

Beyond the Game: Comparing Political News Coverage and Twitter Discussions during the 2022 FIFA World Cup

SANDRINE CHAUSSON , YOUSSEF AL HARIRI 

University of Edinburgh, United Kingdom

AXEL BRUNS 

Queensland University of Technology, Australia

WALID MAGDY , BJÖRN ROSS 

University of Edinburgh, United Kingdom

The 2022 FIFA World Cup in Qatar was as much a political event as a sporting one, marked by controversies over the death of construction workers, concerns over LGBTQ+ rights, and allegations of corruption. Drawing on 573,927 tweets from UK-based accounts and 15,812 excerpts from 66 UK newspapers, we examine how political and apolitical attention to the competition was distributed across mainstream and social media, how it varied across various Twitter sub-networks, and how UK journalists navigated both arenas. Combining Topic Modelling, LLM-based classification, Social Network Analysis, and close reading, we find that political content accounted for 29.0% of tweets and 18.5% of news excerpts—a gap driven largely by Twitter discussion of issues only loosely tied to the tournament, particularly the Mahsa Amini protests in Iran. Consistent with Wright et al. (2015), we find that political talk routinely surfaced within seemingly apolitical conversations about teams, players, and matches—from debates over footballers’ working conditions to readings of Morocco’s run as a symbolic challenge to European footballing power. At the same time, political conversations on Twitter were quite segmented, with different sub-networks focusing on different political topics. Moreover, while the UK press emphasised LGBTQ+ rights and migrant worker

Sandrine Chausson: sandrine.chausson@ed.ac.uk

Date submitted: 2026-05-08

deaths, Twitter hosted more sustained debate about perceived bias in Western media coverage of Qatar, especially within a community of users with ties to Arab countries and Islam. These insights offer unusually direct empirical support for counterpublic theory. Journalists, meanwhile, mostly used Twitter to push content but also engaged in disputes with one another over bias, accuracy, and the propriety of attending the event, reflecting a shift toward more assertive personal positioning consistent with broader trends in journalistic branding. We close by arguing that the framework of online “third spaces”—apolitical online fan environments into which politics occasionally intrudes—fits the World Cup poorly: the tournament’s scale, commercialisation, and political salience make sustained apoliticism untenable, and a meta-debate over whether to “keep politics out of football” played out openly across both platforms. The concept may apply more productively to smaller, more bounded fan communities.

Keywords: *Third Spaces, World Cup, Political Talk, Digital Journalism, Counterpublic*

Introduction

Social media platforms such as X (previously known as Twitter) provide a space for users to react to world events, from political crises and natural disasters (Meraz and Papacharissi, 2013) to cultural and sporting events (Highfield et al., 2013). In the context of large sports competitions such as the FIFA World Cup, the most important football¹ competition worldwide, conversations online reflect social media users’ support for different teams or players (Fan et al., 2019) but also—to the extent that users voice support or criticism of host countries and national teams, or the overall staging of the event—soft power dynamics at play in such competitions (Meier et al., 2021). FIFA World Cups are also widely covered by the news. The predictability and frequency of games, their ultimate importance in determining teams’ dominance, the celebrity status of athletes, and the emotional engagement of football fans, provide plenty of opportunities for reporting. Moreover, the competition shines a spotlight on

¹We use the term “football” to refer to association football, or “soccer”. References to similar sports are always qualified, e.g. “American football”.

hosting and participating countries, and the culture, people, and government they represent, thus setting the stage for a more political form of coverage.

Finally, many journalists are active on social media, and the accounts of news organisations are among the most widely followed on most platforms (Baftiu and Dodds, 2023). In 2010, Hermida defined “ambient journalism” as a framework for journalists and their audiences who use social media as a source of stories and a news delivery medium (Hermida, 2010): like ambient music, the Twitter newsfeed is always on, always in the background, and draws increased attention and engagement from audiences as significant events unfold. In 2022, 25% of verified accounts belonged to journalists and news outlets and for ~83% of younger journalists, Twitter was the first or second social media platform they use the most in their job (Twitter News, 2022) ².

The result is a complex interplay between the news and social media. Professional journalists retain an important role in curating information and shaping public opinion (Baftiu and Dodds, 2023; McGregor and Molyneux, 2018), while their audiences share—and thereby amplify—news and related information, critique journalistic coverage choices and crowdsource alternative voices to greater attention (Meraz and Papacharissi, 2013). On platforms such as Twitter, news coverage by professional journalists and news sharing and discussion by ordinary users are therefore both intertwined and potentially divergent, as users engage with professional news coverage as gatewatchers and news curators (Bruns, 2018) in ways that will not always align with the editorial choices of news organisations.

Potentially, this contrast in views and interpretations was especially pronounced during the 2022 FIFA World Cup, which took place in Qatar and was covered variously by political, sports, and entertainment journalists and attracted highly engaged and knowledgeable football fan communities, but also more politically-minded audiences. After all, while Qatar was the first Middle Eastern country to ever host the competition, its nomination was marred by allegations of corruption (Brannagan and Reiche, 2022), reports of exploitative conditions for migrant construction workers, some of whom died in infrastructure projects related to the World

²Such patterns may have changed substantially in recent years, following the acquisition and radical reshaping of Twitter by Elon Musk. Here, however, we address Twitter in the context of an event which occurred before the substantial exodus of journalists and ordinary users from the platform.

Cup (Amnesty International, 2019) and concerns about the country’s treatment of women, LGBTQ+ minorities, and other vulnerable groups. The choice of Qatar as a host was only made more controversial by the need to move the competition to November to accommodate Qatar’s hot climate, and the perceived incompatibility between alcohol-heavy football fan cultures and the core place of Islam (and sobriety) in Qatari culture and law. From this, we formulate our overarching research question:

In the context of the 2022 FIFA World Cup in Qatar, how much attention was given to the competition itself and how much to the politics surrounding it?

This question is itself divided into the following sub-questions:

- **RQ1:** What political aspects of the 2022 FIFA World Cup were discussed on Twitter and in the news respectively?
- **RQ2:** How were political discussions distributed across different groups of Twitter users?
- **RQ3:** How did journalists contribute to both political and apolitical conversations on Twitter during the World Cup?

In this paper, we address these questions with a corpus of ~570K tweets from UK-based Twitter users and ~16K UK newspaper excerpts. Our focus on the UK is motivated by the fact that, not only is football the country’s most-watched sport (Fleck, 2022), but British football following also widely exceeds that of other English-speaking countries after adjusting for population (Statista Research Department, 2024). Our analysis relies on the TwiXplorer tool (Al Hariri et al., 2024) for semantic analysis and Social Network Analysis, an LLM-based prompting approach for tweets classification, and close reading of specific tweets and user profiles to complement our more quantitative results.

Our findings show that, while political content was more prominent on Twitter than in UK newspapers (i.e. ~29.0% vs. ~18.5%), the difference can mostly be attributed to issues that do not directly relate to the World Cup itself or Qatar’s hosting of the event (i.e. the Palestine conflict and protests in Iran). This content is much more elaborate than simple “hashtag hijacking”, however, and mirrors offline forms of activism at the World Cup, thus highlighting the competition’s role as a stage for political messaging. In general, political content on Twitter tended to be segmented, with different clusters of users focusing on different political issues.

While journalists on Twitter mostly engaged with sports coverage and used the platform to “push” journalistic content, some also used the platform to take position, debate with others and critique each other, with journalistic values such as bias and accuracy at the heart of these interactions. What emerges is a process of collective sense-making, where fundamentally opposed interpretations of the World Cup—as either “sportswashing” or an opportunity to break stereotypes surrounding middle-eastern countries—can be seen to have collided. The latter was more present on Twitter, particularly among British users with ties to Arab countries and/or Islam, with sustained conversations around bias in Western media coverage of the event and accusations of double standards. These political interpretations of the World Cup were not always welcomed by football fans though, who occasionally pushed back on all forms of politicisation and urged others to “keep politics out of football”.

These insights contribute to various strands of existing literature. Firstly, they speak to debates about the relationship between online and offline activism, suggesting that Twitter’s amplificatory potential remains contingent on visibility established through physical, real-world action. Secondly, they advance our understanding of “political talk” (Wright et al., 2015) on Twitter, demonstrating that in the context of the 2022 World Cup political content was not marginalised by football-related content but often actively intertwined with it. Third, our insights reveal tensions and complementarities between institutional news media and social media platforms with respect to political content, demonstrating how racialised or religious minorities on Twitter formed a counterpublic that challenged dominant framings in the mainstream UK press. Relatedly, our analysis provides unusually concrete empirical evidence of the structural dynamics counterpublic theory describes (cf. Kaiser and Rauchfleisch, 2019), demonstrating through Social Network Analysis how distinct discursive communities engaged with the World Cup in relative isolation from one another. Fourth, our paper offers an original empirical and theoretical contribution to the literature on “third spaces”, arguing that the scale, commercialisation, and political salience of the FIFA World Cup makes it structurally incompatible with the apolitical dynamics that third space theory describes, and that the concept is more productively applied to smaller, more bounded fan communities. Finally, through its analysis of journalists’ tweets, the paper contributes to research on journalistic identity and social media practice, documenting an evolution towards more assertive forms of professional self-positioning on Twitter consistent with broader trends towards journalistic branding and individualisation (cf. Hanusch and Bruns, 2016; Hurcombe, 2024).

1. Related Work

1.1. *Sports and Political Talk*

Wright et al. (2015, p.74) point to the “everydayness” of political talk, where reflections on “power, its uses and ramifications” are often “interweaved with conversations that do not have a political character”. Large international competitions such as the Olympics or FIFA World Cups, often described as “sports mega-events”, perfectly illustrate this interweaving in the context of sports as they present opportunities for participating countries to showcase their public image, geopolitical standing, and touristic appeal (Meier et al., 2021; Grix, 2012; Knott et al., 2017). The case of the 2022 FIFA World Cup, which was hosted in Qatar, is particularly interesting. The strategic importance of the competition with respect to Qatar’s efforts to build a positive image (Dubinsky, 2023) and further its long-term economic and social aspirations (Scharfenort, 2012) are well documented and have often been described as “sportswashing” (Dubinsky, 2023).

1.2. *Political Discourse on Twitter*

Adapting earlier work by Oldenburg and Brissett (1982) on the concept of “third places”, i.e. informal public gathering spaces beyond the home or workplace such as barbershops, village pubs or community centres, Wright et al. (2015) point to some forms of online communities on social media platforms as “third spaces”: apolitical environments in which political talk can emerge when affected by political events. While Twitter as a whole is too large and varied to fit the definition of a third space, we speculate that third spaces may in effect have emerged on Twitter during the 2022 World Cup as football fans’ shared their experiences and impressions of the event. Determining whether the types of audiences and discourses really fit the definition of “third spaces” though requires a proper characterisation of the prominence and quality of political talk on Twitter during the event.

Although the majority of Twitter users are not primarily engaging on this platform for political reasons (Papacharissi, 2010), Twitter and in particular its hashtags provide a highly effective mechanism for the ad hoc assembly of large publics Bruns and Burgess (2015) engaging in political information sharing, debate, and activism. This has been evident in several high-profile cases from #Kony2012 (von Engelhardt and Jansz, 2014) through #MeToo (Trott, 2020)

to #BlackLivesMatter (Freelon et al., 2016), and various other smaller-scale contexts. When this happens, the duration and depth of engagement with political topics by ordinary and otherwise politically disinterested users reveals the extent to which a particular issue has been able to capture and hold their attention, and its focus—especially in contrast with journalistic coverage decisions—points to aspects of the issue which are particularly salient to the public.

The relationship between online and offline activism in these contexts is complex, however. While research consistently finds a positive association between social media use and political participation (Boulianne, 2018), online activist content on its own does not always translate into broader engagement. Instead, online mobilisation tends to be most effective when it connects to and amplifies offline organisational infrastructure and visible physical action (Bennett and Segerberg, 2023; Laer and Aelst, 2010), suggesting that digital and physical activism are better understood as complementary than as alternatives.

Past studies make valuable contributions on the topic of political discourse in the context of the 2022 FIFA World Cup on Twitter. For instance, Biswas et al. (2023) focused on conversations surrounding migrants' and workers' rights and facilities during the competition, while Hassan and Wang (2023) and Farrag et al. (2025) analysed political discussions on Twitter at scale through the lens of soft power and public opinion. However, they provide limited insight into the relative prominence of political discourse on Twitter compared to other types of content during the competition, into the distribution of political discourse across different sub-networks, or into the engagement of professional journalists and commentators, on the one hand, and “ordinary” football fans and Twitter users, on the other.

1.3. Traditional Media and Sports Mega-Events

While social media platforms enable new forms of communication, traditional news media outlets remain central actors in the media ecosystem. Their content is widely shared on social media, too, and continues to exercise considerable agenda-setting influence, though no longer exclusively (Brosius et al., 2019). A prominent question in the context of media and journalism studies is that of “newsworthiness”: what makes some stories more likely than others to attract media attention? In the context of the 2022 FIFA World Cup, several news values that determine newsworthiness may compete for attention. The football competition itself

may produce news that is predictable in scheduling, frequent in matches, familiar in themes, timely in outcomes, superlative in scope, elite in participants, and even positive in match results—though not usually for the English team. However, the surrounding political and human rights debate may emphasise conflict on various political issues, negativity about Qatari and FIFA policies, personalisation of affected communities, and consonance with existing attitudes about Qatar (Swart and Hussain, 2023). Journalism is likely to balance these considerations of newsworthiness in its coverage of sporting action at the World Cup itself and the political controversies surrounding it. And various journalistic specialisations (or newsbeats) such as sports, entertainment, and political journalism may make very different choices as they do so.

1.4. Interactions between Social Media and the News

While questions relating to political discourse and newsworthiness can be investigated in the context of the news and Twitter separately, contemporary journalism research shows important interconnections between these two forms of media (Chadwick, 2017; Hermida, 2010). Journalists use social media platforms to curate content, promote stories, or engage with audiences, among other purposes (Bruns, 2018). Audiences, on the other hand, follow, amplify, engage with, respond to, critique, and offer alternative views and sources in response to journalistic content. The stories and issues that audiences choose to interact with in turn reflect the “share-worthiness”—as opposed to newsworthiness—of this content (cf. Trilling et al., 2016). These mechanisms might lead to dynamics of convergence and divergence between UK newspapers’ coverage and Twitter users’ discussions of the 2022 FIFA World Cup.

Another key theoretical framework for interpreting these dynamics is that of counterpublics. Building on Habermas’ theory of the public sphere (Habermas, 1989) as a space for open and rational deliberation, Fraser (1990) points out that marginalised communities have historically often been excluded from dominant public spheres. As a result, she argues, these communities tend to form counterpublics: parallel discursive spaces where they develop their own identities, interpretations and interests. Expanding on this, Warner (2005) argues that counterpublics form through the circulation of discourse among its members and are constantly oriented outwards, seeking to challenge dominant narratives rather than simply retreat from them, and platforms like Twitter are crucial to the formation of counterpublics that can effectively challenge mainstream media framings at scale (Jackson et al., 2020).

Taken together, the literature reviewed above points to several gaps that this paper addresses. While existing studies of the 2022 FIFA World Cup on Twitter have examined specific political dimensions of the event, none has systematically compared the prominence of political versus apolitical content across both Twitter and the news, examined how political discourse was distributed across different sub-networks of Twitter users, or looked at the role of professional journalists specifically. More broadly, while the concepts of third spaces (Wright et al., 2015) and counterpublics (Fraser, 1990; Warner, 2005) have been applied to online political discourse, their applicability to the specific context of large-scale sports mega-events and their associated fandoms remains underexplored empirically. Finally, as Twitter has transformed into X, and data access for researchers has been terminated, scholars have been scrambling to document past activities both as historical records of what the platform once was and as analyses that may inform our understanding of the gap that the dismantling of Twitter has left in the social media landscape (Weller et al., 2026). In what follows, we will address these gaps by drawing on a corpus of approximately 570K tweets from UK-based accounts and 16K UK newspaper excerpts, combining semantic analysis, Social Network Analysis, LLM-based content classification and close reading of data.

2. Data Collection

We source our Twitter data from a larger multilingual dataset, which was collected using the Twitter API v2 between November 11th 2022 and January 18th 2023 from 181 keywords, including the name of the tournament (“*FIFA World Cup*”) in multiple languages and the (nick)names of participating teams in their respective local languages (see Appendix A for more information on the collection of this bigger dataset). For this study, however, we focus exclusively on a subset from this larger dataset containing English-language tweets (as labelled by Twitter itself) published between the November 20th and December 20th 2022, i.e. the active phase of the World Cup plus two days³, and matching the following subset of English-language keywords: “*world cup*”, “*world cup qualifiers*”, “*football*”, “*doha*”, “*fifa*”, “*qatar*”, “*khalifa international stadium*”, “*al janoub stadium*”, “*ahmed bin ali stadium*”, “*al bayt stadium*”, “*al thumana stadium*”, “*stadium 974*”, “*husail stadium*” and “*education city stadium*”. Applying these filters yields a total of ~15M tweets (before filtering, see below).

³The two days were added to account for latency between the last day of the World Cup and the publication of tweets relating to it.

