

Segmentação e Caracterização Temporal de Sequências de Grafos: Uma Análise de Partidas de Futebol

João Lucas Lage Gonçalves, *Estudante, UFMG*, Wagner Meira Jr., *Professor, UFMG*,
and Adriano César Machado Pereira, *Professor, UFMG*

Resumo—Este trabalho apresenta uma nova forma de segmentação e caracterização de partidas de futebol baseada em técnicas de Redes Complexas e detecção de *Concept Drift*. Inicialmente, cada intervalo de jogo é modelado como um grafo que representa as interações entre jogadores, permitindo calcular métricas estruturais (densidade, clustering, centralidade e tamanho médio do caminho) ao longo do tempo. Em seguida, aplicamos um algoritmo de detecção de *Concept Drift* com o método KSWin para identificar pontos de transição tática e comportamental, dispensando dependência exclusiva de eventos clássicos, como gols ou substituições. Os resultados mostram que, em média, três pontos de mudança são detectados por partida. A análise do ranking das métricas de rede e do desvio padrão interno a cada time sugere que diferentes estratégias, desde estilos de jogo mais coesos até abordagens mais diretas, podem conduzir a êxitos em campo, reforçando a inexistência de uma única “fórmula” de sucesso.

Index Terms—Grafos Redes Complexas Concept Drift
Segmentação Temporal

I. INTRODUÇÃO

O uso de grafos e redes complexas (RC) [1] para modelar interações no mundo real tem se consolidado em diversas áreas do conhecimento. Essas técnicas permitem representar sistemas formados por entidades interconectadas e têm sido amplamente aplicadas em domínios como redes de computadores [2], redes sociais (SNA) [3] e biologia [4]. Em muitos desses casos, os sistemas modelados são dinâmicos, ou seja, transformam-se ao longo do tempo, com mudanças que podem ser desencadeadas por uma série de fatores internos ou externos. Muitas vezes, essas transformações ocorrem sem que suas causas sejam imediatamente identificadas.

Recentemente, uma área em que a aplicação de redes complexas vem ganhando destaque é o esporte, em particular o futebol [5]. Este é um esporte extremamente dinâmico, em que diversos fatores contextuais — como local da partida, placar atual, tempo decorrido e equipe adversária — têm influência significativa. Ademais, diferentemente de outros esportes com interrupções frequentes, o futebol mantém um fluxo de jogo quase ininterrupto. Nesse contexto, redes complexas são utilizadas para representar a dinâmica entre os jogadores em campo. Para representar a dinâmica temporal do futebol, empregamos sequências de grafos, na qual cada grafo representa o comportamento de uma equipe em um intervalo de 2 minutos de forma a capturar a evolução ao longo do tempo.

Aplicando conceitos de SNA, usamos métricas como centralidade, *clustering* e caminho médio para caracterizar as formas de jogo [6], [7]. Essa forma pode ser estudada em dois níveis principais: o tático, que envolve aspectos estruturais e estratégicos, como a organização e a configuração do time em campo, e o comportamental, que reflete a postura e a atitude mais direta ou mais paciente na construção de jogadas. A análise conjunta das métricas permite uma visão de como cada equipe caracteriza sua forma de jogo ao longo da partida.

Para detectar transições nesses padrões sem depender unicamente de eventos clássicos, como gols ou substituições, aplicamos técnicas de detecção de *Concept Drift (CD)* [8] para segmentar as partidas de futebol sem depender de interrupções claras, como gols, substituições e intervalo. Com o avanço das tecnologias de coleta e processamento de dados no futebol, em especial o *tracking* (que fornece a posição dos 22 jogadores e da bola a cada *frame*) e os eventos (registrando passes, carregadas de bola, finalizações etc.), torna-se possível integrar esses registros para criar modelos de grafos que capturem a evolução temporal de cada equipe ao longo de uma partida. Desse modo, podemos caracterizar não apenas a forma de jogo das equipes em um determinado instante, mas também acompanhar como essas configurações táticas e comportamentais se transformam ao longo do tempo.

Neste trabalho, apresentamos uma metodologia inovadora para a caracterização e segmentação temporal de partidas de futebol. Essa metodologia habilita treinadores e analistas de desempenho esportivo a identificar momentos-chave que influenciam o desempenho das equipes, o que permite uma preparação pré-jogo com uma forma de jogo adaptada às mudanças de contexto dentro de uma partida, e assim realizar ajustes mais precisos durante as mesmas. Os objetivos principais deste trabalho são criar uma nova estratégia de segmentação de partidas de futebol baseada em variações nas formas de jogo das equipes e assim caracterizá-las e compará-las em diferentes fases ao longo das partidas.

