

De torcedor para torcedor: Uma análise sobre as interações digitais no futebol brasileiro

Henrique de Souza Campos¹

¹Universidade Federal de Viçosa - Campus Florestal (UFV-CAF)
Florestal – MG – Brazil

Abstract. *Football is one of Brazil's greatest passions, generating intense digital engagement. This study investigates interactions on the official YouTube channels of Série A teams in the Brazilian Championship, modeling a network that connects users and videos based on comments from the most popular videos. The analysis revealed engagement patterns among fans of different clubs, highlighting possible connections and rivalries. Additionally, the study applies complex network concepts to understand social interactions in the digital sports environment.*

Resumo. *O futebol é uma das maiores paixões do Brasil, gerando intenso engajamento digital. Este trabalho investiga interações nos canais oficiais dos times da Série A do Campeonato Brasileiro no YouTube, modelando uma rede que conecta usuários e vídeos a partir dos comentários nos vídeos mais populares. A análise revelou padrões de engajamento entre torcedores de diferentes clubes, destacando possíveis conexões e rivalidades. Além disso, o estudo aplica conceitos de redes complexas para compreender as interações sociais no ambiente esportivo digital.*

1. Introdução

O futebol, mais do que um esporte, é uma verdadeira paixão nacional. Ele atravessa fronteiras regionais, sociais e culturais, sendo importante na união de milhões de brasileiros em torno de seus times e jogadores, principalmente. Em um país onde o futebol é tão comentado e debatido no dia a dia, as interações digitais sobre os times da série A do Campeonato Brasileiro ganham cada vez mais destaque, na medida em que as redes sociais tem crescido, facilitando as interações. Este artigo propõe-se a analisar as interações entre os torcedores nas plataformas digitais, especialmente em comentários do Youtube, utilizando dados sobre comentários em vídeos oficiais dos clubes nessa rede social, para construir uma rede de relacionamentos que revele padrões de rivalidade e engajamento entre os torcedores dos times. A paixão por futebol e as dinâmicas de interações digitais são os principais estimulantes desse estudo.

2. Coleta de dados

A coleta de dados para este projeto foi realizada a partir de vídeos disponíveis no YouTube nos canais oficiais dos 20 times que participam da Série A do Campeonato Brasileiro. A exceção foi o Cuiabá, que não possui um canal oficial na plataforma e, portanto, foi excluído da coleta. A escolha por esses canais tem como objetivo obter um conjunto de dados representativo das interações dos torcedores com o conteúdo produzido pelos próprios times, o que permite uma análise mais detalhada e autêntica sobre o engajamento

nas redes sociais. Vale ressaltar que para a coleta dos dados, foi utilizada a API do próprio YouTube.

Inicialmente, foram feitas uma listagem e uma busca manual dos canais oficiais dos times no YouTube, para organizar a busca e a coleta dos dados.

| | | |
|----|---------------------|-----|
| 1 | Atletico Goianiense | Sim |
| 2 | Atletico Paranaense | Sim |
| 3 | Atlético Mineiro | Sim |
| 4 | Bahia | Sim |
| 5 | Botafogo | Sim |
| 6 | Corinthians | Sim |
| 7 | Criciúma | Sim |
| 8 | Cruzeiro | Sim |
| 9 | Cuiabá | Não |
| 10 | Flamengo | Sim |
| 11 | Fluminense | Sim |
| 12 | Fortaleza | Sim |
| 13 | Grêmio | Sim |
| 14 | Internacional | Sim |
| 15 | Juventude | Sim |
| 16 | Palmeiras | Sim |
| 17 | Red Bull Bragantino | Sim |
| 18 | São Paulo | Sim |
| 19 | Vasco da Gama | Sim |
| 20 | Vitória | Sim |

Figure 1. Listagem dos times e se possuem canal oficial

Para garantir que o conjunto de dados fosse suficientemente robusto, a coleta foi limitada aos 100 vídeos mais visualizados de cada canal. De cada vídeo, foram coletados até 100 comentários, o que permitiu a obtenção de um número substancial de dados sem comprometer a escalabilidade do projeto. Para vídeos que tinham menos de 100 comentários disponíveis, a coleta foi limitada ao total de comentários presentes. Essa abordagem visava otimizar o processo, garantindo que a análise fosse feita com um número representativo de interações, ao mesmo tempo em que limitava a quantidade de dados para manter a viabilidade do estudo.

