Making Sense of Post-match Fan Behaviors in the Online Football Communities

Yucheng Wang
City University of Hong Kong
Hong Kong, China
Columbia University
New York, USA
yuchwang7-c@my.cityu.edu.hk

Zhicong Lu
City University of Hong Kong
Hong Kong, China
zhicong.lu@cityu.edu.hk

ABSTRACT

Professional sports have large fan bases that congregate in online sports fan communities. The sports community is suitable to be a sandbox for studying offline context's effects on online community behavior. By now, prior works did not present a detailed study on the offline-online connection by examining detailed community discussion content. To fill this gap, this work presents a comprehensive study of online communities' comments about football (soccer) matches, grounded in the data from Premier League teams' *Reddit* online communities during the 2020-2021 season. We propose a metric "gap score" to quantify offline events' effects by measuring the gap between fans' prematch expectations and actual match results. Using this metric, we investigated how team performance impacted comments' sentiment, discussion topics, and the pattern of comments' votes. The findings highlight the close connection that exists between offline events and online discussions and reveals both theoretical and practical implications for online communities.

CCS CONCEPTS

• **Applied computing → Law, social and behavioral sciences**; • **Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

Online Communities, Fan Behaviors, Professional Sports, Topic Modeling, England Premier League, Football, Soccer

ACM Reference Format:
Yucheng Wang and Zhicong Lu. 2023. Making Sense of Post-match Fan Behaviors in the Online Football Communities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3581310

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581310

1 INTRODUCTION

A common usage of online communities is for professional sports, particularly football. Fans of such sports participate in online discussion forums and exhibit the intrinsic characteristics of fandom culture. Nowadays, people enjoy sharing their feelings to seek emotional common ground with others in a live discussion environment in scenarios such as watching TV series [60]. Similar in sports match discussions, fans join the online discussion to feel connected to a larger group, no matter what the end match result is. Previous work has presented research in this area by examining how social media dynamic responds to the co-presence scenario in FIFA World Cup. They analyzed the live-tweeting and found user activity, discussion contents, language use, and other reactions are changed with the events [41, 72]. Nevertheless, there has been growing recognition of the disordered community discussion atmosphere that exists in the sports community. Some argue that the same dilemmas faced by online political discussion now exist within sports communities [4]. Initially considered a beneficial instrument to facilitate online engagement and like-minded people connection, social media has turned into an "impoverished land" where polarization and extremism are indulgent. Offensive comments without proper moderation may have negative impacts on players, other community users, or even community moderators [16]. One prior study that characterized the online sports community conducted on the National Basketball Association (NBA) Reddit community investigated aspects including user activity, user loyalty, and so on [82]. The authors identified one of the limitations of their research as not considering features of comments themselves, like sentiment or passion, which is part of the direct motivation of the present research.

In recent years, considerable literature has examined the theme of the online community as its representativeness in studying human behavior patterns on the Internet. Online communities provide a venue for large-scale discussions that connect people using the topics and issues they are concerned about. Users with varying interests post comments that reflect their thoughts and feelings. Sundaram, etc. [66] analyzed the significance of understanding online community dynamics in online social networks and suggested several key applications of community dynamics comprehension, including implications in community moderation and behavioral prediction. These implications may be expanded to inspirational community design, which solves the problems of the current online community. Current problems center around the growing trend for people to become immersed in online forums, thus leading to exposure to the virtual world with a mixed bag of comments. As Herring and colleagues noted, harassment often arises in spaces known for their freedom, lack of censure, and experimental nature [31]. The football community, with the following factors involved: the sense of identification of fans, the strong connection with the offline context, and potential conflict between different groups, creates a complex discussion environment suitable for conducting the study mentioned above on understanding online community dynamics.

This study sets out to understand the dynamics of online fan communities through the lens of one of the most popular professional sports, football (soccer). The sports communities usually have a strong connection with the offline context. Previous research has shown that "the online community does not only exist in the virtual world." It closely connects with offline events [82]. Indeed, in the early stages of online communities, scholars proposed that online interactions could not be understood without considering their offline context [76]. Recent studies on the relationship between online communities and offline events have used sports as a sandbox several times [14, 80, 82]. Besides its representativeness of the close connection with its corresponding offline context, the sports community also has a strong fandom culture. Previous research has investigated fan behaviors in politics or the entertainment industry [36, 52, 77]. Sports fans share similar traits when supporting their favorite teams and players in the same way that people support partisanship in politics and artists in entertainment. These fans look for a common ground to discuss the teams they support, and the use of online communities is currently the most mainstream way to do so.

Motivated by the above, this work investigates online fan behavior via an in-depth analysis of comments' emotions, content, and feedback. Emotional comments can turn into trolls, toxic comments, or even anti-social behavior under certain circumstances. Previous studies have found that emotion is one of the causes of online trolling, and the emotion of comments is a direct reflection of the emotions of the users [12]. Other research contended that emotional comments would receive more attention and may, in turn, shape readers' minds in different ways [42]. Especially in sports forums, these emotional behaviors manifest as excessive praise after a team wins and abuse after a team loses. Therefore, to explore the emotional fluctuations posed by the match results, we were curious to understand to what extent football fans were emotionally affected by football matches and what they talked about in the course of their online discussions. It is also essential to understand whether the discussion content varies with match results or other elements. Meanwhile, the feedback, the user vote of each comment is also essential to help create a sense of the dominant views and community atmosphere. Related research on community feedback argues that there is a vicious circle when users are downvoted and pass their negative emotions on [13]. It is also one of the potential triggers for anti-social behavior. Therefore, we formulate our research questions as follows:

- **RQ1**: To what extent are football fans emotionally affected by match results?
- **RQ2**: What are fans discussing during and after a match, and how is discussion content affected as a result?
- **RQ3**: What are the characteristics and patterns of comment votes?

To answer these questions, we first need a metric to better quantify the effects of offline events. This work proposes the "gap score" to measure the gap between fans' prematch prospects and feelings after the match and uses this metric as an indirect way to measure the emotional effect of real-world events. To conduct the analysis, we compiled a dataset with over 177k posts and 3.7 million comments, as well as corresponding offline match statistics and supporting metadata from FiveThirtyEight[1], a popular forecasting website. In *r/soccer* on Reddit, around 3.5 million users subscribe to this subreddit, and the top teams in the England Premier League also possess over 100k subscribers. We conducted a quantitative analysis of the dataset to answer related questions using techniques such as hierarchical regression analysis. To supplement the analysis, we leveraged Natural Language Processing (NLP) techniques for sentiment analysis (RQ1) and topic modeling (RQ2). The findings revealed that fans' sentiments were largely correlated with team performance and that there were four main categories of discussion topics: game process, season performance, squad discussion, and team member turnover. We also found that community users preferred to vote for comments with extreme personal emotions rather than neutral comments. We also elaborate on possible directions in community design, along with moderation mechanisms.

Based on these findings, the identification of the facets of online sports communities can assist Human-Computer Interaction (HCI) research by helping others understand the dynamics of the online sports community under the impact of offline context and fandom culture through the lens of emotion, content, and community feedback. Making sense of the dynamics of the online community may have implications for other current HCI topics, including trolling, anti-social behavior, and fandoms.

2 RELATED WORK

Within this research, we drew inspiration from prior work on online-offline connections in online communities, sports fan behavior, and dynamics in online communities.

2.1 Online-offline Connections in Online Communities

The web-based social media system is a critical component of our daily lives and enable us to transmit information, engage in discussions, and form communities on the Internet [26]. Offline context can be an essential factor influencing user behavior online and users' online experience can also shape their offline behaviors. It is important to understand online-offline relationships in the context of online communities. Early studies have argued that offline context is indispensable when understanding online interactions [76]. Several studies evaluated the external impact of offline events on social media groups [29, 49, 64, 70]. McCully [49] contended that there was a counter-intuitive impact of offline interaction on online participation that offline relationships undermines the sustainability of the community. Guan et al. [29] analyzed the user behavior triggered by hot social events on the micro-blogging website Sina Weibo. Troudi et al. [70] proposed a multidimensional way of analyzing real-world events from social media sources from the dimensions of spatio-temporal, velocity, popularity, and sentiment. There

[1] https://fivethirtyeight.com/

has also been other research focusing on the impact of specific events, such as natural disasters or social movements, on online communities [14, 25, 68, 73]. While these real-world events give rise to the behavior of users in the virtual world, the discussion, and consequent online events, have an impact on the real world [47].

Several studies have examined how online interactions shape users' offline behaviors. Kavanaugh and etc. [37] found that Internet use strengthened users' social contact and community engagement. Erete [23] found that community-based online conversations influenced how they protect themselves against the crime. Some offline events like parades are also closely related to online communities, as social media is usually deeply embedded in our real-life [51, 61]. Unlike previous work, this research contributes a more comprehensive understanding of the offline context's impact on online fan behavior through the lens of a sports community. Detailed features of community behavior like discussion emotions, content, and community feedback were studied.

2.2 Sports Fan Behavior

Fans are critical components of the constitution of professional sports culture. The existence of fans might lead to real-world incidents if an unhealthy fan culture is nurtured. Football fandom is a particularly unique culture, given its long history and enormous influence on the world compared with other sports. The fandom sometimes causes extreme malignant events and leads to the birth of the term — 'football hooliganism' [20, 21]. Past research has studied how football hooliganism, which became a world phenomenon, was located in the social structure [20, 21]. Spajj [65] compared six western European football clubs to determine the context of this culture and the group's identity. Despite previous mainly work on football fan culture from its origin, some recent works discussed the fandom with the contemporary technology [32, 55]. Hopkins and Treadwell [32] investigated this phenomenon in the early global media age. Poulton [55] analyzed the market triggered by football hooliganism from the aspects of derivative products. Regarding recent research on sports fans' behaviors, Zhang et al. [82] provided first large-scale characterization on analyzing the impact of team performance on the online fan communities of NBA teams. There are also related works focusing on social media dynamics in the World Cup, analyzing the live-tweeting and found user activity and discussion contents online are changed with the events [41]. Vasconcelos et al. compared the difference of emotional reactions and language use between 2014 and 2018 World Cup Final [72]. Fan et al. measured sentiment level trends over the tournament progression of the England National Team in the 2018 World Cup to reveal how fans associate them with a successful team [24]. Other aspects of sports fan behaviors like fan attraction, fan satisfaction, and fan identification have also been studied [6, 18, 22]. However, most research tended to employ qualitative methods like interviews or surveys instead of numerical and quantitative analyses. Compared with previous works, the present research connects football fan behavior with user interactions in online communities and conducted an analysis at a much larger scale based on the data collected from Reddit.