Moreover, we used the Lexis+ platform⁴ to download a comprehensive list of 37,345 articles (again, before filtering) containing at least one of these keywords in either their body or their headline. These articles were published between November 20th and December 20th 2022 by 66 different UK newspapers, the top three of which are *The Sun* (England), *The Daily Mirror*, and *The Independent UK* (see Table 8 in Appendix A for a full breakdown of the dataset by newspaper).

3. Filtering and Pre-Processing

We carry out a number of filtering and pre-processing steps to improve the quality and focus of our Twitter and news data for this research. First of all, given the difference in length between tweets and newspaper articles, we transform all newspaper articles into excerpts composed of their headline and their first paragraph. This unit of text is the one used throughout our analysis, and provides enough context to reveal the subject matter and tone of each article while keeping them to relatively the same length as tweets. We then filter out excerpts and tweets that do not contain any of our collection keywords, and drop duplicates.

We observe that, beyond exact duplicates, our Twitter dataset also contains a lot of “copy pasta” content, i.e. near duplicates that are often the product of bots or a few coordinated accounts. While this content would be interesting to analyse in its own right, its analysis is beyond the scope of this study and we therefore filter it out using the multi-stage deduplication pipeline proposed by Manku et al. (2007). That is, tweets are first cleaned by removing URLs, mentions, and non-alphanumeric characters. We then generate 64-bit fingerprints using Simhash (Charikar, 2002), partition these fingerprints into four 16-bit bands, compare fingerprints sharing at least one identical band, and filter out tweets with fingerprints that differ by only 3 bits or less (i.e. Hamming distance less than or equal to 3).

Moreover, to bring our corpus of English-language tweets more in line with the UK focus of this study, we filter out tweets from non-UK accounts. While one solution to achieve this would be to only keep accounts that declared a UK location as part of their Twitter profile, only a small and arguably non-random minority of Twitter accounts declare a location in the first place. Instead, we therefore use an approach based on Social Network Analysis to identify

⁴<https://plus.lexis.com/uk/>

users that are likely not UK-based. That is, we leverage the Social Network Analysis (SNA) feature of the TwixPlover tool (Al Hariri et al., 2024), itself built on top of the Gephi toolbox (Bastian et al., 2009), to represent users as nodes connected by directed edges, where the weight of these edges is equal to the number of times user A has retweeted user B in our dataset. To reduce computational complexity, we only retain nodes with a weighted degree of 5 and edges with a weight greater than 1. We then use the Leiden algorithm (Traag et al., 2019) to identify clusters of densely connected accounts (see Appendix B for hyperparameter tuning).

Finally, we identify the distribution of locations declared by users in each cluster. Locations were retrieved using a dictionary method to map the free-form location descriptions to standardised country names⁵. For each cluster, if the most common location is the UK, then we flag the entire cluster as relevant to our analysis. If on the other hand the most common location is not the UK, then we only flag it as relevant if it contains at most twice as many accounts from this top location as UK accounts. The latter allows us to consider international and cross-national clusters which, although they are not strictly speaking UK-based, see a lot of engagement from UK accounts. Our filtered Twitter dataset then includes 1) all tweets from accounts having declared a UK location, as well as 2) tweets from accounts that have not declared a location but which belong to one of the clusters we flagged as “relevant”. The filtered dataset omits all tweets from accounts having declared a non-UK location and from accounts without a location but that belong to “irrelevant” clusters. Figure 1 below summarises the location-based filtering strategy for Twitter accounts.

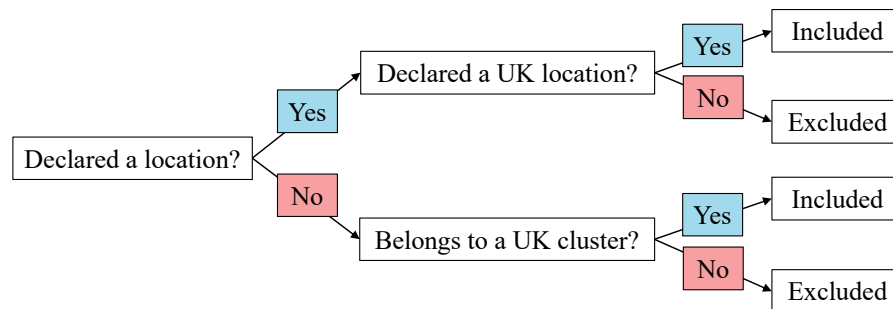


Figure 1. Twitter accounts location-based filtering strategy

⁵While we take users’ location at face value, note that this information can be faked.

We validate this approach by inspecting the most common words used in user descriptions for each cluster, their 10 most retweeted accounts, and 10 random accounts (see Table 9 in Appendix B for details). Two “relevant” clusters dominate the network: 1) a UK-based “Football” cluster, structured around football clubs, football leagues, and football news and statistics, and 2) a more international “World Cup” cluster structured around the official accounts of the FIFA World Cup and various national teams. Other “relevant” clusters are structured around accounts with ties to Arabic countries and/or Islam (“Arabic/Muslim”); Iranian nationals living abroad (“Iranian”); mainstream news media accounts and news aggregators (“News media”); and an EA “FIFA” gaming community (“Gaming”). “Irrelevant” clusters, on the other hand, bring together accounts from North America (“American”, “K-pop”, “Pop culture”, “Collectibles”); East Asia (“K-pop”); South and South-East Asia (“Indian”, “Tech”, “NFT/Crypto”); and Africa (“NFT/Crypto”) (see Table 1).

Table 1: Number of accounts per cluster with labels and descriptions. Clusters highlighted in red were filtered out based on our location criteria. Only clusters with more than 10K accounts are shown.

Cluster	Label	Description	Nb accounts
1	Football	Football leagues, clubs, stats and news	187,078
2	World Cup	FIFA World Cup official account, national teams	162,027
3	American	American football & soccer, US national team, American sports news	67,637
4	K-pop	K-pop fans accounts from US and East Asia	60,026
5	Arabic/Muslim	Account with ties to Arabic countries, Islam, Palestine activism	58,726
6	NFT/Crypto	Football-themed NFT and crypto promotional accounts	52,698
7	Iranian	Iranian diaspora	41,361
8	News media	Mainstream news accounts and aggregators	38,591
9	Indian	Indian football/cricket/sports fans	35,599
10	Pop culture	Pop culture, music, entertainment accounts	26,710
11	Gaming	FC Ultimate Team cards trading, promotion, fan accounts	25,438
12	Tech	Promotional pages for tech brands sponsoring World Cup	16,796
13	Collectibles	Sports cards promotional accounts	14,573

Lastly, we run an additional language check on our remaining Twitter data and filter out tweets that contain less than 5 tokens after removing URLs, hashtags and mentions, as these are unlikely to contain enough meaningful information for our analysis. After applying these various pre-processing and filtering steps, our final dataset contains 15,812 news excerpts and 573,927 tweets respectively (see Appendix C for a breakdown of volume per filtering step).

4. RQ1: Twitter and News Coverage

During the 2022 FIFA World Cup, what did UK newspapers report on, and in what manner, compared to UK users on Twitter? We address this question using Topic Modelling and political content classification.

4.1. Topic Modelling

First, we use the *all-mpnet-base-v2* SBERT model⁶ (Reimers and Gurevych, 2019) to encode tweets and news excerpts as 384-dimensional embeddings, and then run K-means clustering on these embeddings with an artificially high value of K, i.e. K=300. For each cluster, we then randomly sample 30 datapoints (i.e. tweets or excerpts), identify the 30 most common words in the cluster, and prompt an LLM (Gemini 2.5 Flash Lite) to generate a short (fewer than 5 words) description of each cluster from these top words and random sample (see prompt in Table 2). We use this information (top words, sample tweets and LLM-generated descriptions) to manually label all 300 clusters based on the topic they represent. The set of labels is iteratively revised to reflect the desired level of abstraction and to highlight the categories that are of particular interest to our analysis⁷. This step also allows us to validate the quality of the clusters, which we judge to be very high: i.e. most clusters are very coherent, with the exception of a few clusters which mixed very different topics and which we grouped together as a single “Other” topic. As expected given the high value of K, topics are often duplicated across clusters but these are simply merged during the iterative labelling step.

This approach, which can essentially be described as “hypersegment/merge”, has the advantage of leveraging the computational efficiency of K-means without relying on the clustering algorithm itself or automated hyperparameter-tuning metrics⁸ to identify an “objectively” ideal partition. Instead, the “ideal” partition is identified by us, the researchers, based on our research objectives yet informed by the actual content of our data (for top words, descriptions and volumes for each topic see Appendix D). This approach was preferred to using HDB-

⁶<https://huggingface.co/sentence-transformers/all-mpnet-base-v2>

⁷For instance, while we merged clusters that discuss individual players (e.g. Lionel Messi, Kylian Mbappe, Cristiano Ronaldo, Harry Kane, etc.) into a single topic “Players”, we decided to keep mentions to Qatar as a country, as the host of the World Cup, or as a national team separate.

⁸Silhouette scores, Davies-Bouldin indices and Calinski-Harabasz indices did not reveal a clear elbow and tended to disagree with each other.

SCAN, another clustering algorithm that works well on large datasets, as the latter produced a large proportion of “unclustered” data points in our experiments which was difficult to reduce through hyper-parameter tuning, and which included the majority of the news data.

Table 2: Prompts used to generate descriptions of semantic clusters and to classify tweets and excerpts based on whether they are political using Gemini 2.5 Flash Lite.

Task	Prompt
Semantic cluster description	<p>CONTEXT: I have data related to the 2022 FIFA World Cup. I embedded the data using SBERT and then clustered it. Here is a sample of data from a single cluster. More specifically:</p> <ul style="list-style-type: none"> - “Top words” contains the 30 words from the cluster with the highest TF-IDF scores; - “Random docs” contains a random sample of 30 tweets from the cluster. <p>TASK: Return a label of AT MOST a 5 words that summarises/describes the topic of the cluster.</p> <p>NOTE: Topic labels need to be specific, e.g. more specific than “football”, “politics” or “FIFA World Cup”.</p> <p>CONSTRAINTS: Do NOT return any other text than the label itself (i.e. no explanation, no comments, no formatting).</p>
Political content classification	<p>CONTEXT: Here are [tweets/news excerpts] related to the 2022 FIFA World Cup.</p> <p>TASK: Classify [tweets/excerpts] as either “Political” or “Apolitical”. Political tweets discuss power, its uses and its ramifications, e.g. in the context of:</p> <ul style="list-style-type: none"> - Governments, political interest groups, lobbies, regulating bodies, political/economic elites - Ideology, religion, partisan issues - Laws, regulations, policies - Power, money, influence - Activism, social movements - International relations, geopolitics, armed conflict - Minority rights, civil rights, human rights - Inequality or discrimination based on protected characteristics (e.g. gender, race, disability, sexual orientation, religion) <p>CONSTRAINTS: Return all labels on a single line separated by commas in the same order as the corresponding input [tweets/excerpts]. Do NOT return any other text than the comma-separated labels (i.e. no explanation, no comments, no formatting).</p>

This analysis revealed a number of “noisy” clusters, which we discard from the rest of our analysis. These include content related to other sports (e.g. American football, Australian football, Gaelic football, or cricket); remaining “copy pasta” and non-English content that was not caught by our filtering pipeline; coincidental mentions of Qatar or football; and commercial content (e.g. ads for cryptocurrencies, NFTs, betting platforms, football merch and equipment⁹). Figure 2 shows the resulting clustering in 2-dimensions.

⁹While the presence of this content in our dataset is interesting, highlighting the commercial nature of the event, its treatment is beyond the scope of this paper.

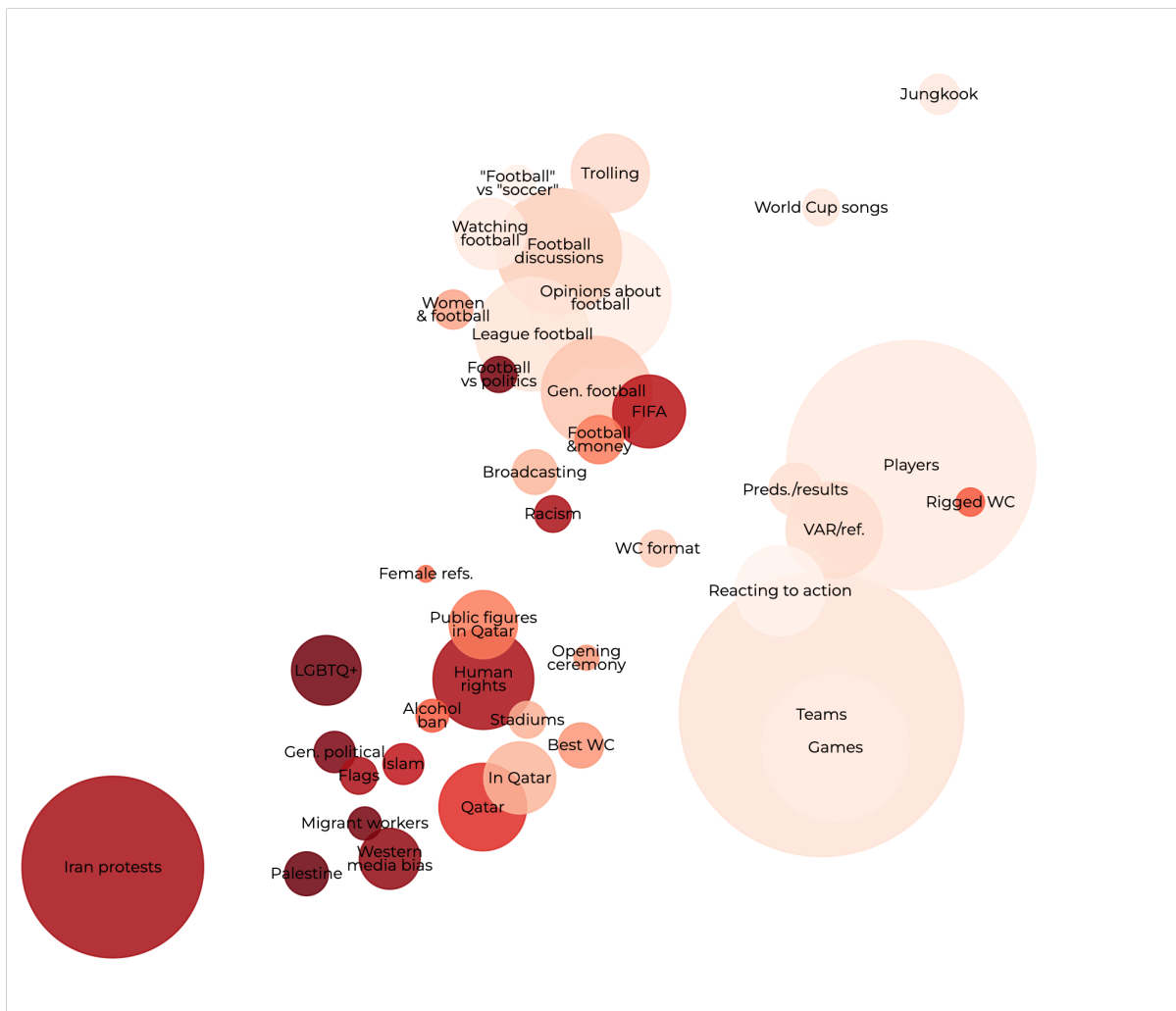


Figure 2. Projection of topics from news and Twitter corpora in semantic space. The deeper the red colour of the topic, the higher its proportion of political content, as determined using political content classification. The topic's position was calculated from its centroid in two-dimensional space, and its area is proportional to its prevalence. Topics that appear closer together are semantically more similar.

Many of the topics that emerge from this analysis relate to **football** (65.3%). Namely, we find discussions of different aspects of the World Cup competition itself, such as national teams (19.6%), players (15.9%), games (6.2%), predictions and results (0.7%), or reactions to action on the field such as referee decisions, goals or penalties (2.1%). We also find a

number of topics relating to football beyond the World Cup context. These cover discussions of British football clubs and league football (30.9%), people expressing their love (or otherwise) of watching or playing football (4.8%), sharing their football knowledge or challenging that of others, or making fun of Americans for calling football “soccer” (1.6%).

Another set of topics relate to the **World Cup** itself as an event (5.7%), for instance in the context of the opening ceremony, particularly Jungkook’s performance of “Dreamers” (0.4%), World-Cup-related events and attractions in Qatar such as fan zones (1.3%), the quality of the World Cup infrastructure and logistics (0.3%), celebrities and public figures attending the competition (1.2%), the quality (or rather lack thereof) of the World Cup broadcasting (0.5%) and the (un)availability of alcohol in stadiums (0.3%).

Finally, a third category of topics is much more **political** in nature. These include the corruption of the FIFA organisation (1.4%), racist football fan cultures (0.3%), Islam (0.4%), Qatar’s stance on LGBTQ+ rights (1.2%), and its controversial human rights record (2.6%), particularly in the context of the death of migrant workers during the construction of World Cup stadiums (0.3%). Almost as a direct reaction to the latter, however, we also find accusations of bias in Western media covering the World Cup (1.0%), whose scrutiny over Qatar’s human rights record is denounced as a racist double standard in light of the record of previous World Cup hosts (e.g. Russia). Lastly, our topic modelling also reveals higher-level debates relating to the role of politics in football (0.3%): in short, should we “keep politics out of football”, or are the political implications of football too important to ignore, arguably more important than the game itself? Interestingly, we also find political topics that do not seem directly related to the World Cup, namely protests unfolding in Iran at the time of the World Cup in response to the death of Mahsa Amini in police custody (8.6%), as well as displays of solidarity with Palestine (0.5%).