Cada comentário coletado continha informações cruciais para a análise, como o título do vídeo, o ID do vídeo, o autor do comentário, o conteúdo do comentário e o número de likes que cada comentário recebeu. Esses dados permitiram que fosse construída uma rede de interações entre os torcedores e os vídeos de seus respectivos times.

A escolha dessas variáveis foi fundamentada na necessidade de entender não apenas o volume de interações, mas também as relações entre os torcedores e o tipo de conteúdo que gera mais engajamento.

A coleta foi realizada utilizando a API oficial do YouTube, o que permitiu acessar os dados de maneira estruturada e automatizada. No entanto, vale destacar que, durante o processo de coleta, ocorreram algumas limitações. Em determinadas situações, falhas na própria API do YouTube impediram a coleta de vídeos e/ou comentários, o que resultou na ausência de alguns dados específicos. Essas falhas ocorreram principalmente em vídeos que estavam com problemas de acesso ou que foram removidos, o que é uma limitação comum ao se trabalhar com APIs de plataformas de grande escala como o YouTube. Essas falhas, embora menores, impactaram parcialmente o número total de comentários coletados, mas não comprometem a qualidade geral do conjunto de dados, que permanece representativo para os objetivos do estudo.

Após a coleta, os dados foram organizados em arquivos CSV separados por time, facilitando a gestão e a análise subsequente. Cada arquivo continha os dados dos vídeos e comentários específicos de cada time, permitindo que as informações fossem processadas de maneira eficiente.

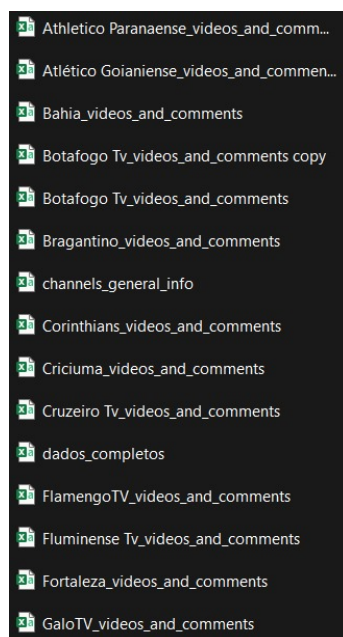


Figure 2. Organização dos dados coletados

Posteriormente, para a construção da rede de interações, todos os dados foram reunidos em um único arquivo chamado dadosprocessados.csv, no qual foi adicionada uma coluna extra denominada "time" para identificar a qual time cada vídeo estava associado. Essa etapa foi crucial para integrar as informações de maneira coesa e possibilitar a análise da rede de interações entre torcedores de diferentes times.

```

# Inicializar uma lista para armazenar os dados
dados = []

# Ler cada arquivo CSV e adicionar a coluna "time"
for time, arquivo in arquivos_csv.items():
    # Lê o arquivo CSV
    df = pd.read_csv(arquivo, names=["video_title", "video_id", "comment_author", "comment_text", "comment_likes"])

    # Adiciona o nome do time como uma nova coluna no DataFrame
    df["time"] = time

    # Adiciona o DataFrame à lista
    dados.append(df)

# Concatenar todos os DataFrames da lista em um único DataFrame
dados_completos = pd.concat(dados, ignore_index=True)

# Salvar o DataFrame completo em um novo arquivo CSV (opcional)
dados_completos.to_csv("dados_completos.csv", index=False)

```

Figure 3. Código usado para concatenar todos os dados em uma única tabela

Esse conjunto de dados, por sua organização e abrangência, constitui a base para o desenvolvimento das análises subsequentes, permitindo que padrões de comportamento e interações entre torcedores sejam explorados e visualizados de maneira eficaz.

