Published in *Information Discovery and Delivery* by Emerald. Guzmán, E.M., Zhang, Z. and Ahmed, W. (2021), "Towards understanding a football club's social media network: an exploratory case study of Manchester United", *Information Discovery and Delivery*, Vol. 49 No. 1, pp. 71-83. https://doi.org/10.1108/IDD-08-2020-0106.

Understanding a Football Club's Social Media Network: An Exploratory Case Study of Manchester United

Abstract

Purpose: Social media, particularly Twitter, has been extensively utilised by football fans forming dedicated online communities, and has become an essential part of communication and marketing strategies for most football clubs. However, there is a lack of in-depth understanding of the stakeholders and their interactions on such online platforms, such as the patterns of interaction among different stakeholders, the most influential actors, and the topics of interest in their communications.
Research design: We analysed the social networks derived from over two million tweets collected during football matches played by Manchester United. We applied social network analysis to discover influencers and sub-communities, and performed content analysis on the most popular tweets of the prominent influencers.
Originality. Compared to previous research we discovered a wide range of influencers and denser networks characterised by a smaller number of large clusters. Interestingly, our study also found that bots appeared to become influential within the network.
Findings: Sub-communities can be formed around current affairs that are irrelevant to football, perhaps due to opportunistic attempts of using the large networks and massive attention during football matches to disseminate information. Furthermore, the popularity of tweets featuring different topics depends on the types of influencers involved.
Practical Implications: Our methods can help football clubs develop a deeper understanding of their online social communities. Our findings can also inform football clubs on how to optimise their communication strategies by utilising various influencers.

Keywords: social media, network analysis, Manchester United, football, content analysis

Introduction

Social media has become an essential part of the communication strategies for many organisations, because it has contributed to revenue growth and public relation management for many businesses. Sports brands are investing significant time and resources to drive engagement and relationships with their customers online. In particular, the role of social media in the football industry has attracted increasing attention from researchers and practitioners alike (McCarthy *et al.*, 2014). The financial success of a football club largely depends on the effective management of a strong fan base. Compared to other businesses, fan bases embed complex community relationships beyond typical transaction-based customer relationships (McCarthy *et al.*, 2014). With the extensive use of social media in the football industry including clubs, players, and their fans, it is becoming increasingly important to develop an understanding of their interactions within these online spaces.

The analysis of the composition of these stakeholders on social media, their interaction patterns and information needs has been an underexplored subject. We argue that it is important to develop a deep understanding of the social networks formed by these stakeholders, as well as their communication. Social network analysis can be used to discover influential actors, the presence of sub-communities, and interesting interaction patterns. A content analysis of their communication will help reveal information needs and provision among different stakeholders. This information could be used as intelligence because it may provide football clubs the ability to strategically target influential actors for the purpose of information dissemination and/or social listening.

In this study, we conduct an exploratory analysis of the social network surrounding of Manchester United (MU) on Twitter and follow this up with content analysis. MU is among one of the most successful football clubs in the world and the most valuable football club worldwide in revenue terms (Ozanian, 2019).

Our goal is to answer three research questions:

> RQ1. What are the patterns of interaction among different stakeholders in the MU network on Twitter?
>
> RQ2. Who are the most influential actors within this network and how do sub-communities emerge?
>
> RQ3. What are the topics of interest in communication and how are they manifested in the different stakeholders?

Our work provides multiple contributions to the literature. Firstly, this is the first empirical work that examines social network structures and content sharing in social media communities focused on a specific football club. Earlier work instead, has been based around football events (Yan *et al.*, 2019), or other sports (Clavio *et al.*, 2012; Meng *et al.*, 2015). Henceforth, our findings bring insights to football clubs on how to understand their stakeholders' interactions, which is potentially useful to their marketing and communication strategies. Secondly, our method combines social network analysis and content analysis, where the latter is informed by findings from the former.

Our analysis revealed that MU's networks are typical scale-free networks where degree distributions follow a power-law. This is characterised by a network containing a very large number of actors, but sparse connections, most of which are concentrated on a small set of actors. The networks were influenced not only by football clubs and footballers, but also by a diverse range of actors including bots, journalists, artists (including actors, singers, etc.), and politicians. Further, we found that within these networks, distant communities can form around current affairs irrelevant to football, particularly politics. By analysing the most popular tweets involving highly influential actors, we discovered a wide range of different topics that attract significant popularity in the network. Furthermore, different types of influencers may be involved in communicating particular types of content. These findings are useful for football clubs to optimise their communication strategies for maximum impact.

Related Literature

We discuss related work from three perspectives: the broad categories of social media research in sports, studies of social media network analysis in sports, and studies of social media content analysis in sports.

Categories of social media research in sports

Filo *et. al*. (2015) identified three categories of social media research in the sports domain: strategic, operational, and user-focussed. Research on the strategic use of social media focuses on the role of social media in brand development. Questionnaires and interviews are often used to assess the relevance of social media to brand equity, managers' attitudes towards social media, and to evaluate how social media can be integrated into a brand's communication strategy. For example, McCarthy *et al.* (2014) investigated the issues and benefits associated with managing social media using a case study based approach focused on four well-known

UK football clubs. Other studies of a similar nature include those for football (Anagnostopoulos et al., 2018; Manoli, 2020), baseball, basketball and ice hockey (Abeza et al., 2019).

Social media research has also been conducted to support operational decisions and to explore how brands and their fans use social media on a daily basis. Herein, sentiment analysis has been widely used to analyse large sports events such as the World Cup (Lucas *et al.,* 2017) and Olympic Games (Kassens-Noor *et al.,* 2019).

The user-focused category of studies include research on fan profiling and market composition. Twitter data has been used to characterise football supporters (Pacheco *et al.*, 2016), to identify nexus between politics and sports (Hayat *et al.*, 2016), and to examine the relationships between players and fans in the context of University American football (Yan *et al.*, 2018).

Social network analysis in sports

Social Network Analysis (SNA) is a sociological approach to analyse patterns of relationships between social actors in order to discover hidden social structures within a network (Wasserman and Faust, 1994). A network is viewed as a graph, with the actors as nodes and their relationships as edges. Most of the analytical methods can be divided into those that aim to quantify the relevance of a node, discover sub-communities, or understand information propagation.

The detection of prominent nodes in a network has been widely employed in the context of politics, misinformation and emergency response. For example, Bovet and Makse (2019) used social network analysis to study how fake news on Twitter influenced the 2016 US election. Rowe and Pitfield (2018) explored the characteristics of a social activist network based on an activist group's Twitter followers. In regard to previous research which examined information propagation through social networks, Ahmed *et al.* (2020) examined misinformation content shared around a conspiracy theory linking 5G to COVID-19 and highlighted how prominent users helped spread the conspiracy. Khajeheian and Kolli (2020) utilised SNA to develop an understanding of how the game Pokémon Go promoted the social relationship of users on Twitter. Community detection is a task of identifying groups or clusters of nodes that interact with each other more often than those outside the group (Ozer *et al.*, 2016). For example, Komorowski *et al.* (2017) detected clusters of Twitter users that share similar political orientation, sociodemographic attributes and/or polarity of their messages, based on their Twitter interactions (e.g., retweet, reply, follow).