4.2. Political content classification

While the very nature of this last category of topics is political, other topics cannot really be classified as completely apolitical. After all, as Wright et al. (2015) point out, “political talk” can emerge in contexts that are not strictly speaking political. To gain a more fine-

grained understanding of “political talk” in our data, we therefore classify a sample¹⁰ from each topic as either “Political” or “Apolitical” by prompting Gemini 2.5 Flash Lite (see Table 2 for prompts). More specifically, in line with recent work on the political classification of social media posts (Bidewell et al., 2026), we define “political” content broadly as relating to governments, political interest groups, lobbies, regulating bodies, political or economic elites, ideology, religion, partisan issues, laws, regulations, policies, power, money, influence, activism or social movements, international relations, geopolitics or armed conflict, or discrimination or inequality related to protected characteristics (e.g. gender, race, religion, etc.). The approach achieves a macro-average F1-score of 0.92 (0.93 for tweets and 0.90 for news excerpts), based on a random sample of 150 manually annotated datapoints for each corpus (300 in total¹¹). Results from these classification tasks are shown in Table 3¹².

Overall, coverage both in the news and on Twitter predominantly focused on the football competition, i.e. national teams, players and games (~58% and ~43% of the news and Twitter corpora respectively). The Moroccan and Argentinian teams received a lot of attention, as the first African team to reach the semi-finals and the winners of the competition respectively. Related to the latter, Lionel Messi was the most discussed player of the tournament by far, with claims that he is the “GOAT” (Greatest Of All Time) and that he “completed football” with his 2022 World Cup win. Saudi Arabia’s unexpected victory against Argentina in the group stage also attracted attention. The Qatari team, on the other hand, was mocked both by the news and Twitter users for being the first team to be eliminated from the tournament—a first for a host country.

¹⁰For each source (news vs Twitter), the sample contains 10% of the topic capped between 50 and 300 datapoints (topics with fewer than 50 datapoints were classified in their entirety). Topics with fewer than 5 datapoints were ignored.

¹¹One author annotated all 300 datapoints and another author annotated a subset of 100 datapoints from these 300 to validate inter-annotator agreement. The Cohen-Kappa was equal to 0.8.

¹²See Appendix E for a similar approach for fine-grained sentiment classification per topic.

Table 3: Frequency of topic and proportion of political content per source. Asterisks stand for the average proportion of political content across the news and Twitter (***: more than 75%; **: more than 50%; *: more than 25%). Positive ratios mean the metric is greater for the news than Twitter (i.e. more frequent, more political) and inversely.

	Topic	Frequency			Political		
		News	Twitter	Ratio	News	Twitter	Ratio
**	Female referees	0.2%	0.1%	3.60	55.6%	77.0%	-1.39
	League football	10.7%	3.2%	3.36	8.2%	13.3%	-1.62
*	Broadcasting and press	1.6%	0.5%	3.30	22.0%	34.1%	-1.55
**	Alcohol ban	0.9%	0.3%	3.27	42.0%	73.4%	-1.75
	World Cup format	1.0%	0.3%	2.96	14.0%	28.7%	-2.05
*	Football and money	1.3%	0.6%	2.13	44.0%	55.7%	-1.27
*	Stadium/infrastructure	0.7%	0.3%	1.89	34.0%	38.9%	-1.14
*	Public figures in Qatar	2.2%	1.2%	1.83	38.0%	55.0%	-1.45
	Teams	36.3%	21.0%	1.73	9.7%	16.3%	-1.69
***	LGBTQ+ rights	2.2%	1.3%	1.72	98.0%	99.0%	-1.01
***	Migrant workers deaths	0.4%	0.3%	1.44	96.0%	98.5%	-1.03
*	In Qatar	1.8%	1.4%	1.29	22.0%	35.3%	-1.61
***	Flags	0.4%	0.4%	1.12	82.0%	89.3%	-1.09
	Players	16.8%	16.5%	1.02	6.9%	6.0%	1.15
*	Women and football	0.4%	0.4%	-1.01	42.0%	43.9%	-1.05
	Games	5.4%	5.7%	-1.07	5.1%	7.3%	-1.43
*	Opening ceremony	0.1%	0.2%	-1.16	47.6%	23.0%	2.07
	World Cup songs	0.3%	0.4%	-1.21	9.3%	10.2%	-1.10
	Other	4.1%	5.0%	-1.22	12.4%	13.2%	-1.07
	General football content	2.5%	3.2%	-1.27	26.8%	19.5%	1.37
***	Racism	0.3%	0.4%	-1.28	92.7%	86.3%	1.07
**	Qatar	1.5%	2.0%	-1.36	70.0%	70.0%	-1.00
	VAR/ref. controversies	1.8%	2.5%	-1.37	14.0%	22.0%	-1.57
***	General political content	0.3%	0.5%	-1.67	100.0%	92.2%	1.08
*	Best World Cup	0.3%	0.5%	-2.16	56.8%	21.5%	2.64
	Football discussions	1.9%	4.2%	-2.17	24.0%	13.7%	1.76
***	Human rights	1.1%	2.7%	-2.55	94.0%	80.0%	1.18
***	FIFA corruption	0.5%	1.4%	-2.77	92.0%	72.3%	1.27
***	Islam	0.1%	0.4%	-5.10	69.2%	91.0%	-1.31
***	Western media bias	0.2%	1.0%	-5.11	89.7%	97.7%	-1.09
***	Football vs politics	0.1%	0.4%	-5.74	100.0%	94.5%	1.06
	Watching WC/football	0.2%	1.4%	-5.80	6.9%	5.6%	1.23
	Opinions about football	0.9%	5.4%	-6.02	8.8%	4.2%	2.09
***	Iran protests	1.2%	9.0%	-7.43	82.0%	94.7%	-1.15
***	Palestine	0.1%	0.5%	-8.54	100.0%	98.4%	1.02
	Predictions and results	0.1%	0.8%	-10.09	9.0%	7.3%	1.23
	Reacting to action	0.2%	2.2%	-11.64	3.6%	4.3%	-1.21
	Football trolling	0.1%	1.7%	-12.89	21.1%	17.0%	1.24
**	Rigged World Cup	0.0%	0.2%	-15.90		65.3%	
	Jungkook	0.0%	0.4%	-16.02		8.3%	
	“Football” vs “soccer”	0.0%	0.4%	-25.87		2.4%	

Although predominantly apolitical, all of these football-related topics occasionally became politicised (see Table 4 for paraphrased examples). In fact, the sheer volume of these topics means that, although only a minority of their content is political, they are often the greatest source of political talk in our data. For instance, Morocco’s victories were seen as a symbolic disruption to the power dynamic between Arab-Muslim and European countries. We see this in the context of Morocco’s defeat of Belgium (and ensuing reports of violence in Brussels) and Morocco’s highly anticipated yet ultimately disappointing semi-final game against France, Morocco’s former colonial occupier. On Twitter, the Argentinian team and Lionel Messi were the targets of allegations that the World Cup was “rigged” in their favour. Finally, Qatar’s abysmal performance in the competition only helped delegitimise the country’s selection as host of the 2022 World Cup.

Table 4: Paraphrased examples of tweets and news excerpts from seemingly apolitical football-related topics, i.e. Teams, Players and Games.

Topic	Paraphrased examples
Teams	Belgium is only 2nd in FIFA rankings because Gianni Infantino is from there.
	Morocco has united Africa and the Arab World!!! They’re the first African Muslim country to make it to a World Cup semifinal Honestly, how much did they pay Portugal to lose the match?
Players	An Mbappe vs Messi final would be lovely for their Qatari owners in the Qatar world cup, wouldn’t it?
	Harry Kane, God said you could either be a great football player or a politician. You picked football. So play football and shut up. The FIFA Mafia fixed the match for Messi
Games	Widespread vandalism and violence as Moroccan fans celebrate their win against Belgium France are used to all this pomp and attention, that’s why I am supporting Morocco tonight Qatar World Cup defeat proves there are some things in sport you can’t pay for

In parallel to the World Cup coverage, the news also dedicated a lot of coverage to league football and British football clubs (~11% vs. only 3% on Twitter). More so than people on Twitter, the news also discussed the appointment of female referees Stéphanie Frappart, Neuza Back, Karen Diaz and Salima Mukansanga to officiate men’s games at the World Cup as a historic first, the broadcasting of the event, alcohol bans, future editions of the World Cup format and money in the context of football. While discussed less frequently on Twitter, though, these topics all tended to be framed politically more often on Twitter than in the news. On the other hand, the news politicised the opening ceremony (by reflecting on the controversial circumstances which led to Qatar’s nomination) and the question of whether this World Cup is

“the best ever” (specifically as criticism of Gianni Infantino, president of FIFA, claiming that it is) to a greater extent than Twitter users. The news also engaged more politically with the higher-level debate of whether football is something to be loved or hated.

On Twitter, significantly more attention was given to Jungkook’s performance of his song “Dreamers” at the opening ceremony, something that can almost entirely be attributed to a dedicated and active Jungkook fan community on the platform. Moreover, people on Twitter used the platform in ways that are almost unique to its affordances as a social network: they shared their enthusiasm about football, reacted to action on the field, trolled each other based on football knowledge, made predictions or japed at Americans calling football “soccer”. Lastly, with the exception of LGBTQ+ rights and the death of migrant workers during the construction of World Cup stadiums, many of the most political topics (e.g. human rights, FIFA, Western media bias, Islam, the football vs. politics debate, racism) were discussed on Twitter more than in the news.

The overrepresentation of political content on Twitter compared to the news is particularly dramatic in the context of tweets relating to Palestine or the Mahsa Amini protests in Iran. While this content might easily be dismissed as unrelated noise, it highlights the role of the World Cup as an international stage, which activist groups both online and offline tried to use to attract attention to their cause. For instance, a large proportion of posts expressing solidarity with Palestine were in response to the waving of Palestinian flags in stadiums by members of the audience or by the Moroccan team after their victory against Spain. Tweets about the Iran protests, on the other hand, namely include accusations that the Iranian regime is using football for propaganda, that the Iranian team is “having a good time” in Qatar while Iranian civilians were being killed, praise to the same Iranian team for their symbolic refusal to sing the Iranian national anthem, criticism of FIFA for letting Iran participate in the tournament, or descriptions of victims of the Iranian regime with mentions of the football teams they supported.

Overall, from the frequency of each topic and its proportion of political content, we can estimate that 29.0% of our Twitter data is political, as opposed to only 18.5% of our news data. Setting aside the topic of the Iran protests however, which represents the largest chunk of political content in our Twitter dataset (9.0% out of the 29.0%), Twitter users and the news

were very aligned in terms of the topics they framed politically most often: these were national teams, human rights, Qatar, LGBTQ+ rights, and FIFA and players. It is worth noting that these quantitative results are contingent on our filtering strategy though, which excluded large amounts of “copy pasta” tweets related to both political and apolitical topics (e.g. Jungkook, Iran protests), and also on our decision to disregard commercial topics from the analysis (e.g. football merch and collectibles, betting, NFTs and crypto).

RQ2: Comparing across Twitter communities

We have so far treated Twitter as a monolith. The reality of communication on Twitter is very different though, with users’ experiences being shaped by the sub-networks they are embedded in. To gain a more accurate understanding of how political conversations were distributed across the platform, we therefore leverage cluster labels from our earlier Social Network Analysis (see “Pre-processing and filtering” section). Figure 3 shows the remaining (i.e. post-filtering) clusters from this analysis after using the Force Atlas 2 layout algorithm (Jacomy et al., 2014).

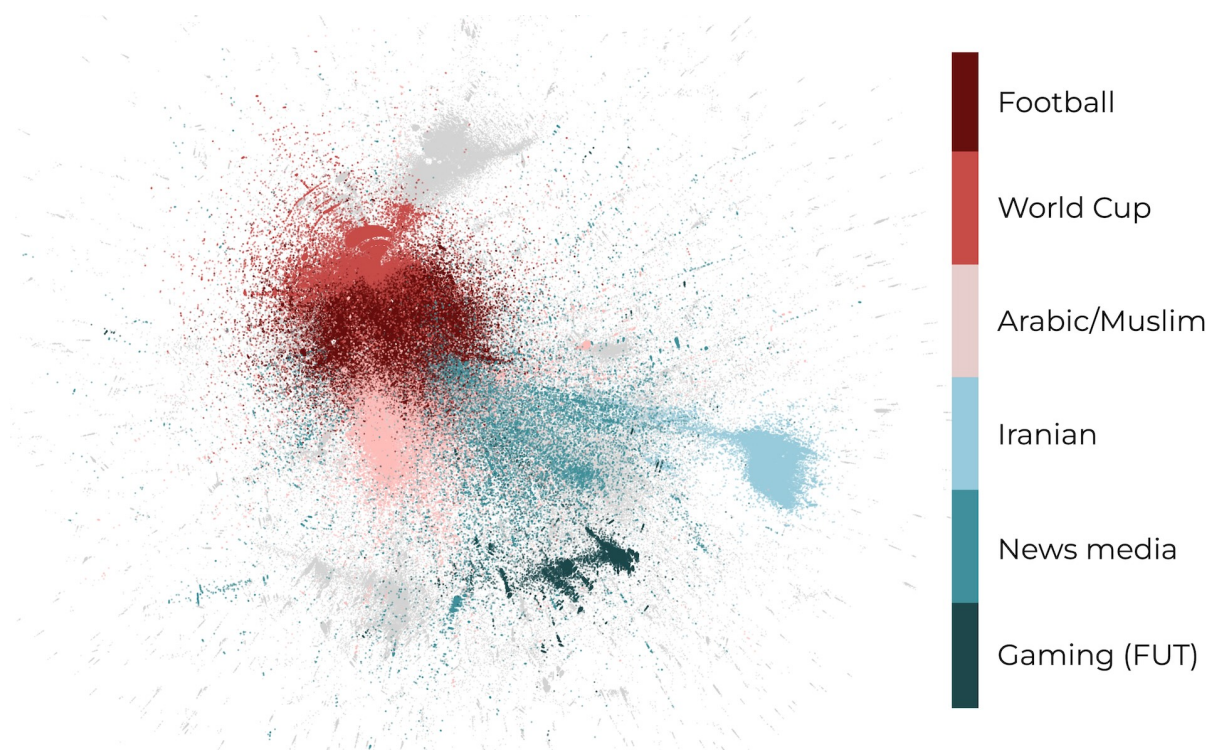


Figure 3. Social network of accounts in the Twitter dataset based on users’ retweets.

We calculate the prevalence of each topic per cluster by cross-referencing cluster labels with the topic labels from our Semantic Analysis. From this, we also calculate the (normalised) Shannon Entropy of each topic, where a value closer to 0 means that the topic is highly concentrated in one or a few clusters, while a value closer to 1 means that the topic is discussed across all communities (see Appendix F for equations). The prevalence of each topic per cluster and associated (normalised) entropy values are shown in Table 5.

Two topics stand out for their very low entropy: 1) protests in Iran, which were most discussed by the Iranian cluster, and 2) Palestine, which was most discussed by the Arabic/Muslim cluster. In other words, despite a large mobilisation on Twitter by the Iranian and pro-Palestinian communities, their efforts to raise awareness regarding the political and humanitarian situation in their respective countries by relating them to the World Cup via hashtag hijacking and comments on World Cup-related content, these issues largely did not capture the attention of the wider Twitter community discussing the World Cup.

Interestingly, political topics are systematically more community-specific than apolitical ones, as evidenced by their lower entropy scores: e.g. the death of migrant workers in the lead up to the World Cup (discussed by the News and Arabic/Muslim clusters mainly); Islam, the flying of rainbow and Palestinian flags in stadiums, and bias in Western media coverage of the World Cup (Arabic/Muslim cluster); the World Cup being rigged in favour of Lionel Messi (Football cluster); or female referees, LGBTQ+ rights and more general political content (News cluster). In fact, the only apolitical topic with an entropy lower than 0.85 is that of Jungkook's "Dreamers" performance, discussed most actively by a subset of dedicated fans from within the World Cup cluster. On the other end of our entropy scale, we mostly find apolitical football-related topics. A notable exception is the topic of human rights, which was discussed across all clusters. This might be due to the more high-level nature of this topic, which relates to all sorts of human rights issues of interest to different clusters (e.g. migrant workers, LGBTQ+ rights, Palestine, etc.).

Overall, the Iranian cluster was by far the most political, with 88% of its posts relating to politics, followed by the Arabic/Muslim cluster (37%), and the News cluster (34%). On the other hand, the World Cup (16%), Football (17%) and Gaming (19%) clusters were the least political.

Table 5: Prevalence of each topic t per cluster and (normalised) Shannon Entropy scores $H(t)$: a score close to 0 means t is discussed in few clusters, and close to 1 means it is discussed equally across clusters. Asterisks signify how frequently a topic was discussed politically on Twitter (***: more than 75%; **: more than 50%; *: more than 25%).