2.3 Dynamics in Online Communities

The complex and rich interactions between users in online communities provide evidence to understand the user patterns and dynamics of online interpersonal communication. One essential topic is the community norms that group members develop and how they, in turn, influence the ways that members interact with each other and participate in the community [34]. One previous work conducted in weight loss community has found that different norms can effectively regulate user behaviors into different patterns of similar types of communities [11]. Other research focused on topics such as how to help newcomers become involved in the community [40]. Besides positive developments, online communities may evolve in a negative direction. Literature on online community dynamics has also explored polarization and extremism within these communities. [17, 19]. Online community polarization is frequently discussed in research on politics. One early study pointed out that conservative and liberal blogs formed distinct user groups with little overlap [1]. Groups with distinct ideologies have also been identified in several studies on various online social media platforms such as Twitter and YouTube [15, 71]. Further, the implications of online social media on polarization have also been examined. Levendusky [45] showed how online media polarized and aggravated the extremeness of viewers. Garrett et al. [28] provided evidence that the use of biased news sites promotes inaccurate beliefs.

Until now, research on online community dynamics has still been the primary concern of the social sciences and the field of Human-Computer Interaction. In recent years, researchers leveraged more advanced techniques to conduct an analysis at a larger scale. Kubin et al. [43] ran a systematic review of online political polarization and spurred a rapid increase in research in this field. Yarchi et.al. evaluated polarization from the aspects of interactional polarization, positional polarization, and affective polarization [78]. Waller and Anderson developed a neural-embedding methodology to quantify online polarization at a large scale in online communities by analyzing similarities in community membership [75]. Even though the majority of the published research is limited to the scope of politics, the contrariety between rival teams in sports may also give rise to the unique behaviors found in their online communities [30]. This research unfolds the discussion about the dynamics of fan behaviors by evaluating features of comments and users' responses to other comments in online sports communities.

3 DATA PREPARATION

This research was primarily based on the Reddit community, and the data, including posts and comments collected from the Reddit football community, formed the foundation of the project. Some derived statistics published by FiveThirtyEight[2] were also included as controlled factors of the research.

3.1 Reddit

Reddit is one of the most popular websites and has also become one of the leading data sources in recent HCI research [56]. This

[2] https://fivethirtyeight.com/

project's primary dataset is derived from Reddit data and consisted of posts, comments, and other corresponding statistics. Micro-communities are referred to as "subreddits", and enable users to create posts, make comments on posts, and upvote or downvote posts and comments. The displayed vote count is the difference between the number of upvotes and downvotes. Users' Karma refers to the score users receive based on their community involvement (e.g., when they post and comment on Reddit and receive up or downvotes); thus Karma is a representation of a user's reputation. Generally speaking, most comments receive more upvotes than downvotes, with only around 4% of comments having negative votes ($\frac{150207}{3792471}$ to be specific in our datasets). Due to this bias, posting more comments will usually accumulate more Karma for a user. If a user has high Karma, it means that the user publishes a large volume of posts and comments that are of high quality.

3.1.1 Overview of the Reddit Football (Soccer) Community. The England Premier League consists of 20 teams. Clubs in the last three positions of the ranking table will be degraded to the Football League Championship, and winners of the Football League Championship will fill the vacant position next year. For the 2020-2021 season, the twenty teams were Manchester City, Manchester United, Liverpool, Chelsea, Leicester City, West Ham United, Tottenham Hotspur, Arsenal, Leeds United, Everton, Aston Villa, Newcastle United, Wolverhampton Wanderers, Crystal Palace, Southampton, Brighton & Hove Albion, Burnley, Fulham, West Bromwich Albion, and Sheffield United (ordered by team standing). Among them, six "super clubs," which occupied the top six positions of the ranking table for the bulk of the 2010s decade, were known as the "Big Six" (i.e., Arsenal, Chelsea, Manchester United, Manchester City, Liverpool, and Tottenham Spurs) due to their excellent performance on the pitch and significant financial influence off the pitch, thus leading to a huge gap between the majority of the Premier League and the "Big Six." Among the 20 teams, all five of the "Big Six" except Manchester City had an overwhelming number of posts and comments compared to other teams within the target time period.

The England Premier League starts in early September and ends in late May the following calendar year. There are two transfer windows during the year. The summer transfer window for the premier league usually lasts from June to August and the winter one lasts throughout January. Figure 1 illustrates the user activity distribution in all twenty clubs' subreddits during the 2020/2021 season from 1 Sept. 2020 to 31 Aug. 2021. Two summits can be seen around early September and May. The first peak starts at the end of August, near the end of the transfer window. Widespread discussion centering on the influential trades or accompanying rumors dominates the community during this time. As the season unfolds in September, the number of comments stays at a high level. The other peak occurs around May, near the end of the season. Another reason accounting for the high user activity at this time can be attributed to the UEFA Champions League, the most valued continental-level competition for European football clubs. As two EPL teams, Chelsea and Manchester City, both made it to the final, it led to more posts from Premier League fans and maintained the high user activity seen in late May. It is noteworthy that the steep decrease during March was mainly due to damage that occurred to the Pushshift database. Partial data of some days during that period were deprecated in the database.

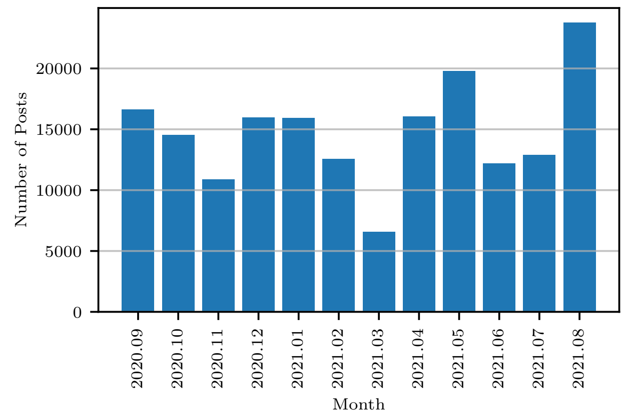

Figure 1: User activity aggregated across team subreddits by month

3.1.2 Data Collection. Balancing sample size, language, and other factors, this work focuses on posts about the England Premier League. We only collect data from the official subreddits of each team as the source: *r/MCFC*, *r/reddevils*, *r/LiverpoolFC*, *r/chelseafc*, *r/lcfc*, *r/Hammers*, *r/coys*, *r/Gunners*, *r/LeedsUnited*, *r/Everton*, *r/avfc*, *r/NUFC*, *r/WWFC*, *r/crystalpalace*, *r/SaintsFC*, *r/BrightonHoveAlbion*, *r/Burnley*, *r/fulha-mfc*, *r/wbafootball*, and *r/SheffieldUnited* (ordered by team standing). The data from the Reddit community was collected from pushshift.io [5]. First, we scraped all the posts from the target subreddits using PSAW (Python Pushshift.io API Wrapper) [5] by searching within the specified subreddits. For comment scraping, we searched for all the comments in a subreddit within the one-year period from September 2020 to August 2021, which coincided with the Premier League season cycle.

All posts were collected during this period for all 20 teams in the Premier League at that time. However, due to the overwhelming quantity of comments, only selected comments were collected and taken into consideration according to the research questions. Since the study focused on the effects of the events on match day, this research collected three days of comments starting from the day of each Premier League match. A total of 177,899 posts and 3,792,471 comments were included in the dataset.

3.2 FiveThirtyEIght

FiveThirtyEight is a popular forecasting website. It developed a club soccer prediction system based on a revised version of ESPN's Soccer Power Index(SPI) ratings, which provided their best estimate of a football team's overall strength [35]. The offensive and defensive ratings represented the goals a team expected to score or concede. These two ratings, in turn, generated the overall SPI rating. On the grounds of the SPI ratings of both teams in a match, the website publishes the predicted result of the match.

The SPI rating is affected by two factors before a season begins: the SPI rating at the end of the previous season and the evaluated performance in the transfer market according to the market value during the summer transfer window. The rating is adjusted after each match. By design, a team's rating will not necessarily rise after winning a match, as other factors, like the opponent's strength, may also affect the rating. One method of measuring team match performance, which mitigates the randomness of the football match, integrates three metrics, including adjusted goals, shot-based expected goals, and non-shot expected goals, to produce the final composite offensive and defensive score of the team performance. In this research, the match prediction data for all the matches in the England Premier League during the 2020-2021 season was included.

4 METHOD

For our study, we integrate several methods to facilitate the analysis. For RQ1, We first use hierarchical linear regression to analyze the matches' impacts on user sentiment. The independent variables and metrics to measure the sentiment is described. We also use structural topic modeling to examine the discussion content for RQ2. For part of RQ3, a politeness measurement method is provided.

4.1 Independent Variables

A hierarchical regression analysis was conducted to evaluate our research questions. A hierarchical regression requires controlling other variables while analyzing the predictor variable. By accounting for the change in variance that occurs after adding each group of variables into the model, outside effects can be controlled and one can determine how target variables affect the dependent variable [46]. The independent variables in the study were grouped in team information, the match results, and related indicators about the match.

4.1.1 Team Information. The rank at the end of the season was used to identify the performance of each team during the 2020-2021 season, which is one of the most significant variables that could identify variations between teams. Another variable (***the big six***), which is a dummy variable, was also used to capture whether the team was among the six biggest teams (big six (1, *is big six*; 0, *not big six*)). This was important to determine because the big six had larger fan bases than the other teams and, thus the potential for more posts about them. Another variable identified whether a team was a top or bottom team (i.e., the top five teams were the top team, and the five lowest-ranked teams were bottom teams). The top five teams in 2020-2021 season are Manchester City, Manchester United, Liverpool, Chelsea, and Leicester City. The bottom five teams are Brighton, Burnley, Fulham, West Bromwich Albion, and Sheffield United.