SNA has been widely used to study team dynamics in sports (Mclean *et al*., 2018), or other topics such as the European football loan system (Bond *et al.*, 2019). However, SNA's application to social media use in a sport context has been limited (Clavio *et al.,* 2015). Yan *et al.* (2019) studied the social network of Twitter users that posted during the 2017 UEFA Champions League Final. Tweets containing certain hashtags were collected before and after the match, as well as during the half-time, and a social network was created for each of the three sets of tweets. The authors then applied the ForceAtlas2 algorithm (Jacomy *et al.,* 2014) to visualise the networks, and statistical metrics such as eigenvector centrality to discover prominent nodes and clusters. The study showed that the Twitter networks had the effect of promoting large sport entities such as clubs and players, while individual citizens had limited influence. It also showed that emergent game dynamics helped shape the structures of the networks as they created clustered discussions. As we shall show later, our study used similar methods but discovered different insights based on our dataset.

Other SNA studies related to sports normally follow similar approaches: a collection of social media posts matching certain criteria, usually based on hashtags or user accounts, are collected. Then nodes and edges are identified based on users and their interactions to form a network. Statistical measures are then used to identify prominent nodes, clusters, or interactions. Clavio *et al.* (2015) studied an American college football team's Twitter community and revealed that fan accounts composed the largest percentage of the network which was characterised with sparse reciprocal interactions. Typically, media accounts frequently interact with each other while fans interact primarily with other fans. Naraine *et al.* (2019) studied the Twitter community of an NBA team and discovered a 'tight knit' network of subcommunities that are not solely associated with the team or even basketball, but also associated to the topical keywords extracted from the network. In the cycling domain, Lamirán-Palomares *et al.* (2019) and Baviera-Puig (2020) used SNA to discover opinion leaders during an international cycling event. They showed that different network metrics may be used to identify different kinds of influencers. They argued that such results could be used as intelligence for companies interested in sponsoring cyclists and teams.

Content analysis in sports

Content analysis (Krippendorff, 2004) is a method of interpreting data through the systematic classification process of coding to identify recurring patterns in data. It is often followed by statistical analysis of the patterns that are identified. Achen *et al.* (2020) analysed posts collected from users on Twitter and Facebook that followed several US professional sports leagues, such as, the NBA, NFL, and MLB. A content analysis showed that both networks were used most often for player and personnel promotion and fans interacted most often with such content. Meng *et al.* (2015) analysed online comments posted by NBA teams on their Facebook and Twitter accounts during the off-season. They revealed four types of communication that were often used to engage fans: informing (e.g., news), marketing (e.g., promotion, sales), personalising (e.g., initial contact, direct response) and activating (e.g., group involvement, gathering feedback). Winard *et al*. (2018) analysed content from FIFA's Twitter account and revealed that FIFA did not fully utilise the potential of Twitter as a channel of communication with football fans.

Coche (2017) studied Twitter profiles of athletes to analyse how they present themselves on social media using images and photos. Geurin and McNary (2020) examined U.S. Olympic athletes' Instagram posts covering a six-week time period around the 2016 Rio Olympic Games in order to discover if any athletes violated Rule 40 of the Olympic Charter, designed to prevent athletes from posting about non-Olympic sponsors.

From a fans' perspective, Stavros *et al.* (2014) analysed online comments by fans of NBA teams to understand their motivations of engagement. These were found to be generally categorised into passion (e.g., affection for the team), hope (e.g., ambition), esteem (e.g., where comments demonstrate their specialist knowledge and insight) and camaraderie (e.g., defending the team). A similar study (Schubert and Seyffert, 2017), in the context of table tennis, discovered eight categories of fan motives when communicating with the International Table Tennis Federation (ITTF) on its Facebook page.

Gaps in Current Research

Our literature review suggests that there is a lack of previous empirical research on social media communication among different stakeholders in the football domain. Although, related work has been conducted in other sports it is unclear whether the findings are transferable. A part of our methods are similar to those in Yan *et al.* (2019) but our study reveals different findings. Furthermore, while previous work primarily used social network analysis and content analysis separately we argue that combining the two may help discover further insights. For instance, it may be particularly useful to analyse the content created by the most influential actors in the social media network because this helps understand how different kinds of information is spread through the network.

Methodology

Our approach is informed by the theory of explanatory models of information diffusion (Li *et al.*, (2017). Such models represent interactions between individuals in 'real' society using a network connecting nodes (representing individuals) and edges (representing relations), and hypothesise that information is spread through these interactions within such networks. Explanatory models aim to study properties about the information diffusion process, such as, the main factors affecting information diffusion and the influencing nodes. Specifically, we adopt the 'influence models' in social networks (Chaudhury *et al.*, 2012; Sadri *et al.*, 2018) that focus on individual influence, community influence, and influence maximisation (*Li et al.*, 2017). Such models attempt to explain information diffusion through influence search. Individual influence refers to actors (influencers) that play a role as a bridge of information diffusion in a network. Community influence refers to actors that form cohesive groups based on interest, such that the information diffusion within the community is more intense than outside the community. Influence maximisation is focused on where to start (e.g., which influencer, or community) the information diffusion to maximise information propagation through a network. Thus, built on these theories, our approach first identifies influencers and communities in the social network. It then adopts a content analysis to develop a deeper understanding of the type of information diffused through the network. Figure 1 presents an overview of the flow of our methodology. The different components are discussed in detail in the following sections

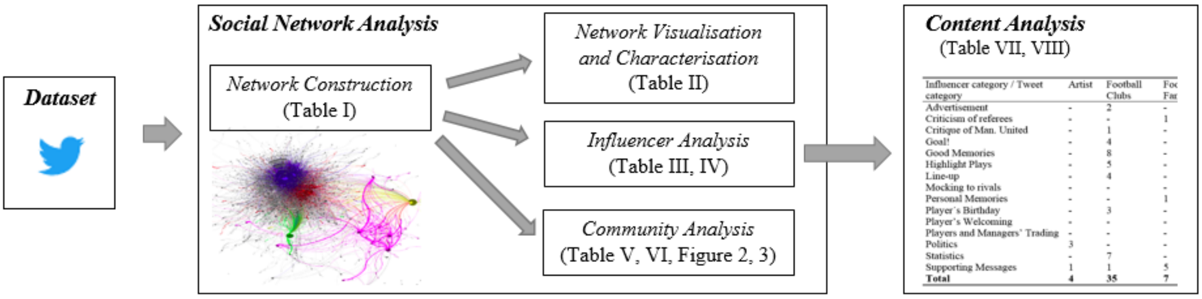

Fig. 1. Overview of our methodology. ***Bold italic*** and *italic* labels correspond to the ***section*** and *sub-section* headings in our **Methodology** section. Tables and Figures in brackets show how each part of results relates to each part of our methodology. Grey arrows indicate how each part of our methodology informs other parts.