Topic	Frequency per cluster							H(t)
	Football	World Cup	Arab/Muslim	Iranian	News	Gaming	Others	
*** Iran protests	0.90%	1.45%	2.87%	83.79%	2.33%	2.05%	1.82%	0.30
*** Palestine	0.06%	0.08%	4.50%	0.25%	0.36%	0.01%	0.18%	0.37
Jungkook	0.10%	1.68%	0.17%	0.07%	0.12%	0.33%	0.41%	0.69
*** Migrant workers	0.14%	0.12%	0.54%	0.10%	1.35%	0.01%	0.25%	0.70
*** Islam	0.18%	0.20%	2.03%	0.58%	0.27%	0.04%	0.37%	0.72
*** Flags	0.15%	0.19%	1.75%	0.28%	0.52%	0.05%	0.25%	0.73
*** Western media bias	0.56%	0.29%	4.03%	1.15%	1.95%	0.28%	0.53%	0.78
** Rigged World Cup	0.36%	0.24%	0.08%	0.00%	0.04%	0.15%	0.08%	0.81
** Female referees	0.05%	0.08%	0.06%	0.00%	0.15%	0.00%	0.10%	0.81
*** Gen. political content	0.11%	0.10%	0.67%	0.84%	0.94%	0.03%	0.79%	0.82
*** LGBTQ+ rights	0.88%	0.60%	2.27%	0.53%	4.45%	0.59%	1.14%	0.83
Opening ceremony	0.12%	0.24%	0.31%	0.01%	0.15%	0.01%	0.17%	0.85
Best World Cup	0.68%	0.59%	0.94%	0.04%	0.27%	0.13%	0.28%	0.86
Predictions and results	0.64%	1.45%	0.43%	0.04%	0.29%	1.21%	0.75%	0.87
** Public figures in Qatar	0.93%	0.64%	1.74%	0.44%	3.07%	0.35%	1.45%	0.88
*** Football vs politics	0.34%	0.17%	0.62%	0.15%	0.83%	0.14%	0.23%	0.88
World Cup format	0.28%	0.39%	0.29%	0.02%	0.31%	0.09%	0.56%	0.88
League football	3.74%	1.28%	1.81%	0.13%	4.05%	4.88%	4.91%	0.89
* Stadium/infrastruct.	0.26%	0.24%	0.51%	0.05%	0.82%	0.18%	0.43%	0.89
* In Qatar	0.92%	0.89%	2.42%	0.39%	2.72%	0.40%	1.78%	0.89
*** Racism	0.29%	0.21%	0.78%	0.11%	0.66%	0.14%	0.34%	0.89
* Broadcasting/press	0.40%	0.24%	0.79%	0.09%	1.23%	0.49%	0.66%	0.89
** Alcohol ban	0.20%	0.16%	0.60%	0.07%	0.52%	0.16%	0.34%	0.89
Football and money	0.53%	0.30%	0.48%	0.07%	1.17%	1.01%	0.87%	0.90
VAR/ref. controversies	3.08%	3.01%	1.29%	0.05%	1.72%	3.53%	1.97%	0.90
Football discussions	5.67%	2.94%	2.99%	0.30%	3.73%	7.42%	3.18%	0.91
** Qatar	1.94%	1.17%	3.57%	0.84%	4.29%	0.89%	1.59%	0.91
Opinions on football	6.55%	7.37%	3.73%	0.25%	3.28%	6.65%	3.57%	0.91
** FIFA corruption	1.34%	1.03%	1.46%	0.59%	2.80%	3.80%	1.14%	0.91
Players	22.38%	16.29%	10.36%	0.87%	8.60%	14.98%	13.49%	0.92
Reacting to action	2.32%	2.99%	1.42%	0.13%	1.19%	2.25%	2.15%	0.92
“Football” vs “soccer”	0.39%	0.46%	0.35%	0.04%	0.25%	0.61%	0.23%	0.92
Games	4.78%	8.18%	5.00%	0.25%	5.44%	4.46%	6.98%	0.92
Watching football	1.54%	1.57%	0.87%	0.06%	1.04%	1.18%	1.32%	0.93
World Cup songs	0.26%	0.42%	0.29%	0.04%	0.43%	0.19%	0.50%	0.93
Football trolling	2.22%	1.23%	1.39%	0.27%	1.69%	2.80%	0.98%	0.93
* Women and football	0.42%	0.24%	0.34%	0.06%	0.66%	0.31%	0.45%	0.93
*** Human rights	1.78%	2.08%	6.31%	3.64%	5.06%	1.88%	1.74%	0.93
Teams	22.06%	25.20%	20.08%	1.54%	19.37%	16.77%	22.01%	0.94
Gen. football content	3.75%	2.89%	2.81%	0.48%	2.79%	3.04%	3.09%	0.96
Other	4.57%	6.51%	5.10%	1.06%	5.08%	5.53%	4.88%	0.96

6. RQ3: Journalists on Twitter

We have so far looked at the coverage of the 2022 FIFA World Cup on Twitter and in the UK written press independently. The question then is, how did UK journalists engage with both political and apolitical conversations on Twitter in the context of the 2022 FIFA World Cup? We address this question by focusing on users that declared a UK location and used any of the following terms in their Twitter profile description: “*journalist*”, “*reporter*”, “*editor*”, “*correspondent*”, “*writer*”, “*columnist*”, “*contributor*”, “*broadcaster*”, “*anchor*”, “*producer*”, “*photojournalist*” or “*presenter*”. These keywords were selected by manually identifying common keywords in user descriptions from an existing dataset of UK journalists’ Twitter accounts (Toprak et al., 2022), and then validating the quality of the matches on a subset of the sampled data (see Appendix G for more details). This approach gives us 123 accounts, for a total of 3,101 tweets.

We stress here that these are the descriptions chosen by users themselves in setting up their accounts; this necessarily means that professional journalists who chose not to identify themselves as such in their profiles will have been excluded from this selection of accounts, while the Twitter accounts of news influencers or “newsfluencers” (Hurcombe, 2024), including independent football commentators, professional influencers on YouTube, Substack, or other platforms, and various other para-journalistic actors will have been included if they identified themselves as ‘journalists’ or using any of the other descriptors we drew upon for this selection. Our selection is thus essentially self-nominated; this is deliberate and enables us to include not only professional journalists in a conventional sense, but also others who claim a journalistic role.

Using labels from our earlier SNA, we see that these journalist accounts are predominantly associated with the Football (29.6%) and mainstream News (15.4%) clusters. Another 10.3% are associated with the Arabic/Muslim cluster and 10% with the World Cup cluster. Only a very small minority are associated with the Iranian (1.5%) and Gaming (0.8%) clusters. The majority of journalist accounts though (32.4%) are not associated with any of the main clusters. To determine what journalist accounts discussed on Twitter during the World Cup, we inspect the topic labels generated in our earlier Topic Modelling analysis for the subset of tweets posted by journalists. Moreover, to determine how journalists approached these topics,

we repeat our earlier binary political content classification on a sample of their tweets. We complement these more quantitative insights with a close reading of tweets published by a random sample of 10 journalist accounts.

Similarly to the news and Twitter users in general, journalists on Twitter mainly discussed the FIFA World Cup tournament itself: that is, national teams, players and games. They made predictions about match outcomes, discussed results and preparation in the lead-up to the World Cup, and teams' updates before, during, or after games. For instance, journalists discussed the health of Karim Benzema and Christopher Nkunku, both of whom were excluded from the French team just before the start of the competition due to injury. While some journalists showed sympathy with the players, others demanded reforms prioritising players' health needs. In parallel to their World Cup coverage, journalists also posted a substantial amount about football personalities at UK Football Clubs (e.g. players, managers, coaches, and owners). Arguments arose about bids for the ownership of domestic football clubs like Manchester United FC and Newcastle United FC. Some journalists criticised government-funded bids, describing these as "sportswashing". Journalists on Twitter did not discuss league football as much as newspapers (5.9% vs 10.7%), however.

Journalists dedicated more attention than the news though to the FIFA organisation, Islam, the debate about politics vs football, and bias in the Western media. Indeed, some of the most significant and controversial tweets from our sampled UK journalists criticised the stances or claims of other news sources. The BBC attracted the most debate, following a claim by The Guardian that it boycotted the Qatar 2022 World Cup opening ceremony. Some journalists promoted that claim: for instance, Robert Carter from 5Pillars UK tweeted "The BBC has never boycotted a World Cup opening event until now". Others, however, refuted it, like the English broadcaster and former footballer Gary Lineker who tweeted that the opening ceremony "was shown live in its entirety on @BBCiPlayer, BBC Sport website and red button". Still, the frequency of the Western media bias topic in journalists' tweets remains small and less than among Twitter users in general. Journalists also tweeted substantially less than other Twitter users about protests in Iran, although some journalists acclaimed the symbolic silence of the Iranian team during their national anthem in response to the protests.

Table 6: Frequency of topic and proportion of political content in journalists’ tweets with comparisons to news and Twitter as a whole. Asterisks stand for the proportion of political content in journalist tweets (***: more than 75%; **: more than 50%; *: more than 25%). Positive ratios mean the metric is greater for journalists’ tweets than the other dataset (i.e. more frequent, more political than in the News/Twitter) and inversely.

Topic	Frequency			Political		
	Journalists	vs News	vs Twitter	Journalists	vs News	vs Twitter
Teams	23.9%	-1.52	1.14	8.6%	-1.13	-1.91
Players	15.8%	-1.06	-1.04	3.0%	-2.32	-2.02
League football	5.9%	-1.83	1.84	5.7%	-1.44	-2.33
Football discussions	5.9%	3.03	1.39	4.0%	-6.00	-3.42
Games	5.2%	-1.04	-1.11	5.8%	1.14	-1.25
Other	4.8%	1.18	-1.03	20.8%	1.69	1.58
Opinions about football	3.9%	4.31	-1.40	4.3%	-2.03	1.03
*** LGBTQ+ rights	3.6%	1.65	2.84	97.2%	-1.01	-1.02
General football content	3.1%	1.23	-1.04	7.5%	-3.56	-2.59
VAR/referee controversies	2.9%	1.59	1.17	10.5%	-1.34	-2.10
** Qatar	2.5%	1.67	1.23	58.7%	-1.19	-1.19
* In Qatar	2.4%	1.36	1.75	31.0%	1.41	-1.14
** Public figures in Qatar	2.3%	1.07	1.96	58.6%	1.54	1.06
Reacting to action	2.3%	12.36	1.06	2.9%	-1.25	-1.52
*** Human rights	1.7%	1.63	-1.57	80.8%	-1.16	1.01
** FIFA corruption	1.3%	2.54	-1.09	71.8%	-1.28	-1.01
*** Iran protests	1.3%	1.08	-6.86	82.1%	1.00	-1.15
Watching World Cup	1.2%	5.23	-1.11	0.0%		
Football trolling	1.2%	9.37	-1.38	5.6%	-3.79	-3.06
* Stadium/infrastruct.	1.0%	1.60	3.01	38.7%	1.14	-1.00
*** Western media bias	0.9%	4.60	-1.11	92.6%	1.03	-1.05
* Women and football	0.9%	2.26	2.25	37.0%	-1.13	-1.19
Broadcasting and press	0.8%	-2.01	1.64	20.8%	-1.06	-1.63
* Alcohol ban	0.6%	-1.43	2.29	27.8%	-1.51	-2.64
*** Islam	0.5%	6.09	1.19	81.3%	1.17	-1.12
*** Migrant workers deaths	0.5%	1.32	1.90	93.8%	-1.02	-1.05
* Football and money	0.5%	-2.70	-1.27	28.6%	-1.54	-1.95
* World Cup format	0.4%	-2.46	1.20	33.3%	2.38	1.16
* Best World Cup	0.4%	1.47	-1.47	27.3%	-2.08	1.27
*** Flags	0.4%	-1.08	1.03	81.8%	-1.00	-1.09
*** Football vs politics	0.4%	6.05	1.05	90.9%	-1.10	-1.04
World Cup songs	0.3%	-1.09	-1.32	0.0%		
Predictions and results	0.2%	2.70	-3.74	0.0%		
“Football” vs “soccer”	0.2%	12.36	-2.09	0.0%		
** Female referees	0.2%	-1.46	2.47	60.0%	1.08	-1.28
** General political content	0.2%	-1.62	-2.70	60.0%	-1.67	-1.54
*** Racism	0.2%	-1.66	-2.12	80.0%	-1.16	-1.08

Compared to other Twitter users, journalists posted more about controversial issues such as the death of migrant workers during the construction of stadiums, LGBTQ+ rights, women's rights (e.g. "Why are there no Qatari women watching the football?") as well as alcohol bans. They discussed the last-minute—and therefore ostensibly exceptional—nature of FIFA's decision to ban from stadiums both alcohol sales and the "OneLove" armband shortly before the start of the competition. Journalists spoke of "troubled times" for FIFA's sponsorship contracts with Budweiser and reflected on the early engagement of both the English and Welsh teams in the "OneLove" campaign, whether players taking the risk of sanctions from FIFA would actually help the LGBTQ+ cause, and the reactions of these players—particularly Harry Kane—to the ban. Moreover, journalists with a more dedicated interest in LGBTQ+ activism (e.g. Heather Paterson, LGBTQ+ columnist for Exposed Magazine) used their platform to share stories from Qatari LGBTQ+ people and activist organisations.

Qatar, as the host country, was the object of regular criticism. While some journalists argued that the World Cup was an opportunity for Qatar and Arab-Muslim countries to open their culture to the world, others argued that Qatar's desire to host was motivated by the desire to "sportswash" allegations of corruption and human rights abuse. Professional associations with Qatar were occasionally framed in a derogatory light, as seen in tweets like "Wouldn't be surprised if they live and work in Qatar" in response to perceived racism in a picture published by The Times. Some journalists boycotted Qatar altogether and criticised those who attended the event or worked with Qatari media outlets such as Al Jazeera or beIN Sports. A small number of journalists did not tweet about the World Cup except to criticise its political and social aspects. Some journalists advocated for reforms within FIFA to prevent future World Cup events being held in countries with contentious human rights records. Conversely, supporters of Qatar as a host country did not refute these claims outright, but rather asserted that the situation is not as bad as reported in the media. As demonstrated by the promoted documentary "Gary Neville in Qatar", some journalists contacted authorities in Qatar to address concerns regarding migrants' working conditions. Other journalists demanded moral consistency instead of stereotyping, pointing to the human rights records of previous and future FIFA World Cup host countries.

Overall, journalists seem to have dedicated more attention to the experience of attending the World Cup in Qatar than either the news or the rest of Twitter. This might reflect the

role journalists took during the World Cup as correspondents on the ground, which Twitter enabled in a more spontaneous and informal way than their official reporting work. This includes coverage of the tourist and football fan experience in Qatar, and the quality of stadiums—but also, perhaps relatedly, the death of migrant workers in the process of building this infrastructure—and the presence of rainbow and Palestinian flags in the audience. Journalists also commented on cultural aspects of the event, such as wearing local clothing, or incorporating national flags into outfits. We find tweets, both positive and more critical, about Lionel Messi wearing the traditional Qatari bisht, a men’s cloak popular in Arab countries (mainly in the Gulf, Levant, and Iraq). While this was interpreted as a sign of highest respect in the Arab world, it was criticised by some journalists as another form of sportswashing. All in all, ~20% of journalists’ tweets were political: lower than the 29% figure for Twitter users in general, but slightly higher than the 18.5% for news.

7. Discussion

Overall, British newspapers, Twitter users and journalists were mostly aligned in terms of how much they discussed the 2022 Qatar World Cup politically, with some differences between the aspects they focused on the most. While the news focused on LGBTQ+ rights and the death of migrant workers during the construction of World Cup infrastructure, users online discussed bias in Western media coverage of the competition to a much greater extent. The relative prominence of the LGBTQ+ topic in our Twitter and news corpora is particularly interesting in light of football’s fraught history of homophobia (Cashmore and Cleland, 2012), which successive campaigns in the UK have sought to address among other forms of discrimination (e.g. Stonewall’s Rainbow Laces campaign, Football v Homophobia, Kick It Out). This history might in part explain the difficulty that Qatar’s stance on LGBTQ+ rights posed to the UK football community: it came in contradiction to the industry’s recent efforts to vocally assert its commitment to positive change.

The most important difference between our two corpora though takes the form of political content on Twitter on issues that were not directly related to the World Cup, such as protests happening in Iran at the time of the World Cup. Interestingly, this content is much more elaborate than mere “hashtag hijacking”. Activists’ tweets formulate explicit links between the World Cup and the situation in Iran: they contrast the joyful atmosphere of the

World Cup to the suffering of the people of Iran, point to the political role football plays in legitimising the Iranian regime, or use football to humanise its victims, and build on top of offline forms of activism at the World Cup (e.g. anti or pro-regime flags, posters, t-shirts in the audience, or the Iranian team refusing to sing their national anthem). While these efforts might still have attracted some attention, our results show that the protests were only minimally discussed by other Twitter users. Moreover, when they did talk about Iran, this tended to be in response to offline rather than to online activism. One contributing factor to this though might have been the network dynamics of the Twitter platform itself, given the relative isolation of the Iranian activist accounts from other clusters in our data. This isolation might itself reflect the genuine “outsider” status of most of these accounts with respect to football fandoms on Twitter.

In fact, our Social Network Analysis and the Shannon Entropy measures derived from it reveal a more general pattern of “segmentation” in the context of political conversation on Twitter, with different clusters of users each focusing on different political topics. Content relating to LGBTQ+ rights or the death of migrant construction workers in Qatar in the lead up to the competition is mostly confined to a cluster of users that retweeted mainstream news accounts. Accusations of bias in Western media’s coverage of the World Cup and messages of solidarity with Palestine are predominantly posted by UK-based users with Arabic and/or Muslim cultural ties. Users most engaged with the Football cluster cared about the integrity of the World Cup as a sports competition, and denounced what they saw as a form of rigging in favour of certain teams. This segmentation lends empirical support to counterpublic theory’s conception of political discourse as constituted by a plurality of separate, parallel discursive arenas (Fraser, 1990; Warner, 2005; Jackson et al., 2020).