4.1.2 Match Results. The results of the match can be essential driving factors of fan behaviors, so a number of variables were included to capture the results of the match, including winning and not losing that are dummy variables when included in the regression (i.e. winning (1, *win*; 0, *tie or lose*); not losing (1, *win or tie*; 0, *lose*)). Another important indicator was the number of goals of both teams, with which the match's characteristics could be further captured. However, the final scoreline may not reflect football fans' impressions of the team's performance due to the limited information brought by the low-scoring nature of the football match [35]. Thus, we propose a new metric to measure team performance in section 4.1.4.

4.1.3 Match Information. Apart from the match results, we hypothesized that fans' emotions and behaviors could be further explained by the gap between their expectation and the final match results. To quantify fan expectations of each match, we included the match prediction data from the FiveThirty Eight database for all the matches in the England Premier League 2020-2021 season. Among all the supporting data used in the prediction, we chose the following variables: 1) The SPI of both teams, which measures whether the match was a well-balanced or if there was a great disparity between the two teams (*mean*: 74.4; *std.*: 9.5; *max*: 94.2; *min*: 51.6); 2) The winning probability is the most direct index to measure the fans' expectations before a match starts, which is based on the projected match scores calculated based on the SPI of both teams [35]. Losing a match with a high probability might lead to some peculiar fan behaviors (*mean*: 38.1%; *std.*: 18.7%; *max*: 92.9%; *min*: 1%); and 3) The importance of the match, which was a value ranging from 0 to 100 that measured the importance of the match (*mean*: 34.8; *std.*: 27.5). The difference in teams' strengths cannot fully represent the significance of a match since other factors such as the team standings may also lead to varying match importance, especially near the end of the season (e.g., matches that decide whether a team can win the champion or qualify for into the European Champions League or avoid relegation are usually considered to be key matches).

4.1.4 Measuring the Gap Between Pre-match Expectations and Match Results. As illustrated by the FiveThirtyEight team, a match's final scoreline (i.e., score) often disagrees with football fans' impressions of a team's performance due to the low-scoring nature of a football match [35]. The FiveThirtyEight team proposed a better metric to estimate team performance that uses three parameters, i.e., adjusted goals, shot-based expected goals, and non-shot expected goals. All three parameters were calculated and inferred according to the on-field statistics and football match-related rules. By averaging the results from three parameters, one composite offensive score and one composite defensive score can be generated to measure the team's offensive and defensive performance, respectively.

This metric proposed by FiveThirtyEight provides a more reasonable index that reveals a team's quality of play. However, it does not capture fans' impressions of a match as fans will judge the team's performance not only by on-pitch statistics but also by expectations they have in mind. For example, since fans of underdogs will not have high expectations when their team plays league leaders, they are unlikely to be too disappointed if their team underperforms. Correspondingly, fans of the "Big Six" are likely to have complaints even when they win against lower-ranked teams. Another underlying element is match importance. Fans will not have high expectations about their team's performance before trivial matches when a team rotates its lineup and retains its strength. Therefore, our new metric includes the winning probability, losing probability, and match importance provided by FiveThirtyEight to quantify fans' prematch expectations and match results, which is a more reasonable metric to estimate fans' impressions of a match. For simplicity, we refer to this measure as the "gap score".

In this new metric, the composite offensive score and defensive score provided by FiveThirtyEight are denoted as S_o and S_d.

Table 1: Independent Variables for the Hierarchical Regression Analysis

Variable	Definition	Source
Team Information		
end Season Rank	The team's rank at the end of the season	FiveThirtyEight
big six	If a team is among the recognized "Big Six" in the EPL	N/A
top team	If a team is ranked in the top 5 at the end of the season	N/A
bottom team	If a team is ranked in the bottom 5 at the end of the season	N/A
Match Results		
winning	If a team wins the match	FiveThirtyEight
not losing	If a team wins or ties the match	FiveThirtyEight
goals	The number of team goals during the match	FiveThirtyEight
opponent goals	The number of opposing team's goals during the match	FiveThirtyEight
Match Information		
team SPI	A team's SPI before the match	FiveThirtyEight
opponent SPI	The opponent team's SPI before the match	FiveThirtyEight
winning probability	A team's winning probability estimated by FiveThirtyEight	FiveThirtyEight
match importance	The importance of the match to the team estimated by FiveThirtyEight	FiveThirtyEight

The probability of winning, losing, and tying, also calculated by FiveThirtyEight, are denoted as P_w, P_l and P_t. These three probabilities are summed to 1, i.e., $P_w + P_l + P_t = 1$. We also denote the importance of the match as i. Another variable, R, represents the match result:

$$R = \begin{cases} 1 & \text{win} \\ 0 & \text{tie} \\ -1 & \text{loss} \end{cases} \tag{1}$$

We denote the adjusted score difference as δ, the importance coefficient as μ, and the match result bonus as β. The following equations are used to calculate the *gap score*:

$$\begin{cases} \delta = \frac{P_l S_o - P_w S_d}{S_o + S_d} \\ \mu = e^{\frac{i}{100} - 1} \\ \beta = 0.2R \end{cases} \tag{2}$$

$$gap\ score = \mu(\delta + \beta) \tag{3}$$

Instead of simply subtracting the goals conceded from the goals scored, the adjusted score difference adds the coefficient P_l to the offensive score and P_w to the defensive score. After adding the coefficient, the goals scored will weigh more if the pre-match expected probability of losing is high. Consistent with common sense, people will place a higher value on the goals if the team is the underdog. Correspondingly, goals conceded will have a higher weight if people take it for granted that the team will win. We divide the result by $S_o + S_d$ to scale the value between -1 and 1.

In practice, the matches in 2020-2021 EPL season have a δ minimum of -0.435, a maximum of 0.439, and a standard deviation of 0.148. μ represents the importance coefficient which is a value from e^{-1} to 1. Since critical matches tend to play a greater influence on fans' behaviors, a more significant importance coefficient will contribute to the greater absolute value of the emotional fluctuation, i.e., the *gap score* and vice versa, people will not pay much attention to trivial matches, subsequently leading to a *gap score* close to 0. To achieve this, we apply the exponential function to the linear transformation of the initial match importance i. Within our dataset, μ covers the whole range from e^{-1} to 1, with a mean of 0.546 and a standard deviation of 0.146. For a significant proportion of fans, who did not watch the match, the immediate result significantly shaped their impression of the match. β was designed as a bonus to differentiate the final calculated *gap score* when the match result was a win, tie, or loss, respectively. The final *gap score* of games calculated within the dataset had a minimum value of -0.59, a maximum value of 0.40, and a standard deviation of 0.15.

4.2 Metrics to Measure Fan Sentiment

The sentiment is the most direct feature representing the mainstream attitudes of a current discussion. To further explore the sentiment of football fans, we used VADER (Valence Aware Dictionary and sentiment Reasoner), a lexicon and rule-based sentiment analysis tool for social media, to generate the sentiment score of each comment [33]. VADER is widely used and performs well in the social media domain [9]. The model provides the proportion of positive or negative sentiment within a sentence without any training [9]. The resulting score is a number normalized within the range of -1 (most extreme negative) and 1 (most extreme positive) [33]. A threshold of 0.05 and -0.05 is set to classify sentences as either positive, neutral, or negative. To provide an overview of the sentiment score of the dataset, the mean sentiment score was -0.11 and the standard deviation was 0.46, with 26.6% of comments having a negative score, 43.5% having a positive score, and approximately 30% of comments having a score of 0, which could be due to some comments only containing images or URLs.

Besides directly measuring the mean sentiment score, we also computed the correlation between the "gap score" and the vote-weighted sentiment score. The vote of each comment was counted as a coefficient, and we computed the weighted arithmetic mean of the sentiment score. We denote the sentiment score as s_i and its vote as v_i. The total sample size is n_{sample}. The vote-weighted mean of the sentiment score was defined as:

$$Vote\ Weighted\ Mean\ of\ Sentiment\ Score = \frac{\sum_{i=0}^{n_{sample}} v_i s_i}{\sum_{i=0}^{n_{sample}} v_i} \quad (4)$$

4.3 Hierarchical Linear Regression

To evaluate the contributions of match results on user emotions, we employed hierarchical linear regression, which is commonly found in social sciences research [46]. By calculating the difference of the adjusted R^2 of the regression model after adding each group of variables to the model, the analysis can control the effects of other factors and examine the contribution of the predictors to the result [53]. Therefore, we choose this model to help evaluate the factors driving the community sentiment towards match results. We first defined a time period according to the match time and filtered all the comments of both match teams' subreddits. Various periods, including 6 hours and 12 hours since the match started, were evaluated. The average sentiment score of a match is denoted as $Sent_m$ for short in the following model. The full linear regression model was:

$$\begin{aligned} Sent_m \sim & \beta_0 + \beta_1 end\ season\ rank + \beta_2 big\ six + \beta_3 top\ team \\ & + \beta_4 winning + \beta_5 not\ losing + \beta_6 goals \\ & + \beta_7 opponent\ goals + \beta_8 team\ SPI + \beta_9 opponent\ SPI \\ & + \beta_{10} winning\ probability + \beta_{11} match\ importance \end{aligned} \quad (5)$$

4.4 Structural Topic Modelling

To examine the discussion topics of the fan community and the effects of team performances, utilized Structural Topic Model (STM), a general framework for topic modeling with document-level covariate information. We used this topic modeling rather than using Latent Dirichlet Allocation [8] because it analyzes covariates' inference and qualitative interpretability on topical prevalence [57]. Prior research has employed STM to understand content sense-making using the data collected from social media [27, 59].

The topic modeling was run on the *Big Six*'s dataset. We include comments within 12 hours after the kick-off time of each match to inspect the community discussion topics during and after the match. A series of preprocessing steps were performed on the dataset before loading the data into the topic model. We first filtered out non-English and empty comments and converted each word to lowercase and lemmatized inflected words. We also included n-grams(n=2,3), removed stop words, tokenized the sentences in the dataset, and removed the names of the players and teams.The names of the coaches who manage England Premier League teams or other big clubs were replaced by the word "manager." All the names of EPL teams' stadiums are substituted by the word "stadium". Lastly, we removed all the words that were not nouns to improve the model performance [48].

