From Tweets to Teams: Analyzing Social Media Dynamics Among Football Fans

Akshay Reddy Narra
akshayreddy@vt.edu
Virginia Tech
Blacksburg, Virginia, USA

Samhitha Pentaparthy
samhithap@vt.edu
Virginia Tech
Blacksburg, Virginia, USA

Sivakumarreddy Sangu
sivakumarreddy@vt.edu
Virginia Tech
Blacksburg, Virginia, USA

ABSTRACT

This project aims to delve into the structural dynamics of fan communities within a dataset particularly focusing on the English Premier League fan base on Twitter. By constructing a network from user interactions such as retweets, mentions, each user is represented as a node with interactions serving as edges between nodes. This approach facilitates the application of community detection algorithms that identify cohesive partitions signifying distinct fan communities that frequently interact. Complementing the network analysis, sentiment analysis on tweets is conducted to assess the sentiments expressed towards the football teams in English Premier League. This analysis helps in understanding the emotional ties and attitudes prevailing within the communities, thereby contributing significantly to the comprehension of community dynamics. Moreover, the influence of geographic location on expressing sentiment towards the club is identified. The availability of location data within the dataset provides a unique opportunity to explore if users within the same geographic region are inclined to form cohesive communities or share similar sentiment patterns. This geographic analysis is crucial as it might reveal regional biases or preferences that could influence community dynamics. The integration of network analysis, sentiment analysis, and geographic information enables a comprehensive examination of the factors driving community formation and dynamics. By understanding these elements, the analysis seeks to uncover not just how fan communities are formed but also the underlying reasons behind their interactions and emotional investments. This holistic approach offers valuable insights into the structure and behavior of fan communities in digital social networks, particularly in the context of sports fandom on platforms like Twitter.

KEYWORDS

Sentiment Analysis, Network Analysis, Community Detection, Geographic Sentiment Distribution, Social Media Analytics, Football Engagement, English Premier League, NRCLex, NetworkX, Geopy, Twitter Data, Emotion Analysis, Graph Theory, Fan Communities, Data Visualization, Social Network Dynamics, Influential Nodes in Social Networks, Sports Analytics

1 INTRODUCTION

The digital age has transformed how fans interact and express their emotions particularly within the realms of sports fandom. The English Premier League (EPL) being one of the most followed sports leagues globally has a vast and diverse fan base on social media platforms like Twitter. Understanding the structure of these fan communities and the dynamics that govern their interactions is not only intriguing but crucial for multiple stakeholders, including marketers, sociologists, and the sports teams themselves. The insights garnered can inform targeted marketing strategies, enhance fan engagement, and even influence the management of the teams.

This project focuses on exploring the community structure within the dataset of Twitter users discussing the English Premier League. By examining the interactions among fans—such as retweets, and mentions-we can identify distinct fan communities or clusters, which are crucial for understanding how information and sentiments spread within the network. Moreover, the emotional undertones in these interactions, discerned through sentiment analysis, reveal the fans' perceptions and attitudes towards teams, matches, and events. Such emotional connections are key to comprehending how cohesive or divided fan communities might be.

Additionally, the influence of geographic location on exhibiting the sentiment is another focal point of this study. It is hypothesized that fans in closer geographic proximity may exhibit more cohesive behavior and similar sentiment patterns, influenced by regional cultural factors and accessibility to games and fan-related events.

The plan for approaching this problem involves a methodical examination using network analysis to construct and analyze the community network, sentiment analysis to evaluate the emotional content of communications, and geographic analysis to determine the role of location in exhibiting sentiment. This comprehensive approach will allow us to unravel the complex web of interactions and influences that shape the fan communities of the English Premier League on Twitter, providing valuable insights into the digital sociology of sports fandom.

2 RELATED WORK

Our project intersects with several dynamic fields, including social media analytics, sentiment analysis, and network science, with a specific focus on the football community. Previous studies within these domains provide a solid foundation for understanding fan interactions and sentiments on platforms such as Twitter.

A significant contribution to the field is an analysis of tweets regarding the Turkish football leagues in 2013 and 2018, which elucidates the shifting patterns of fan engagement and interaction over different seasons [1]. This study offers a longitudinal view of how sentiments and discussions around football evolve, reflecting broader trends in social media usage and fan behavior.