A large proportion of political conversations in our data relate to less obviously political topics, however, such as games, players or national teams. Intertwined with conversations on players’ health, ownership bids of football clubs, or the performance of national teams, we find reflections on footballers’ working conditions, on the propriety of governments’ implication in football, and on national histories of conflict. These results are very much in line with prior literature on “political talk” as a pervasive phenomenon that can emerge in seemingly apolitical contexts (Wright et al., 2015) and have concrete implications for research: i.e. that these contexts should not be disregarded if we are to understand contemporary democratic life

in its fullest sense. Moreover, political conversations we identified in the context of national teams' performance reflect the more symbolic geopolitical dimension of sports mega-events such as the World Cup. This resonates with a wider sociological literature on sports, and football in particular, as a space where national identities are reinforced and where power dynamics between nations are enacted (Hobsbawm, 1992; Bairner, 2001), and is in line with prior work on the politicization of FIFA World Cups specifically (Meier et al., 2021; Hassan and Wang, 2023; Swart and Hussain, 2023).

Overall, a significant amount of debate took place in the news and on Twitter on how to interpret Qatar's hosting of the World Cup. Should the competition be viewed as a blatant sportswashing attempt, an effort to sweep under the rug political and humanitarian accusations made against Qatar? Or should it be seen as an opportunity for the world of football to open up to Middle Eastern countries and for Western audiences to see past overwhelmingly negative stereotypes of these cultures? Indeed, rather than political accusations against Qatar being "eclipsed" by football-related content, our results show they were actively challenged, mainly through the counter-narrative of Qatar as the victim of racist double standards. This narrative was most prominent on Twitter, particularly within a cluster of users with ties to Arab countries and/or Islam, highlighting the platform's role in enabling the circulation of contestatory discourse among counterpublics (Warner, 2005; Jackson et al., 2020).

Given the central role of news organisations in both these interpretations, i.e. as platforms that either enable sportswashing or perpetuate prejudice against Middle Eastern countries, these different accounts of the World Cup played out most interestingly in the context of journalists' activity on Twitter. Journalists criticised each other for their decisions to attend (or not) the World Cup, their associations with Qatari organisations or their claims about Qatar, occasionally offering different interpretations of events. Issues of bias and accuracy were central to these interactions, and journalists occasionally used Twitter to clarify their views or those of their affiliated media outlet, or to refute allegations against them.

Such practices remain in line with earlier observations of journalists' loyalty towards and support for their home institution, in an effort to protect its integrity and promote its journalistic output (cf. Molyneux and Holton, 2014; Opgenhaffen and Scheerlinck, 2014). While these results might at least in part be influenced by institutional social media policies prohibit-

ing journalists from criticising their own institution, they seem to show that, in spite of the growing precarity and individualisation of journalistic employment, substantial attachment to established institutional structures remains.

However, these observations also advance somewhat beyond the cautiously professional strategies of social media and especially Twitter use that earlier studies had observed (Baftiu and Dodds, 2023), and perhaps reflect the evolution both of the journalistic profession as such and of its approach to social media as a component of its professional toolkit. As the need to establish a personalised journalistic brand (cf. Brems et al., 2016; Hanusch and Bruns, 2016) has become ever more important in an increasingly precarious industry, journalists are perhaps becoming more proactive in seeking out ways to establish and communicate their own professional identity, almost in the manner of influencers or “newsfluencers” (Olausson, 2017; Hurcombe, 2024). A more robust approach to arguing their point on Twitter would be a logical part of this journalistic branding strategy.

Finally, it is important to note that some commentators online and in the news pushed back on political conversations in the context of the World Cup altogether, and a distinct debate regarding the place of politics in football emerged. This debate often opposed “true” football fans, who advocated for politics to be kept out of football, against less football-focused individuals who denounced this dismissal of politics as short-sighted, selfish, or as a form of denial. Such debates do not simply indicate that spaces for football fandoms such as World Cup-related discussion on Twitter can act as third spaces in Wright et al.’s (2015) sense of the term—i.e. as primarily apolitical spaces for fandom communities into which political themes occasionally intrude, and which must then find ways to defuse the destructive potential of political conflict for their communal cohesion.

Rather, they represent an active contestation amongst ordinary fans as well as professional commentators over whether football fandom is or indeed should be an apolitical third space in the first place: while some participants continue to resist this politicisation of their fandom and encourage others to focus exclusively on the football itself, others recognise the inescapably political nature of the sport, its rich and powerful administrative bodies, and its decision-making processes about hosting rights, tournament rules, media coverage, and other aspects. The necessary consequence of this recognition is that an entirely apolitical approach

to World Cup fandom is rendered impossible, and that attempts to ignore the politics of the tournament require fans to deliberately ignore substantial aspects of the event.

This raises considerable questions for the overall theoretical framework of third spaces: in essence, it means that the FIFA World Cup is simply too large, too important, too commercialised, and too politicised an event to have any chance of being considered as enabling the apolitical third spaces of fandom that the theory envisages; this is as true for the 2022 Qatar World Cup as it is for its 2018 predecessor in Russia or the subsequent 2026 edition in Canada, the United States, and Mexico, but unlikely to be limited to these particularly controversial World Cups only. If genuine third spaces do remain in fandoms relating to professional football, they are instead more likely to be found in the more limited and bespoke environs of fan communities dedicated to specific teams or players, and perhaps especially for footballers without global or even national profiles.

This, in fact, would bring the theory of (online) third spaces closer again to its original inspiration in the theory of (offline) third places, as initially formulated by Oldenburg and Brissett (1982), with its emphasis on explicitly local, small-scale, and cozy places like bars, coffee shops, hair salons, religious venues, and indeed (social and amateur rather than professional and commercial) sports clubs. By contrast, the global public conversation about football that the World Cup produced, that Twitter hosted, and that journalists, influencers, and ordinary users from all over the world engaged in was perhaps always already too large and too public to engender the dynamics that characterise both third places and third spaces.

Conclusion

In this paper, we investigated how a major sporting event, the 2022 FIFA World Cup, was covered in UK newspapers and on Twitter, and the activity of journalists on Twitter throughout the event. Our results show that, overall, the UK news and the British Twittersphere were relatively aligned in terms of how much political attention they gave to the World Cup (18.5% of news excerpts vs 29% of tweets). One key difference between the two platforms though came from dedicated groups of activists on Twitter leveraging the World Cup to bring attention to their cause, especially in the context of the Mahsa Amini protests in Iran. These forms of political “appropriation” of the World Cup, which we show are much more elaborate than

simple “hashtag hijacking”, account for most of the difference in the proportion of political content between our two corpora (8.0% out of 10.5%).

Generally, political conversations on Twitter remained quite segmented, with different clusters of users focusing on different political issues. This segmentation is consistent with the predictions of counterpublic theory, but our analysis provides unusually direct quantitative evidence of this dynamic by combining fine-grained topic modelling with Social Network Analysis. On a higher level, two dominant political interpretations of the World Cup emerged: one saw the competition as a sportswashing attempt, while the other saw it as an opportunity to make football more inclusive of Arabic or Muslim cultures. All these forms of politicisation were occasionally received with backlash from football fans that preferred to “keep politics out of football”, a stance which itself was criticised in the name that “some things are more important than football”. This meta-debate about the place of politics in the context of the World Cup, and the heated reactions it provoked, lead us to conclude that the concept of “third spaces” does not apply well for the analysis of World Cup football discussions on Twitter, although we suggest it might apply better to smaller fan communities online. Finally, our analysis sheds light on the complex interplay between journalism and social media, namely showing how news coverage itself became an important object of debate amongst journalists on Twitter.

These divergent trends demonstrate the multifaceted nature of the 2022 FIFA World Cup as a controversial sporting, political, and commercial mega-event, and the approaches taken by diverse media and their participants in covering and discussing this event. The 2026 FIFA World Cup, hosted by an uneasy alliance between Mexico, Canada, and the United States, is likely to offer another opportunity to observe these complex intersecting dynamics in action.

Acknowledgments

This work was supported by the Turing’s Defence and Security programme through a partnership with the UK government and Dstl in accordance with the framework agreement between GCHQ & The Alan Turing Institute and the School of Informatics at the University of Edinburgh. Sandrine Chausson’s research is funded by the UKRI Centre for Doctoral Training in Natural Language Processing (grant EP/S022481/1). Axel Bruns’s research is funded by the Australian Research Council through the Australian Laureate Fellowship project Determining

the Drivers and Dynamics of Partisanship and Polarisation in Online Public Debate.

Data and Code Availability

We publicly release the code for all parts of our analysis as a GitHub repository under an open-source licence: <https://github.com/s-l-chausson/beyond-the-game>. Given our data's potential inclusion of personal information, we do not provide direct access to it. Instead, as is standard with Twitter data and required by Twitter's Terms & Conditions, we include a list of all tweet IDs alongside a datasheet for the dataset in the repository. Since X/Twitter's decision to revoke free research access to their API, we cannot guarantee researchers will be able to access the data freely.

References

- Al Hariri, Y., Chausson, S., Ross, B., and Magdy, W. (2024). Twixplorer: An interactive tool for narrative detection and analysis in historic Twitter data. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, CSCW '24*, page 83–86. ACM.
- Amnesty International (2019). Reality Check: Migrant Workers' Rights in Qatar. <https://www.amnesty.org/en/latest/campaigns/2019/02/reality-check-migrant-workers-rights-with-two-years-to-qatar-2022-world-cup/>.
- Baftiu, D. and Dodds, T. (2023). Adapting to Twitter: The entanglement of journalistic values and online personas. *Journalism Studies*, 24(10):1295–1315.
- Bairner, A. (2001). *Sport, nationalism, and globalization : European and North American perspectives*. SUNY series in national identities. State University of New York Press, Albany, N.Y.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1):361–362.
- Bennett, W. L. and Segerberg, A. (2023). *The logic of connective action: digital media and the personalization of contentious politics*, page 287–338. Edward Elgar Publishing.

- Bidewell, T., Deligianni, A., Elmas, T., Llewellyn, C., and Ross, B. (2026). Gendered communication patterns of political elites on Truth Social. In *Proceedings of the 18th ACM Web Science Conference 2026*, United States. Association for Computing Machinery (ACM).
- Biswas, M. R., Khan, S., and Shah, Z. (2023). Twitter users discussions about migrants rights and facilities during FIFA World Cup 2022. In *2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*, page 1–5. IEEE.
- Boulianne, S. (2018). Twenty years of digital media effects on civic and political participation. *Communication Research*, 47(7):947–966.
- Brannagan, P. M. and Reiche, D. (2022). *Qatar and the 2022 FIFA World Cup: Politics, Controversy, Change*. Springer International Publishing.
- Brems, C., Temmerman, M., Graham, T., and Broersma, M. (2016). Personal branding on Twitter: How employed and freelance journalists stage themselves on social media. *Digital Journalism*, 5(4):443–459.
- Brosius, H.-B., Haim, M., and Weimann, G. (2019). Diffusion as a future perspective of agenda setting. *The Agenda Setting Journal*, 3(2):123–138.
- Bruns, A. (2018). *Gatewatching and News Curation*. Peter Lang US.
- Bruns, A. and Burgess, J. (2015). Twitter hashtags from ad hoc to calculated publics. In Rambukkana, N., editor, *Hashtag publics: The power and politics of discursive networks [Digital Formations]*, volume 103, page 13–27. Peter Lang Publishing, United States of America.
- Cashmore, E. and Cleland, J. (2012). Fans, homophobia and masculinities in association football: Evidence of a more inclusive environment. *The British Journal of Sociology*, 63(2):370–387.
- Chadwick, A. (2017). *The Hybrid Media System: Politics and Power*. Oxford University Press New York.
- Charikar, M. S. (2002). Similarity estimation techniques from rounding algorithms. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, STOC02*, page 380–388. ACM.

- Dubinsky, Y. (2023). Clashes of cultures at the FIFA World Cup: Reflections on soft power, nation building, and sportswashing in Qatar 2022. *Place Branding and Public Diplomacy*, 20(2):218–231.
- Fan, M., Billings, A., Zhu, X., and Yu, P. (2019). Twitter-based birging: Big data analysis of English national team fans during the 2018 FIFA World Cup. *Communication & Sport*, 8(3):317–345.
- Farrag, D. A., Fadol, Y., Hariri, Y. A., and ElAlfy, A. (2025). Harnessing big data for soft power: A sentiment analysis of FIFA Qatar 2022. *Event Management*, 29(8):1185–1203.
- Fleck, A. (2022). Infographic: Which sports do Brits love the most? Statista Daily Data. <https://www.statista.com/chart/28160/most-followed-sports-in-the-uk>.
- Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, (25/26):56.
- Freelon, D., McIlwain, C., and Clark, M. (2016). Quantifying the power and consequences of social media protest. *New Media & Society*, 20(3):990–1011.
- Grix, J. (2012). ‘Image’ leveraging and sports mega-events: Germany and the 2006 FIFA World Cup. *Journal of Sport & Tourism*, 17(4):289–312.
- Habermas, J. (1989). *The structural transformation of the public sphere*. Polity Press, Oxford, England.
- Hanusch, F. and Bruns, A. (2016). Journalistic branding on twitter: A representative study of Australian journalists’ profile descriptions. *Digital Journalism*, 5(1):26–43.
- Hassan, A. A. M. and Wang, J. (2023). The Qatar World Cup and Twitter sentiment: Unraveling the interplay of soft power, public opinion, and media scrutiny. *International Review for the Sociology of Sport*, 59(5):679–704.
- Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. *Journalism Practice*, 4(3):297–308.

- Highfield, T., Harrington, S., and Bruns, A. (2013). Twitter as a technology for audience and fandom: The #eurovision phenomenon. *Information, Communication & Society*, 16(3):315–339.
- Hobsbawm, E. J. (1992). Introduction. In *Nations and Nationalism since 1780*, pages 1–13. Cambridge University Press, Cambridge.
- Hurcombe, E. (2024). Conceptualising the “newsfluencer”: Intersecting trajectories in online content creation and platformised journalism. *Digital Journalism*, 13(9):1523–1534.
- Jackson, S. J., Bailey, M., and Welles, B. F. (2020). *#HashtagActivism*. The MIT Press. MIT Press, London, England.
- Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6):e98679.
- Kaiser, J. and Rauchfleisch, A. (2019). Integrating concepts of counterpublics into generalised public sphere frameworks: Contemporary transformations in radical forms. *Javnost - The Public*, 26(3):241–257.
- Knott, B., Fyall, A., and Jones, I. (2017). Sport mega-events and nation branding: Unique characteristics of the 2010 FIFA World Cup, South Africa. *International Journal of Contemporary Hospitality Management*, 29(3):900–923.
- Laer, J. V. and Aelst, P. V. (2010). Internet and social movement action repertoires. *Information, Communication & Society*, 13(8):1146–1171.
- Manku, G. S., Jain, A., and Das Sarma, A. (2007). Detecting near-duplicates for web crawling. In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, page 141–150, New York, NY, USA. Association for Computing Machinery.
- McGregor, S. C. and Molyneux, L. (2018). Twitter’s influence on news judgment: An experiment among journalists. *Journalism*, 21(5):597–613.
- Meier, H. E., Mutz, M., Glathe, J., Jetzke, M., and Hölzen, M. (2021). Politicization of a contested mega event: The 2018 FIFA World Cup on Twitter. *Communication & Sport*, 9(5):785–810.

- Meraz, S. and Papacharissi, Z. (2013). Networked gatekeeping and networked framing on #egypt. *The International Journal of Press/Politics*, 18(2):138–166.
- Molyneux, L. and Holton, A. (2014). Branding (health) journalism: Perceptions, practices, and emerging norms. *Digital Journalism*, 3(2):225–242.
- Olausson, U. (2017). The celebrified journalist: Journalistic self-promotion and branding in celebrity constructions on Twitter. *Journalism Studies*, 19(16):2379–2399.
- Oldenburg, R. and Brissett, D. (1982). The third place. *Qualitative Sociology*, 5(4):265–284.
- Opgenhaffen, M. and Scheerlinck, H. (2014). Social media guidelines for journalists: An investigation into the sense and nonsense among Flemish journalists. *Journalism Practice*, 8(6):726–741.
- Papacharissi, Z. A. (2010). *A private sphere*. Digital Media and Society. Polity Press, Oxford, England.
- Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Scharfenort, N. (2012). Urban Development and Social Change in Qatar: The Qatar National Vision 2030 and the 2022 FIFA World Cup. *Journal of Arabian Studies*, 2(2):209–230.
- Statista Research Department (2024). Level of interest in soccer in selected countries worldwide as of September 2022. <https://www-statista-com.eux.idm.oclc.org/statistics/868932/soccer-interest-worldwide-countries/>.
- Swart, K. and Hussain, U. (2023). *FIFA 2022 World Cup in Qatar: Under the Gaze of Western Media*, pages 199–222. Springer Nature Singapore, Singapore.
- Toprak, M., Boldrini, C., Passarella, A., and Conti, M. (2022). Journalists’ ego networks in Twitter: Invariant and distinctive structural features. *Online Social Networks and Media*, 30:100207.

- Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1).
- Trilling, D., Tolochko, P., and Burscher, B. (2016). From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly*, 94(1):38–60.
- Trott, V. (2020). Networked feminism: Counterpublics and the intersectional issues of #metoo. *Feminist Media Studies*, 21(7):1125–1142.
- Twitter News (2022). How many people come to Twitter for news? As it turns out, a LOT. https://blog.x.com/en_us/topics/insights/2022/how-many-people-come-twitter-for-news.
- von Engelhardt, J. and Jansz, J. (2014). Challenging humanitarian communication: An empirical exploration of Kony 2012. *International Communication Gazette*, 76(6):464–484.
- Warner, M. (2005). *Publics and Counterpublics*. Zone Books. Zone Books, New York, NY.
- Weller, K., Wang, Y., Peters, Y., and Gruber, J. B. (2026). Can 20 years of Twitter be preserved?: What is lost, what remains, and what is in between. *M/C Journal*, 29(2).
- Wright, S., Graham, T., and Jackson, D. (2015). Third space, social media, and everyday political talk. In *The Routledge companion to social media and politics*, pages 74–88. Routledge.

Appendix

A. Data Collection

We used the Twitter API v2 Filtered Stream endpoint (Academic Research access tier) to collect the tweets matching a set of keywords, as well as replies to and quotes of tweets with these keywords. Moreover, if a collected tweet was a reply or quoted another tweet, the original tweet was also retrieved. To improve robustness, a primary pipeline performed real-time data collection via filtered stream while a second recovery pipeline performed periodic backfilling using the recent search endpoint (up to 7 days). The data also includes retweets of and replies to tweets containing the keywords.

The keywords we used for the collection of our larger multilingual 2022 FIFA World Cup Twitter dataset were selected through an iterative process combining manual curation and data-driven expansion. We began with a seed set of English-language keywords representing the official name and common references to the FIFA World Cup (e.g., *FIFAWorldCup*, *WorldCup2022*). This dataset was then manually reviewed to identify high-frequency terms, hashtags, and entity references. Terms exceeding a predefined frequency threshold and judged to be contextually relevant were incorporated into the query set (“snowball” phase). This process enabled the inclusion of multilingual and colloquial expressions (e.g., team nicknames, localised tournament references) observed in real-time discourse. This approach is consistent with established practices in social media data collection. The expansion process continued iteratively until saturation was reached, i.e. where newly added terms contributed minimal additional data. To ensure structured coverage, query terms were organized into four categories:

1. Official and commonly used names of the FIFA World Cup across languages and formats.
2. References to the host location (e.g., Doha), capturing location-based discourse.
3. Official stadium names, which serve as contextual anchors for match-related discussions.
4. Names of participating national teams, including official names, common nicknames and multilingual variants

To maximise multilingual coverage across participating audiences, the query set incor-

porated terms in the official languages of all participating national teams. These non-English keywords included both official translations of tournament-related terms and colloquial and fan-generated expressions identified during the snowball phase. We deliberately excluded country names such as *Qatar, England, Saudi Arabia*) (although we kept *Qatar* as a secondary filter when creating our smaller subset of English tweets from our larger multi-lingual dataset) because empirical observations from the initial dataset indicated that such terms were strongly associated with non-sport discourse, including geopolitical and economic discussion. We also deliberately excluded player-specific terms (e.g., names of prominent athletes such as Lionel Messi, Kylian Mbappé, Cristiano Ronaldo) from the query design. While mentioned a lot in football-related tweets, their inclusion disproportionately amplified content associated with globally famous players, leading to representation bias across teams.

Table 7: Keywords used to collect the wider Qatar 2022 FIFA World Cup Twitter dataset from which our own dataset is derived. Note that this original data also includes retweets of and replies to tweets containing the keywords, as well as original tweets when quotes or replies were caught with our keywords.

#worldcup2022	Coupe du monde 2022	Coppa del Mondo Calcio 2022	TeamEngland	EITri	RedDevils
#qatar-	Coupe_du_Monde_Qatar_2022	Mondial	ThreeLions	El Tri	TeamBelgique
#worldcup-	Coupe_du_Monde	Mondial2022	equipedefrance	AtlasLions	TeamBelgium
#fifa-	Coupe_du_Monde_2022	2022 جام جهاني	Fiersdetrebleus	The Atlas Lions	TeamBrazil
#doha-	CoupeDuMonde2022	جام _ جهاني	FrenchTeam	DimaMaghrib	TeamBresil
#football-	verdensmesterskab	جام _ جهاني _ قطر	Les Bleus	TeamMorocco	Seleção
#dohaqatar-	VM i fodbold 2022	2022 جام _ جهاني _ قطر	LesBleus	Oranje	TeamCameroon
#fifa2022-	VM_2022	mistrzostwaświata	listedesbleus	SamuraiBlue	TeamCameroun
FIFAWorldCup	weltmeisterschaft	Mistrzostwa Świata	TeamFrance	TeamNetherlands	LesLionsIndomptables
FIFAWorldCup22	Fußballweltmeisterschaft	MistrzostwaŚwiatawPłceNożnej	LaCeleste	Bialo-czerwoni	TeamSpain
fifaworldcup2022	Fußballweltmeisterschaft2022	Mundial22	TeamUSA	TeamPoland	La Furia Roja
Qatar2022	Fußballweltmeisterschaft_2022	Cwpan y Byd	USMNT	TeamPortugal	LaFuriaRoja
QATARFIFA2022	copa2022	cwpanybyd	OnlyForward	Os Navegadores	TeamSwitzerland
QatarWorldCup2022	Copa do Mundo 2022	كأس _ العالم	Nationalelf	OsNavegadores	Les Lions Indomptables
WorldCup	Copa_do_Mundo	كأس _ العالم _ قطر 2022	Nationalmannschaft	TeamQatar	INDOMPTABLES
WorldCup2022	Copa_do_Mundo_2022	موندیال _ قطر_2022	TeamGermany	TeamSaudi	ALLEZLESLIONS
WorldcupQatar2022	Wkvoetbal	كأس _ العالم_2022	TeamGhana	TeamSenegal	TeamCanada
RoadToQatar	Wereldkampioenschap voetbal	TeamCroatia	BlackStars	TeamArgentina	TeamCosta Rica
fifaworldcupqualifiers	Wereldkampioenschap 2022	RodHvide	Black Stars	TeamArgentine	TeamCostaRica
FIFAWorldCup2022Qatar	Svjetsko prvenstvo 2022	fordanmark	TeamIran	La Albiceleste	LosTicos
World Cup	copa del mundo 2022	herrelandsholdet	Team Melli	LaAlbiceleste	Los Ticos
roadto2022	CopadelMundo	skuldervedskulder	TeamMelli	TeamAustralia	kickasti
Doha	CopadelMundo2022	TeamDenmark	TeamJapan	Socceroos	Rossocrociati
Education City Stadium	CoppadelMundo	TeamEcuador	SoloPalante	Belgica	TeamTunisia
Lusail Stadium	MundialQatar2022	LaTri	TeamWales	DEVILTIME	المنتخب التونسي
Stadium 974	Mundial Qatar 2022	ElBancoDeLaTRI	Dragons	Die Roten Teufel	منتخب تونس
Al Thumama Stadium	VamosEspaña	EaglesofCarthage	Samurai Blue	Lions of Teranga	
Al Bayt Stadium	FiestaTicaEnCatar	Eagles of Carthage	TeamKorea	LionsOfTeranga	
Ahmed bin Ali Stadium	EstamosEnCatar	TeamUruguay	Taegeuk Warriors	TeamSerbia	
Al Janoub Stadium	Сверско првенство	La Celeste	TaegeukWarriors	Orlovi	
Khalifa International Stadium	Сверсконпвенство	La Tri	TeamMexico	DieRotenTeufel	

Table 8: Number of articles per newspaper in our UK newspaper corpus.

The Sun (England)	5381	Yorkshire Evening Post	392	Grimsby Telegraph	165
Daily Mirror	3991	Sunderland Echo	361	Harrogate Advertiser	125
The Independent (UK)	3511	The Evening Standard (LDN)	361	The Observer	122
The Times (LDN)	3121	Daily Post (North Wales)	354	Lancaster Guardian	120
Daily Star Online	1812	Birmingham Evening Mail	354	The Plymouth Herald	119
Daily Record & Sunday Mail	1299	The Independent - Daily	351	Dewsbury Reporter	102
The Guardian (LDN)	1270	The Herald (Glasgow)	338	Burnley Express	96
birminghammail.co.uk	1073	Evening Times (Glasgow)	319	Sunday Mercury	78
The Daily Telegraph (LDN)	915	Evening Chronicle	304	Western Telegraph	72
The Express	870	The Northern Echo	299	The Sunday Herald (GLA)	71
walesonline.co.uk	787	Evening Gazette	275	Wales on Sunday	64
The Sunday Times (LDN)	671	Aberdeen Press & Journal	275	Stornoway Gazette	58
The People	590	Sunday Express	234	Gloucestershire Echo	46
Yorkshire Post	525	coventrytelegraph.net	227	The Sentinel	34
South Wales Echo	518	Aberdeen Evening Express	217	Birmingham Post	29
Belfast Telegraph	515	Hull Daily Mail	216	Exeter Express and Echo	24
Manchester Evening News	513	Western Daily Press	215	Hampstead & Highgate Express	20
BBC	448	Nottingham Post	212	Scunthorpe Telegraph	14
Scotsman	425	Derby Telegraph	210	Mail on Sunday (LDN)	11
Liverpool Echo	424	Lancashire Evening Post	207	Islington Gazette	9
Belfast Telegraph Online	405	Leicester Mercury	203	Times Higher Education	8
The Star (Sheffield)	396	The Sunday Telegraph (LDN)	190	Chorley Guardian	5

B. Social Network Analysis

The Leiden algorithm, which we used to identify clusters of users from our retweet data as part of our Social Analysis, requires a single hyperparameter: resolution, which controls how fine-grained the clusters are. We tuned the resolution hyperparameter with values ranging between 0.01 and 5. While the modularity score obtained with the different resolution values indicated that lower resolution led to better clustering, we also took into account the visual rendering of the clustering to decide on a final resolution value. In the end, we used a resolution of 0.6 which gave a modularity of ~ 0.808 . While using even lower resolution values gave higher modularity scores, we decided against this on visual inspection of the results as this essentially translated into just one or a few all-encompassing clusters, which did not reflect the actual level of granularity in the sub-networks in our data.

Table 9: Top words, top locations and most retweeted accounts per cluster. Clusters highlighted in red were filtered out based on our location criteria. Only clusters with more than 10K accounts are shown.

Cluster	Label	Top 30 words	Top 3 locations	Most retweeted accounts
1	Football	Football, fan, Sport, Arsenal, love, life, MUFC, God, Messi, Music, Manchester United, new, view, ManUtd, enthusiast	UK: 2,255 Nigeria: 1,334 Ghana: 1,334	FIFA World Cup Stats (@alimo_philip) Troll Football Football Tweet 🗣️ (@TrollFootball) Premier League (@premierleague) Manchester United (@ManUtd)
2	Word Cup	love, BT, fan, fan account, Football, Sport, BTS_twt, life, ARMY, OT7, Music, de, live, lover, enthusiast	USA: 581 UK: 407 India: 363	FIFA World Cup (@FIFAWorldCup) Road to 2022 (@roadto2022en) Portugal (@selecaoportugal) OnsOranje (@OnsOranje) HNS (@HNS_CFF)
3	American	Sport, fan, Football, Host, Alum, love, soccer, Podcast, new, Former, High School, writer, life, thing, team	USA: 3,365 Canada: 219 UK: 118	FOX Soccer (@FOXsoccer) TSN (@TSN_Sports) NBC Sports Soccer (@NBCSportsSoccer) ESPN (@espn) US Soccer Men's National Team (@USMNT)
4	K-pop	BTS, fan account, BTS_twt, ARMY, OT7, love, Jungkook, life, 아포방포, 방탄소년단, BTS ARMY, Bangtan, BTS fan, JK, stan	USA: 299 Korea: 284 Japan: 180	BTS Charts & Translations (@charts_k) Jungkook SNS (@Jungkook_SNS) chart data (@chartdata)
5	Arabic/Muslim	Muslim, new, love, world, football, fan, life, Sport, RT, Cricket, endorsement, FreePalestine, Views, Journalist, politics	UK: 457 Pakistan: 378 USA: 213	Farid Khan (@FaridKhan) PALESTINE ONLINE 🇵🇸 (@OnlinePalEng) Khaled Beydoun (@KhaledBeydoun) Qatar Airways (@qatarairways) Ashok Swain (@ashoswai)
6	NFT/Crypto	NFT, Crypto, Sport, Web3, BNB, love, world, Football, fan, BTC, Community, DM, lover, Bitcoin, Free	Nigeria: 232 Indonesia: 110 USA: 80	Stake.com (@Stake) YokaiSwap (@yokaiswapnervos) TOPGOAL (@TopGoal_NFT) Official Catcoin 🐱 (@officialcatcoin)
7	Iranian	از, به, زن زندگی, زندگی آزادی, و, Iranian, Life Freedom, Freedom, که, در, مهسا امینی, من, ایران, people	USA: 87 UK: 60 Iran: 43	🇮🇷 (@fromIranLand) Rev Group Iran (@Rev_Group_iran) Omid Memarian (@Omid_M)
8	News media	sport, new, view, fan, love, Football, politics, UK, world, Writer, Editor, Journalist, Author, Proud, music	UK: 812 USA: 374 Ireland: 74	The New York Times (@nytimes) Visegrád 24 (@visegrad24) Reuters (@Reuters) The Guardian (@guardian) NEXTA (@nexta_tv)
9	Indian	Sport, India, fan, Cricket, New, endorsement, Indian, RT, Love, world, फ़, Hindu, Football, स, य	India: 843 USA: 56 UK: 43	RVCJ Media (@RVCJ_FB) Matchday (@matchday) World Index (@theworldindex) தல அரவிந்த் (@aravinth43AK) The Field (@thefield_in)
10	Pop culture	fan account, BTS, love, Jin Jin, stan, life, time, fan, de, OT7, music, new, enthusiast, follow, ARMY	USA: 61 UK: 30 Brazil: 18	Pop Base (@PopBase) ENHYPEN WEVERSE (@enhypenweverse) Vai Desmaiar (@vaidesmaiar) Pop Tingz (@ThePopTingz)
11	Gaming (FUT)	fan, football, love, Sport, FIFA, life, FC, family, one, new, Free, follow, player, home, team	UK: 324 USA: 52 Ireland: 11	Stephen R Power (@racingblogger) Mystery Football (@MysteryFootba11) Donk 🍌 (@DonkTrading) boxtobox (@boxtoboxUK) Fut Sheriff (@FutSheriff)
12	Tech	stufflistingsarmy, Lover, Love, Cricket, fan, Crypto, life, Sport, Contest Lover, Tech, Contest, India, Music, Tech Enthusiast, Football	India: 219 Indonesia: 25 USA: 18	Toshiba TV Global (@ToshibaTVGlobal) Sportskeeda Football (@skworldfootball) iQOO India (@IqooInd)
13	Collectibles	Sport, fan, Football, PC, new, love, Go, Card, Lion, Dad, Sports Card, 49ers, Sales, Life	USA: 131 UK: 12 Canada: 6	Detroit Lions (@Lions) Dave & Adam's (@dacardworld) Steel City Collectibles (@SCCTradingCards)

C. Filtering steps

Table 10: Number of tweets/news excerpts at each filtering step.

Filtering step	Twitter	News
Original collection	15,024,413	37,345
Copy pasta removal/deduplication	13,771,912	29,995
Keyword matching	11,328,058	15,818
Social Network Analysis: degree filter	1,466,766	-
Social Network Analysis: cluster filter	697,393	-
English language	622,936	-
Short posts	573,927	15812

D. Topics summary and breakdown

Table 11: Number of datapoints (tweets and news excerpts) per category, topic and sub-topic with most frequent terms.