3. Limitações

Essa seção aborda um pouco sobre alguns fatores que limitaram a coleta dos dados, mas que não impediram uma análise robusta da rede. Vale ressaltar também que o estudo coletou dados dos times que estavam na série A do Campeonato Brasileiro de futebol do ano de 2024.

Apesar dos esforços para coletar um conjunto de dados representativo sobre a interação dos torcedores nos canais oficiais dos times da Série A do Campeonato Brasileiro, algumas limitações impactaram a abrangência da análise.

A principal limitação está relacionada à API do YouTube, que apresentou restrições na coleta de dados. Embora a metodologia adotada tenha previsto a extração dos 100 vídeos mais visualizados de cada time, houve casos em que alguns vídeos não puderam ser acessados ou recuperados devido a falhas ou limitações impostas pela própria API. Isso pode ter afetado a representatividade dos vídeos mais populares para certos times, potencialmente reduzindo a diversidade do conteúdo coletado.

Além disso, o Cuiabá, um dos 20 clubes da Série A, não possui um canal oficial na plataforma do YouTube. Como o critério de seleção dos vídeos foi baseado exclusivamente nos canais oficiais dos times, o Cuiabá não pôde ser incluído na análise, o que representa uma limitação na comparação entre os times.

Essas restrições, embora não inviabilizem a análise da rede de interações entre torcedores, devem ser consideradas ao interpretar os resultados, pois podem influenciar a representatividade da amostra e a identificação de padrões de engajamento entre diferentes torcidas.

4. Desenvolvimento

Para modelar a rede de interações entre torcedores e conteúdos dos times, foram considerados dois tipos principais de nós: usuários e vídeos. Cada nó de usuário representa uma pessoa que fez um comentário em algum dos vídeos coletados, enquanto os nós de vídeo representam os conteúdos publicados nos canais oficiais dos times. As conexões entre esses nós, chamadas de arestas, representam a relação entre um usuário e um vídeo específico, ou seja, sempre que um usuário comenta em um vídeo, uma aresta é criada

entre eles. Dessa forma, a estrutura da rede reflete como os torcedores interagem com os conteúdos dos seus times e, potencialmente, com vídeos de diferentes clubes.

Os times aparecem na rede como um atributo associado aos nós dos vídeos. Cada vídeo possui um atributo indicando a qual time ele pertence, permitindo análises sobre como os usuários interagem com conteúdos de diferentes clubes e identificando sobreposições entre torcidas. Essa estrutura possibilitou a investigação do engajamento cruzado entre torcedores de diferentes times, fornecendo insights sobre rivalidade e envolvimento dos usuários com múltiplos clubes.

Para facilitar a visualização da rede, foi gerado um arquivo no formato .gexf, que permitiu a análise gráfica utilizando o software Gephi. Com essa ferramenta, foi possível explorar a estrutura da rede, identificar padrões de conexão e destacar grupos de usuários que interagem com múltiplos times.

A rede resultante da construção contou com um total de 37.901 nós e 44.687 arestas. Esses números refletem a quantidade de usuários e vídeos na rede, assim como as interações (comentários) entre eles, indicando uma rede de grande escala que representa uma significativa quantidade de interações no cenário digital dos torcedores de futebol.

Além da construção da rede, foram realizadas diversas análises quantitativas para melhor compreender sua estrutura. O grau médio da rede foi calculado para indicar o número médio de conexões por nó, refletindo o nível de interação entre usuários e vídeos. O coeficiente de clusterização foi analisado para verificar se existem padrões de agrupamento na rede, ou seja, se há uma tendência de formação de comunidades mais densas de interação. Também foi calculado o PageRank, uma métrica de centralidade que identifica os nós mais influentes na rede, possibilitando a identificação de usuários ou vídeos que exercem maior impacto nas interações. Além disso, a assortatividade foi medida para avaliar se há tendência de conexões entre nós com características semelhantes, como usuários comentando predominantemente em vídeos de um mesmo time ou diversificando sua participação entre diferentes clubes.

Essas análises forneceram uma visão detalhada da estrutura da rede, permitindo uma compreensão mais profunda sobre os padrões de engajamento e a interseção entre torcidas no ambiente digital.