As the training of *STM* demands preset numbers of topics, we used the *searchK* function to evaluate model performance with topic numbers ranging from 3 to 20 [57]. *STM* measures model performance based upon the following four metrics: held-out likelihood [74], semantic coherence [50], residual [67], and lower bound. Semantic coherence measures the co-occurrence of the words in the same document to ensure that the topics generated by the model are semantically coherent [50], which is one of the mainstream methods to measure the quality of the topic. However, Roberts et al. pointed out that semantic coherence alone will easily reach a high score when the number of topics is low and the topic is dominated by the most common words [58]. Held-out likelihood evaluates the probability of the held-out portion to reveal the generalization of the model [50]. For the final training, we chose 15 topics as this model reached a relatively high semantic coherence and held-out likelihood but maintained low residuals and a lower bound (Figure 2) with 15 topics. Meanwhile, the *gap score* was set as the target covariant of the model separately. The model analyzed the effects of the covariant on the topical frequency after training. Section 5.2 discusses the result of topic modeling in detail.

4.5 Politeness Measurement

To answer **RQ3**, we measured the relationship between sentence politeness and the votes it received. One user complained about his experience in *r/soccer*: *"If your opinion isn't exactly in line with the majority, they'll downvote you and make you feel dumb for thinking that way. I asked a question and was very careful to say "not taking anything away from you or disagreeing, I'm just interested in why so many people think this way" and got downvoted for literally just trying to understand WHY the popular opinion was what it was."* [3] This experience implies that polite word use in online communities should be encouraged even though it is not in line with the mainstream values of the group. This is also a sign indicating a community's inclusiveness, i.e., comments in an sincere tone are expected to receive more votes than other comments. In section 5.3.2, we evaluate whether comments using polite words are preferred in online discussions.

In this work, we use R's Politeness package to detect politeness features in the dataset [79]. The package extracted linguistic politeness markers based on the previous literature on politeness study [10, 44]. A total of 36 features were used to detect politeness. We set the metric to "Average" to count the prevalence of features as a percentage of the word count of each sentence. Then, we summed the frequencies of all features to generate the metric. After excluding comments comprised of images, URLs, or other content that was linguistically incomprehensible, comments without any sign of politeness comprised around 4% of the remaining comments. The overall politeness of the dataset had a mean of 0.249 and a standard deviation of 0.246. Afterward, we further categorized all the comments with a politeness score equal to or larger than 0.7 as the "polite" group and treat other comments as the control group. Note that, as we defined before, the politeness score is the sum of all the frequency of politeness markers, which is a number larger than 0 and may exceed 1 in some marginal cases.

5 FINDINGS

This section presents the results of our research questions. We found that fan sentiment is significantly affected by the game results, as well as the gap score. Meanwhile, we conclude four categories of discussion topics and identify the topic distribution varies with different game results. For the last research question, we conclude

[3] https://www.reddit.com/r/SoccerNoobs/comments/97wtf4/does_anyone_else_find_rsoccer_toxic/

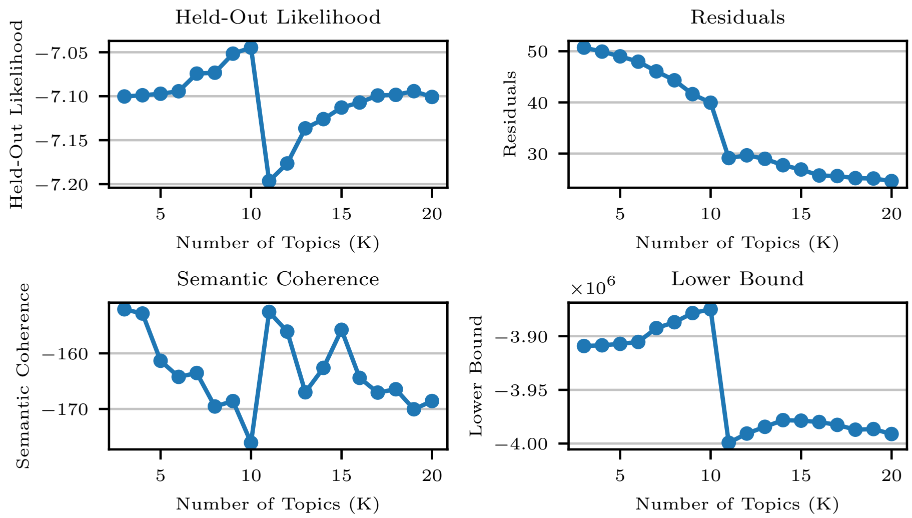

Figure 2: Held-out likelihood, lower bound, residuals, and semantic coherence scores for models with 3–20 topics.

that users tend to vote for comments with extreme emotions. Meanwhile, being polite does not mean you will get a higher number of votes.

5.1 Match Effects on Fan Sentiment

Intuitively, a sports team's performance is closely related to fans' sentiment levels. Nevertheless, as it is still unclear to what extent these two factors relate to each other, we analyzed this relationship from the perspective of match results.

5.1.1 Fans' Sentiment is Significantly Affected by Match Results. Firstly, having a general idea of the match results-fan sentiment relationship is important (Figure 3). The comments we analyzed were within six hours of the kick-off time for each match, and high-voting comments were defined as those in the third or higher quartile of the corresponding subreddits.

When a team won a match, the average sentiment scores for all comments and selected high-voting comments were significantly higher than those when the team lost. The patterns in the sentiment scores while winning and losing were relatively similar across all teams except for Burnley, which had much lower negative sentiment score for all comments and high voting comments when the team lost. As expected, when a team won a match, the sentiment scores were slightly lower for all comments than high voting comments. Thus, fans tended to support those comments that expressed a more positive sentiment. When a team lost, the sentiment scores for all and high voting comments was somewhat similar, suggesting that this may be a way to determine whether a community prefers extreme comments or not. If users in a community tend to vote for those comments with lower sentiment scores when the team loses, chances are that they will prefer criticism rather than rational reflection after a loss. We can infer that these fan communities have a relatively low tolerance for the team's poor performance. We will further discuss fans' preference for voting in the section 5.3.1.

For all teams except Burnley, there were more positive sentiment comments than negative sentiment ones, regardless of the match's results (Figure 4). When there are wins, there are more positive comments (which are near to .7 vs .3), but when there are losses, the sentiment is more mixed (around .5) because fans will always have something negative to say. When Burnley wins, comments with positive sentiments comprise about 90 percent of all comments, and the negative ones account for around 10 percent. By contrast, when it loses, the proportion of positive comments and negative comments reverses, with there being over 60 percent negative comments and less than 40 percent positive comments.

Having a general idea of the match results-fan sentiment relationship, we conducted a Hierarchical Linear Regression Analysis to further examine football fans' emotional behavior. Having the impression that match results may have a significant impact on community sentiment, the adjusted R^2 value shown in the bottom of Table 2 is 0.409 and 0.423 for the 6-hour and 12-hour analysis, respectively, demonstrated that the selected variables explained the emotions of fans (Table 2). Overall, winning a match leads to a higher sentiment score, and so do more goals scored and fewer goals conceded ($r = .33,\ p < .0001$). By controlling other features, adding the group of match results (i.e. from Reg.1 to Reg.2) into the regression model led to an increase in the adjusted R^2 from 0.058 to 0.440 for the six-hour data and 0.057 to 0.427 for the 12-hour data. For other features for controlling, though their effects on user sentiment are not significant, some observations can be concluded. For example, the team's SPI is negatively correlated with the overall sentiment score, while the opposing team's SPI is positively correlated with the sentiment. As the SPI ratings can be interpreted as the team strength, if a fan's team is weaker and the opposing team is stronger, fans will have low expectations and will be less likely to post negative comments. By contrast, higher expectations can lead to a gap between fact and expectation, which is a strong trigger for words with negative emotions. To conclude, the sentiment of

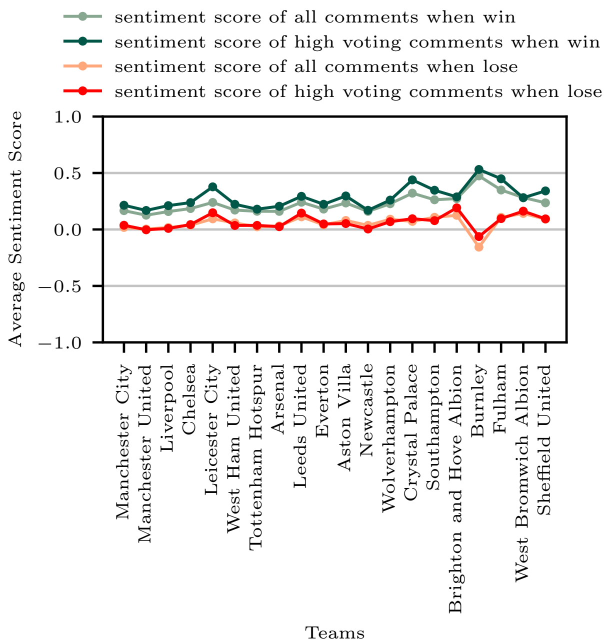

Figure 3: The average sentiment score in the team subreddits for different match outcomes.

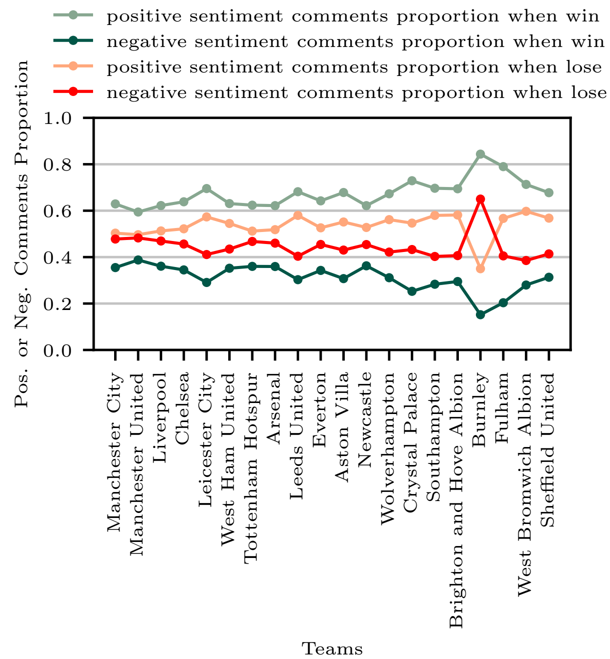

Figure 4: The proportion of positive & negative comments in team subreddits when a team wins or loses a match.

fans is strongly affected by the game results. We will also see how this result echoes the relationship between the gap score and fan sentiment in the next section.

