção anterior, aplicouse a métrica de **Contribuição Relativa** (RC_P). A vantagem dessa métrica perante $\overline{WPA}$ é a normalização do impacto pelo contexto coletivo, permitindo o agrupamento por posições de maneira simples. Calculando essa métrica para cada jogo, agrupando ela pela média da posição, obtemos a distribuição que pode ser observada na Figura 7.

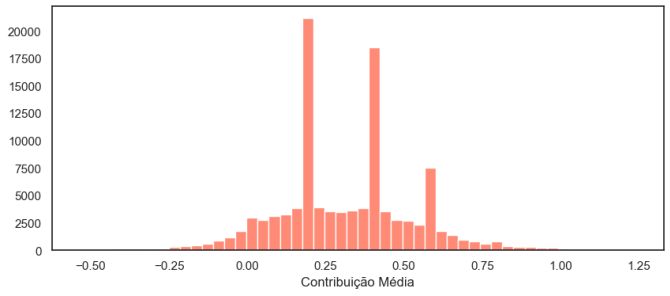

Fig. 7. Distribuição da métrica de Contribuição Relativa Média por posição.

Diferente de $\overline{WPA}$, a distribuição de $\overline{RC_P}$ tem três picos de concentração. O mais importante é que, não há outliers na distribuição da métrica.

Dessa forma, é possível criar uma metodologia baseada na métrica para encontrar os ”buracos” de elenco. Tal abordagem será descrita na próxima subseção.

F. Metodologia para detecção de ”buracos” de elenco

Usando a métrica descrita anteriormente podemos criar uma métrica baseada na diferença da Contribuição média de uma posição em um time com a média da Contribuição média dessa posição na liga na mesma temporada, comparando pares equivalentes. Observando a figura 8 podemos perceber que a média da liga é 0 na distribuição de δ_P, dessa forma podemos calcular δ_P para cada uma das equipes. Caso o valor seja positivo, então a métrica indica uma vantagem desse elenco perante o resto da liga, e não é necessário melhorar a posição.

Caso o valor seja negativo, então isso é um indicativo de ”buraco” de elenco detectado, indicando um possível ponto de melhora para a equipe em uma posição.

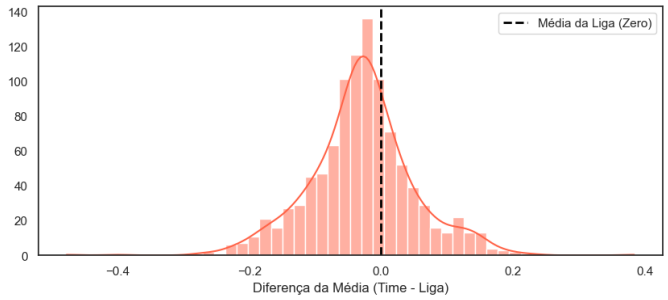

Fig. 8. Diferença entre a Contribuição Relativa do time e a média da liga (Detecção de ”Buracos”).

Dessa forma, um exemplo de uso é analisando duas das posições possíveis de uma equipe, os Pivôs e os Armadores. É comum dividir-se os jogadores de uma equipe entre *”bigs”* e *”guards”* sendo as duas posições mais representativas destes arquétipos os *pivôs* e os *armadores*. Dessa forma, é possível construir um *scatterplot* contendo $\delta_{Pivô}$ em um eixo e $\delta_{Armador}$ no outro, podemos visualizar esse gráfico na Figura 9.

Nesta visualização (Figura 9) foram destacados os extremos de cada quadrante. Podemos validar essa metodologia observando os pontos de melhora de uma equipe e comparando com seus elencos da temporada seguinte e vendo a localização dessa equipe na temporada seguinte.

Em 2022, a equipe do Oklahoma City Thunder teve o pior $\delta_{Pivô}$ do período analisado. De acordo com a metodologia desenvolvida o time deveria adquirir novos pivôs para suprir essa deficiência de elenco, o que de fato ocorreu, a equipe contratou um novo pivô saindo de lesão (Dario Šarić), mas que em sua última aparição na liga esteve em um elenco com contribuição acima da média da posição, e adquiriu um novo pivô via *draft* (Jaylin Williams). O resultado foi uma escalada de $\delta_{Pivô}$ da equipe, onde em uma temporada a equipe era a com menor contribuição relativa da liga, para praticamente alcançar a média na temporada seguinte, como pode ser observado na Figura 10.