Dataset

We selected Manchester United (MU) as our case study in this work and collected a Twitter dataset covering one month between November and December of 2018. MU is currently the most valuable football club worldwide in revenue terms (Ozanian, 2019). The Twitter Streaming API was used during this time to collect any tweets containing one of the hashtags, #manutd, and #manchesterunited (all case in-sentitive). A total of 2,293,156 tweets were collected from 587,713 distinct Twitter users. These were further filtered to only contain tweets generated on the day when MU played a Premier League match, as per Chakrabarti and Punera (2011), who found that interesting tweets and user interactions are mostly concentrated on the days of sports events. There were a total of 4 matches as listed in Table I, and we refer to tweets collected for each of these matches as ‘sub-datasets’.

Table I. Matches played by Manchester United during the data collection period.

Opponent	Date	Away/Home
Manchester City	11 Nov 2018	Away
Crystal Palace	24 Nov 2018	Home
Southampton	01 Dec 2018	Away
Arsenal	05 Dec 2018	Home

Social network analysis

Network Construction

Following previous studies (Sadri *et al.*, 2018), we identify unique Twitter users as nodes of a network, and their interactions as edges in the network. Unique Twitter users are identified as those that created a tweet (including retweet), were a target of reply, or mentioned (i.e., using ‘@’) in a tweet. Interactions can be either a mention, or reply (treated indifferently). Note that either ‘mention’ or ‘reply’ will result in a directed edge, with the user mentioning or replying to others being the source node, and those mentioned or replied to by the tweet as target nodes. Further, each edge is weighted by the frequency of the interactions between two nodes in the dataset, following Sadri *et al* (2018)’s theory of capturing strength of interactions within such networks. We applied this procedure to each sub-dataset depending on the match played to create four networks for analysis.

Network Visualisation and Characterisation

We apply the ForceAtlas2 layout algorithm (Jacomy *et al.*, 2014) in Gephi[1] to plot each network. ForceAtlas2 visualises a network by taking into account the intensity of interactions, such that nodes with more frequent interactions are placed closer to each other.

To describe the global characteristics of each network, we apply the degree centrality and density measures commonly used in social network analysis (Cheong and Cheong, 2011). Degree centrality measures the average number of direct links that a node has in a network. An ‘out-degree’ of a node is the number of its outgoing edges, while an ‘in-degree’ of a node is the number of its incoming edges. The out- and in-degree of the network is then the average of the out- and in-degrees of its nodes. Density refers to the proportion of connections that exists within a network over its maximum number of possible connections. A network of more densely connected nodes indicates more extensive communication between them (Scott, 2012).

Influencer Analysis

To capture ‘individual influence’, we identify the most prominent nodes or actors in each network. We use two popular measures for this purpose: eigenvector centrality and PageRank, which are typical methods for studying influencing nodes in a network (Li *et al.*, 2017). Eigenvector centrality evaluates the relevance of nodes within networks. Scores are calculated based on the idea that connections to high-scoring nodes contribute more to the score of the node than connections to low-scoring ones. The calculation of eigenvector centrality discounts directions or weights of edges.

PageRank is a variant of eigenvector measure which was originally developed by Google to rank web pages based on their hyperlinks. It starts with a network of nodes with equal weights

1 https://gephi.org/. Last accessed: September 2020

which are refined in a recursive process that calculates the score of a node depending on scores of other nodes that it is connected with. The process ends when scores of all nodes 'converge' i.e., the difference between the score of a node in the previous iteration and the current iteration is below a certain threshold. Compared to eigenvector centrality PageRank treats in- and out-links separately. A variant of PageRank called weighted PageRank (Grover and Wason, 2012), used in our study, also allows edge weights to be considered.

For each network, we apply the above measures to the nodes in the network and identify influencers as top *N* nodes ranked by each measure. We then adopt a manual coding approach to categorise these influencers based on their Twitter bio.

Community Analysis

To capture 'community influence', we identify cohesive groups (i.e., communities) of nodes that are highly interconnected to each other compared to other nodes within a network. This can be solved by network optimisation algorithms that try to maximise a benefit function called modularity (De Meo *et al.*, 2011). Modularity measures the strength of sub-structures in a network and is defined as the fraction of the edges within the sub-structures compared to the expected fraction of edges in a randomly distributed scenario (Gach and Hao, 2014). A low modularity indicates the absence of real communities. In contrast, a high modularity implies the presence of highly connected communities.

A popular algorithm utilising the concept of modularity is the Louvain algorithm (Blondel *et al.*, 2008). It begins with each node in its own community and incrementally merges them with their neighbour nodes to form larger communities, only if the merge operation optimizes the modularity scores of the network. This is repeated until there are no more modularity-optimising changes. The Louvain algorithm is considered one of the fastest algorithms for community detection from large networks and is widely used in other social network analysis research (e.g., Gach and Hao, 2014).

For each network, we apply the above algorithm to discover sub-communities within themselves. We then analyse these communities by their content and structure. For content, we apply Named Entity Recognition (NER) to the tweets belonging to each community and analyse the most frequent Named Entities (NE) mentioned in each community. NER identifies and classifies names of entities such as people, organisations and locations. For example, in 'Manchester United will be playing against Sheffield United next Wednesday', 'Manchester United' and 'Sheffield United' are organisations, while 'Wednesday' is a date. The extracted NEs may be indicative of the topic of interest within each community.

For structure, we measure the average in- and out-degrees, and density of each community, and compare them with the overall network.

Content analysis

To address 'influence maximisation', we propose to understand the topics of interest disseminated during these football matches by conducting a content analysis of the most popular tweets of the most prominent actors in the network. The intuition is that such topics may encourage engagement and information diffusion.

First, based on the influencer categories we identify from our influencer analysis before, we select, for each influencer category, the two that are the highest ranked by their PageRank scores (as we shall show later, the rankings by PageRank and eigenvector centrality are largely indifferent).

Second, we select from our dataset tweets either generated by this set of influencers, or contain mentions of them. In the following, we will say that the influencers are 'involved' in these tweets. We then rank these posts based on their number of retweets and select the top 10 tweets for each influencer (for certain influencers, there can be a total of less than 10 tweets in our dataset. In this case, all of their tweets will be selected). An exception of these influencers is MU's Twitter account, for which we select the top 25. On the one hand, MU is, as our results will show, the most influential actor by either eigenvector centrality or PageRank. On the other hand, we are interested in what content of communication often explicitly involves the club. Such content may be compared with the most popular content related to other influencers, to identify correlation or divergence.

Finally, we apply a coding process on these tweets to categorise them into different topics. While the NER analysis applied before as part of the community analysis may reveal topic patterns from a quantitative point of view (as it is applied to the entire dataset based on community groups), the analysis here is qualitative and provides a complementary perspective (as it is conducted manually and from the influencers point of view).

Results

Social network analysis

Table II. Overall network characteristics.