II. REFERENCIAL TEÓRICO

Em 2010, [9] foi pioneiro ao aplicar grafos para avaliar o desempenho de jogadores, onde cada jogador era modelado como um nó e os passes entre eles como arestas. Logo em seguida, [10] introduziu redes de passes para representar as interações durante o jogo, representando a partida como uma única rede estática que considerava apenas as interações

diretas, como passes. Embora as redes de passes tenham sido o foco inicial, essas abordagens apresentam limitações. A principal delas é não conseguir capturar as variações dinâmicas que ocorrem ao longo do tempo, devido ao caráter estático das redes utilizadas, que representam toda a partida em uma única rede agregada. Em contraponto, algumas abordagens tentam ir além das redes estáticas. Por exemplo, [11] propôs a construção de uma rede para cada sequência de passes, onde cada sequência é modelada separadamente. Já [12] ampliou essa ideia ao incorporar dados de eventos, como passes, dribles e interceptações, e dados de *tracking*, o que permitiu uma representação mais detalhada da movimentação e interações dos jogadores.

O presente trabalho mistura todas essas abordagens ao construir grafos representando sequências de 2 minutos de tempo decorrido, usando dados de eventos como passes, carregadas de bola, e a posição média dos jogadores durante esse período, obtida a partir de dados de *tracking*. Além disso, foram calculadas as métricas apresentadas nesses trabalhos para cada um dos grafos, caracterizando a forma de jogo das equipes em diversas fases do jogo.

Para avaliar a evolução das características das equipes baseadas em redes de passes, [13] introduz um *framework* que analisa a consistência da "identidade tática" das equipes ao longo de um campeonato, permitindo observar variações na escalabilidade das redes que impactam no desempenho final das equipes. Nesse trabalho foi feita a análise de consistência e identidade tanto dentro de uma partida, comparando os valores entre intervalos, quanto ao longo das partidas.

Em [14] é implementado *framework* automatizado denominado SoccerCPD para identificar alterações em formações e papéis dos jogadores durante as partidas. Ele utiliza técnicas baseadas em algoritmos de detecção de pontos de mudança não supervisionados, que consideram a variação nas métricas de centralidade dos jogadores e na estrutura da rede de passes, permitindo a identificação de mudanças significativas na tática da equipe, porém se limitando a detecção de mudanças em formações e papéis previamente conhecidos. Nesse trabalho, utilizamos técnicas de *CD* baseadas em janelas adaptativas para detectar os pontos de mudanças na partida, baseadas em métricas de distância de grafos para segmentar as partidas.

As técnicas de detecção de *CD* baseadas em janelas adaptativas ajustam o tamanho da janela de dados de acordo com as mudanças observadas nos dados. O objetivo é detectar quando a distribuição dos dados muda. Essas técnicas são particularmente úteis em fluxos de dados, onde os padrões podem evoluir ao longo do tempo. Uma das técnicas é a *KSWIN* [15], que usa o teste de Kolmogorov-Smirnov (KS) [16] para detectar mudanças nas distribuições de dados. *KSWIN* mantém uma janela deslizante de dados recentes e compara uma amostra dos dados mais recentes com uma amostra dos dados não tão recentes da janela. O teste KS quantifica a distância entre as duas distribuições e, se essa distância for maior do que um limiar, um *CD* é detectado. Uma vantagem do *KSWIN* é que ele não assume nenhuma distribuição específica dos dados.

III. METODOLOGIA

A metodologia - ilustrada na Figura 1 - proposta neste trabalho consiste em uma abordagem sistemática para análise de partidas de futebol utilizando grafos e técnicas de detecção de mudanças. O processo inicia-se com a coleta e preparação dos dados de *tracking* e eventos das partidas, seguido pela construção de grafos direcionados e ponderados em intervalos de 2 minutos, caracterização desses grafos através de métricas de redes complexas, e aplicação de técnicas de *CD* para identificar transições significativas no comportamento das equipes. Esta abordagem permite uma análise quantitativa e objetiva das mudanças táticas e comportamentais das equipes ao longo do jogo.

A. Dados

Os dados utilizados neste estudo foram fornecidos pela PFF FC em colaboração com o SALab – Speed – DCC – UFMG e consistem em informações detalhadas de 190 partidas do Campeonato Inglês da temporada 2022/2023. O conjunto de dados é composto por dois tipos principais de informações: dados de *tracking* e dados de eventos. Os dados de *tracking* contêm as coordenadas espaciais dos 22 jogadores em campo e da bola em cada *frame* da partida, onde os *frames* têm uma frequência de aproximadamente 30Hz. Já os dados de eventos descrevem ações específicas realizadas durante os jogos, como passes, carregadas de bola (dribles inclusos) e finalizações, identificando o tipo de evento, o jogador responsável e outras informações associadas.

Antes da análise, os dados passaram por um processo de preparação para garantir sua qualidade e adequação ao escopo do trabalho. Primeiramente, foram identificados e filtrados os *frames* inválidos, que incluíam registros com dados faltantes ou posições inconsistentes de jogadores e da bola. Além disso, foram removidas jogadas de bola parada, como escanteios, laterais, faltas e cobranças de pênalti, já que essas situações introduzem dinâmicas muito diferentes das observadas no fluxo contínuo do jogo. Foi feita também a filtragem para apenas tipos de eventos de interesse: passes (95%), carregadas de bola e dribles (2.7%), e chutes (2.3%).