Além disso, outro reflexo que também evidencia a efetividade da métrica é o aumento no número de vitórias da equipe. Na temporada 2021-2022 a equipe teve um desempenho de 24 vitórias e 58 derrotas para na temporada seguinte alcançar 40

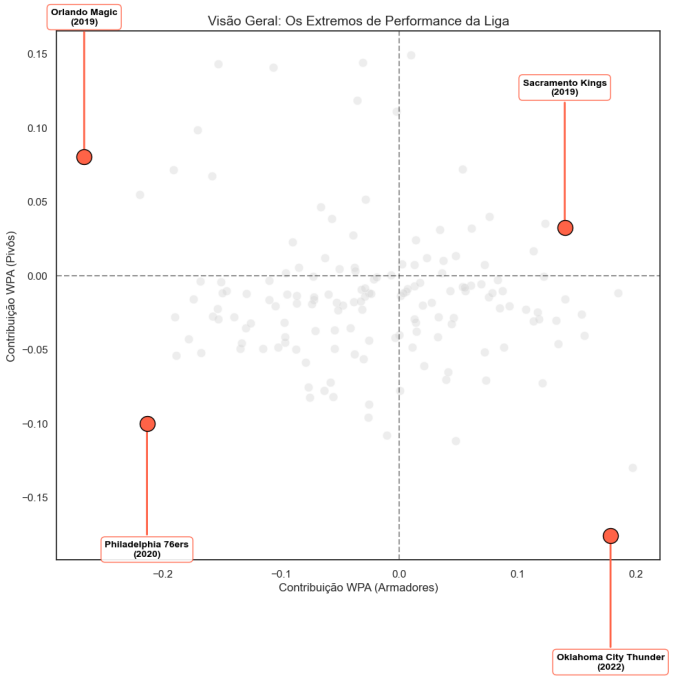

Fig. 9. Análise comparativa de desempenho por posição: Pivôs vs. Armadores.

vitórias e 42 derrotas. Em 2023-2024 ocorreu outro salto no número de vitórias, terminando a temporada com um *score* de 57 vitórias e 25 derrotas, nesta temporada ocorreu a estreia de Chet Holmgren pela franquia, considerado hoje um dos melhores pivôs da liga.

Um segundo exemplo é o caso do Orlando Magic de 2018-2019, que tinha um valor de $\delta_{Pivô}$ muito acima da média, porém um $\delta_{Armador}$ muito baixo. Para a temporada seguinte, 2019-2020, a equipe contratou e adquiriu via *draft* armadores, deslocando o ponto da equipe para a direita, o que indica um aumento da contribuição média dos armadores, como também observado na Figura 10.

Porém, diferente do caso anterior, o número de vitórias da equipe sofreu uma queda, saindo de 42 vitórias e 40 derrotas na temporada 18-19 para 33 vitórias e 40 derrotas na temporada seguinte. Isso pode ter acontecido devido a outras mudanças ocorridas no elenco de uma temporada para a outra, uma vez que times raramente realizam mudanças em somente uma posição entre duas temporadas.

A abordagem descrita tem principalmente dois possíveis usos, diagnóstico de elenco, buscando identificar pontos de melhoria no planejamento dos jogadores que irão compor um elenco, mas também no scouting de adversários antes de uma partida, identificando posições onde uma equipe tem vantagem em uma posição quando comparada com seu adversário.

VI. Trabalhos Futuros

A metodologia de diagnóstico estrutural desenvolvida neste trabalho estabelece uma nova perspectiva para a avaliação de elencos na NBA, abrindo caminho para diversas investigações subsequentes.