Network (match against)	Total nodes	Total edges	Density	Average in-degree	Average out-degree
Manchester City	23,523	34,478	<0.001	1.46	1.46
Crystal Palace	51,524	92,096	<0.001	1.78	1.78
Southampton	81,028	131,375	<0.001	1.62	1.62
Arsenal	99,861	200,093	<0.001	2.01	2.01

We begin with summarising the characteristics of the social networks. This relates to RQ1. In terms of the size of these networks, Table II shows that MU's match against Manchester City created the smallest network (to be referred to as the 'Manchester City network', likewise for others), while the Arsenal network was significantly larger. The incremental growth of the network size with respect to time could be related to the progress of the league and the UEFA Champions League. For example, by the time of the match against Arsenal, the two teams had won a series of matches in the group stage of the UEFA Champions League, and Arsenal were the leader in the Champions league group. This could have drawn an increasing number of fans, including those of the opponent teams in the UEFA Champions League. The Arsenal network is arguably the most active because it has the highest average in- and out-degrees while the Manchester City network is the least active among the four. All networks feature very low density which is also lower compared to previous research of real-world networks such as a web crawl (Melancon, 2006). We also notice that the actor with the highest in-degree is the same in all four networks and corresponds to MU's Twitter official account (@ManUtd).

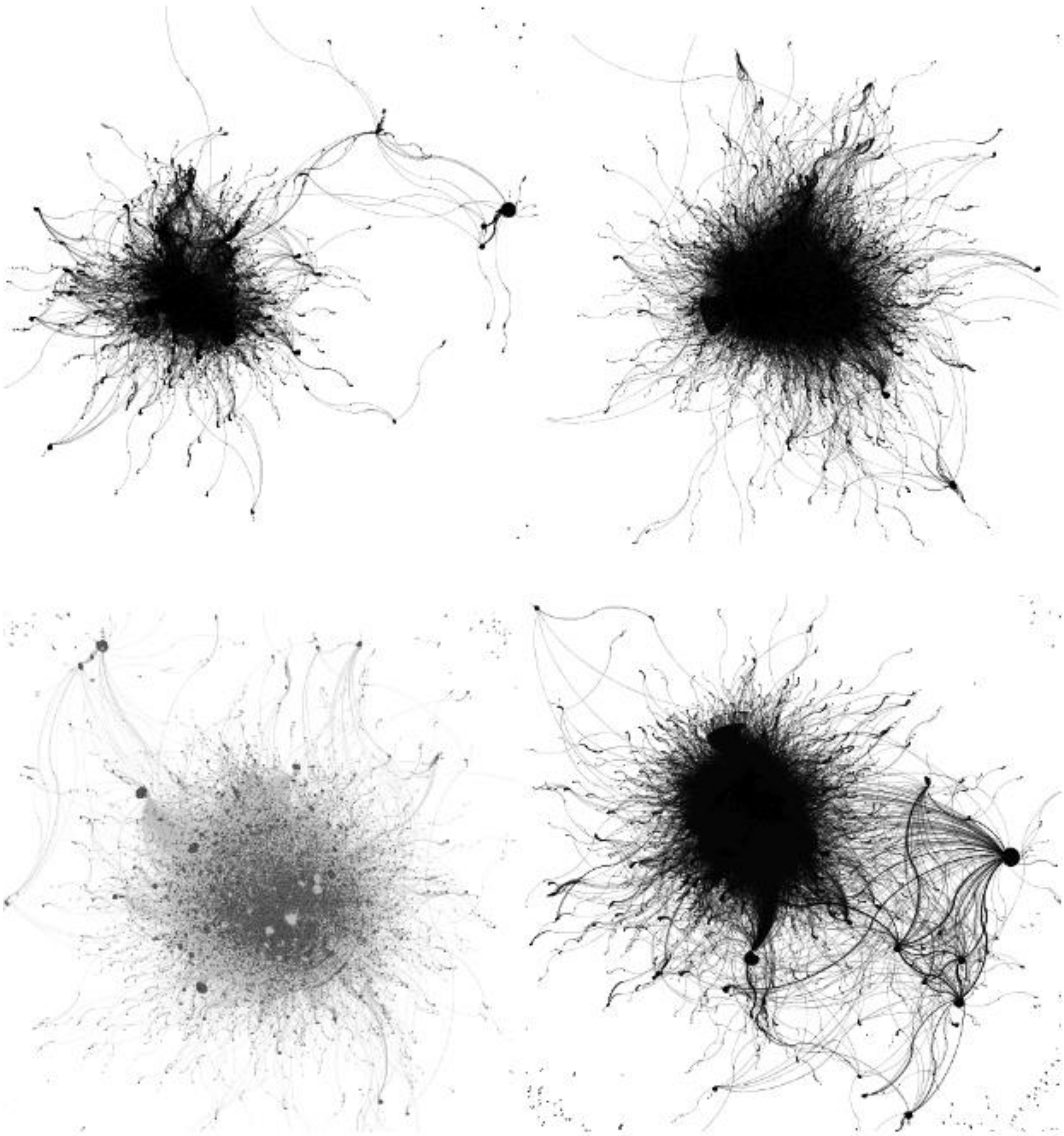

Fig. 2. Visualisation of the four networks. In clockwise order from the top left: match against Manchester City, Crystal Palace, Southampton, and Ars

In terms of interactions within these networks, Figure 2 shows that the Manchester City and Southampton networks share similarities as they are both characterised by a large cluster of actors that are connected to some distant, smaller groups which form clearly different communities. Similar patterns are also noted for both Crystal Palace's and Arsenal's networks which are composed of a principal core that includes most of the interactions between actors with smaller pockets of discussion taking place around this core. As our sample of tweets relates to match-days core clusters could represent discussions surrounding the match including, build-up, the match itself, and after-discussion.

Influencer Analysis

Here we present our findings about influencers from these networks. This relates to RQ2. We observe similar distributions of both the eigenvector centrality and PageRank scores, i.e., they follow a power-law, which is indicative of the scale-free property of most real networks (Sadri *et al.*, 2018). Thus for both measures, we select the top 10 highest ranked nodes from each network for further analysis. Due to the similarity between the two measures we notice a high degree of overlap in the influencers discovered by both measures (100% for the Manchester City and Arsenal networks, 80% Crystal Palace, and 60% Southampton). Therefore, we merge the results from both measures together for analysis.

In terms of the different categories of influencers, Table III summarises their statistics over all the networks. Note in particular that 'average out-degree' indicates the number of tweets containing explicit interaction (i.e., mention or reply) with other actors in the network, while 'average tweets' indicates the overall number of tweets (with or without interaction with other Twitter users) posted by an actor.

Table III. Summary of Influencers identified from all four networks

Type	#Unique users	Average in-degree	Average out-degree	Average tweets	Average followers (thousands)
Academics/Politicians	5	916	1	22	3,176
Artists	3	1,636	24	51	92
Journalists	4	1,880	10	32	263
Football clubs	3	17,370	21	25	13,799
Football fans	2	635	2	2	0.2
Football players	8	4,413	1	2	4,737
News/Humour bots	10	2,829	16	30	417
Organisations	3	2,572	7	21	14,693
Total	38	3811	10	24	3,663

We see a rather diverse range of actors that may have influenced the Twitter communications during the four matches. Among them, football clubs and players had the highest average in-degree suggesting that they were the main targets of interaction. In addition, we also discovered highly influential actors who were academics, politicians, artists, and journalists. Although, only two football fans were found to be highly influential, arguably, some of the above individuals were also football fans. However, we treat them as separate categories in this study[2].

2 If a prominent politician was also a football fan s/he is coded as a politician. Only fans who are not public figures or prominent individuals are coded as fans.