5.1.2 Effects of the Gap Between Pre-match Expectations and Match Results. Following the results from the regression analysis, we speculated that fan behavior may be largely attributed to the difference between fans' expectations of the team before the match and the team's actual performance. The gap score computation revealed that the gap score are positively correlated with fan sentiment ($r = .58,\ p < .01$) for significance at the 0.01 level, which means if the team's performance exceeds the fan expectation, the overall sentiment will be higher. When using the vote-weighted mean of sentiment score, the gap score is also positively correlated with weighted fan sentiment ($r = .54,\ p < .0001$). Meanwhile, it was also positively correlated with the proportion of comments with a positive sentiment score ($r = .57,\ p < .0001$) and negatively correlated with the negative sentiment comments' proportion ($r = -.57,\ p < .0001$). The result, in turn, proves the reasonability of our definition of the *gap score*.

5.2 Characteristics of Discussion Content

This section presents the four types of content that were present in the community posts and how these topic types change based on the outcome of a match.

5.2.1 Discussion Content Types. The structural topic modeling revealed fifteen topics that were common in the dataset, and we listed the top 8 topics with the highest topic proportion (Table 3). The topics are displayed in descending order of topic proportion. We also grouped these topics into four groups: match process, season performance, squad discussion, and team member turnover.

Topics directly concerning the football match itself, **Match Process** ($Topic_{11}$), takes up the largest proportion of discussion contents in the community (e.g., match events, tactics, and team or player performance). For **Season Performance** ($Topic_{12}$) and **Season Prospects** ($Topic_3$) are similar to team performance-related topics, but they were more concerned with performance throughout the season. Fans were especially concerned about the team's position in the ranking tables, e.g., from Chelsea's subreddit *r/chelseafc*, "*The Everton loss is big, because one of their matches in hand is against Man City on Wednesday and they're not winning that. So if we beat Newcastle, we'll most likely be 5 points ahead and officially above them despite them having another match in hand.*" Such commentary combined past team performance and the current standings to illustrate the team's prospects this season and how to achieve a desired position in the standings. Comments like this should be considered positive and encouraging as they have inspiring suggestions for the team and can help those unfamiliar with the team quickly understand the team's current situation. **Club Decision** ($Topic_7$) and **Transactions & Contracts** ($Topic_4$) focused on the club's decisions about member turnover and other business-related issues. Lots of attention was paid by fans to rumors regarding player and coach signings and major player renewals, especially in the off-season. Despite our dataset limiting the valid time period to the match day, discussions regarding the transfer market still occupied community activity. Another category containing the topics

Table 2: The Hierarchical Linear Regression Analyses of the sentiment level based on the match results. The analysis used comments selected six hours and twelve hours from kick-off of the match. The value are the standard beta coefficients (* denotes p < 0.05, ** denotes p < 0.01, and * denotes p < 0.001; p-values are reported without a Bonferroni correction in all regression tables)**

	12h			6h		
Predictor Variable	Reg. 1	Reg. 2	Reg. 3	Reg. 1	Reg. 2	Reg. 3
Team Information						
End Season Rank	0.158*	0.373***	0.346***	0.141*	0.354***	0.320***
Big Six	-0.210***	-0.201***	-0.176***	-0.220***	-0.212***	-0.184***
Top Team	0.162**	0.172***	0.209***	0.148***	0.156***	0.194***
Match Results						
Won		0.317***	0.326***		0.321***	0.332***
Lost		0.177***	0.192***		0.187***	0.202***
Number of Goals		0.143**	0.147**		0.140**	0.144**
Number of Opponent Goals		-0.144**	-0.167***		-0.142**	-0.164***
Match Information						
Team SPI			-0.148*			-0.165*
Opponent SPI			0.238***			0.244***
Winning Probability			0.139*			0.148*
Match Importance			-0.075*			-0.075*
Adjusted R^2	0.057	0.427	0.451	0.058	0.440	0.465

of the **Coaching** ($Topic_1$) and the **Squad** ($Topic_{14}$) involved daily discussions about the team line-up, including the manager and players. Though arranging the lineups is closely connected with match strategy, these two themes were classified into distinct topics because the squad rotation may not be match-specific. The topic may contain comparisons and discussions between players, not limited to the tactical deployment of lineups. Consequently, many accusations and attacks on players will also appear in this topic. The last topic centers on the online community ecology itself ($Topic_{10}$). Terms like 'post,' 'comment,' 'opinion,' 'Reddit,' and 'subreddit' were frequently used. These online football community discussion topics can be grouped into the following four categories: match process, season performance, squad discussion, and team member turnover.

5.2.2 Discussion Topic Distribution Varies as the Team Wins and Loses. Leveraging **stm**'s ability to estimate the relationship between metadata and topics, we use **stm** to estimate the underlying implications of the gap between pre-match expectations and match results [57], which can be leveraged to . As the *gap score* was in the range from -1 to 1, we converted it to a binary variable to analyze how the change in topic proportion shifts from one to another (Figure 5), To specify, scores greater than 0 indicated a positive gap (i.e., the team performance satisfied the initial fan expectations) whereas scores less than 0 indicated that the team's performance disappointed fans.

As expected, topics involving personnel changes appeared to be accompanied more by comments about poor team performance, as the disappointing performance may give rise to discussions among fans about the management of the club (**Club decision**($Topic_7$)) or player contracts (**Transactios and Contracts**($Topic_4$)). It is worth mentioning that both topics are categorized into the same group "Team Member Turnover" and these are the only two topics that had a negative gap score. However, discussions about the **squad** ($Topic_{14}$) were positive-leaning.

Another observation is that **Match Process**($Topic_{11}$) was frequently used when the team won a match, which contradicts our prediction the discussion around the game process will appear more often when the team loses. Zhang et al. [82] drew the conclusion that fans are more active when top teams in the NBA lose. As the topics are drawn from the subreddit of *Big Six*, which can be considered strong teams in the league, we previously considered that higher user activity and the discussion around the match itself would be more frequently used when the team loses the match. We infer that the reason for our results may be that fans will focus more on matches when a team wins, while many other topics, like player transactions, may be involved when the team loses. Thus, the overall sentiment level was positively correlated with the match result. Hence, we can conclude that the negative word usage (i.e., topics frequently used when the team underperforms) may accompany the discussion of the lineup changes, and the positive comments (i.e., topics frequently used when the team overperforms) may take up a larger part in the discussion around the match itself.

5.3 Patterns of Vote Commenting

Voting on a comment imparts a series of potential effects on the future behavior of users [13]. It can also reflect the identities of different communities. Here we explore the inclusiveness and polarization of the community by inspecting the voting patterns in our dataset.

5.3.1 Comments with Extreme Emotions are more welcomed than Neutral Comments. This section discusses the distribution of comments from the perspective of sentiment and votes. Figure 7 first displays the distribution of all comments. Note that to reach the

Table 3: List of top 8 topics sorted by their topic proportion (i.e., the percentage of posts on a topic across all posts). The keywords were based on the probability likelihood of each word in the specified topic and the frequency and exclusivity, which measures the exclusivity that balances the word frequency [2, 7].

Topic(ID)	Description	Metric	Keywords	Topic Proportion
Match Process T_{11}	Words related to the match, including match events, match tactics and etc.	Prob.	goal, score, position, pass, line, midfield, attack, quality, defender, midfielder	11.5%
		Frex.	press, box, space, shot, pace, possession, assist, cross, attack, goal	
Season Performance T_{12}	Team performance during the season that is particularly concerned with the position in the standings table	Prob.	team, point, league, start, match,win, form, performance, result, drop, title	11.4%
		Frex.	win, beat, draw, point, cup, europa, facup, fixture, team, deserve	
Club Decision T_7	Club decisions, including the business and management	Prob.	club, money, support, care, decision, owner, board, competition, sport, business	8.3%
		Frex.	owner, superleague, fifa, fsg, revenue, ownership, debt, billionaire, model, support	
Community T_{10}	Words relevant to the online community itself	Prob.	user, post, comment, talk, question, mate, opinion, base, mention, concern	7.8%
		Frex.	comment, concern, action, video, subreddit, thread, moderator, bot, downvote	
Coaching T_1	Discussion about the coaching team	Prob.	manager, world, level, imagine, turn, age, coach, plan, experience, talent	7.7%
		Frex.	manager, experience, year, career, coach, imagine, academy, age, talent, level	
Season Prospects T_3	Discussion regarding the season and the championship	Prob.	season, chance, end, squad, injury, sense, trophy, rate, compare, championship	7.0%
		Frex.	compare, season, average, bundesliga, park, injury, end, second, bus, premier	
Squad T_{14}	Discussion about the lineup	Prob.	striker, bench, option, injure, fit, choice, winger, lineup, formation, backup	6.3%
		Frex.	bench, fit, lineup, worry, starter, rotation, pair, prefer, backup, option	
Transactions & Contracts T_4	Discussion related to the transfer market, contract renewal, possible signings, etc.	Prob.	sign, summer, pay, deal, transfer, contract, loan, sell, wage, window	0.063
		Frex.	deal, transfer, contract, wage, price, offer, fee, agent, budget, sign	

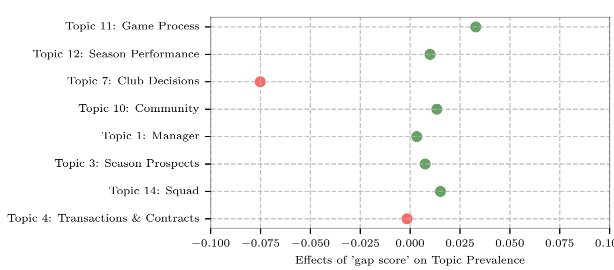

Figure 5: Effects of 'gap score' on Topic Prevalence. All these values are tested with statistical significance, with $p < .001$.

relatively accurate proportion distribution, we exclude all the comments with the sentiment score equal to 0. These comments are comprised of images, URLs, or other content that is hard to judge the sentiment. Even though a Shapiro–Wilk test indicated that the data was not normally distributed, the positive and negative comments, respectively showed a pattern similar to a normal distribution, with data near the mean being more frequent in occurrence than data far from the mean [63]. That means users tend to have an established positive or negative feeling before posting a comment, which leads to the summit distributed near 0.5 and -0.5 separately instead of 0.