Topic	Football		Top words
	Twitter	News	
“Football” vs “soccer”	1,656	2	soccer, called, sport, rugby, american
British football clubs	15,060	1,580	
Transfers and contracts	1,777	335	club, transfer, united, chelsea, manchester
Club vs. international football debate	3,106	52	players, team, play, teams, european
English football club fan discussions	3,568	2	club, england, league, team, football_scot
Football clubs, matches, and players	2,824	235	club, england, bellingham, league, game
Premier League clubs return after World Cup	3,118	65	club, league, chelsea, united, arsenal
Scottish football post-World Cup break	667	891	celtic, rangers, beale, club, city
Female referees	320	36	frappart, female, referee, stephanie
Football discussions	19,888	286	
Football match discussions and rules	3,555	65	game, play, ball, players, know
Complaints and opinions about play	3,771	66	fans, people, players, know, game
Opinions and discussions	2,703	0	know, football__tweet, playerzpot, good, play
Opinions and discussions about “football Twitter”	3,465	4	football__tweet, twitter, tweet, tweeting
Opinions on football quality	3,130	0	team, game, play, good, win
World Cup upsets and unpredictability	3,264	151	fifa, teams, team, group, tournament
Football trolling	7,839	19	
Football fans and their knowledge debated	3,279	2	people, know, fans, watch, football__tweet
Football vs American Football	2,948	0	know, game, play, sport, watch
Trolling Americans for American football	1,612	17	american, soccer, americans, usa, america

Games	30,312	857	
Cameroon vs Serbia, Brazil matches	1,431	35	cameroon, fifaserbia, brazil, switzerland
Croatia beats Brazil on penalties	2,126	0	croatia, fifabrazil, croatiavsbrasil, crobra
-	1,627	75	croatia, brazil, fifaargentina, vs, final
Croatia vs Morocco third-place match	1,034	37	croatia, morocco, vs, place
Ecuador beats Qatar in opener	1,469	39	ecuador, qatar, game, valencia, match
Argentina vs France World Cup Final	1,855	55	france, argentina, fifafinal, vs
France vs Morocco semi-final	1,543	65	france, morocco, final, tunisia
Match results	2,539	7	qatar2022, fifavs, group, qatar, argentina
Match results and upsets analysis	2,420	98	game, argentina, match, goal, final
Match scores and game progress	1,921	6	fifahalf, game, score, vs, match
Match times	2,054	88	qatar, stadium, fifaal, live
Match updates	1,990	0	fifa, fifa, brazil, coinerr_err, bra
Match updates and results	1,875	105	2022, qatar, fifa, fifaqatar2022, vs
Morocco vs Belgium, ensuing riots	612	22	belgium, morocco, fifabrussels, riots
Senegal beats Qatar	1,176	19	senegal, qatar, fifafrican, africa
Saudi Arabia historic win against Argentina	3,202	68	saudi, arabia, argentina, win
Match schedules and scores	1,438	138	2022, fifa, vs, argentina, france
General football content	12,365	368	
Extra time, game duration, countdowns	1,623	52	time, minutes, fifaadded, stoppage, extra
FIFA president, coaches, players controversies	3,045	237	fifa, fifainfantino, man, fans, qatar
Football knowledge and player discussions	3,010	0	know, football__tweet, dey, play, team
Personal relationship with football	2,771	10	play, playing, time, life, know
Youth football development and academies	1,916	69	players, academy, team, club, play
Opinions about football	23,632	110	
Games anticipation and disappointment	2,484	0	game, fifa, adidas match, time, england
Emotional football fan reactions	1,865	16	game, love, feel, watching, crying
Excitement and drama	2,428	2	game, , match, final, drama
-	1,940	0	fifagame, match, going, fifa, final
Excitement and opinions on football matches	2,117	10	game, match, best, greatest, ve
Football love and passion expressed	3,974	0	love, game, know, play, time
Football is/isn't the greatest sport	2,858	3	sport, love, game, best, greatest
Love and Enjoyment of Football	3,046	0	watch, watching, love, game, fan
Watching, enjoying, and debating World Cup	2,920	79	watch, watching, game, fifa, love
Players	77,683	2,475	
Argentina key players	2,071	95	fifamartinez, alvarez, argentina, valencia
Discussions about football players careers	3,044	1	player, play, football__tweet, good, guy
French players sickness and injuries	1,305	187	france, benzema, fifadeschamps, karim
Giroud breaks France goal scoring record	819	19	giroud, olivier, france, fifathierry, henry
Goalkeepers making penalty saves	858	16	fifasave, saves, penalty, saved, martinez
World Cup Golden Boot & Glove races	543	31	golden, boot, mbappe, kylian, glove
Maguire performance & criticism	633	112	maguire, harry, england, stones
Kane penalty miss against France	1,339	324	kane, harry, england, penalty, eng
Marcus Rashford goals	927	99	rashford, marcus, england, fifafoden
Mbappe performance	4,653	225	mbappe, fifakylian, france, messi

Messi	19,146	495	
Messi leads Argentina to World Cup	2,912	68	messi, argentina, lionel, argentinavsfrance
Messi completes football legacy	2,874	46	messi, lionel, completed, greatest, game
Messi wins World Cup 2022 final	2,673	1	messi, ronaldo, fifa, argentina
Messi wins FIFA World Cup 2022	1,983	203	messi, argentina, lionel, win, fifafinal
Messi's World Cup Legacy and Records	1,904	79	messi, lionel, fifaargentina, goals, player
Messi's World Cup win in Qatar	1,969	54	messi, qatar, argentina, lionel
Messi GOAT debate	896	25	messi, greatest, lionel, player, time
Messi GOAT debate is settled	1,618	17	goat, messi, fifa lionel, ronaldo, debate
Messi's World Cup Win and GOAT Status	2,317	2	messi, arg, lionel, final
Messi vs Ronaldo GOAT debate	3,146	36	messi, ronaldo, player, better, fans
-	1,818	9	goat, goats, debate, ronaldo
Modric, Lukaku	851	58	luka, modric, lukaku, croatia, romelu
Neymar injury and performance	1,515	109	neymar, brazil, ankle, neymarjr
Player goal records	2,145	83	goals, fifascored, goal, player
Player performance and opinions	4,623	136	player, best, good, play, fifaplayers
Player performances and match events	4,119	17	fifa, man, game, usmnt, fifa
Player performances and tournament updates	2,933	51	player, fifa, best, 2022
Player retirements and tributes	1,250	61	international, retirement, hazard, pele, retire
Players general discussions	3,861	61	player, man, play, guy, good
Qatar World Cup player and coach analysis	3,968	118	qatar, team, keeper, player
Ronaldo	8,068	85	
Performance and legacy	3,587	0	fifacristiano, bts_bighit, fifa, player, best
World Cup legacy debate	2,786	68	ronaldo, fifacristiano, player, fans, fifa
World Cup achievements and legacy	1,695	17	cristiano, ronaldo, portugal
Player and team updates	3,408	1	fifaqatar, qatar, intermediaries, bletchley
Young stars shine in World Cup	640	46	gakpo, gavi, cody, youngest, pedri
Predictions and results	3,552	11	
Match predictions and reactions	1,750	0	adidas, moow2earn, budweiser, france, goal
Predictions and scores	93	0	binary_x, bnx, brazil, netherlands
Win predictions and polls	1,709	11	fifawin, team, brazil, argentina, fifa
Reacting to action	10,436	28	
Amazing goals	2,515	17	goal, richarlison, brazil, scored, goals
Goals and penalties	2,828	8	goal, fifaqatar, valencia, penalty
Penalty shootouts and decisions	1,844	3	penalty, fifapenalties, shootout, argentina
World Cup match reactions and predictions	3,249	0	fifa, fifa, eng, game, let
Teams	95,678	5,281	
African teams	3,253	18	african, africa, nigeria, africans, team
Argentina	15,208	337	
Argentina wins World Cup final	2,620	0	argentina, france, worldcupfinal
-	1,985	2	argentina, congratulations, fifamessi, france
-	1,649	0	arg, fifafra, argfra, vs
-	1,178	0	argentina, win, thesandboxgame, france
Argentina win reactions	3,469	143	argentina, team, win, fifa, game
Argentina World Cup Journey	2,479	42	argentina, win, final, messi
Messi leads Argentina World Cup triumph	1,828	150	argentina, messi, maradona, lionel, france

Australia knockout qualification	1,455	49	australia, socceros, aus, denmark, tunisia
Belgium	1,780	84	
Belgium's golden generation underperforms	1,063	65	belgium, fifa fifa, lukaku, team, roberto
Belgium's FIFA World Cup Exit	717	19	belgium, canada, fifacroatia, belcan
Brazil	2,804	86	brazil, team, brazilian, brazilianspring
Canada	1,743	49	
Canada's first World Cup appearance	1,108	22	canada, canmnt, team, canadasocceren
Canada's World Cup matches and results	635	27	canada, croatia, davies, goal, alphonso
Costa Rica	1,105	35	rica, costa, spain, fifajapan, germany
Croatia	2,389	81	croatia, fifabrazil, team, croatian, serbia
England	6,919	2,317	
Performance analysis	3,680	487	england, fifa eng, win, team, france
England squad	1,592	1,325	england, sterling, raheem, wales, qatar
Southgate's England team debate	1,647	505	southgate, gareth, england, fifamanager, foden
France	4,508	162	
France World Cup run	2,091	8	france, fifaengland, win, fra
Performance discussion	2,417	154	france, fifafrench, team, england, final
Germany	1,757	111	germany, german, team, germans, flick
Ghana	3,561	31	ghana, ghanablackstars, portugal, blackstars
Iran	2,493	64	iran, fifaengland, wales, usa
Japan	5,018	100	
Japan upsets Germany, tops group	2,247	25	japan, fifagermany, jpn, gerjpn
Japan fans cleaning stadiums and team	1,939	46	japan, japanese, fifa fans, team, samurai
Japan's shock wins over Spain/Germany	832	29	japan, spain, germany, fifagroup, rica
Morocco	11,052	252	
Morocco's historic World Cup run	3,595	2	morocco, france, win, congratulations
-	3,275	160	morocco, moroccan, fifateam, moroccans
Morocco beats Spain and Portugal	2,545	39	morocco, fifaspain, portugal, mar
Morocco reaches World Cup Semifinals	1,637	51	morocco, african, fifafrica, team, semi
Netherlands	4,333	120	
Matches and results	3,203	120	netherlands, fifadutch, argentina, usa
-	1,130	0	ned, fifanedarg, nedusa, arg
Poland	1,477	47	poland, mexico, lewandowski, argentina, saudi
Portugal	4,321	176	
Portugal's FIFA World Cup journey	2,330	25	portugal, fifaghana, switzerland, uruguay
Ronaldo dropped, Portugal eliminated	1,991	151	portugal, ronaldo, cristiano, santos, ramos
Qatar	5,028	82	
Qatar team performance and criticism	4,252	41	qatar, team, win, fifagame
Qatar first host eliminated from WC	776	41	qatar, host, fifanation, opening, lose
Senegal	1,494	85	senegal, england, netherlands, ecuador, mane
South Korea	3,134	73	
South Korean team	1,330	35	korea, korean, south, asian, fifateam
South Korea advances past Portugal	1,804	38	korea, south, fifauruguay, ghana, brazil
Spain	2,549	59	spain, fifagermany, spanish, enrique
USA	1,754	70	usa, fifaengland, usmnt, team, game
Wales	2,192	573	wales, england, welsh, fifabale, usa

Various teams	4,351	220	
Team selections and analysis	2,181	87	fifaatt, gametime, xi, game
Teams and performance	2,170	133	fifagroup, fifa, teams, brazil, team
VAR/referee controversies	11,669	267	
Controversial ref. decisions and officiating	3,367	95	referee, fifaref, fifa, refs, referees
Controversial penalties and ref. decisions	2,845	44	penalty, fifapenalties, messi, fifa, argentina
Referees, fouls, and goalkeeper incidents	2,691	87	fifaball, goal, fifa, game, kick
VAR controversies and referee decisions	1,575	33	var, fifa fifa, penalty, goal, game
VAR offside decisions and controversy	1,191	8	offside, fifaoffsides, goal, var, qatar
Watching World Cup/football	6,499	35	
Anticipating or reflecting on games (positive)	2,350	7	day, today, home, game, night
Viewing schedules and sleep	1,791	22	fifa watch, day, fifa, sleep, today
Watching football , opinions and enjoyment	2,358	6	watch, watching, game, watched, people
Women and football	1,899	59	women, men, female, woman, girls

World Cup				
Topic	Description	Twitter	News	Top words
Alcohol ban		1,242	127	alcohol, qatar, beer, ban, fans
Best World Cup		4,390	59	
	People loving Qatar World Cup	2,564	37	qatar, best, fifa greatest, thank
	Final described as best ever	1,826	22	best, final, fifagreatest, game, match
Broadcasting/press		2,321	239	bbc, itv, fifacoverage, commentary, qatar
In Qatar		6412	258	
	Fan experience and observations	3,191	190	qatar, fans, fifadoha, fifa
	Qatar travel and jobs	3,221	68	qatar, doha, hiring, dubai
Jungkook		2,051	4	
	Jungkook's "Dreamers" FIFA WC song	16	0	bts_twt, jungkook, dreamers, fifastreaming
	-	270	0	jungkook, dreamers, bts_twt, fifasong
	-	8	0	bts_twt, jungkook, dreamers
	-	490	1	jungkook, fifa, bts_bighit, activities
	-	437	0	jungkook, reamers2022, dreamers, bts_twt
	-	174	0	dreamers, jung, , kook, soundtrack
	-	192	3	fifa, jk, bts, song, bh
	-	423	0	videogaming, fortnine, xboxone
	YouTube chart error for Dreamers	41	0	counted, youtube, redone_official
Opening ceremony		778	21	ceremony, opening, qatar, fifafifa, 2022
Public figures in Qatar		5,656	324	
	Grant Wahl death in Qatar during World Cup	937	52	wahl, grant, journalist, qatar, dies
	Nora Fatehi and female football fans	1,825	178	fifafifa, nora, fatehi, qatar, norafatehi
	Phyna representing Nigeria at World Cup	23	0	phyna, phynation, qatar, sa
	Leaders, performers, and fans in Qatar	2,871	94	qatar, doha, president, al
Stadium/infrastructure		1,628	96	stadium, qatar, stadiums, fifa974, lusail
World Cup format		1,578	146	fifa, 2022, 2026, fifa2025, teams
World Cup songs		1,667	43	song, tonica, fifafifa, waka, songs

		Political		
Topic	Description	Twitter	News	Top words
FIFA		6,743	76	
	FIFA corruption and bias accusations	2,854	1	fifa, fifafifacom, know, game, think
	FIFA corruption and human rights	3,889	75	fifa, corrupt, fifacorruption, thejusticedept, rights
Flags (Palestine, rainbow)		1,683	59	flag, palestinian, qatar, palestine, flags
Football and money (FIFA finances and WC cost)		2,809	187	money, fifa, fifaqatar, million, 000
Football vs politics		1,654	9	politics, political, people, sport, politicians
Gen. political content		2,138	40	fifaww3, putin, wwiii, ukraine
Human rights		12,905	158	
	Human rights protests at FIFA World Cup	3,831	2	fifa, fifacom, people, expellusa, balenciaga
	Qatar human rights and boycott discussions	3,553	1	qatar, kagutamuseveni, people, rights, country
	Qatar human rights record criticism	5,521	155	qatar, fifa, fifarights, people
Iran protests		42,350	178	
	England's support for Mahsa Amini	438	0	kneel4mahsa, irn, mahsaamini, england, voice
	Iran executions during World Cup	1,101	6	iran, mashhad, islamic, regime, rahnavard
	Iran human rights abuse protest	155	0	barrage, normalize, bullets, authorities, situations
	Iran human rights abuses Qatar World Cup	569	8	rights, human, hrwfint, fiacat_org, penalreformint
	Iran human rights protest Qatar	594	0	saytheirnames, irnusa, voice, names
	Iran protester death sentence Mahsa Amini	2,631	0	sentenced, majidrezarahnava, tortured
	Iran protests death sentences 2022	3,257	6	saytheirnames, sentenced, death, irnusa
	Iran protests and FIFA World Cup	3,365	0	iran, fifairanian, mahsaamini, islamic, people
	Iran protests during FIFA World Cup	4,281	108	iran, iranian, team, people, islamic
	Iran Protests Javanroud FIFA World Cup Crimes	3,968	0	islamic, crimes, javanroud, republic, mahsaamini
	Iran protests, children killed by regime	3,476	0	saytheirnames, irnusa, killed, regime
	Iran regime kills children	1,241	0	saytheirnames, children, dehgolan, regime, iran
	Iran women's rights protests in Qatar	1,492	22	women, qatar, iranian, woman, iran
	Iranian fan protests and repression	2,237	5	saytheirnames, sarina, sahar, khodayari
	Iranian footballer arrested for protests	2,157	13	arrested, iranian, voriaghafouri, ghafouri, voria
	Iranian footballer Voria Ghafouri arrested	1,076	0	voria, voriaghafouri, iranian, kidnapped, ghafouri
	Iranian victims of regime violence	2,286	10	saytheirnames, islamic, anzali, regime
	Iranian youth killed, football fans	5,284	0	saytheirnames, islamic, murdered, engirn
	Iran protests FIFA World Cup inclusion	2,742	0	fifabts_bighit, fifa, fifacom, cristiano
Islam		2,122	13	islam, muslim, qatar, muslims, islamic
LGBTQ+ rights		5,902	317	
	LGBTQ+ rights and homophobia in football	1,440	38	gay, lgbtq, lgbt, fifa, homophobic
	One Love armband	1,929	235	armband, fifa, armbands, wear, onelove
	Qatar LGBTQ+ rights controversy	2,533	44	qatar, lgbtq, gay, lgbt, rights
Migrant workers deaths		1,335	60	workers, migrant, qatar, worker, died
Palestine (solidarity with)		2,461	9	palestine, palestinian, israel, israeli, palestinians
Qatar		9,656	222	
	Qatar World Cup controversies and jokes	3,709	22	qatar, country, people, going
	Qatar World Cup controversy and hosting	3,393	164	qatar, host, fifa, hosting, fifacountry
	Qatar's global investments and energy deals	2,554	36	qatar, gas, lng, money, oil
Racism		1,681	41	racist, racism, black, people, players

Rigged World Cup (in favour of Messi)	1,018	2	rigged, messi, argentina, fifa, fifawin
Western media bias	4,742	29	qatar, rights, human, people, country