Ademais, depois de todos os cálculos citados acima, foi feita uma análise específica para entender melhor sobre a rivalidade entre os times e acerca do engajamento cruzado. Para isso foi calculada e construída uma matriz de sobreposição, que indica quantos usuários comentaram em vídeos de mais de um time. Uma alta sobreposição entre dois times pode sugerir uma maior rivalidade ou maior engajamento cruzado entre as torcidas desses times.

5. Resultados

Essa seção mostra alguns resultados obtidos nos cálculos realizados na seção anterior, como grau médio dos nós, número de componentes, coeficiente de clusterização, assortatividade, Pagerank. Além de graficos para representar, por exemplo, a distribuição do coeficiente de clusterização, distribuição do tamanho dos componentes, distribuição do grau (Log-log), distribuição do grau dos nós e, por fim, a matriz de sobreposição, com um mapa de calor(heatmap).

Inicialmente, vamos mostrar algumas visões da rede completa formada, as imagens foram geradas pelo software Gephi.

As imagens 4 e 5 mostram a rede, destacando os times, que são atributos dos nós vídeos. A imagem 4 mostra uma visão mais ampliada da rede, enquanto a 5 uma vista mais aproximada da rede, mostrando mais detalhes da dinâmica de interação.

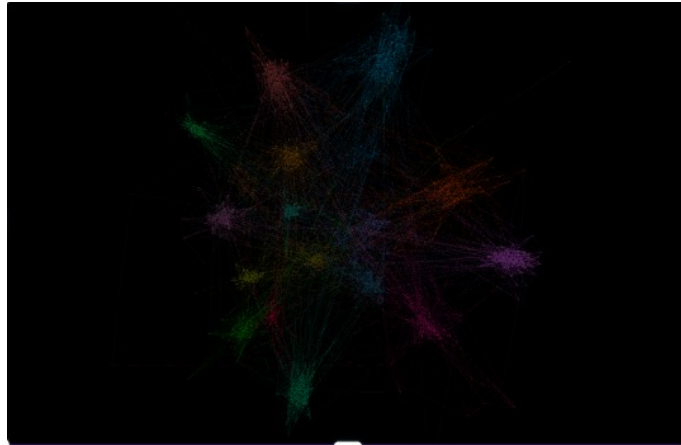


Figure 4. Visão ampliada - rede com destaque nos times



Figure 5. Visão aproximada - rede com destaque nos times

A seguir, as imagens 6 e 7, mostram as visões ampliadas e aproximadas, respectivamente, da rede, destacando apenas os usuários e vídeos, sem analisar os times.