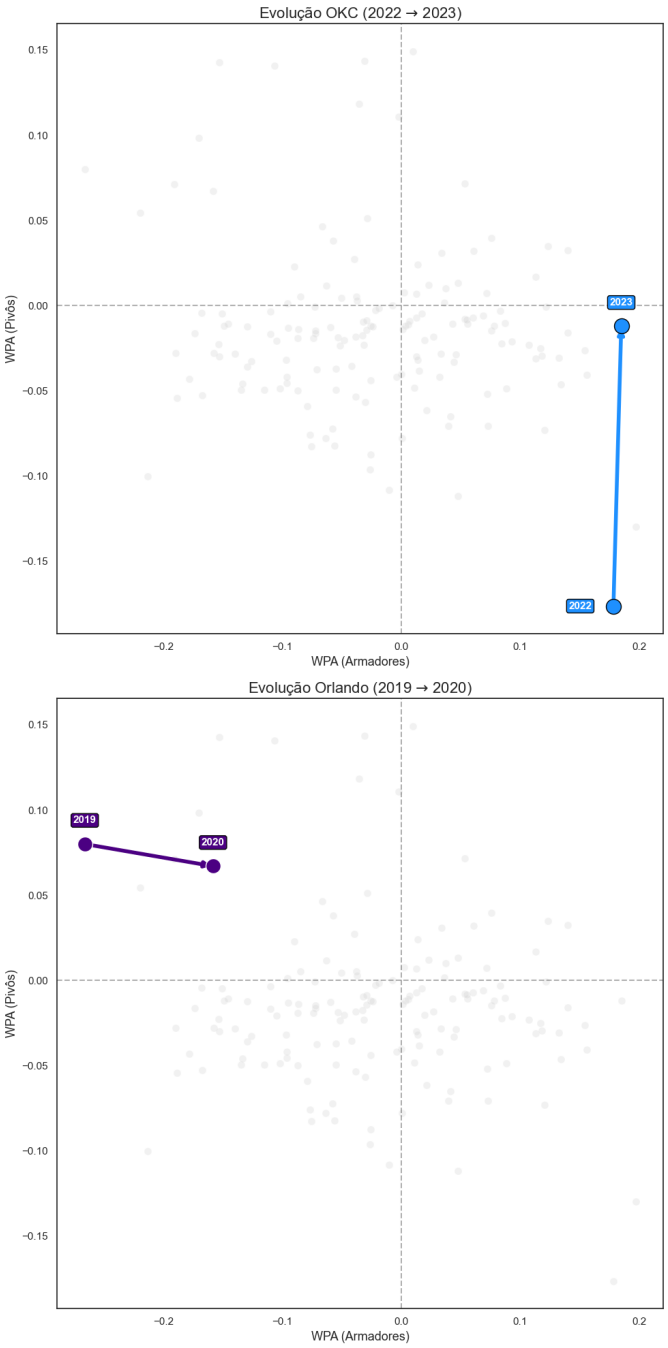

Fig. 10. Evolução da Contribuição Relativa entre temporadas: Estudos de caso do Oklahoma City Thunder e Orlando Magic.

Primeiramente, sugere-se a aplicação das métricas de Contribuição Relativa (RC_P) e Desvio Posicional (δ_P) como variáveis preditoras (features) em modelos de aprendizado de máquina. O objetivo seria quantificar o ganho marginal de acurácia na previsão de resultados de partidas ou no número total de vitórias na temporada ao incorporar o equilíbrio estrutural do elenco aos modelos tradicionais.

Uma segunda linha de pesquisa promissora reside na interseção com a análise econômica (Salary Cap Management). Futuros trabalhos podem cruzar os indicadores de "buraco" ($\delta_P < 0$) e "virtude" ($\delta_P > 0$) com os salários dos atletas, criando métricas de eficiência financeira. Isso permitiria identificar não apenas quais posições precisam de reforço, mas quais estão entregando o pior retorno sobre o investimento.

Por fim, a metodologia pode ser refinada através da substituição das posições clássicas por agrupamentos não-supervisionados (Clustering). Em vez de utilizar as categorias "Armador" ou "Pivô", futuros estudos poderiam aplicar algoritmos como K-Means para agrupar jogadores por perfis estatísticos de comportamento em quadra. Essa abordagem é particularmente interessante uma vez que as equipes da liga têm cada vez mais explorado o conceito do *positionless basketball*, ignorando posições clássicas pré-definidas para valorizar papéis em quadra. O cálculo de δ_P específico para esses arquétipos funcionais, poderia gerar diagnósticos de elenco ainda mais granulares.