Figure 6 exhibits the comments distribution for the subreddits of all teams. For the vote comments, all teams, besides the bottom five teams, have a similar pattern with a slight difference. The pattern displays a U-like distribution. The extremely positive or negative comments will receive somewhat higher votes than the neutral comments. Among EPL's ***Big Six***, all six teams' subreddits spontaneously have a much higher end on the positive side than on the negative side. The mid-table teams presented a more typical U-shape distribution with both ends having a closer value to each other. For some bottom teams, the vote comments failed to follow this pattern. The reason can probably be attributed to their small sample size due to the low user activity. Nevertheless, the general result is that 70% (14/20) communities have the most votes for comments expressing extremely positive feelings.

Similarly, if we only focus on comments with negative emotions, the public tends to vote for those with more extreme sentiments. To further verify the observation described above, we labeled the comments with an absolute value equal to or greater than 0.8 as "extreme" and the comments with those absolute values equal to or smaller than 0.2 as "neutral." We use the Mann-Whitney U test to evaluate whether these two groups have a significant voting difference. The result also verifies that comments with extreme sentiment expression received higher average votes than neutral comments ($U = 2.53 \times 10^{11}$, $p < .0001$). The result may imply that users are more likely to relate to comments holding a similar opinion with a more straightforward expression. Neutral comments and comments with implicit feeling expressions may have less resonance. The tendency to vote for comments with extreme emotions might be another piece of evidence proving the polarization and irrationality of fan behavior, which we cannot conclude now since emotional comments do not always mean extreme behaviors.

5.3.2 Polite Comments Will not Receive a Higher Vote than Other Comments. As one critical dimension of human communication, politeness is a powerful component that affects the speakers' social goals [79]. As being polite tends to carry a friendly message, it is encouraged and deserves to have better feedback in the course of online discussion [10, 44, 79]. For our dataset, we compared the votes between the comments in the "polite" group and the control group. We used the Mann-Whitney U test did not reveal a significant difference in the average sentiment score of these two groups ($p = 0.35$). Contrary to our previous expectations, having a good sense of politeness will not necessarily lead to better feedback. The finding also corresponds to the user's poor experience in section 4.5 that you may be downvoted if your opinion goes against the mainstream even if you have a good manner.

6 DISCUSSION

The sports community is an event-driven online discussion space. How fans behave online after a match is a way to understand people's behavior in the face of something meeting or not meeting their expectations. Prior research examining the offline-online connection of online communities studied how online experiences shaped users' offline behaviors, and they found the integration between online and offline supports passion-centric activities [54]. For those studying the impact of offline events on online discussions, researchers mainly use online communities as a tool and focus on specific events, like natural disasters or social movements [25, 68, 73]. In the context of sports matches, which may have a continuous influence on user behavior, prior research focused on the high-level description of the community and analyzed it from the aspect of user activity, fan loyalty, or community conflict [82, 83]. Unlike previous work, this work contributes a more comprehensive understanding of offline contexts' impacts on online fan behavior. The findings align with previous findings on the types of situations that lead to higher user engagement. Moreover, our findings provide a more comprehensive understanding of user behavior on emotion, content, and feedback.

6.1 Theoretical Implication

Understanding user behaviors under the effect of offline events is important to comprehend the benefits and potential threats of such kinds of online discussions.

On the positive side, online forum discussions present a pattern that helps uninformed users involved. It is essential to keep community users abreast on what is happening in the community. Prior research identified feeling connected to a larger community as one of the motivations of live tweeting, which is similar to the motivation of online communities [60]. Moreover, the format of live-tweeting is based on ongoing events in the TV series, which can also be regarded as an event-driven discussion, just like the sports community [3, 41, 72]. Therefore, we can integrate match broadcasting and online communities to create a new experience of watching live sports events. Current live match broadcasting mainly depends on the video platform, which lacks a mature discussion environment for online communities. New match viewing modes can be designed, based on the real-time online discussion platform, to combine the match viewing experience and online discussion experience. We can also further study how users behave in an integrated environment to inspire more interactive discussion ways. Via such methods, people can feel connected in the online community while watching a match. Meanwhile, after the match ends, for those users who do not watch the match in this case, the match's strong impact on the community helps these users get involved. According to the results of RQ1 in section 5.1.1 and RQ2 in section 5.2.2, the fans' emotions and discussion content vary widely after each match so that uninformed users can quickly grasp the key events in the match, which is something that cannot be understood only by watching the news or statistics. We can further help uninformed fans to understand the match process by applying some community design techniques which can be borrowed from previous research. Related work investigating designs to help understand online discussions in group chats or live tweeting proposed methods like annotating the media or discussion content [62, 81]. The characteristics of the distinct behaviors under different offline events' effects can be combined with these mentioned methods to facilitate the user connection inside the community.

The significant effects of offline contexts on online discussion also have problems. It is inevitable for communities to have an inclined standpoint. When uninformed users regard the community as the main source of information about a match, some biased ideas caused by the mainstream position of the community will inevitably be instilled in these users. These biases may lead to an echo chamber or even extremism in the community. Our findings about the emotional behavior of fans (RQ1) and their tendency to vote for more emotional comments (RQ3) both prove the potential trend for extremism. However, it is difficult to say which one of emotional or neutral commenting is necessarily better. The passion itself is intrinsically part of the sports fan culture. It is arbitrary to define emotional fan behavior as unreasonable behavior. Nevertheless, there is no doubt that potential extremism can be a hidden danger that leads to conflict in online groups. Proper guidance or moderation might be needed to maintain the balance of neutral or emotional behaviors in such a fan community.

6.2 Practical Implication

Our findings also shed light on the design of online sports communities. As mentioned above, proper guidance is necessary for online sports communities to function properly. One of the guidance types in online communities is online community norms, which is essential to online community design [34]. Previous work has pointed out that different norms can effectively steer user behaviors into different patterns of similar types of communities [11]. Unlike other communities, the sports community should consider the impact of offline matches when designing community norms. In *r/RedDevils*, the official subreddit of Manchester United on Reddit, moderators include regulations concerning transfer rumors and brigading or trolling on other teams' subreddits in addition to common community norms. We can also derive some norms from our findings, like not excessively praising some players after a match win. However, the problem is that there is still no well-defined criteria to distinguish the boundary between praising and being excessively adulating, or criticizing and being abusive. While strong emotions are an essential part of sports fan culture, more work is needed to regulate emotions in forums to ensure that those emotional comments are not extreme.

Another significant part of community design is the moderation system. Kiene etc. interviewed several online community moderation teams on how technology brings about the challenging and critical sides of moderating online content at scale [39]. Several bots and APIs have been implemented to facilitate the moderation: "UB3R-B0T" for automatic word detection and filtering, "Dyno Bot" for managing the moderation logs, "ModMail" for mediating communication with community members [38, 39]. As suggested by our findings, we can integrate the offline context, which can be match events or other news, into these moderation systems as early warnings of possible infringement. Some straightforward features such as user activity, comments' sentiment, or discussion topics and indirect features like the "gap score" proposed by this work can also serve as indicators of a user's behavior pattern. They can either assist utilities the moderators can refer to or be part of an automated moderation system.

6.3 Limitations

This work investigated the football fan activities in online communities using the England Premier League as the primary data source. Thus, the scope is within one league in one country. Nevertheless, the football events were not limited to the EPL, England, or club matches. National team matches, such as the World Cup, may have more profound implications not only on the football fans but also on others who may not be faithful fans of football clubs, thus triggering more controversy in the field akin to political issues [69]. Still, the England Premier League was used as the subject of this study due to its popularity, relatively long duration, and high exposure to online social media. We tried to generalize the investigation to national team matches, but the discussion data on Reddit was disorganized. We also focused on the comments on match day instead of the whole season to find the implications of matches on fan behaviors. Results such as specific discussion topics may not be generalized to the discussion during other periods. Meanwhile, this work only focuses on one sport, one online community source (Reddit), and text comments. Future work can be expanded to other sports communities with different characteristics. For example, images and videos are a primary form of discussion and controversy in the Formula 1 subreddit. We also need to consider that the chosen time period in this work is during the pandemic when more and more people turn to online communities.

7 CONCLUSION

This paper examined post-match fan behaviors in online football (soccer) communities. The findings elaborated on the significant impact of team performance on fan behaviors in online communities, including the role of emotions and specific discussion topics. Patterns that users prefer to vote for emotional comments were explored to construct the landscape of post-match fan behaviors. This work also highlighted the significance of offline events having a significant impact on online fan performance, as well as implications for future online community design.

REFERENCES

[1] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*. 36–43.

[2] Edoardo M Airoldi and Jonathan M Bischof. 2016. Improving and evaluating topic models and other models of text. *J. Amer. Statist. Assoc.* 111, 516 (2016), 1381–1403.

[3] Riku Arakawa and Hiromu Yakura. 2021. Reaction or Speculation: Building Computational Support for Users in Catching-Up Series Based on an Emerging Media Consumption Phenomenon. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–28.

[4] Ethan Bauer. 2020. Is Social Media Ruining Sports? https://www.deseret.com/indepth/2020/2/21/21136415/sports-social-mediatwitter-toxic-sean-clifford-fans-nick-saban

[5] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 830–839.

[6] Colleen C Bee and Mark E Havitz. 2010. Exploring the relationship between involvement, fan attraction, psychological commitment and behavioural loyalty in a sports spectator context. *International Journal of Sports Marketing and Sponsorship* 11, 2 (2010), 37–54.

[7] Jonathan Bischof and Edoardo M Airoldi. 2012. Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. 201–208.