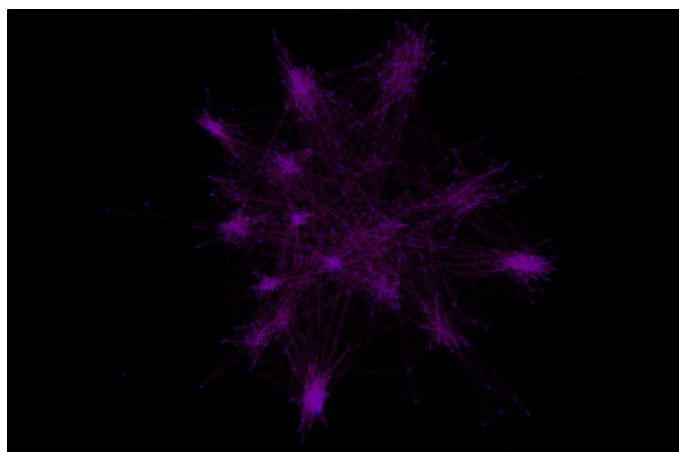


Figure 6. Visão ampliada - rede usuários e vídeos

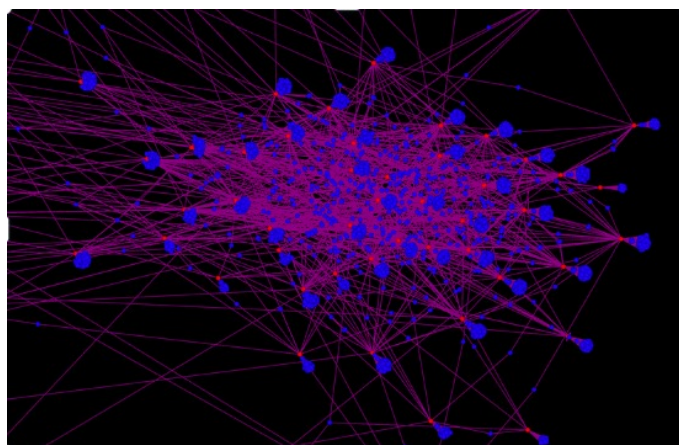


Figure 7. Visão aproximada- rede usuários e vídeos

A partir dos cálculos realizados na rede construída, foi possível obter diversas métricas que ajudam a entender a estrutura e a dinâmica de interação entre usuários e vídeos.

O grau médio da rede foi de 2.358090815545764 , o que indica que, em média, cada nó da rede (usuário ou vídeo) está conectado com pouco mais de 2 outros nós. Esse valor sugere uma rede relativamente dispersa, na qual a maioria dos usuários interage com poucos vídeos, mas ainda há alguns casos de usuários que comentam em múltiplos vídeos.

Em relação à distância média dos graus, não foi possível realizar o cálculo corretamente, devido ao fato de que na rede modelada, os usuários estão conectados apenas aos vídeos que eles comentaram, e não a outros usuários diretamente. Isso resulta em muitos componentes desconectados, ou seja, subconjuntos de rede que não possuem ligações entre si. Logo, muitos pares de nós não possuem um caminho que os conecte, impossibilitando o cálculo da distância entre eles. O problema ocorre porque o cálculo da distância média considera todas as conexões da rede e se existem pares de nós sem caminho entre eles, a distância entre esses pares é infinita, inviabilizando o cálculo.

O coeficiente de clusterização obtido na rede foi 0. Vale ressaltar que esse coeficiente calcula o quão conectados estão os vizinhos de cada nó. O resultado igual a 0 reforça a formação de uma rede bipartida. A rede é bipartida devido ao fato de não haver a formação de triângulos, pois usuários se conectam apenas aos vídeos que eles comentaram, não ocorrendo conexão entre usuários e outros usuários ou entre vídeos distintos, que fechariam a formação de triângulos. Para que o coeficiente de clusterização fosse significativo, seria necessário haver conexões diretas entre usuários que comentam nos mesmos vídeos, algo que não foi modelado nesta abordagem. A imagem 8 mostra o gráfico da distribuição do coeficiente de clusterização.

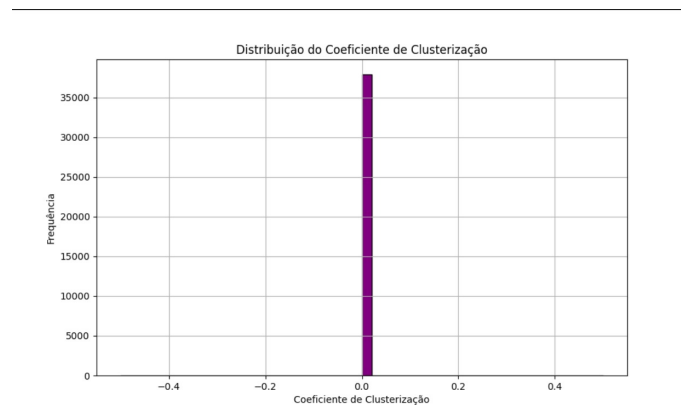


Figure 8. Distribuição Coeficiente Clusterização

A análise da assortatividade da rede revelou um coeficiente negativo, indicando que nós com graus altos tendem a se conectar com nós de grau menor. Esse comportamento é característico de redes heterogêneas, onde poucos vídeos altamente populares concentram um grande número de interações, enquanto a maioria dos vídeos recebe apenas alguns comentários.