In terms of methodological refinement, substantial work has been done on tailoring sentiment analysis tools for sports-related content. One study in particular developed specialized techniques for pinpointing emotions within football-specific tweets [2]. This research is pivotal in breaking down the complex emotional responses fans exhibit towards game results, player performance, and other football-related events, providing a granular analysis of sentiment within sports contexts.

Network dynamics within specific football clubs have also been examined, with one notable exploratory case study focusing on Manchester United [3]. This research delved into how fans are interconnected through social media, highlighting influential nodes and the flow of information within the community, thereby offering insights into the structural aspects of online fan interactions.

Expanding on these themes, several conference papers have utilized libraries such as NetworkX for constructing and analyzing network graphs and NRCLex for emotional analysis, to explore the nuances of fan interactions and sentiments across different geographies and times. These studies underscore the capabilities of advanced computational tools in dissecting complex datasets to reveal underlying patterns in fan behavior and community formation.

Moreover, the integration of geographic sentiment analysis has shown that fans within proximal regions tend to form more cohesive community clusters, reflecting localized fan bases that resonate with global sports phenomena like football. This geographical aspect adds a significant layer of depth to understanding how fan communities develop and evolve in a digitally connected world.

This background sets a comprehensive stage for the current project, aiming to leverage advanced analytics to dissect the interactions, sentiments, and community structures of football fans on Twitter. By combining longitudinal and geographical analyses with cutting-edge network science methods, the project seeks to uncover the layers of interaction that define sports fandom in the digital age.

3 APPROACH

Our study begins by constructing a social network derived from user tweets on Twitter about football especially the English Premier League (EPL). The network consists of node represents an individual user and edge denoting the connection between the user based on mentions in their tweets. We employed graph-theoretical metrics including Degree Centrality, Closeness Centrality, Betweenness Centrality, and Clustering Coefficient to delve into the network's structure and characteristics, enhancing our understanding of its complexity and connectivity.

To gain a deeper understanding of the user dynamics and shared interests within the network, we employed the Louvain method for community detection. This algorithm allowed us to identify tightly knit groups of users based on their interactions and connections within the network.

We focused our analysis on determining the English Premier League (EPL) teams that gathered the most attention and engagement within each community. By examining the tweets with users mentions related to various EPL teams, we sought to identify the teams that predominantly captured the interest of users within each group.

To probe the emotional landscape within these football communities, we conduct sentiment analysis using the NRC Lexicon to categorize tweets by emotions such as fear, anger, trust, surprise, and joy, among others. This enables us to pinpoint the dominant emotional responses tied to specific football teams, offering insights into the prevailing sentiments within fan bases and how these might drive community interactions and alignments.

Furthermore, our project considers the role of geographic location in shaping community structures and dynamics. Leveraging geotagged data from user profiles and tweets, we assess if proximity influences the formation of uniform sentiment patterns. In this component of our project, we aim to analyze and visualize sentiment patterns across different continents and football clubs using advanced techniques.

Our approach involves employing visual mapping to provide a comprehensive understanding of the positive, negative, and neutral sentiments expressed by users in various geographical regions. In addition to that, we conducted a detailed analysis of the sentiment trends specific to each football club within each continent. By examining the sentiment expressed towards individual clubs, we can uncover insights into how opinions and perceptions vary across different geographical areas.

Ultimately, our approach integrates findings from network structure, sentiment analysis, and geographic analysis to provide a holistic view of how fan communities around football teams emerge and evolve based on shared sentiments and geographic proximity. This comprehensive analysis of network, sentiments, and geographic contexts, aims to reveal the underlying dynamics that influence community formations and interactions within the digital sphere of football fandom.

4 EXPERIMENT

Our study utilizes a rich dataset sourced from Twitter, containing tweets that mention various football teams, complete with metadata like content, user ID, timestamps, and hashtags. This dataset capturing tweets during the English Premier League in 2020, serves as an ideal basis for observing sentiment fluctuations and network dynamics amid real-time events.

We constructed an undirected graph using NetworkX, with nodes representing a Twitter user and edge denoting the connection between the users based on similar mentions in their tweets.