Furthermore, bots played an important role because they represented 10 out of 38 identified influencers. These were identified based on information contained in their Twitter bio, posting frequency, and a review of their latest and most popular posts. They mainly shared updates regarding football clubs without personal opinions. Surprisingly, the average out-degree was low for all influencers compared to their in-degrees. Their influence was, therefore, due to being mentioned or replied to by other actors. In other words, they are very well known on these networks.

Table IV. Individual influencers' Twitter screen name identified from all four networks - identities for public figures and organisations are shown; otherwise they are anonymised.

	Manchester City	**Crystal Palace**	**Southampton**	**Arsenal**
Academics	(2 anonymised)	-	(1 anonymised)	-
Artists	-	Tosin Akingba	Ken Olin	@activist360
Journalists	-	Liam Canning, Andy Mitten, Daniel Storey, Samuel Luckhurst	-	-
Football clubs	Manchester City, Manchester United	Manchester United	Manchester United	Manchester United, Arsenal FC
Football fans	-	(1 anonymised)	-	(1 anonymised)
Football players	Rio Ferdinand	Juan Mata Garcia	David de Gea	Anthony Martial, Alan Shearer
Humour bots	Naija Fans Challenge	-	Losing the love for Football	Troll Football
News bots	United Xtra, Man United Pidgin	United Xtra, Simply Utd	Twitter Moments	United Xtra, B/R Football
Organisations	UEFA Champions League, Heineken	-	Paddy Power	Premier League
Politicians	-	-	Bernie Sanders, David Axelrod	-

In terms of the identities of these influencers Table IV lists some of them for each network. 'Artists' refer to its broader sense of individuals specialising in the creation and performance of artistic work, including singers, film actors, painters etc. Journalists were classed as anyone who is part of the editorial process of gathering and disseminating news, thus, can include news reporters, presenters, columnists etc.

One of the academics is a US former federal prosecutor and university lecturer who regularly posts and shares messages against Trump's social policies. For politicians, Bernard Sanders (@BernieSanders) is an American politician known for his opposition to the Trump administration; while David Axelrod (@davidaxelrod) is a US political consultant, known for his participation in Barack Obama's presidential campaigns. As we shall discuss later, the influence of these individuals were likely due to the emergence of a separate sub-community that discussed political matters in the Manchester City and Southampton matches.

Regarding artists, Tosin Akingba (@venusakingba) is a fan of MU and was mentioned often by other MU fans. Ken Olin (@kenolin1) shared a lot of political messages and was therefore, likely involved in the sub-community mentioned above. The third artist is an actor whose profile no longer exists by the time of this analysis.

Regarding journalists, Liam Canning (@LiamPaulCanning) is a freelance writer specialising in European football, particularly the Premier League. He is also an MU fan. Andy Mitten (@AndyMitten) is an author and founder of the best-selling magazine ‘United We Stand’, an independent MU fan-based publication. Similarly, Daniel Storey (@danielstorey85) is an English journalist specialised in football and was named ‘Football Writer of the Year’ in 2016. Finally, Samuel Luckhurst (@samuelluckhurst) is the Chief MU writer for the Manchester Evening News.

The bots were mentioned because they are mostly specialised in MU’s news. United Xtra (@utdxtra), Man United Pidgin (@ManUtdInPidgin), and Simply Utd (@SimplyUtd) shared regular updates regarding matches, results, players, trades and interviews, or any relevant news related to MU. Interestingly, two of the humour bot accounts, i.e., ‘Naija Fans Challenge’ (@NFFCshow) and ‘Losing the love for Football’ (@StevenUtd_) were suspended by the time of this analysis. We could only imagine this might be attributed to their monotonous and repetitive posting behaviours being detected by Twitter. Both anonymised football fans are MU followers who often shared their ideas regarding MU line-ups, managers, and players.

Regarding organisations, Heineken (@Heineken) was influential because it was the UEFA Champions League’s primary sponsor. Finally, Paddy Power (@paddypower), a bookmaker that specialises in sports betting, appeared to be influential in the match against Southampton.

Community Analysis

In this section we present our findings regarding the sub-communities detected from each network. This relates to RQ1. For each network, the top five largest communities by the number of actors were selected for further analysis. These communities ranged from just over 1,000 actors to over 26,000 actors and collectively represented more than 55% of their source networks.

In terms of the overall patterns of these communities, Table V shows their network statistics and Figure 3 presents the structure of these communities within their home networks. Compared to their source networks described by Table II before we cannot see strong patterns of difference in terms of their average in- and out-degrees because the communities can have values that are higher or lower than their home networks. In terms of density, we notice that many communities have higher (or at least equal) density compared to their source network. However, they are still very low compared to those reported for other real-world networks (Melancon, 2006). Figure 3 shows a lack of clear boundaries between these communities except for the cases of the Manchester City and Southampton networks.

In terms of the topics discussed within these communities our NER analysis of the content created by each of the communities revealed interesting findings for the Manchester City and Southampton matches. While for the other two matches the most frequently mentioned NEs, across the top five communities, had a high degree of overlap and were predominantly related to football clubs, managers, and players. For this reason, we only show results for the Manchester City and Southampton matches in Table VI.

Table V. Statistics of the top five (#1~5, ranked by the number of users) communities detected from each network.

		#1	#2	#3	#4	#5
Manchester City	Avg. in-degree	1.62	1.42	1.04	2.02	1.06
	Avg. out-degree	1.43	1.49	1.04	2.20	2.10
	Density	<0.001	0.001	0.001	0.002	0.003
Crystal Palace	Avg. in-degree	2.21	1.86	1.68	1.80	2.57
	Avg. out-degree	2.06	1.91	1.69	1.93	2.47
	Density	<0.001	<0.001	<0.001	0.001	0.002
Southampton	Avg. in-degree	2.10	1.99	1.04	1.09	1.12
	Avg. out-degree	1.97	2.00	1.00	1.09	1.04
	Density	<0.001	<0.001	<0.001	<0.001	0.001
Arsenal	Avg. in-degree	2.65	2.14	3.37	1.43	1.53
	Avg. out-degree	2.51	2.30	2.80	1.30	1.84
	Density	<0.001	<0.001	0.001	<0.001	0.001

Table VI. The most frequent five named entities mentioned by each community during the Manchester City and Southampton matches

		Manchester City match		
#1 (blue)	#2 (red)	#3 (yellow)	#4 (pink)	#5 (green)
MUFC	MUFC	United States	MUFC	MUFC
ManchcesterDerby	Pogba	FBI	MCIMUN	ManchesterDerby
MCIMUN	Anthony	Trump	City	United States
GGMU	Martial	Fox News	Jose Mourinho	Mourinho
Sanchez	Mourinho	United Nations	GGMU	Trump
		Southampton match		
#1 (blue)	#2 (red)	#3 (yellow)	#4 (pink)	#5 (green)
MUFC	MUFC	United States	United States	SOUMUN
Jose	Jose	George	George HW Bush	Matic
SOUMUN	Southampton	Bush	Bush	MUFC
Mourinho	MOURINHO	Herbert Walker	America	ManUtd
Southampton	SOUMUN	George HW Bush	Geroge HW	Southampton