[8] David Blei, Andrew Ng, and Michael Jordan. 2001. Latent Dirichlet Allocation. In *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), Vol. 14. MIT Press. https://proceedings.neurips.cc/paper/2001/file/296472c9542ad4d4788d543508116cbc-Paper.pdf

[9] Venkateswarlu Bonta and Nandhini Kumaresh2and N Janardhan. 2019. A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology* 8, S2 (2019), 1–6.

[10] Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*. Vol. 4. Cambridge university press.

[11] Stevie Chancellor, Andrea Hu, and Munmun De Choudhury. 2018. Norms matter: contrasting social support around behavior change in online weight loss communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[12] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1217–1230.

[13] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2014. How community feedback shapes user behavior. In *Eighth International AAAI Conference on Weblogs and Social Media*.

[14] Te-Lin Doreen Chung, Olivia Johnson, Adrienne Hall-Phillips, and Kyuree Kim. 2021. The effects of offline events on online connective actions: An examination of# BoycottNFL using social network analysis. *Computers in Human Behavior* 115 (2021), 106623.

[15] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, Vol. 5. 89–96.

[16] Christine L. Cook, Jie Cai, and Donghee Yvette Wohn. 2022. Awe Versus Aww: The Effectiveness of Two Kinds of Positive Emotional Stimulation on Stress Reduction for Online Content Moderators. 6, CSCW2, Article 277 (nov 2022), 19 pages. https://doi.org/10.1145/3555168

[17] C Dewey. 2013. How twitter makes the political echo chamber worse. *Washington Post* 3 (2013).

[18] Peter Dolton[1] and George MacKerron. 2018. Is football a matter of life and death–or is it more important than that? (2018).

[19] Siying Du and Steve Gregory. 2016. The echo chamber effect in Twitter: does community polarization increase?. In *International workshop on complex networks*

and their applications. Springer, 373–378.
[20] Eric Dunning. 2000. Towards a sociological understanding of football hooliganism as a world phenomenon. *European journal on criminal policy and research* 8, 2 (2000), 141–162.
[21] Eric Dunning, Patrick J Murphy, and John Williams. 2014. *The roots of football hooliganism (RLE sports studies): An historical and sociological study*. Routledge.
[22] Brendan Dwyer. 2011. Divided loyalty? An analysis of fantasy football involvement and fan loyalty to individual National Football League (NFL) teams. *Journal of Sport Management* 25, 5 (2011), 445–457.
[23] Sheena L Erete. 2015. Engaging around neighborhood issues: How online communication affects offline behavior. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1590–1601.
[24] Minghui Fan, Andrew Billings, Xiangyu Zhu, and Panfeng Yu. 2020. Twitter-based BIRGing: Big Data analysis of English national team fans during the 2018 FIFA World Cup. *Communication & Sport* 8, 3 (2020), 317–345.
[25] Anjalie Field, Chan Young Park, Antonio Theophilo, Jamelle Watson-Daniels, and Yulia Tsvetkov. 2022. An analysis of emotions and the prominence of positivity in# BlackLivesMatter tweets. *Proceedings of the National Academy of Sciences* 119, 35 (2022), e2205767119.
[26] Tim Finin, Anupam Joshi, Pranam Kolari, Akshay Java, Anubhav Kale, and Amit Karandikar. 2008. The Information Ecology of Social Media and Online Communities. *AI Magazine* 29, 3 (Sep. 2008), 77. https://doi.org/10.1609/aimag.v29i3.2158
[27] Diana Fischer-Preßler, Carsten Schwemmer, and Kai Fischbach. 2019. Collective sense-making in times of crisis: Connecting terror management theory with Twitter user reactions to the Berlin terrorist attack. *Computers in Human Behavior* 100 (2019), 138–151. https://doi.org/10.1016/j.chb.2019.05.012
[28] R Kelly Garrett, Brian E Weeks, and Rachel L Neo. 2016. Driving a wedge between evidence and beliefs: How online ideological news exposure promotes political misperceptions. *Journal of Computer-Mediated Communication* 21, 5 (2016), 331–348.
[29] Wanqiu Guan, Haoyu Gao, Mingmin Yang, Yuan Li, Haixin Ma, Weining Qian, Zhigang Cao, and Xiaoguang Yang. 2014. Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. *Physica A: Statistical Mechanics and its Applications* 395 (2014), 340–351.
[30] Taylor M Henry and Thomas P Oates. 2020. "Sport Is Argument": Polarization, Racial Tension, and the Televised Sport Debate Format. *Journal of Sport and Social Issues* 44, 2 (2020), 154–174.
[31] Susan Herring, Kirk Job-Sluder, Rebecca Scheckler, and Sasha Barab. 2002. Searching for safety online: Managing" trolling" in a feminist forum. *The information society* 18, 5 (2002), 371–384.
[32] Matt Hopkins and James Treadwell. 2014. *Football hooliganism, fan behaviour and crime: Contemporary issues*. Springer.
[33] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.
[34] Koteswara Ivaturi and Cecil Chua. 2019. Framing norms in online communities. *Information & Management* 56, 1 (2019), 15–27.
[35] Jayboice. 2018. How our club soccer predictions work. https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work
[36] Jiwon Kang, Minsung Lee, Eunil Park, Minsam Ko, Munyoung Lee, and Jinyoung Han. 2019. Alliance for My idol: Analyzing the K-pop fandom collaboration network. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.
[37] Andrea Kavanaugh, John M Carroll, Mary Beth Rosson, Than Than Zin, and Debbie Denise Reese. 2005. Community networks: Where offline communities meet online. *Journal of Computer-Mediated Communication* 10, 4 (2005), JCMC10417.
[38] Charles Kiene and Benjamin Mako Hill. 2020. Who uses bots? A statistical analysis of bot usage in moderation teams. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–8.
[39] Charles Kiene, Jialun Aaron Jiang, and Benjamin Mako Hill. 2019. Technological frames and user innovation: Exploring technological change in community moderation teams. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
[40] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an" Eternal September" How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1152–1156.
[41] Jae Won Kim, Dongwoo Kim, Brian Keegan, Joon Hee Kim, Suin Kim, and Alice Oh. 2015. Social media dynamics of global co-presence during the 2014 FIFA World Cup. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2623–2632.
[42] Susann Kohout, Sanne Kruikemeier, and Bert N Bakker. 2022. May I have your Attention, please? An eye tracking study on emotional social media comments. *Computers in Human Behavior* (2022), 107495.
[43] Emily Kubin and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45, 3 (2021), 188–206.
[44] Robin Lakoff. 1973. The logic of politeness: Or, minding your p's and q's. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, Vol. 9. Chicago Linguistic Society, 292–305.
[45] Matthew S Levendusky. 2013. Why do partisan media polarize viewers? *American Journal of Political Science* 57, 3 (2013), 611–623.
[46] Mitzi Lewis. 2007. Stepwise versus Hierarchical Regression: Pros and Cons. *Online Submission* (2007).
[47] Peilei Liu, Jintao Tang, and Ting Wang. 2013. Information current in twitter: Which brings hot events to the world. In *Proceedings of the 22Nd International Conference on World Wide Web*. 111–112.
[48] Fiona Martin and Mark Johnson. 2015. More efficient topic modelling through a noun only approach. In *Proceedings of the Australasian Language Technology Association Workshop 2015*. 111–115.
[49] Wyl McCully, Cliff Lampe, Chandan Sarkar, Alcides Velasquez, and Akshaya Sreevinasan. 2011. Online and offline interactions in online communities. In *Proceedings of the 7th international symposium on wikis and open collaboration*. 39–48.
[50] David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing*. 262–272.
[51] German Neubaum, Leonie Rösner, Astrid M Rosenthal-von der Pütten, and Nicole C Krämer. 2014. Psychosocial functions of social media usage in a disaster situation: A multi-methodological approach. *Computers in Human Behavior* 34 (2014), 28–38.
[52] So Yeon Park, Nicole K Santero, Blair Kaneshiro, and Jin Ha Lee. 2021. Armed in ARMY: A case study of how BTS fans successfully collaborated to# MatchAMillion for black lives matter. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[53] Elazar J Pedhazur, Fred N Kerlinger, et al. 1982. *Multiple regression in behavioral research*. Holt, Rinehart, and Winston.
[54] Bernd Ploderer, Steve Howard, and Peter Thomas. 2008. Being online, living offline: the influence of social ties over the appropriation of social network sites. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 333–342.
[55] Emma Poulton. 2007. 'Fantasy football hooliganism'in popular media. *Media, Culture & Society* 29, 1 (2007), 151–164.
[56] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society* 7, 2 (2021), 20563051211019004.
[57] Margaret E. Roberts, Brandon M. Stewart, and Dustin Tingley. 2019. stm: An R Package for Structural Topic Models. *Journal of Statistical Software* 91, 2 (2019), 1–40. https://doi.org/10.18637/jss.v091.i02
[58] Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. Structural topic models for open-ended survey responses. *American journal of political science* 58, 4 (2014), 1064–1082.
[59] Maria Y. Rodriguez and Heather Storer. 2020. A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data*. *Journal of Technology in Human Services* 38, 1 (2020), 54–86. https://doi.org/10.1080/15228835.2019.1616350 arXiv:https://doi.org/10.1080/15228835.2019.1616350
[60] Steven Schirra, Huan Sun, and Frank Bentley. 2014. Together alone: Motivations for live-tweeting a television series. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2441–2450.
[61] Andreas Schwarz. 2012. How publics use social media to respond to blame games in crisis communication: The Love Parade tragedy in Duisburg 2010. *Public Relations Review* 38, 3 (2012), 430–437.
[62] David A Shamma, Lyndon Kennedy, and Elizabeth F Churchill. 2009. Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM workshop on Social media*. 3–10.
[63] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.
[64] Xin Shuai, Xiaozhong Liu, Tian Xia, Yuqing Wu, and Chun Guo. 2014. Comparing the pulses of categorical hot events in twitter and weibo. In *Proceedings of the 25th ACM conference on Hypertext and social media*. 126–135.
[65] Ramón Spaaij. 2006. *Understanding football hooliganism: A comparison of six Western European football clubs*. Amsterdam University Press.
[66] Hari Sundaram, Yu-Ru Lin, Munmun De Choudhury, and Aisling Kelliher. 2012. Understanding community dynamics in online social networks: a multidisciplinary review. *IEEE Signal Processing Magazine* 29, 2 (2012), 33–40.
[67] Matt Taddy. 2012. On estimation and selection for topic models. In *Artificial Intelligence and Statistics*. PMLR, 1184–1193.
[68] Emma F Thomas, Nathan Leggett, David Kernot, Lewis Mitchell, Saranzaya Magsarjav, and Nathan Weber. 2022. Reclaim the Beach: How Offline Events Shape Online Interactions and Networks Amongst Those Who Support and Oppose Right-Wing Protest. *Studies in Conflict & Terrorism* (2022), 1–23.