Além disso foram feitos outros tratamentos de forma a ter uma padronização dos dados para serem um fluxo contínuo de tempo e retirar inversão de lados de campo a fim de desconsiderar o intervalo real de jogo, foi feito também o cálculo da posição (x,y) média de cada jogador em cada intervalo de 2 minutos e agrupamento dos eventos por jogador, receptor (no caso de carregadas o próprio jogador e em caso de chutes o gol), intervalo e partida.

B. Modelagem Temporal

Para modelar as interações entre os jogadores durante a partida, foi utilizada a biblioteca *NetworkX*¹ do *Python*, amplamente reconhecida para manipulação e análise de grafos. A modelagem foi realizada em intervalos fixos de dois minutos, durante os quais foram construídos dois grafos direcionados e ponderados, um para cada equipe. Os nós desses grafos

¹<https://networkx.org/>



Figura 1. Fluxograma da metodologia do trabalho

Grafo da partida 4436 no intervalo 45 - Crystal Palace

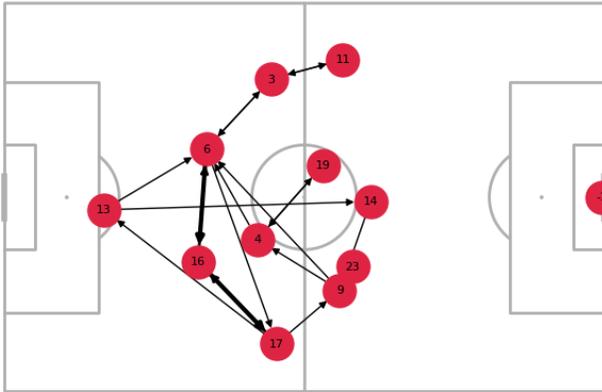


Figura 2. Grafo exemplo

representam os jogadores presentes em campo no intervalo analisado, enquanto as arestas descrevem as interações entre eles, através de passes, carregadas de bola e chutes.

Os passes foram modelados como arestas que partem do jogador que realiza o passe e apontam para o jogador que o recebe, com pesos correspondentes ao número de passes realizados entre o par de jogadores durante o intervalo considerado. As carregadas de bola, por outro lado, foram representadas como arestas que saem do jogador com posse da bola e retornam para ele mesmo, ou seja, auto-arestas que indicam o número de vezes que o jogador conduziu a bola em campo. Já os chutes foram modelados como arestas direcionadas que partem do jogador que realiza a finalização e apontam para um nó adicional fixo que representa o gol. Esse nó especial foi incluído no grafo como uma forma de contabilizar todas as tentativas de finalização de uma equipe, sendo o peso dessas arestas determinado pelo número de chutes realizados pelo jogador no intervalo de dois minutos.

Além das conexões entre os jogadores, cada nó do grafo foi enriquecido com informações sobre o posicionamento médio do respectivo jogador durante o período analisado. As coordenadas (x,y) dos jogadores no campo foram calculadas como a média de sua posição ao longo do intervalo de dois minutos, com base em dados de *tracking*.

C. Caracterização das equipes

A análise de redes complexas oferece um conjunto de métricas que permitem caracterizar quantitativamente as propriedades estruturais e dinâmicas de sistemas interconectados. Nesse contexto, elas servem como uma forma de modelar as

interações entre os jogadores e compreender tanto aspectos táticos quanto comportamentais das equipes durante uma partida de forma quantitativa.

Para cada grafo construído, foram calculadas as métricas, gerando séries temporais que descrevem a evolução das características estruturais de cada equipe ao longo da partida. O cálculo foi realizado separadamente para cada time em intervalos consecutivos de 2 minutos, permitindo uma análise temporal detalhada das mudanças nas propriedades das redes. As métricas selecionadas foram:

1) *Densidade*: Mede a proporção entre o número real de arestas presentes e o número máximo de arestas possíveis em um grafo completamente conectado. No contexto das redes de eventos do futebol, uma densidade elevada sugere que a maior parte dos jogadores está interagindo entre si, revelando uma equipe mais coesa e uma forma de jogo mais coletiva com a bola. Já uma densidade baixa pode indicar segmentação ou concentração das jogadas em subgrupos específicos, demonstrando menor integração global entre os atletas.

2) *Clustering*: Reflete a probabilidade de que, se dois jogadores estão conectados, também haja uma ligação entre eles e um terceiro jogador, formando um triângulo. No futebol, essa métrica ajuda a identificar pequenos grupos colaborativos. Um valor de *clustering* mais alto sugere interações intensas em pequenos subgrupos, indicando maior coesão local e facilidade para a formação de triangulações e trocas de passes curtos. Por outro lado, um valor de *clustering* mais baixo indica que esses pequenos grupos de jogadores são menos frequentes, apontando para um estilo de jogo mais disperso e com menor concentração de interações locais. Ao quantificar a presença desses subgrupos, o clustering mostra como núcleos de jogadores contribuem para a dinâmica coletiva da equipe.