Other				
Topic	Description	Twitter	News	Top words
Mix of relevant topics		2,892	6	qatar, qatar, team, love
-		2,571	0	fifa, fifa, messi , game, final
-		2,700	0	fifabts_bighit, fifa, proud, love, thank
-		735	261	qatar, newsbreak, markets, sp500, stock
-		4,068	0	qatar, fifaqatari, chizzy, doha
-		3,416	251	fifa fifa, final, game, goal,
-		3,522	1	qatar2022, fifaqatar, qatar, game,
Congratulations		2,750	9	proud, congratulations, team, , thank
EA FIFA game		2,566	21	fifa, zcloaknetwork, zk, game, tutorial
Mentions of crying		1,190	59	fifatears, crying, ronaldo, suarez, messi

Commercial				
Topic	Description	Twitter	News	Top words
Football cards		1	0	panini, waxstat, hobbybox, price, hobby
Football podcasts		836	5	podcast, episode, podcasts, listen, fifaspotify
Gen. promo content		1,576	16	fifa adoptmepets, animallovers, rescuedogs, best
Jerseys, merch		1,718	21	infobox, fullname, birth_date, description, taksim
Merch and football-related products		5,837	142	
Christmas shopping and football gifts		1,018	63	christmas, gift, santa, fifachristmas2022, december
World Cup jerseys and merchandise		2,796	57	fifaqatar, fifa, 2022, opensea
World Cup merchandise and sales promotions		2,023	22	taksim, fifazftw18, ufc280, sale, yellowstone
NFTs, crypto, betting		6,662	16	
FIFA World Cup Crypto Tokens		1,489	3	fwctoken, exchanges, fwcommunity_q2t, token
FIFA World Cup giveaways and prizes		641	1	giveaway, comeearthhq, prizes, prize, fifawin
Metaverse World Cup NFT and Token Giveaways		53	0	kaosland, neer, selfie, comeearthhq, komunitas1
NFT and crypto World Cup giveaways		996	1	fifaproject, giveaway, airdrop, unitos_io
NFTs, crypto, and football betting		1,206	4	nft, ultiball,ulti, 1e, ultiversedao
World Cup scores and winner predictions		1,546	7	prediction, predictions, fifapredict, win, vs
World Cup NFTs and Crypto Collectibles		731	0	nft, nfts, 8650, souffl3, aptosnft
Streaming		8,012	223	
FIFA World Cup streaming platforms		1,748	41	fifafifa, watch, jiocinema, streaming, live
World Cup live stream links		2,058	154	live, vs, stream, 2022, fifa
World Cup live streams and predictions		2,942	23	coinstats, today, live, join, day
World Cup streaming service deals		1,264	5	bulan, vidio, fifadevice, private, 2u
Tech ads		22	0	
Hisense FIFA World Cup predictions contest		20	0	hisensesa, hisense_ind, perfectmatch
iQOO giveaway FIFA World Cup		2	0	iqooin, iqoogameoffones, iqoomegagiveaway

Noise			
--------------	--	--	--

Topic	Description	Twitter	News	Top words
American football		14,674	65	
	College football conference rivalries and playoffs	715	3	college, michigan, state, game, ohio
	NFL Fan Merchandise and Collectibles	131	0	nffans, topfanscorner, nfl, hurry, price
	Mike Leach dies, college football legend	235	1	leach, mike, mississippi, coach, college
	College football awards and championships	1,150	0	thank, great, congratulations, team, congrats
	College football coaching carousel hires	749	9	college, coach, deion, auburn, hbcu
	College football coaching hires and searches	502	1	coach, head, assistant, school, college
	College football player awards and recognition	710	7	college, school, gpa, team, player
	College Football Playoff Rankings	626	2	college, playoff, michigan, ohio, tcu
	College football recruitment offers and news	1,076	0	offer, blessed, agtg, receive
	College football season and bowl games	640	6	college, michigan, bowl, state, game
	College football team discussions	593	0	college, clemsonfb, utah, ucf, umich
	College Football Transfer Portal Tracker	114	1	transfer, portal, college, entered, players
	Fantasy Football NFL Waiver Wire	336	0	fantasy, week, nfl, waiver, fantasy dfs
	High School Football Awards and Trading Cards	330	1	2022, 2023, sportscards, nfl, season
	High School Football Playoffs & Championships	1,116	0	game, live, 58tt5tupb5, school
	High School Football State Championships	573	0	state, championship, school, high, division
	NFL and College Football Talk	879	7	nfl, qb, nflnews, superbowl, best
	NFL betting picks and predictions	482	2	nfl, nflnews, superbowl, week, night
	NFL Football News and Updates	190	1	nationalfootballleague, nfl, nflnews
	NFL game streams and schedules	529	11	vs, nfl, live, nflstreams4ktv, streams
	NFL Teams Discussing Current Games	846	5	team, nfl, game, cowboys, win
	SEC college football playoff rankings	257	4	tennessee, georgia, lsu, alabama, college
	Texas high school football playoff games	400	2	td, touchdown, yards, yard, txhsfb
	US High School Football Games	1,401	2	game, night, week, nfl, today
	Utah Utes Football Pac-12 Championship	94	0	utah, byu, goutes, usc, utes
Australian football		1,359	70	australia, socceros, melbourne, aussie
Automated/copy pasta content		12,914	131	
	Sports cards and collectibles	2,383	3	ebay, brunsoncase, gift, brunson, brunsonvsadams
	World Cup Statistics and Records	1,011	3	3071331839, bank, vs, statstate, credit
	Iran protests	8,387	5	
	Baluchistan genocide executions	211	0	balouch, genocide, irnusa, baloch, executed
	Javanrud protests Mahsa Amini regime	10	0	unarmed, mullah, javanrud, irgterrorists, regime
	Kurdistan protests crackdown	3,304	0	javanrud, islamic, javanroud, republic, people
	Protests and Zahedan massacre	579	0	zahedan, walirn, irgc, massacre, people
	Zahedan, Balouchestan, Mahsa Amini	28	0	balouchestan, zahedan, protesters, armed, footbal
	Regime human rights abuses	1,097	0	islamic, deceitful, republic, irnusa, iran
	Protests during World Cup	3,003	5	regime, team, iran, islamic, iranian
	Regime using football for propaganda	34	0	islamic, dictatorships, republic, terrorists, trap
	Iranian footballer arrested for criticism	69	0	kurdish, voriaghafouri, r2pforiran, arrested
	Saman Yasin death penalty protests	52	0	samanyasin, kermanshah, detention, seidi
	Jungkook	171	0	
	Video copyright issue	11	0	bangtantv, unavailable, copyright
	YouTube Chart Issue	2	0	hybeofficialtw, youtube, unitedmasters

-	28	0	teamyoutube, dreamers, unitedmasters
-	13	0	counted, calculating, uploaded, lyrics, official
Promoting “Dreamers”	99	0	bts_twt, jungkook, dreamers, fifasingle
-	2	0	bts_twt, amazed, jungkook, dreamers
Streaming “Dreamers”	16	0	bts_twt, jungkook, dreamers, fifastream
Panini FIFA World Cup Trading Cards	32	1	panini, sportscards, ad, thehobby, prizm
Qatar Airways lost luggage complaints	6	0	qatarairways, resolving, compensation
Youngest players’ careers	924	119	scouting, years, played, date, game
EU/Qatar bribe	1,048	38	parliament, qatar, eu, scandal, corruption
Gaelic football	1,490	106	championship, munster, meathgaa, rathmore
Hashtag hijacking	85	6	newsbreak, sp500, markets, stock, baseball
Not English/UK	4,462	18	
South Asian football fans discuss World Cup	2,192	4	hai, fifabhi, fifa, ki, mein
India football vs cricket debate	1,896	10	cricket, india, indian, fifafifa, indian sports
Thanksgiving football tradition	374	4	thanksgiving, day, happy, family, thankful
Random content	2,162	124	
Football related crime/tragedies	1,097	124	vagina, assaulted, raped, tehran, arrested
Short tweets with no real content	1,065	0	fifaa23_fantasy, chalosaathkhelein

E. Sentiment Analysis

We prompted Gemini 2.5 Flash Lite with the following prompt on a sample of news excerpts and tweets from each topic (10% of each topic capped between 50 and 300 datapoints) using the following prompt:

```
CONTEXT: Here are [tweets/news excerpts] related to the 2022 FIFA World Cup.
TASK: Classify the sentiment expressed in the [tweet/excerpt] as ``Very negative'', ``Somewhat negative'', ``Neutral'', ``Somewhat positive'' or ``Very positive''.
CONSTRAINTS: Return all labels on a single line separated by commas in the same order as the corresponding input [tweets/excerpts]. Do NOT return any other text than the comma-separated labels (i.e. no explanation, no comments, no formatting).
```

We then converted each label into a score between 0 (“Very negative”) and 1 (“Very positive”). The first author then manually annotated a sample of 200 datapoints (100 per source) to evaluate the performance of the model. If we conceive of this task as a multiclass classification task (which doesn’t take into account the proximity between classes), we obtain a macro-average F1-score of 0.83 for Twitter data and 0.86 for news data (see confusion matrices In Figure 4). We use the sentiment scores we obtained to calculate the average sentiment per topic and source (see Table 12 below).

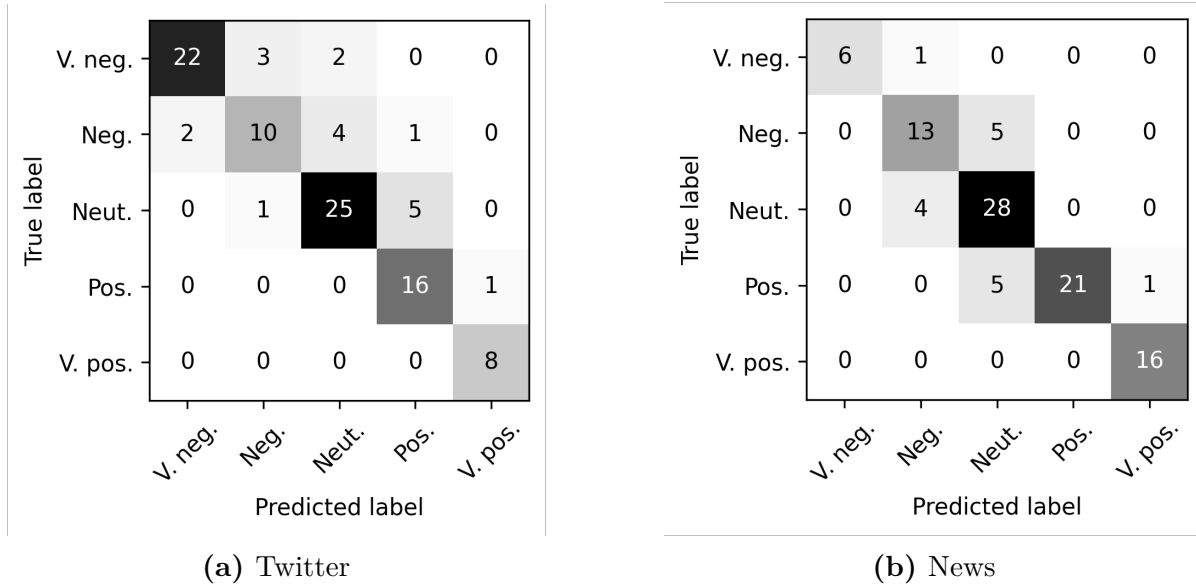


Figure 4. Confusion matrices for fine-grained sentiment classification task for sample corresponding to each corpus (Twitter vs News).

Table 12: Average sentiment per topic and source (news vs Twitter) and ratio of sentiment score per topic between the two corpora. A ratio greater than 1 signifies that the topic is more positive in the news than on Twitter, while a ratio lesser than -1 means it was more positive on Twitter. Asterisks signify the proportion of the topic that is political by averaging across the news and Twitter corpora (***: more than 75%; **: more than 50%; *: more than 25%).

Topic	News	Twitter	Ratio	Topic	News	Twitter	Ratio
*** Iran protests	0.32	0.13	2.58	General football content	0.45	0.48	-1.08
*** Palestine	0.50	0.39	1.27	Teams	0.54	0.58	-1.08
* Football and money	0.50	0.40	1.26	*** Islam	0.44	0.48	-1.09
VAR/ref. controversies	0.38	0.30	1.25	Watching WC/football	0.50	0.55	-1.11
* Women and football	0.54	0.44	1.24	World Cup songs	0.56	0.63	-1.11
Football trolling	0.43	0.36	1.20	*** Flags	0.44	0.50	-1.15
*** Football vs politics	0.36	0.30	1.19	Opinions about football	0.54	0.62	-1.15
British football clubs	0.54	0.46	1.18	Games	0.51	0.59	-1.16
*** Western media bias	0.33	0.28	1.15	Other	0.56	0.67	-1.20
** Alcohol ban	0.51	0.45	1.13	* Stadium/infrastructure	0.40	0.48	-1.21
** Female referees	0.79	0.71	1.12	* Public figures in Qatar	0.42	0.51	-1.23
Football discussions	0.44	0.40	1.12	* Opening ceremony	0.50	0.65	-1.30
Reacting to action	0.70	0.63	1.11	*** Human rights	0.19	0.25	-1.30
*** Racism	0.23	0.21	1.09	* Best World Cup	0.61	0.84	-1.37
* In Qatar	0.54	0.51	1.04	*** FIFA corruption	0.18	0.25	-1.41
World Cup format	0.54	0.52	1.03	*** Migrant workers deaths	0.09	0.13	-1.50
Predictions and results	0.55	0.54	1.01	*** General political content	0.13	0.29	-2.34
Players	0.59	0.60	-1.02	** Rigged World Cup		0.21	
** Qatar	0.39	0.40	-1.02	Jungkook		0.72	
*** LGBTQ+ rights	0.26	0.27	-1.03	“Football” vs “soccer”		0.46	
* Broadcasting and press	0.38	0.40	-1.07				

F. Equations

The news and tweets proportions for a given topic t is given by the following two equations respectively:

$$N_t = \frac{\text{count}(\text{news}, t)}{\sum_{i \in T} \text{count}(\text{news}, i)} \quad (1)$$

$$T_t = \frac{\text{count}(\text{tweets}, t)}{\sum_{i \in T} \text{count}(\text{tweets}, i)} \quad (2)$$

...where T is the set of all topics. The ratio of $News_t$ over $Tweets_t$ when $News_t > Tweets_t$, and inversely of $Tweets_t$ over $News_t$ when $Tweets_t > News_t$, then tells us which topics are under- or over-represented in the news compared to Twitter. We calculate the prevalence $P_{t,c}$ of each topic t for each cluster c by cross-referencing cluster labels with the topic labels from our Semantic Analysis:

$$P_{t,c} = \frac{\text{count}(t, c)}{\sum_{i \in T} \text{count}(i, c)} \quad (3)$$

...where T is the set of all topics. From this we calculate the probability $q_{t,c}$ that topic t appears in cluster c :

$$q_{t,c} = \frac{P_{t,c}}{\sum_{i \in C} P_{t,i}} \quad (4)$$

...where C is the set of all clusters and $P_{t,c}$ is the prevalence of topic t in cluster c , as presented in Equation 3. Using the resulting probability distribution of each topic over all clusters, we can then calculate the Shannon Entropy $H(t)$ (normalised to range from 0 to 1) for each topic t , where lower scores signify that the topic is highly concentrated in one or a few clusters:

$$H(t) = -\frac{\sum_{i \in C} q_{t,i} \log(q_{t,i})}{\log(|C|)} \quad (5)$$

G. Journalist keywords match per cluster

To identify journalists Twitter accounts, we start by inspecting a random sample of 100 accounts from the UK subset of journalist accounts dataset by Toprak et al. (2022) and identifying common keywords in their user descriptions. We do this instead of using the list of accounts in that dataset directly because the dataset was collected in a different year than our own (2018) and potentially covers quite different news beats. The keywords we identify are the following: “*journalist*”, “*reporter*”, “*editor*”, “*correspondent*”, “*writer*”, “*columnist*”, “*contributor*”, “*broadcaster*”, “*anchor*”, “*producer*”, “*commentator*”, “*photojournalist*”, “*news*”, “*media*”, “*press*” and “*presenter*”. We then use these keywords to identify candidate journalist accounts in our own dataset. We sub-sample 20 accounts per keywords to verify the quality of the matches. Based on this check, we remove the keywords “*media*”, “*press*”, “*news*” and “*commentator*”, which led to a lot of false positives. We also require our journalist accounts to have explicitly declared a UK location. This reduces the number of accounts identified but ensures a higher quality in the sample. Table 13 below shows the number of matches per keyword.

Table 13: Number of Twitter accounts per cluster matching each journalist keywords in their user description. Note that the same account can match several keywords.

Keyword	All accounts	Football	World Cup	Arabic/Muslim	Iranian	News	Gaming	Others
journalist	80	32	3	9	0	17	1	18
reporter	36	21	1	0	0	7	1	6
editor	55	24	0	6	0	16	1	8
correspondent	23	10	0	1	1	7	0	4
writer	100	53	2	9	0	15	2	19
columnist	9	1	0	3	0	3	0	2
contributor	21	5	2	2	0	2	0	10
broadcaster	13	9	0	0	0	3	0	1
anchor	1	0	0	0	0	1	0	0
producer	28	10	0	2	2	8	0	6
photojournalist	0	0	0	0	0	0	0	0
presenter	41	23	1	0	0	9	2	6