[69] Alan Tomlinson and Christopher Young. 2006. *National identity and global sports events: Culture, politics, and spectacle in the Olympics and the Football World Cup.* Suny Press.

[70] Abir Troudi, Salma Jamoussi, Corinne Amel Zayani, and Ikram Amous. 2019. Multidimensional analysis of hot events from social media sources. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing.* 2112–2119.

[71] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)* (2018).

[72] Marisa Vasconcelos, Jussara Almeida, Paulo Cavalin, and Claudio Pinhanez. 2019. Live It Up: Analyzing Emotions and Language Use in Tweets during the Soccer World Cup Finals. In *Proceedings of the 10th ACM Conference on Web Science.* 293–294.

[73] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 1079–1088.

[74] Hanna M Wallach, Iain Murray, Ruslan Salakhutdinov, and David Mimno. 2009. Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning.* 1105–1112.

[75] Isaac Waller and Ashton Anderson. 2021. Quantifying social organization and political polarization in online platforms. *Nature* 600, 7888 (2021), 264–268.

[76] Samuel M Wilson and Leighton C Peterson. 2002. The anthropology of online communities. *Annual review of anthropology* (2002), 449–467.

[77] Siqi Wu and Paul Resnick. 2021. Cross-Partisan Discussions on YouTube: Conservatives Talk to Liberals but Liberals Don't Talk to Conservatives. In *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media*, Vol. 15.

[78] Moran Yarchi, Christian Baden, and Neta Kligler-Vilenchik. 2021. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication* 38, 1-2 (2021), 98–139.

[79] Michael Yeomans, Alejandro Kantor, and Dustin Tingley. 2018. The politeness Package: Detecting Politeness in Natural Language. *R Journal* 10, 2 (2018).

[80] Brett Young, Hirotoshi Takeda, and Michael J Cuellar. 2011. Investigating the impact of offline events on group development in an online sports community. (2011).

[81] Amy X Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.

[82] Jason Shuo Zhang, Chenhao Tan, and Qin Lv. 2018. " This is why we play" Characterizing Online Fan Communities of the NBA Teams. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.

[83] Jason Shuo Zhang, Chenhao Tan, and Qin Lv. 2019. Intergroup contact in the wild: Characterizing language differences between intergroup and single-group members in NBA-related discussion forums. *Proceedings of the ACM on Human-Computer interaction* 3, CSCW (2019), 1–35.

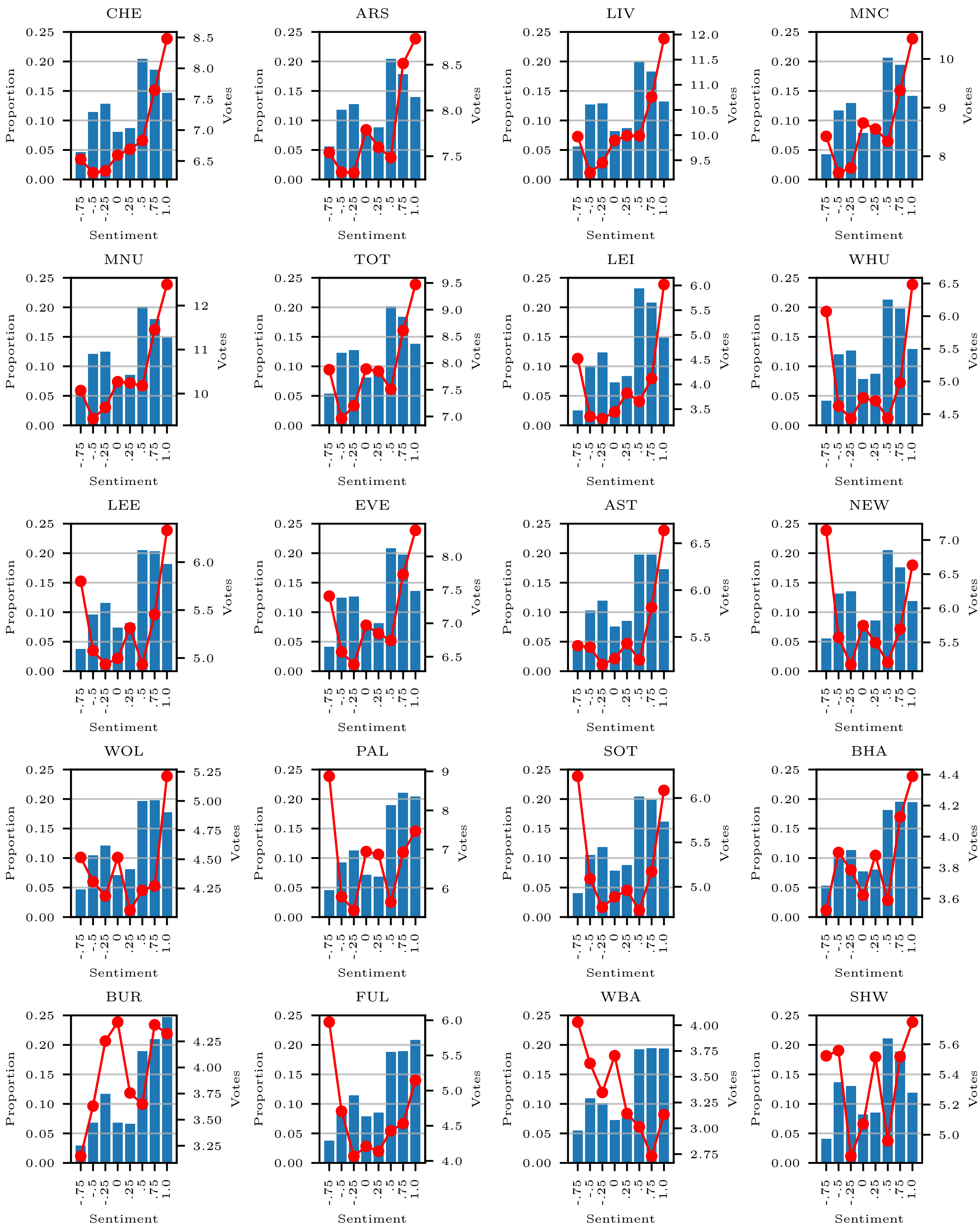

Figure 6: Each teams' distribution of comments based on their sentiment and the mean votes received. The blue bars represent the proportion of comments in each interval of sentiment scores. The red lines represent the average number of votes the comments received. The index k refers to the sentiment interval between $[k - 0.25, k)$.

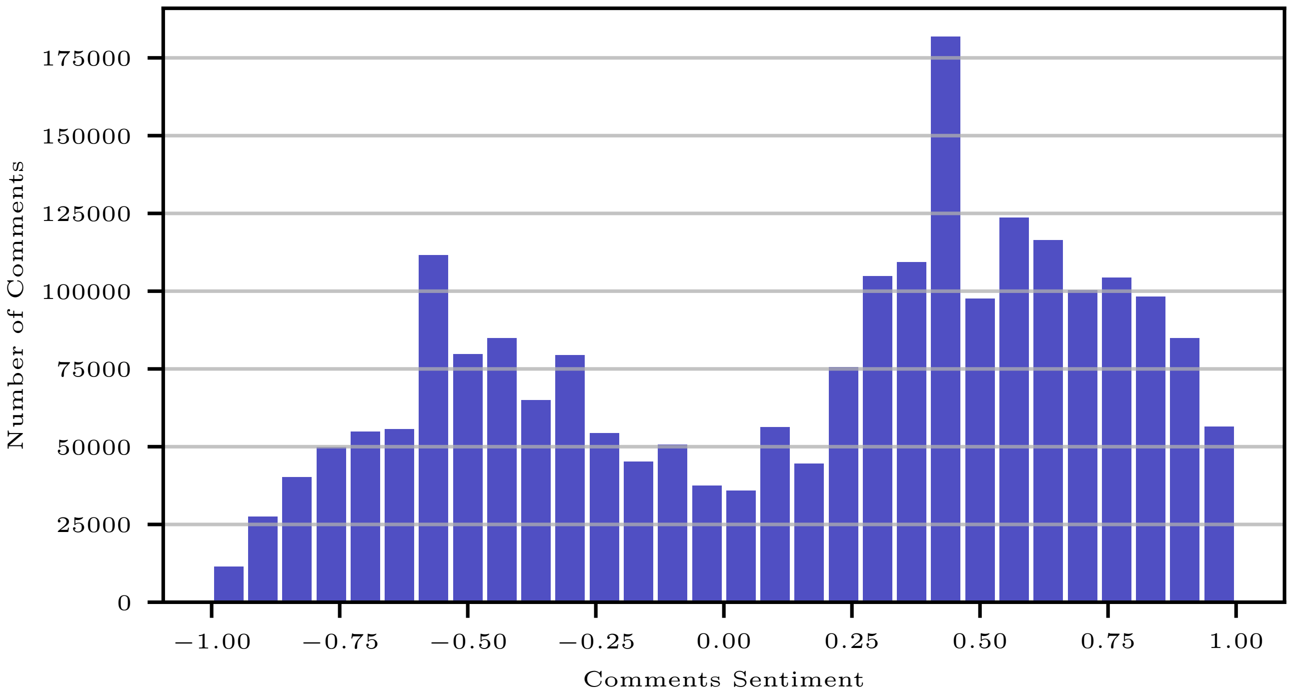

Figure 7: The figure shows the distribution of comments according to their sentiment score. The positive and negative comments, respectively, show a pattern with data near the mean being more frequent in occurrence than data far from the mean.