3) *Centralidade*: A *betweenness centrality* avalia quantas vezes um nó (neste caso, um jogador) aparece nos caminhos mais curtos entre todos os pares de nós em uma rede. Na prática do futebol, isso se traduz na quantidade de jogadas nas quais aquele jogador atua como “ponte” ou intermediário. Ao considerar a mediana da *betweenness centrality* de todos os jogadores, observa-se quanto “típico” é o nível de intermediação de passes dentro da equipe. Um valor de mediana mais elevado sugere que vários atletas desempenham papéis de “ponte” nos caminhos mais curtos, indicando um estilo de jogo coletivo, em que o time não depende exclusivamente de um ou dois jogadores para articular as jogadas. Em contrapartida, uma mediana baixa pode revelar que poucos jogadores concentram grande parte das transições de bola, refletindo uma configuração mais centralizada e possivelmente menos equilibrada.

4) *Caminho Médio para o Gol (CMG)*: Mede a distância média dos caminhos mais curtos entre todos os pares de nós.

Aqui foi adaptada para o futebol calculando-se a distância média do caminho mais curto entre cada jogador na rede e o gol, avaliando assim a eficiência do fluxo de jogo, de modo que valores menores de CMG indicam um estilo mais direto e rápido de troca de passes, enquanto valores mais altos podem refletir estratégias que privilegiam a manutenção da posse de bola ou uma construção de jogadas mais paciente.

D. Segmentação das partidas

O algoritmo de detecção de *CD* proposto nesse trabalho se baseia em métricas de distância entre grafos para segmentar as partidas. A métrica selecionada foi a Distância de Wasserstein [17], que, de forma geral, mensura o “custo” de transformar uma distribuição de probabilidade em outra. Para aplicá-la na comparação entre dois grafos, cada grafo é representado como uma distribuição dos graus de seus nós, permitindo que a Distância de Wasserstein quantifique o quão diferentes são as estruturas de interações em cada intervalo de jogo.

Como entrada para o algoritmo de detecção de *CD* foi utilizado o vetor d que representa a soma das distâncias de Wasserstein calculada por time, como mostrada na equação 1:

$$\mathbf{d} = [(W(GH_i, GH_{i+1}) + W(GA_i, GA_{i+1}))]_{i=1}^{n-1} \quad (1)$$

onde n é o número de intervalos na partida, GH_i representa o grafo do time mandante no intervalo i e GA_i o grafo do time visitante no intervalo i , W representa a distância *Wasserstein* entre grafos do mesmo time nos intervalos i e $i + 1$.

A fim de detectar mudanças nos padrões de jogo, foi empregada a técnica *KSWin* (*Kolmogorov-Smirnov Windowing*) usando como entrada o vetor d . Inicialmente, calcula-se a distância de Wasserstein entre cada par de intervalos da partida, gerando uma série temporal de distâncias. Em seguida, o algoritmo *KSWin* mantém duas janelas deslizantes (ou uma janela dividida em duas partes): uma janela antiga (comportamento histórico) e uma janela recente (comportamento atual). A cada novo valor de distância, esse valor entra na janela recente, enquanto a janela antiga é atualizada para refletir os dados anteriores. Então, o *KSWin* compara estatisticamente as distribuições de valores das duas janelas por meio do teste de Kolmogorov-Smirnov, que examina a maior diferença entre as funções de distribuição acumulada (CDF) de cada amostra. Se essa diferença for maior que um limiar pré-estabelecido, conclui-se que houve uma mudança estatisticamente significativa e marca-se o instante como ponto de mudança. No contexto do futebol, esse ponto indica uma possível transição tática ou comportamental, pois revela que as estruturas de interação entre os jogadores se tornaram distintas o suficiente para serem consideradas um novo padrão de jogo.

E. Avaliação

O principal desafio na implementação de algoritmos de *concept drift* não supervisionados está na avaliação de desempenho [18]. Para contornar isso, utilizamos uma base de dados sintética rotulada, permitindo comparar o desempenho das duas técnicas desenvolvidas. Esse conjunto de dados sintéticos foi criado com base em um método parametrizável que simula

mudanças conceituais em séries temporais. Os dados foram gerados por uma função desenvolvida para criar sequências numéricas divididas em segmentos, onde cada segmento possui características distintas devido à introdução de deslocamentos graduais, simulando os *drifts*.