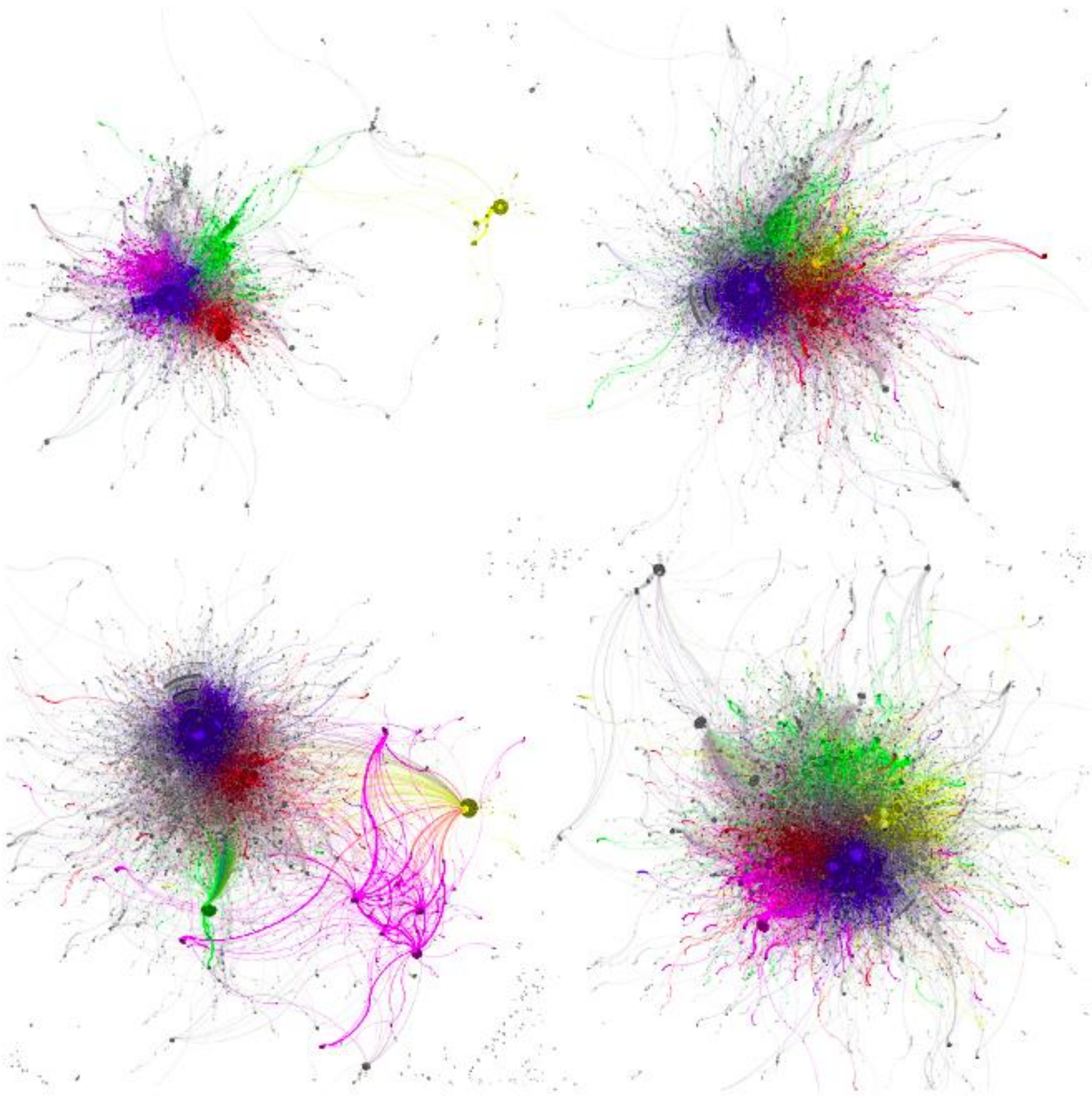

Fig 3. Visualisation of the top five communities within each network. In clockwise order from the top left: match against Manchester City, Crystal Palace, Southampton, and Arsenal. Colour code: blue - #1; red - #2; yellow - #3; pink - #4; green - #5 (this figure should be viewed in colour)

Table VI shows that in both matches, the communities distant from the principal cluster (see Figure 2) concentrated on conversations related to politics. The Manchester City match occurred shortly after an FBI investigation against Trump's election campaign. The Southampton match took place one day after the death of the former U.S. politician George H.W. Bush. These events could have triggered political conversations during matches and they might have been the consequence of people using MU's network and the massive attention during football matches as a means to disseminate information.

Content analysis

In this section we present our findings from the content analysis of tweets involving the most prominent influencers. This relates to RQ3. Table VII shows the distribution of topics discovered from the top 25 most retweeted posts that either mentioned or were generated by MU's Twitter account. In fact, we can observe that most of these tweets appeared to be created by users other than MU but targeted at MU (using @ManUtd). Most of the popular tweets related to MU were based on positive news such as goals (RT @ManU: Goooal! @AnderHerrera draws us level!), birthdays (RT @ManUtd: @AnthonyMartial feeling those birthday vibes), line-ups, statistics or good memories (RT @vancole9: Loved these battles against Arsenal … scoring goals at Old Trafford under the lights for @ManUts) from previous seasons or matches. Interestingly, one tweet was categorised as critique (RT @GaryLineker: If only @ManUtd had a decent keeper), and yet it was among the most popular tweets involving MU.

Table VII. Topics discovered from tweets involving Man. United

Tweet Category	#Tweets
Goal!	4
Good Memories	4
Highlight Plays	4
Line-up	4
Player's Birthday	3
Statistics	3
Advertisement	2
Critique of Man. United	1
Total	**25**

Table VIII shows the distribution of different topics over the different influencer categories. It is particularly worth mentioning the Politicians and Academics group where the tweets about politics attracted significant attention in the network. For football fans tweets showing support represented 70% of their most popular tweets. Whereas for journalists, critiques represented the highest percentage at 50%. For organisations (typically linked to the UEFA Champions League, such as @Heineken and @ChampionsLeague), tweets highlighting plays and recalling memories were the most popular. Only one popular tweet was related to advertising (brand promotion). Instead, advertising tweets from/containing football clubs and journalists were just as popular.

Table VIII. Topics discovered among influencer categories. For example tweets for each category, please see Appendix A in the supplementary material.

Influencer category / Tweet category	Art-ist	Football Clubs	Football Fan	Football Players	News & Humour Bots	Organisa-tions	Politicians & Aca-demics	Journal-ists
Advertisement	-	2	-	-	-	1	-	3
Criticism of referees	-	-	1	-	-	-	-	-
Critique of Man. United	-	1	-	3	3	-	-	10
Goal!	-	4	-	3	-	-	-	-
Good Memories	-	8	-	1	3	4	-	2
Highlight Plays	-	5	-	1	-	4	-	-
Line-up	-	4	-	-	3	-	-	-
Mocking to rivals	-	-	-	-	1	-	-	-
Personal Memories	-	-	1	-	-	-	-	2
Player´s Birthday	-	3	-	6	1	-	-	-
Player’s Welcoming	-	-	-	-	1	-	-	-
Players and Managers’ Trad-ing	-	-	-	-	2	-	-	-
Politics	3	-	-	-	-	-	3	-
Statistics	-	7	-	-	2	2	-	2
Supporting Messages	1	1	5	3	-	1	-	1
Total	**4**	**35**	**7**	**17**	**16**	**12**	**3**	**20**

We also compared popular topics on a per-match basis. About a third of the most popular tweets against Manchester City were about MU's good memories (goals and comebacks from previous matches between the two teams). It is important to remember that MU and Manchester City are traditional rivals not just in football but also in numerous other sports within the United Kingdom. Regarding Crystal Palace's match, about a third of the most popular tweets were critiques about MU's performance in that game and all of them were posted by a broadcaster (@LiamPaulCanning). The messages were focused on how detached the fans were from the team and its players and the type of training the team was doing between matches. Topics from the other two matches were more evenly distributed[3].