The network examined in our study includes a substantial collection of 27,578 nodes and 5,891 edges, suggesting a relatively low density of connections with an average degree centrality of 0.43. This observation indicates that while there are numerous users within the network, each user typically connects with a limited number of others. This sparse connectivity was further analyzed by visualizing the Giant Connected Component (GCC), where nodes were color-coded based on football club affiliations of their tweets as depicted in (Fig 1). This visualization helps to underscore the alignment of fan communities and the distribution of club affiliations within the network, providing key insights into the structure of interactions among fans.

The analysis of the network revealed an average degree centrality of 0.43, highlighting a low level of direct connectivity among users, which implies sparse interaction within the network. The average clustering coefficient stood at a mere 0.0042, indicating that tightly-knit groups or cliques are uncommon in this extensive network. Key nodes identified by high degree centrality, including prominent football clubs like Arsenal and Liverpool, along with influential Twitter accounts such as FabrizioRomano, play pivotal roles in

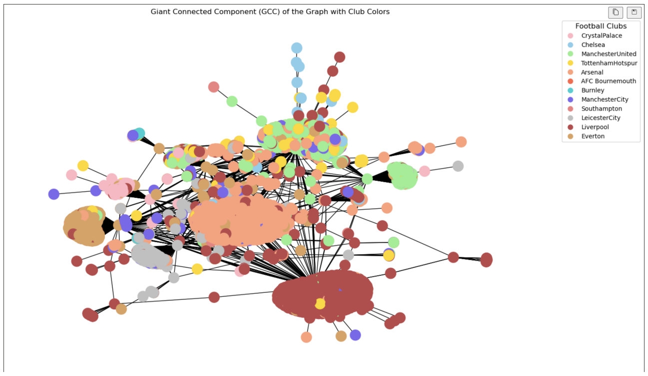

Figure 1: Gaint Connected Component of Network

disseminating content that attracts significant interactions. This central positioning in the network facilitates rapid information flow, making these nodes critical to the spread of information and sentiment across the network.

Closeness centrality measures underscored users like ’Mark’ and ’Arsenal’ for their ability to efficiently propagate information due to their strategic locations within the network, which are pivotal for quick information dissemination. Similarly, betweenness centrality identified nodes such as ’Arsenal’ and ’Liverpool’ as essential connectors bridging different community segments. Their roles are crucial in linking disparate groups and shaping the communication dynamics within the fan community. The overall network structure, characterized by its broad dispersion and low clustering, coupled with the influence of key nodes, significantly affects how information and sentiments are transmitted across various fan segments, thereby influencing the dynamics of online discussions related to football events.

Our PageRank analysis played a crucial role in pinpointing the most influential nodes within our network, essential for the spread of information and driving discussions among football fans on social media. This investigation targeted key opinion leaders whose profound impact shapes fan interactions and influences the sentiment landscape significantly. The analysis highlighted entities like ’Arsenal’, ’LFC’ (Liverpool Football Club), and noted football journalist ’FabrizioRomano’, who emerged as top influencers with PageRank scores of 0.0772, 0.0722, and 0.0540, respectively. These figures emphasize their central role as primary news sources and major engagement hubs within the network (Fig 2).

Additionally, other notable nodes such as ’Everton’, ’LCFC’ (Leicester City Football Club), and media outlets like ’talkSPORT’, along with fan-centric accounts like ’utdreport’, ranked highly, showing extensive engagement across various club-focused communities. Their presence in the top echelons of the PageRank list not only reaffirms their status as crucial information and opinion disseminators but also underscores the network’s reliance on these nodes for communication flow. Understanding this dynamic is essential for grasping how information and sentiments spread across the network, influencing the structure and interactions within the online football fan community. These insights from the PageRank analysis provide a deeper understanding of the influential pathways and content dissemination within the network.

```
Top 10 Nodes by PageRank:

User: Arsenal, Rank: 0.07722237526252372
User: LFC, Rank: 0.072240050730415
User: FabrizioRomano, Rank: 0.054019714880934634
User: Everton, Rank: 0.025712502244691147
User: LCFC, Rank: 0.012122074780688885
User: talkSPORT, Rank: 0.011386565229110612
User: utdreport, Rank: 0.010230048721195708
User: CPFC, Rank: 0.005962562507465992
User: afcstuff, Rank: 0.001794928812887828
User: PepTeam, Rank: 0.0010784557220398796
```

Figure 2: Top 10 Nodes by PageRank

By applying the Louvain method, we uncovered communities that exhibit strong internal connections, suggesting the presence of shared interests or affiliations among users. Once these communities were identified, we further analyzed their characteristics and interactions to gain insights into the specific interests that unite each group.