A distribuição do grau dos nós seguiu um padrão semelhante ao observado em redes de mídia digital, com uma cauda longa, sugerindo que poucos vídeos recebem um volume muito alto de interações, enquanto a maioria recebe uma quantidade menor de comentários. Esse comportamento é consistente com o modelo de redes livres de escala, onde poucos elementos altamente conectados desempenham um papel central na estrutura da rede. A distribuição dos graus do nós está representada na figura 9.

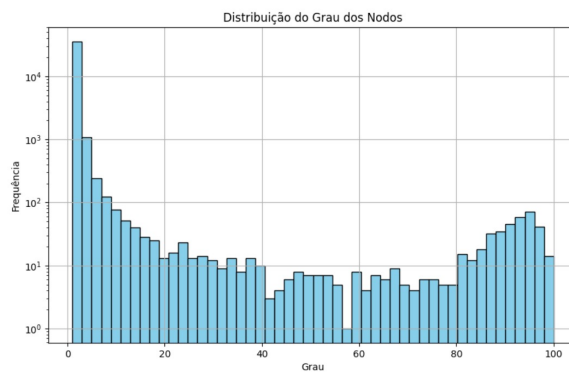


Figure 9. Distribuição do grau dos nós

Alguns outros cálculos foram realizados, e os resultados são mostrados a seguir.

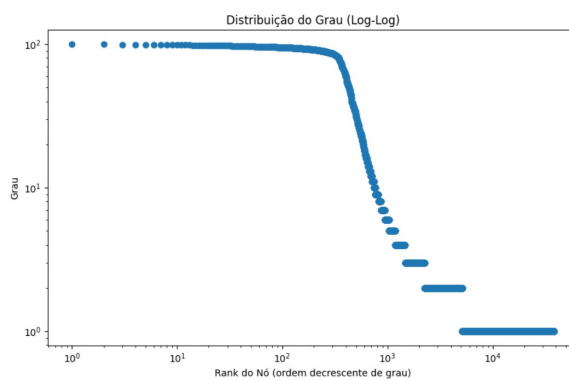


Figure 10. Distribuição do grau Log-Log

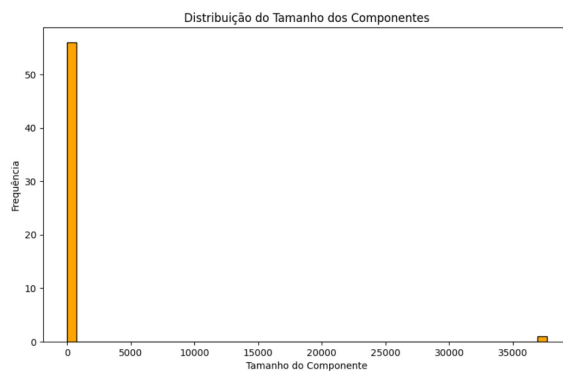


Figure 11. Distribuição tamanho dos componentes

Após esses cálculos, o trabalho seguiu para uma análise mais específica, para entender um pouco mais sobre como a rivalidade dos times se expressa nessa interação entre torcedores a partir de comentários nos vídeos do Youtube. Para isso, foi construída

uma matriz de sobreposição. A matriz de sobreposição fornece uma visão clara sobre a interação dos torcedores entre diferentes canais de times da Série A do Campeonato Brasileiro. Cada célula da matriz representa a quantidade de usuários que comentaram em vídeos de ambos os times correspondentes às linhas e colunas, permitindo identificar padrões de engajamento e rivalidade. Além disso, essa matriz foi complementada por um mapa de calor, que facilita a visualização das interações, onde células com cores mais intensas indicam uma maior sobreposição de usuários, enquanto cores mais suaves representam uma menor interação. A diagonal principal da matriz mostra o número de usuários que comentaram em múltiplos vídeos de um mesmo time, refletindo o nível de fidelidade e engajamento da própria torcida com o canal do clube. Já os valores fora da diagonal indicam interações cruzadas entre torcedores de diferentes times, podendo apontar rivalidades ou interesses compartilhados.