Discussion

With respect to RQ1, our network analysis identifies significant differences in terms of network size depending on different matches. This may be indicative of different levels of interest that could be caused by many factors, such as historical rivalries, the strength of the opponent, the date and time of the match, or generally increasing interests as both the English Premier League and the UEFA Champions League progressed. However, due to the limited sample chosen in this study (to be discussed further in the next section), we could not ascertain these factors and their links to the public interest in the matches.

Across all networks, there is a principal cluster of actors suggesting that actors in the network have predominantly focused on the same topics. This is different from the earlier study by Yan *et al*. (2019), who showed a much more sparse network consisting of a large number of small clusters featuring heterogeneous communication interest both before and during the match. Yan *et al*. (2019) did not analyse the content of the tweets from these clusters, and therefore, we cannot ascertain the reasons for this difference. However, our speculation is that this could be due to the different nature of the events covered by the two datasets. The UEFA Champions League is an annual competition for all European football clubs attracting fans of the most successful clubs across the entire Europe. Even though the final features only two teams it is likely that it attracted a significantly wider fan base of different clubs from different nations that may speak different languages. This may have led to the emergence of diverse small clusters that could be characterised by different football clubs, countries, or even languages. In contrast, our dataset covers matches between two teams in the local English Premier League, which may have comparatively reduced influence. Its audience may be limited to those who have a keen interest in the two teams only especially considering that other matches may be taking place at the same time.

We believe this has an implication on a football club's communication strategy. When a club aims to target its existing fan base it may be more effective to communicate during its own league matches as it is more likely to benefit from a densely connected, large, concentrated network of audience. When a club aims to expand its influence beyond its current fan base it may be more effective to communicate during international events in which it participates because it is more likely to benefit from the wider reach to diverse sub-communities.

With respect to RQ2, our influencer analysis shows that there is a very small set of highly influential actors who are known by a substantial part of the network as indicated by their in-degree and follower count. This is similar to the results from Yan *et al*. (2019). While Yan *et al*. (2019) showed the highly influential actors to be mostly football clubs, players, and fans,

3 Details of these statistics can be found in Appendix B of the supplementary material.

our results discover other groups of influencers such as journalists and independent news bots[4]. Most of the influencers in our dataset did not frequently interact with others in the network directly (as they did not directly mention or reply to other Twitter users), but still tweeted frequently and had a large follower base. This means that content shared by them will have high visibility.

We believe that the implications are many-fold for football clubs. First, they can benefit from working with journalists in their communication strategy because they have a large follower base and may be considered as credible sources of information. Secondly, they may want to analyse the success factors of news bots and incorporate them into their digital strategies (Lokot and Diakopoulos, 2015). However, this needs to be carefully managed because the use of bots may be seen as controversial.

Our community analysis showed that, generally there were no clear boundaries between the sub-communities within a network and the topics were generally coherent. Our NER analysis conducted within the different networks for the four matches showed that distant communities may form around current affairs irrelevant to football and this might be inevitable as it is largely driven by people's interests. Interestingly, our case resonates with previous studies about the nexus between sports and politics (Hayat *et al.*, 2016).

With respect to RQ3, our content analysis based on the most popular tweets involving the most influential actors identified a wide range of topics that attracted significant popularity in the networks. A total of 15 different topics were discovered to involve influencers while 8 were found in the most popular tweets involving MU. A direct implication of this is that MU can consider increasing its tweets on other topics that have been seen to attract popularity such as comments on player/manager trades, player welcome messages, supporting messages, and even important political/current affairs. However, it is important to note that certain topics may not be appropriate such as criticism of referees. Users criticising MU was also among the most popular tweets. This could have two implications. Firstly, it suggests that social media, Twitter in this case, can be a useful channel for listening to the needs of fans. Secondly, it may be useful for the club to consider engaging a reasonable extent of self-criticism which may be seen positively by fans and help build trust.

We also discovered that popular topics from different influencer categories appear to have a strong correlation with the nature of these categories. For instance, football fans' tweets criticising referees and showing support for the team and players were most well-received while those concerning the line-up, play highlights and manager/player trades (if any) were much less so. For journalists, critiques were most well-received suggesting they may be considered to be highly credible on such type of information. An implication of this is that football clubs should engage more proactively and frequently with these different influencer categories which may be a particularly effective channel for information diffusion.

Furthermore, we noticed that among the most popular tweets involving organisations only one was related to an advertisement and this was a retweet including the Twitter handles of the UEFA Champions League and Heineken ('RT: @ChampionsLeague @Heineken Sweet magical twitter'). We further searched our dataset and found no tweets directly created by Heineken. Reflecting on the visibility of these football matches and the fact that advert-related tweets involving football clubs and/or journalists were seen as popular we would suggest that organisations like Heineken should consider working more closely with these entities in their brand promotion strategy on social media perhaps using such events as channels.

---

4 None of the bots in Table 4 were verified by Twitter or had official links with any football clubs. Therefore, it is possible to assume that they are independent.

Since our study is the first in terms of understanding the topics diffused in a football club's social network it is difficult to directly compare the findings against other research which has also performed an analysis of sports related social media content. However, broadly speaking, we can observe some similar patterns. Examples include player promotion (Achen *et al.*, 2020), news broadcast (Meng *et al.*, 2015), and support for the team and players (Stavros *et al.*, 2014).

Finally, relating to the theory of information diffusion, as discussed before, our method followed 'influence models' in order to explain information diffusion and our results confirmed the presence of such influence in a football club's social network. There are individual influencers of different categories that have high visibility in the networks. There are also sub-communities formed with denser connections between smaller groups of actors suggesting the presence of dense, topic-specific information diffusion. Furthermore, there are specific topics that spread more effectively over the network. We also discussed how these findings could be utilised to maximise influence and benefit the football club. Thus, our study serves as an empirical application and verification of 'influence models'. From a theoretical point of view our study proposed a combination of SNA and content analysis, where the latter is informed by the former. As discussed in our results this allowed us to focus on content that is of most interest to the network and therefore helped us to develop an understanding of the diffusion of information and the shape of the network. It could also lead to a further understanding of the potential roles that different stakeholders play in terms of content creation in the network. Our analysis approach can contribute as reference for other domains wishing to undertake similar research.