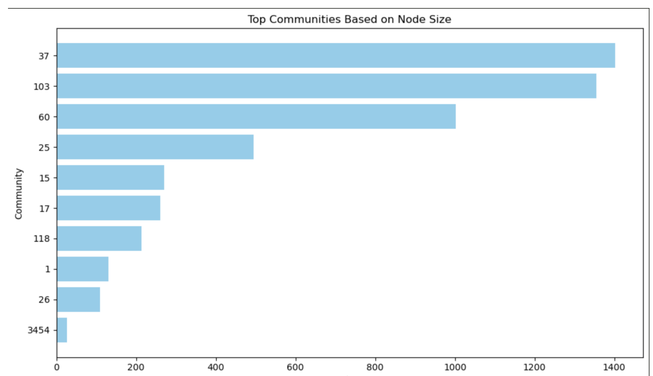

Figure 3: Top Communities by Node Size

As shown in (Fig 3), we visualized the size distribution of the communities, highlighting the most prominent groups within the network. Notably, the predominant community consisted of 1,402 members, indicating a significant concentration of users with shared interests or affiliations.

To gain a deeper understanding of the composition of these communities, we focused on the top 10 communities and examined the dominant football clubs within each group. By creating pie charts for each community, as illustrated in (Fig 4), we were able to visually represent the proportion of users associated with different football clubs. This approach allowed us to identify the most prevalent clubs within each community and understand the tribal nature of the groups.

We also conducted a detailed analysis of the most dominant clubs in the top 10 communities using a bar chart, which displayed the dominant club in each community based on member count, as illustrated in (Fig 5). The visualization utilized a horizontal bar chart plotted in reverse order for clearer community rankings, color-coded to represent different clubs with a corresponding legend.

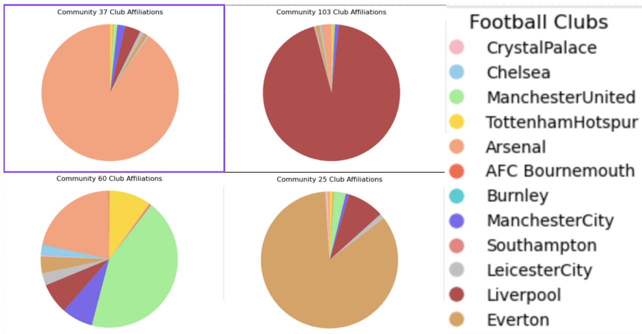

Figure 4: Top four Communities and football clubs in particular community

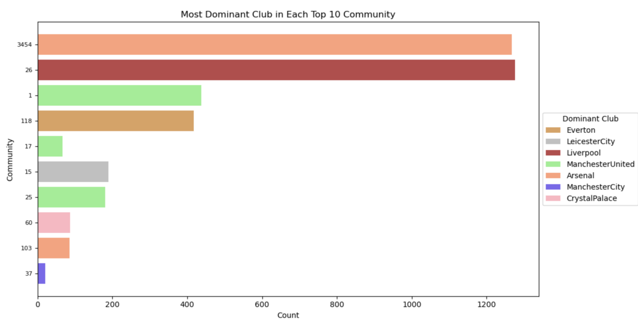

Figure 5: Most Dominant Clubs

This in-depth analysis not only enhanced our understanding of football fan community structures on Twitter but also demonstrated the utility of network analysis in social media studies. The visual representations, particularly the color-coded GCC and bar charts of dominant clubs, provided intuitive insights into network characteristics and the significant influence of specific clubs within communities.

Further in our study, we delve into the nuanced dynamics of football fan communities on Twitter by analyzing sentiment and emotional responses associated with different teams, Our methods include preprocessing tweets to eliminate noise such as URLs, usernames, and non-alphanumeric characters, followed by sentiment analysis using the NRCLex tool.

Our process begins with cleaning and preprocessing the tweets to ensure the data is ready for analysis. This involves tokenizing the text, removing stopwords, and constructing a dataset of cleaned tweets for each football team.