A geração dos dados considera três parâmetros principais, que foram definidos a partir das características dos dados utilizados no trabalho: o número total de amostras, o número de *drifts* e a faixa de valores dentro da qual os dados são gerados. O processo inicia com a definição da quantidade de amostras restantes a serem distribuídas entre os segmentos. Em cada iteração, um segmento é criado com um número aleatório de amostras, respeitando a proporção das amostras restantes, e os valores desse segmento são gerados de forma uniforme dentro do intervalo especificado. Para simular os *drifts*, a partir do segundo segmento, é adicionado um deslocamento aleatório aos valores do segmento atual, deslocando-os em relação ao segmento anterior. Essa mudança cria variações perceptíveis nos dados, correspondentes a alterações nas condições ou distribuições subjacentes. Os pontos onde os *drifts* ocorrem são registrados em uma lista de índices, permitindo identificar com precisão as mudanças inseridas.

Além disso, foi realizada uma otimização de parâmetros para o algoritmo, visando encontrar a melhor configuração para os três parâmetros: *alpha* (a), *window size* (ws), *stat size* (ss). O a corresponde ao nível de significância do teste de Kolmogorov-Smirnov, enquanto ws e ss definem, respectivamente, o tamanho da janela de referência e o tamanho da janela recente para detecção de mudanças. Para isso, utilizou-se o método *grid search* [19], que realiza uma busca exaustiva sobre um conjunto predefinido de valores para cada parâmetro. A tabela I mostra os resultados obtidos, após três execuções da busca, cada uma com um número esperado diferente de *drifts*. Cada configuração foi avaliada com base em quatro métricas:

- **Recall:** avalia a capacidade de identificar corretamente todas as mudanças reais, ou seja, quantos *drifts* existentes foram de fato detectados;
- **Taxa de Falsos Positivos:** indica a proporção de mudanças detectadas que não aconteceram realmente;
- **F1-score:** é a média harmônica de Precisão e Recall, proporcionando uma medida equilibrada entre a qualidade e a completude das detecções;
- **Delay:** é a média de quantos intervalos o algoritmo leva para detectar uma mudança após ela ter realmente ocorrido.
- **Precisão:** indica a proporção de mudanças corretamente detectadas em relação ao total de mudanças reportadas pelo algoritmo. Em outras palavras, mede quantos dos “*drifts*” identificados eram, de fato, verdadeiros.

Em seguida, calculou-se a média de desempenho de cada configuração ao longo das três execuções, e a melhor combinação encontrada foi $a = 0.1$, $ws = 6$, $ss = 3$.

IV. RESULTADOS

A análise das segmentações detectadas pelo algoritmo *KSWin* revelou em média 3.2 pontos de *CD* detectados por partida. Como mostra o histograma da Figura 2, há uma

Parâmetros	Recal	Precisão	FPR	F1-score	Delay
a:0.1,ss:3,ws:6	0.72	0.83	0.17	0.71	4.00
a:0.1,ss:3,ws:10	0.72	0.67	0.33	0.61	5.88
a:0.1,ss:3,ws:12	0.72	0.59	0.41	0.55	6.25
a:0.1,ss:3,ws:15	0.44	0.28	0.72	0.33	9.67
a:0.05,ss:4,ws:10	0.33	0.33	0.67	0.33	7.00

Tabela I

COMPARAÇÃO DA PERFORMANCE DOS TOP 5 PARÂMETROS DO ALGORITMO KSWIN. *a*: ALPHA, *ws*: WINDOW_SIZE, *ss*: STAT_SIZE. FPR: TAXA DE FALSO POSITIVOS

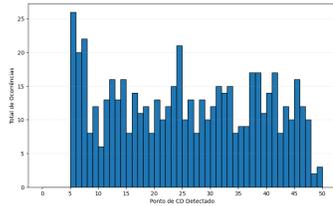


Figura 3. Histograma da distribuição dos pontos de detecção de CD

maior concentração desses pontos entre os intervalos 5 e 10, correspondendo aproximadamente aos minutos 10 a 20 de jogo. Nessa faixa, é comum observar mudanças na postura das equipes inicial para se adaptar ao que o adversário mostra. Além disso, outro pico significativo surge próximo ao intervalo 25, que costuma coincidir com o final do primeiro tempo e o início do segundo, período em que as equipes muitas vezes alteram estratégias ou fazem substituições. Por fim, verifica-se um terceiro pico de ocorrências próximo ao intervalo 40, representando cerca de 80 minutos de jogo, momento em que, em vários casos, há uma intensificação das ações ofensivas ou defensivas na reta final da partida.

Para realizar a caracterização dos padrões de jogo, foram calculadas as métricas de Densidade, Centralidade, *Clustering* e Caminho médio para cada grafo gerado nos intervalos de tempo definidos, resultando em uma série temporal de atributos estruturais. Em seguida, essas métricas foram agregadas em diferentes níveis temporais: por intervalo, por fase (agregando os intervalos pertencentes a cada fase de jogo detectada) e por partida (considerando a média de todos os intervalos ao longo do jogo). Essa abordagem hierárquica possibilita observar tendências tanto em escala micro (intervalos específicos) quanto em escala macro (as fases e a partida inteira), oferecendo uma visão aprofundada da evolução tática das equipes ao longo do jogo.