VII. Conclusões

Os resultados apresentados neste trabalho demonstram o sucesso no desenvolvimento de uma metodologia analítica para a avaliação de impacto de substituições e construção de elencos na NBA. A etapa inicial de preparação de dados permitiu a integração eficaz de fontes distintas (ESPN e hoopR-data), abrangendo as temporadas de 2018 a 2024 e viabilizando a derivação dos *shifts* como unidade de análise.

A análise exploratória revelou padrões fundamentais sobre a dinâmica do jogo. Observou-se a persistência da vantagem dos times da casa na distribuição de vitórias, fenômeno que se manteve mesmo em condições de campo neutro ("bolha" de 2020). Identificou-se também que as substituições seguem padrões temporais rígidos, concentrando-se no início dos quartos e diminuindo drasticamente em prorrogações. Um achado contraintuitivo relevante foi a ausência de correlação direta entre a frequência de trocas (*shifts*) e a probabilidade de vitória, indicando que o determinante para o sucesso é a qualidade e o encaixe dos jogadores selecionados, e não a mera rotatividade do elenco.

No que tange à mensuração de impacto, a investigação conduziu a três conclusões centrais que fundamentam a proposta metodológica final:

- **Limitações do WPA:** A aplicação isolada do Win Probability Added demonstrou vieses de volume e contexto. Observou-se que o WPA médio tende a zero em grandes amostras e é altamente sensível a ruído estatístico, distorcendo a avaliação tanto de estrelas da liga (cujo impacto é diluído) quanto de jogadores de baixa minutagem (sujeitos à alta variância).
- **Eficácia da Contribuição Relativa:**A proposta de normalização através da Contribuição Relativa (RC_P) conseguiu mitigar as distorções supracitadas. A métrica foi capaz de isolar o impacto individual de cada posição do contexto global do jogo, evidenciando de forma clara a participação de cada grupo posicional no desempenho da equipe.
- **Robustez do Diagnóstico de Elenco:**A metodologia baseada na comparação do desempenho da equipe com a média da liga (Desvio Posicional δ_P) provou-se uma ferramenta robusta de gestão. A abordagem permitiu identificar com precisão tanto "buracos" estruturais (deficiências) quanto vantagens competitivas (virtudes) nas rotações, validando-se através de estudos de caso reais onde a correção desses desvios correlacionou-se com a melhora de desempenho esportivo.

Em suma, o trabalho oferece uma estrutura completa de análise, indo da compreensão dos padrões de substituição até uma ferramenta de diagnóstico prática para o planejamento de equipes.

References

[1] Sameer K. Deshpande and Shane T. Jensen, Estimating an NBA player's impact on his team's chances of winning, J. Quant. Anal. Sports, vol.12(2), pp. 51–72, Março 2016.

[2] P. Jalovaara, "Win probability estimation for strategic decision-making in esports", Master's Thesis, Aalto University, 2024.

[3] M. Lewis. Moneyball: The Art Of Winning An Unfair Game. Business book summary. WW Norton, 2003.

[4] D. Oliver, Basketball on Paper: Rules and Tools for Performance Analysis. Potomac Books, 2004.

[5] C. Li, H. Zhang, Y. Zhang, J. Shen, e R. An, "The application of artificial intelligence techniques in predicting game outcomes of professional basketball league: A systematic review", PLoS One, vol. 20, nº 6, p. e0326326, 2025. doi: 10.1371/journal.pone.0326326.

[6] Mark Carey. How data departments have evolved and spread across english football clubs. New York Times - The Athletic, September 3 2024.

[7] Claire Justin, Tom Griggs, and Richard Topping. The business of football: the big data arms race. Financial Times, Semptember 27 2024.

[8] S. Amez, B. Neyt, F. Van Nuffel, e S. Baert, "The right man in the right place? Substitutions and goal-scoring in soccer", Psychology of Sport & Exercise, vol. 54, p. 101898, 2021.

[9] C. A. S. Rüger, "Never change a winning team, or should you? Assessing the impact of coach interventions in player lineups in sports using machine learning models", Master's Thesis, Tilburg University, 2023.

[10] Robson Parmezan Bonidia, Jacques Duilio Brancher, and Rosangela Marques Busto. Data mining in sports: A systematic review. IEEE Latin America Transactions, 16(1):232–239, 2018.

[11] Sports Reference LLC. Basketball Reference. Disponível em: https://www.basketball-reference.com/. Acesso em: 07 dez. 2025.