Utilizing the NRCLex tool, we assessed the emotional content of these tweets, focusing on emotions such as fear, trust, surprise, joy, and anticipation, among others. Our aim was to determine the predominant emotions expressed by football fans and analyze how these sentiments varied across different teams.

The results were quantified and presented in a series of bar charts for each football team, illustrating the distribution of emotional responses. These charts depict the frequencies of various emotions, highlighting patterns such as high levels of positive sentiment or notable instances of anger and trust.

For example, teams like AFC Bournemouth and Arsenal showed significant amounts of positive sentiment among their fanbases, contrasting with other emotions like fear and sadness, which appeared less frequently (Fig 4). The detailed analysis helped in understanding the emotional landscape specific to each team's supporters.

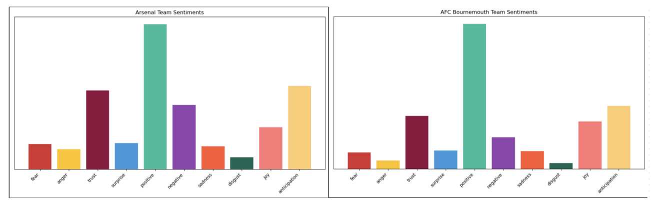

Figure 6: Tweets on various emotions: fear, anger, trust, surprise, positive, negative, sadness, disgust, joy, and anticipation of Arsenal&AFC Teams Sentiments

The processes of tweet preprocessing and emotion analysis with NRCLex culminate in visualizations using bar charts for each team, reflecting the unique emotional landscape of its supporters. These results offer insightful perspectives on how different teams evoke varied emotional responses from their fans, potentially correlating these emotions with community sentiment trends, match outcomes, or significant footballing events. This comprehensive analysis not only highlights the predominant emotions but also underscores the intensity and complexity of fan engagement in online football communities.

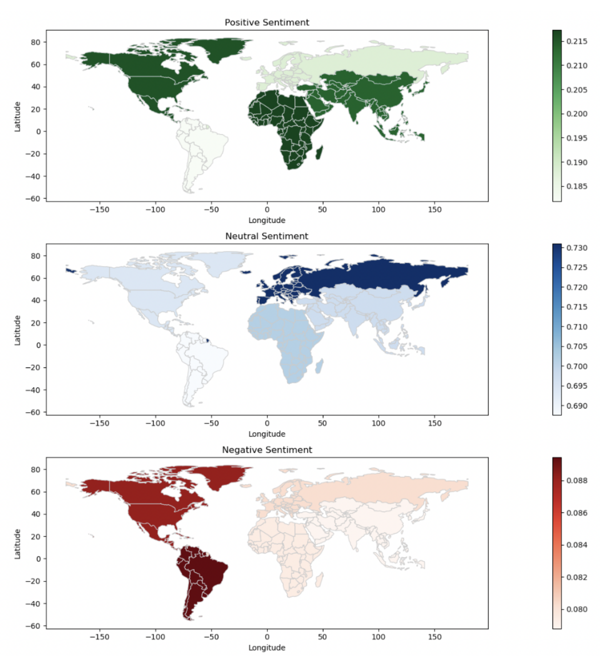

Figure 7: Sentiment distribution across continents

Utilizing geotagged content, we conducted sentiment analysis segmented by continent, preprocessing tweets to remove extraneous elements and employing VADER for sentiment scoring. The results, visually represented through choropleth maps(Fig 7), illustrate sentiment distribution across continents, uncovering notable differences that provide insights into the emotional tone surrounding various football teams.

Our temporal analysis of sentiment provided a detailed view of how fan sentiments evolve in relation to specific events or matches, offering a dynamic perspective on global fan reactions. We presented these changes using a variety of visual formats, including tabular displays, choropleth maps for regional sentiment variations, and line chart and heatmap charts to illustrate the intensity and distribution of sentiments. These tools allowed us to visually track sentiment fluctuations over time, highlighting how major events influence fan emotions across different regions (Fig 8).