A análise dos valores dessas métricas revelou, ainda, correlações com as fases de jogo identificadas pelo algoritmo de *concept drift*, como pode ser observado na Figura 4. Em diversos casos, transições significativas de densidade, centralidade e clustering ocorreram próximas aos pontos de mudança detectados, o que sugere que a segmentação temporal proposta reflete, de fato, mudanças na forma de jogar das equipes.

Para analisar as formas de jogo dos times foi realizado um ranqueamento dos times com base em cada métrica analisada, permitindo identificar quais equipes se destacavam em diferentes aspectos do jogo. A análise da tabela II mostra uma correlação entre as métricas, onde os times tendem a manter posições próximas nos rankings. Os valores de CMG (Cami-

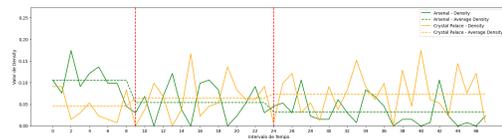


Figura 4. O gráfico apresenta a evolução da métrica de Densidade ao longo dos intervalos de tempo de uma partida. As linhas sólidas mostram a densidade em cada intervalo, enquanto as linhas tracejadas indicam a densidade média de cada time na fase observada. As linhas verticais vermelhas representam os pontos de CD identificados.

Time	den	clu	cen	cmg	cla
Manchester City	1	1	1	1	2
Liverpool	2	2	2	6	7
Chelsea	3	3	5	5	10
Arsenal	4	4	4	7	1
Brighton & Hove Albion	5	5	3	4	6
Tottenham Hotspur	6	7	7	3	5
Manchester United	7	6	6	2	4
Leicester City	8	8	10	8	14
Wolverhampton Wanderers	9	10	12	11	19
West Ham	10	9	8	13	17
Crystal Palace	11	11	9	9	12
Aston Villa	12	12	11	12	11
Newcastle United	13	14	13	15	3
Leeds United	14	16	15	10	13
Fulham	15	13	16	19	8
Southampton	16	15	17	18	20
Everton	17	17	14	16	18
AFC Bournemouth	18	18	18	14	16
Brentford	19	19	19	20	9
Nottingham Forest	20	20	20	17	15

Tabela II

A TABELA EXIBE O RANKING DE CADA EQUIPE EM QUATRO MÉTRICAS DE CARACTERIZAÇÃO DE REDES, ALÉM DE SUA CLASSIFICAÇÃO AO FIM DO PRIMEIRO TURNO DO CAMPEONATO. A COLUNA “DEN” REPRESENTA A POSIÇÃO NO RANKING DE DENSIDADE, “CLU” INDICA O RANKING DE *clustering*, “CEN” MOSTRA A POSIÇÃO EM CENTRALIDADE, “CMG” SE REFERE AO RANKING DE CAMINHO MÉDIO PARA O GOL E “CLA” CORRESPONDE À COLOCAÇÃO DO TIME NA TABELA DE CLASSIFICAÇÃO. QUANTO MENOR O NÚMERO, MELHOR A POSIÇÃO DA EQUIPE NAQUELE CRITÉRIO.

inho Médio Global) apresentam algumas variações que fogem ao padrão das outras medidas, sugerindo que essa métrica capta uma dimensão distinta do comportamento coletivo em campo. Isso sugere que equipes que se destacam em uma determinada característica estrutural de jogo tendem a apresentar desempenho semelhante em outras métricas, indicando uma possível relação entre diferentes aspectos táticos e estratégicos do jogo.

Apesar de não haver uma relação clara entre esses indicadores de rede e o desempenho esportivo dos times na competição, o Manchester City se destaca ao liderar as quatro métricas analisadas, ocupando simultaneamente a segunda posição na classificação da liga. Esse resultado mostra um estilo de jogo dominante, caracterizado por alta coesão coletiva (densidade elevada), grande interação local (clustering alto), facilidade de conexão (centralidade elevada) e caminhos médios maiores, indicando uma postura ofensiva, mas paciente na construção jogadas. Já o Liverpool mantém posições semelhantes em densidade, clustering e centralidade, mas caiu para sexto em CMG, o que sugere que, apesar de ter um perfil tático parecido, opta por um comportamento mais direto em direção ao gol.

Por outro lado, Arsenal e Chelsea exibem desempenhos próximos em quase todas as métricas de rede, porém apresentam uma grande diferença na classificação geral do campeonato, destacando que não há uma “fórmula única” de jogo que garanta bons resultados. Essa ideia se confirma ao observar equipes como Newcastle United, Leeds United e Brentford, que figuram na metade de baixo dos rankings de rede, tendo como característica menor posse de bola, menor coesão entre os jogadores e um comportamento mais direto em direção ao gol, mas ainda assim se encontram na metade superior da tabela classificatória, demonstrando que diferentes estilos de jogo podem levar a desempenhos satisfatórios em campo.