A análise da matriz de sobreposição e do mapa de calor revela algumas tendências marcantes. Clássicos tradicionais do futebol brasileiro, como Atlético-MG x Cruzeiro, Grêmio x Internacional e Botafogo x Vasco, apresentam um número significativo de sobreposições, reforçando a intensa rivalidade entre essas torcidas. Esse fenômeno sugere que torcedores de times rivais acompanham e comentam nos vídeos do adversário, seja para provocar, defender seu clube ou simplesmente acompanhar o debate futebolístico. Por outro lado, times com menor expressão nacional, como Vitória, Juventude e Criciúma, apresentam baixo engajamento, evidenciado por uma menor sobreposição de usuários tanto na diagonal principal (indicando menor fidelidade dos próprios torcedores) quanto fora dela (demonstrando pouca interação com outras torcidas). Esse comportamento pode estar relacionado a uma base de torcedores menor ou a um menor interesse do público geral em consumir conteúdo desses clubes no YouTube.

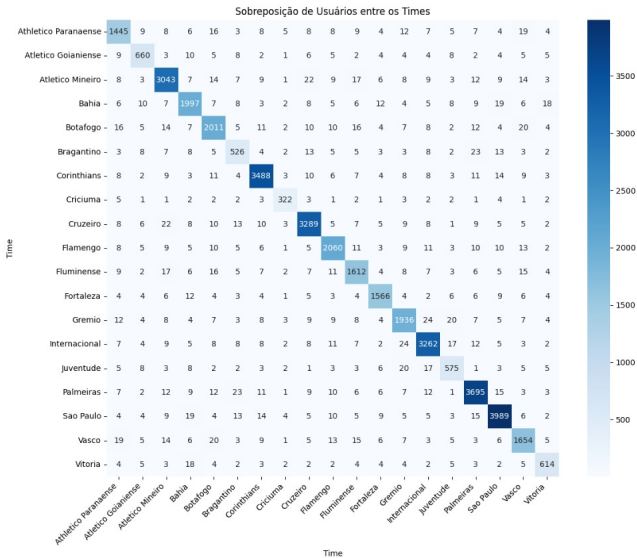


Figure 12. Matriz de sobreposição

6. Conclusão

O presente estudo analisou a interação entre torcedores nos comentários de vídeos do YouTube dos clubes da Série A do Campeonato Brasileiro, modelando uma rede com-

plexa para investigar padrões de engajamento e rivalidade. A construção da rede revelou conexões significativas entre usuários e vídeos, permitindo identificar comportamentos específicos das torcidas. A matriz de sobreposição e o mapa de calor evidenciaram rivalidades históricas, como Atlético-MG x Cruzeiro e Grêmio x Internacional, demonstrando que torcedores frequentemente interagem com conteúdos dos rivais. Além disso, foi possível observar que clubes com menor expressão nacional apresentam um engajamento digital reduzido, tanto em interações internas quanto externas.

Os cálculos realizados forneceram métricas importantes para compreender a estrutura da rede. O grau médio indicou que os usuários interagem, em média, com poucos vídeos, sugerindo um consumo mais direcionado de conteúdo. O coeficiente de clusterização zero demonstrou que a rede é predominantemente bipartida, sem formação de grupos coesos de usuários interagindo entre si diretamente. Já a impossibilidade de calcular corretamente a distância média reforça que a rede possui componentes desconectados, o que é esperado em um sistema onde usuários se vinculam apenas aos vídeos e não entre si.

Por fim, além das descobertas sobre o comportamento das torcidas no ambiente digital, este trabalho foi de grande proveito para aprofundar e aplicar conceitos de redes complexas, proporcionando aprendizado prático na análise de dados e no uso de ferramentas para visualização e interpretação de redes. Futuras pesquisas podem expandir essa abordagem incluindo outras plataformas e analisando como o engajamento das torcidas evolui ao longo do tempo.

7. References

Para a coleta de dados: Google Developers. (2024). "YouTube Data API v3." Disponível em: Youtube Data API v3

Artigos relacionados:

de Alvarenga Miranda, Fernanda. "Marketing digital e o futebol brasileiro: um estudo sobre a interação entre clubes e torcida nas mídias sociais." *Esporte e Sociedade* 22 (2021).

References