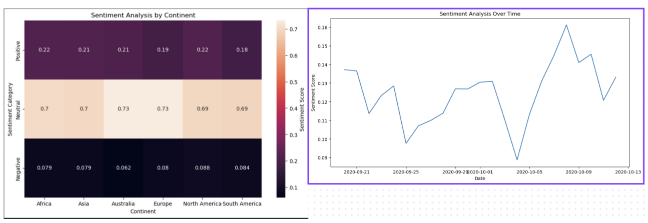

Figure 8: Heatmap and Line chart to detail sentiment intensity

To gain insights into the sentiment associated with different football clubs across continents, we performed a sentiment analysis on the tweets related to each club. The analysis was conducted separately for each continent to identify any geographical variations in sentiment patterns. The sentiment scores for each football club were calculated using the VADER sentiment analysis tool. VADER assigns sentiment scores to text data, ranging from -1 (most negative) to +1 (most positive), based on a lexicon of sentiment-related words and their associated sentiment intensities.

By analyzing the sentiment scores and the color distribution of the bars, we can identify the football clubs that generate more positive or negative sentiment within each continent.During our analysis of sentiment scores by football clubs across continents, we discovered some notable instances of negative sentiment associated with specific clubs in certain regions (Fig 9).

Our analysis revealed that the sentiment expressed towards AFC Bournemouth, a football club based in England, was particularly negative among fans in Asia. The sentiment scores for AFC Bournemouth in the Asian continent were consistently low, indicating a predominance of negative sentiment.

Another interesting finding from our analysis is the negative sentiment associated with Crystal Palace, an English football club, among fans in Australia. The sentiment scores for Crystal Palace in the Australian continent were notably low, indicating a prevailing negative sentiment towards the club. This observation raises curiosity about the factors influencing the negative sentiment towards Crystal Palace in Australia. It could be related to various aspects, such as the club's performance in matches relevant to Australian

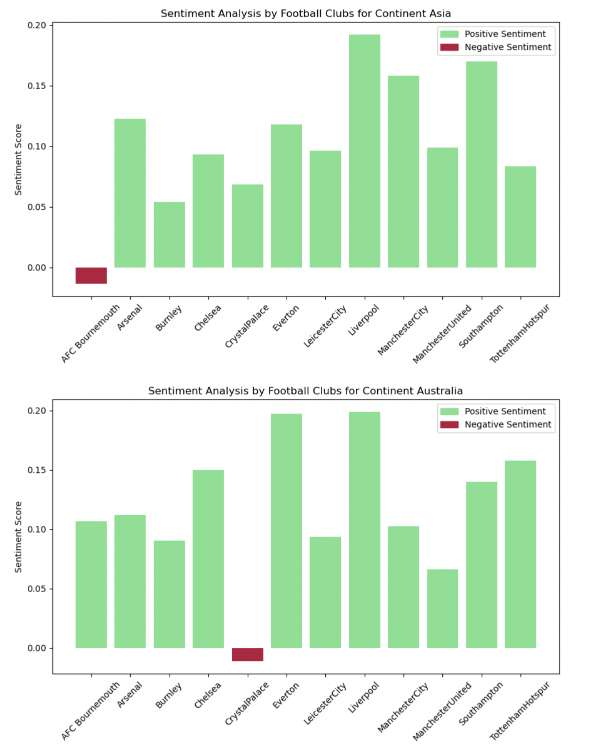

Figure 9: Analysing the Sentiment of AFC Bournemouth & Crystal Palace in Asia and Austrailia

fans, the club's marketing and engagement efforts in the Australian market, or any specific incidents or controversies that may have affected the club's reputation in the region.

These specific findings highlight the importance of conducting a granular analysis of sentiment scores by football clubs across different geographical regions and provides valuable insights into fan sentiment and perception. Our findings confirm that geographic location plays a significant role in shaping community sentiments within the football discourse on Twitter, offering valuable insights for entities in the sports industry engaged in marketing, policy-making, or community management.

In conjunction with the sentiment analysis, we conducted topic modeling for each continent and football club to discern prevalent themes and topics among the fans. This analysis involved preprocessing tweets, constructing a dictionary and corpus from these tweets, and applying the LDA model to uncover the dominant topics within different groups. For instance, topics from the continent of Africa included mentions of prominent football clubs and fan lingo, reflecting specific team allegiances and common discussion points among fans. Similarly, topic analysis for clubs like Arsenal and Chelsea revealed focused discussions around key players, transfer rumors, and match reactions. These insights from topic modeling enrich our understanding of the content and context of discussions.