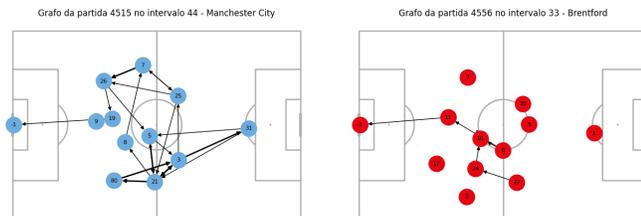


Figura 5. Os círculos representam os jogadores em campo (identificados por seus números de camisa), enquanto o nó -1 indica a meta adversária. No grafo do Manchester City, observa-se maior número de conexões e subgrupos, refletindo sua liderança nas métricas de densidade, clustering e centralidade. Já no grafo do Brentford, há menos passes trocados entre os atletas, menor coesão e uma progressão mais direta ao gol, em concordância com as posições inferiores do time nas mesmas métricas de ranqueamento.

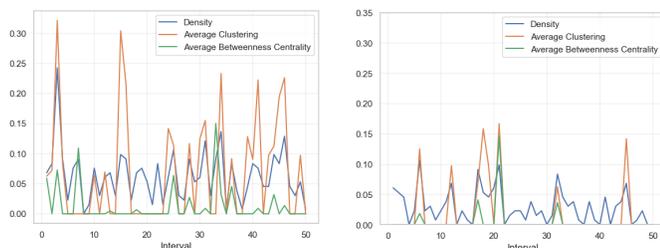


Figura 6. Comparação das métricas de densidade (linha azul), clustering médio (linha laranja) e betweenness centrality média (linha verde) para duas equipes distintas, analisadas intervalo a intervalo. À esquerda, observa-se que o Arsenal apresenta picos mais pronunciados em densidade e clustering, sugerindo uma variação tática considerável ao longo da partida. À direita, o Nottingham Forest exibe oscilações mais discretas, indicando um padrão de jogo relativamente mais consistente.

É possível realizar a análise das métricas de forma mais detalhada, avaliando-se cada intervalo de tempo individualmente, como ilustra a Figura 6. Nessa abordagem, verifica-se que a métrica de Caminho Médio para o gol tende a apresentar as maiores variações entre os intervalos, indicando que as equipes alteram significativamente seu modo de construir jogadas em alguns momentos — por exemplo, passando de um estilo mais direto, com poucos toques até o gol, para uma postura mais paciente, envolvendo maior troca de passes antes da finalização. Por outro lado, a Betweenness Centrality mostra, em geral, as variações mais modestas, o que sugere que o nível de intermediação de passes dentro da equipe permanece relativamente estável durante o jogo.

No que diz respeito aos times, aqueles com valores de métricas menores (por exemplo, baixa densidade ou clustering) tendem a exibir maior consistência dentro das partidas e ao

longo do campeonato, mantendo um padrão de jogo mais uniforme. Em contrapartida, as equipes com métricas mais elevadas frequentemente apresentam oscilações mais acentuadas, refletindo um estilo de jogo mais dinâmico e sujeito a adaptações táticas consideráveis conforme o desenrolar dos 90 minutos.

V. CONCLUSÃO

A metodologia proposta neste trabalho combinou conceitos de Redes Complexas e técnicas de detecção de *concept drift* para segmentar e caracterizar partidas de futebol de forma mais granular e dinâmica. A partir do uso de métricas como densidade, clustering, centralidade, caminho médio e distância de grafo, foi possível evidenciar tanto os momentos de transição na forma de jogo quanto o grau de variação interna de cada equipe ao longo dos 90 minutos.

Os resultados mostraram que há, em média, cerca de três pontos de mudança detectados em cada partida, refletindo transições que nem sempre são identificáveis apenas por eventos clássicos (como gols ou substituições). Observou-se também que as métricas de rede correlacionam-se com essas quebras de padrão: densidade, centralidade e clustering, por exemplo, costumam ter alterações notáveis próximo aos pontos de *concept drift*. Por outro lado, a análise do ranking e do desvio padrão das métricas revelou que não existe um único estilo de jogo vinculado ao sucesso esportivo; equipes podem ter perfis estruturais e comportamentais bastante distintos, e ainda assim atingir resultados expressivos.

Ademais, concluiu-se que a abordagem hierárquica (analisar métricas em nível de intervalo, fase e partida) possibilita enxergar tanto tendências micrológicas (mudanças pontuais) quanto tendências macrológicas (estratégias e comportamentos mais gerais). Exemplos como o Manchester City, posicionado de forma proeminente em quase todas as métricas, e times como o Brentford e o Leeds, que apresentam menor coesão mas resultados satisfatórios na tabela, reforçam a diversidade de caminhos para a eficácia em campo.