5 CONCLUSION

Our comprehensive study of Twitter data centered on football clubs during English Premier League (EPL) has yielded significant insights into the emotional and network dynamics that characterize social media interactions related to football. By leveraging libraries such as NetworkX for network analysis, NRCLex for sentiment analysis and GeoPy for geographical analysis, we have successfully mapped the structure of fan communities and decoded the emotional undertones prevalent among fans of various football clubs.

The use of NetworkX to construct undirected graphs illuminated the community structures within the football fan base on Twitter. our PageRank analysis proved instrumental in pinpointing these influential nodes. We identified major football clubs like Arsenal and Liverpool, alongside prominent figures such as FabrizioRomano, as central pillars within the network. These nodes emerged as vital in facilitating information dissemination and engaging the fanbase, thereby influencing overall community dynamics. This analysis was complemented by our examination of community structures through the Louvain method, which revealed how fans cluster around specific interests or affiliations, forming tightly knit communities within the broader network.

The findings from the centrality and community detection analyses were crucial in illustrating the layers of interaction within the football fan community. We observed how certain clubs and personalities act as bridges, linking various segments of the network and enhancing the connectivity between disparate fan groups. This bridging effect fosters a more interconnected community, enabling quicker and more widespread distribution of news and fan reactions.

Furthermore, our community detection efforts highlighted significant alignments and tribal affiliations within the fan base, showcasing the deep-seated loyalties that define the online interactions of football supporters. By mapping these affiliations and understanding their impact on information flow within the network, we gained valuable insights into the structural and emotional underpinnings of fan interactions on social media.

Utilizing NRCLex, we discovered varied emotional responses across different football clubs. Teams like Arsenal and Manchester City elicited predominantly positive sentiments, indicating strong emotional engagement and satisfaction among their supporters. Conversely, clubs like Southampton displayed a broader emotional spectrum, suggesting a more diverse fanbase reaction to club performances and footballing events.

Our analysis extended to examining the impact of geographic location on fan sentiments. By analyzing geotagged tweets, we observed distinct sentiment patterns across different continents, revealing how cultural and regional factors might influence fan reactions and interactions on social media.

The study also shed light on how sentiments fluctuate over time, correlating these changes with match events or significant football news. This temporal aspect of sentiment analysis can be crucial for understanding how fan emotions evolve in response to the team's on-field performances and off-field activities.

5.1 Suggestions for Future Research

The research could extend these observations over longer periods to capture seasonal variations in fan sentiments and community structures. Such studies could help understand how fan engagement evolves across different phases of a football season or during specific events like transfer windows and major tournaments.

While our study focused on Twitter, extending this analysis to other social media platforms like Facebook, Instagram, and Reddit could provide a more holistic view of fan engagement and sentiment across the social media landscape.

Future studies could incorporate analysis of visual content (images and videos) shared in tweets to gauge emotional reactions and fan engagement. This approach could uncover additional layers of sentiment linked to visual stimuli.

Further research could also explore the impact of real-world events, such as team signings or leadership changes, on community dynamics and sentiments, providing deeper insights into what drives changes in fan behavior.

By advancing these areas of research, we can enhance our understanding of digital fan communities, optimize engagement strategies, and potentially predict trends in fan behavior, contributing valuable knowledge to the fields of sports management and social media analysis.

6 REFERENCES

[**1**] Ozturkcan, S., Kasap, N., Tanaltay, A. and Ozdinc, M., 2019. Analysis of tweets about football: 2013 and 2018 leagues in Turkey. Behaviour & Information Technology, 38(9), pp.887-899.

[**2**] Aloufi, S. and El Saddik, A., 2018. Sentiment identification in football-specific tweets. IEEE Access, 6, pp.78609-78621.

[**3**] Guzmán, E.M., Zhang, Z. and Ahmed, W., 2021. Towards understanding a football club's social media network: an exploratory case study of Manchester United. Information Discovery and Delivery, 49(1), pp.71-83.

[**4**] https://github.com/joymnyaga/Twitter-Sentiment-Analysis-EPL-Teams/blob/master/Sentiment%20Analysis/Arsenal_BarPlot.png

[**5**] https://www.geeksforgeeks.org/emotion-classification-using-nrc-lexicon-in-python/