Em trabalhos futuros, pretende-se explorar métricas adicionais que integrem simultaneamente as fases ofensiva e defensiva, de modo a avaliar se o time se organiza de forma diferente dependendo da posse de bola. Outra possível extensão consiste em adicionar a distância física como uma nova forma de interação entre os jogadores, e assim explorar também o posicionamento dos jogadores durante os intervalos. Além disso, será interessante testar algoritmos de *concept drift* diferentes, permitindo comparar desempenho e sensibilidade na detecção de pontos de mudança. Outra extensão é relacionar os pontos de mudança detectados com eventos visíveis das partidas (por exemplo, substituições ou gols) pode aprofundar ainda mais a compreensão de como as alterações nas formas de jogo das equipes são causadas ou causam esses eventos.

Por fim, a proposta de segmentar e caracterizar partidas de futebol por meio de grafos e *concept drift* mostra-se promissora não só para fins de análise acadêmica, mas também como ferramenta de suporte a treinadores e analistas de desempenho, fornecendo análises que vão além dos eventos tradicionais e permitindo ajustes táticos mais embasados ao longo de uma partida.

REFERÊNCIAS

- [1] Strogatz and S. H. “Exploring complex networks,” *nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [2] Batool, Komal, Niazi, and M. A. “Modeling the internet of things: a hybrid modeling approach using complex networks and agent-based models,” *Complex Adaptive Systems Modeling*, vol. 5, pp. 1–19, 2017.
- [3] Jin, Xin, Wang, and Yaohua, “Research on social network structure and public opinions dissemination of micro-blog based on complex network analysis,” *Journal of Networks*, vol. 8, no. 7, p. 1543, 2013.
- [4] Braun, Pascal, Gingras, and Anne-Claude, “History of protein–protein interactions: From egg-white to complex networks,” *Proteomics*, vol. 12, no. 10, pp. 1478–1498, 2012.
- [5] Lusher, Dean, Robins, Garry, Kremer, and Peter, “The application of social network analysis to team sports,” *Measurement in physical education and exercise science*, vol. 14, no. 4, pp. 211–224, 2010.
- [6] F. M. Clemente, F. M. L. Martins, D. Kalamaras, P. d. Wong, and R. S. Mendes, “General network analysis of national soccer teams in fifa world cup 2014,” *International Journal of Performance Analysis in Sport*, vol. 15, no. 1, pp. 80–96, 2015.
- [7] J. Gama, P. Passos, K. Davids, H. Relvas, J. Ribeiro, V. Vaz, and G. Dias, “Network analysis and intra-team activity in attacking phases of professional football,” *International Journal of Performance Analysis in Sport*, vol. 14, no. 3, pp. 692–708, 2014.
- [8] G. Widmer and M. Kubat, “Learning in the presence of concept drift and hidden contexts,” *Machine learning*, vol. 23, pp. 69–101, 1996.
- [9] Duch, Jordi, Waitzman, J. S., Amaral, and L. A. Nunes, “Quantifying the performance of individual players in a team activity,” *PLoS one*, vol. 5, no. 6, p. e10937, 2010.
- [10] Passos, Pedro, Davids, Keith, Araújo, Duarte, Paz, N, Minguéns, J, Mendes, and Jose, “Networks as a novel tool for studying team ball sports as complex social systems,” *Journal of Science and Medicine in Sport*, vol. 14, no. 2, pp. 170–176, 2011.
- [11] Korte, Florian, Link, Daniel, Groll, Johannes, Lames, and Martin, “Play-by-play network analysis in football,” *Frontiers in psychology*, vol. 10, p. 1738, 2019.
- [12] Buldu, J. M, Garrido, David, Antequera, D. R, Busquets, Javier, Estrada, E, Resta, R, del Campo, and R. López, “Football tracking networks: Beyond event-based connectivity,” *arXiv preprint arXiv:2011.06014*, 2020.
- [13] Garrido, D, Antequera, DR, Busquets, J, L. D. Campo, R, R. Serra, R, J. Vielcazat, S, Buldú, and JM, “Consistency and identifiability of football teams: a network science perspective,” *Scientific reports*, vol. 10, no. 1, p. 19735, 2020.
- [14] Kim, Hyunsung, Kim, Bit, Chung, Dongwook, Yoon, Jinsung, Ko, and Sang-Ki, “Soccercpd: Formation and role change-point detection in soccer matches using spatiotemporal tracking data,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3146–3156.
- [15] C. Raab, M. Heusinger, and F.-M. Schleif, “Reactive soft prototype computing for concept drift streams,” *Neurocomputing*, vol. 416, pp. 340–351, 2020.
- [16] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [17] V. M. Panaretos and Y. Zemel, “Statistical aspects of wasserstein distances,” *Annual review of statistics and its application*, vol. 6, no. 1, pp. 405–431, 2019.
- [18] R. N. Gemaque, A. F. J. Costa, R. Giusti, and E. M. Dos Santos, “An overview of unsupervised drift detection methods,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 6, p. e1381, 2020.
- [19] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 68, no. 1, pp. 49–67, 2006.