Limitations

As mentioned above, one of the limitations in our study is the lack of understanding of the contributing factors to the varying sizes of different networks. This is due to the sample being limited to only four football matches focusing on one team. Another related limitation is the data collection method, which used only two hashtags related to MU. Expanding the data collection by including more football matches and teams and data outside matches will create a much richer and larger dataset which may help address the above mentioned problems and lead to possible further findings. For example, there could be non-match events that can cause an increase of interactions such as the sacking of a manager and transfers. Also, fans may use different hashtags, sometimes as a one-off , based on their country of origin (e.g., #MCITOT for a Manchester City vs.. Tottenham match), and these were not included in our data collection process. However, including such hashtags could potentially lead to a dataset biased towards certain matches or events.

Our analysis of influencers focused on a subset of the most influential actors. This may not necessarily represent the complete picture of influencers. Alternatively, a larger sample focusing on influencers that also interacted directly (i.e., high out-degrees) and frequently with others may lead to further insights.

Although our NER and content analysis helped identify a notion of 'topic' within MU's social networks they do not describe the full context of a topic, such as affective states, and relations with other topics. Applying techniques such as sentiment analysis or topic modelling together with the above techniques may lead to a better understanding of these topics.

Finally, our study empirically applies the 'influence model' of information diffusion theory in the football domain. The generality of our approach and findings also need to be further verified in other domains.

Conclusion

Social media has been extensively used in sports and it is playing an important role in the communication and marketing strategies for many sports brands particularly in football. A better understanding of a football club's online community may generate useful knowledge that can inform its marketing and communication strategies.

In this study, we conducted a social network analysis of Manchester United's online community on Twitter. We discovered a diverse range of stakeholders that are highly influential in the network, the formation of sub-communities that may have focused on distant topics within these networks, and how different topics of content are most effectively communicated when involving different categories of influencers.

Our work is the first empirical study examining the social network structure of a football club on Twitter. As we have discussed, a football club can benefit from our study by implementing communication strategies that utilise different influencers in the diffusion of different topics of content. This will help the effectiveness of their communication as these influencers play different roles in diffusing different content. Although our study revealed interesting findings, it is still limited in several ways as discussed above. Our future work will explore several directions. First, we will compare different football teams to identify if their social networks feature similar patterns. Second, we will expand our method to other sports to investigate if our findings can be generalised to other sports, and domains. This could be accompanied by the use of different methods for the analysis of social networks and text content for comparison against those used in this study. Finally, we will conduct research into developing methods that can capitalise on the properties of these social networks in order to inform sports organisations' communication strategies. For example, our research confirmed the presence of such influencers in the social networks of football clubs. In the future, it will be valuable to predict such influencers and their degree of influence in an early stage, in order for football clubs to benefit from them.

References

Abeza, G., O'Reilly, N. and Seguin, B. (2019). Social Media in Relationship Marketing: The Perspective of Professional Sport Managers in the MLB, NBA, NFL, and NHL. Communication & Sport. 2019;7(1):80-109.

Ahmed, W., López Seguí, F., Vidal-Alaball, J., and Katz, M. (2020). COVID-19 and the "Film Your Hospital" Conspiracy Theory: Social Network Analysis of Twitter Data. Journal of Medical Internet Research, 22(10), e22374

Achen, A., Kaczorowski, J., Horsmann, T. and Ketzler, A. (2020) Comparing organizational content and fan interaction on Twitter and Facebook in United States professional sport, Managing Sport and Leisure, 25:5, 358-375

Anagnostopoulos, C., Parganas, P., Chadwick, S. and Fenton, A. (2018) Branding in pictures: using Instagram as a brand management tool in professional team sport organisations, European Sport Management Quarterly, 18:4, 413-438

Blondel, V., Guillaume, J., Lambiotte, R. and Lafebvre, R. (2008). Fast unfolding of communities in large networks. *Physics and Society*, 2008, 1-6

Bond, A., Widdop, P. and Parnell, D. (2019). Topological network properties of the European football loan system. *European Sport Management Quarterly*, Vol 0, No. 0, 1-24

Bovet, A. and Makse, H. (2019). Influence of fake news in Twitter during the 2016 US presidential election. Nature Communications 10, 7

Chakrabarti, D. and Punera, K. (2011). Event Summarization Using Tweets. *Paper presented at the Fifth International AAAI Conference on Weblogs and Social Media*. Spain

Cheong F. and Cheong C. (2011). Social Media Data Mining: A Social Network Analysis Of Tweets During The Australian 2010-2011 floods. *15th Pacific Asia Conference on Information Systems (PACIS)*

Chaudhury A., Basuchowdhuri P., and Majumder S. (2012) Spread of Information in a Social Network Using Influential Nodes. In: *Tan PN., Chawla S., Ho C.K., Bailey J. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2012*. Lecture Notes in Computer Science, vol 7302. Springer, Berlin, Heidelberg.

Clavio, G., Burch, L. and Frederick, E. (2012). Networked Fandom: Applying Systems Theory to Sport Twitter Analysis. *International Journal of Sport Communication*, Vol 5, 522-538

Coche, R., (2017). How Athletes Frame Themselves on Social Media: An Analysis of Twitter Profiles. Journal of Sports Media, University of Nebraska Press, Vol. 12, No. 1, 89-112

De Meo, P., Ferrara, E., Fiumara, G. and Provetti, A. (2011). Generalized Louvain method for community detection in large networks. *Paper presented at the 11th International Conference on Intelligent Systems Design and Applications*, Spain

Filo, K., Lock, D. and Karg, A. (2015). Sports and social media research: a review. *Sport Management Review*, 18, 166-18

Frederick, E., Lim, C., Clavio, G., Pedersen, P. and Burch, L. (2014). Choosing between the one-way or two-way street: an exploration of relationship promotion by professional athletes on Twitter. *Communication and Sport*, Vol. 2 No. 1, pp. 80-99

Gach, O. and Hao, J. (2014). Improving the Louvain Algorithm for Community Detection with Modularity Maximization. *Paper presented at the International Conference on Artificial Evolution*, France

Geurin, A. and McNary, E. (2020). Athletes as ambush marketers? An examination of Rule 40 and athletes' social media use during the 2016 Rio Olympic Games, *European Sport Management Quarterly*

Grover, N. and Wason, R. (2012). Comparative Analysis Of PageRank And HITS Algorithms. *International Journal of Engineering Research and Technology*, 1(8), 1-15

Hayat, T., Samuel-Azran, T. and Galily, Y. (2016). Al-Jazeera Sport's US Twitter followers: sport-politics nexus? *Online Information Review*, 40(6), 785-797

Jacomy, M., Heymann, S., Venturini, T. and Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One*, 9(6), e98679

Khajeheian, D. and Kolli, S. (2020). Digital games get viral on social media: A social network analysis of Pokémon Go on Twitter. International Journal of Web Based Communities, 16(3), 262-278.

Kassens-Noor, E., Vertalka, J. and Wilson, M. (2019). Good games, bad host? Using big data to measure public attention and imagery of the Olympic Games, *Cities*, Vol. 90, 229-236,

Kim, Y., Seo, S., Ha, Y., Lim, S. and Yoon, Y. (2013). Two Applications of Clustering Techniques to Twitter: Community Detection and Issue Extraction. *Discrete Dynamics in Nature and Society*, 1, 1-8

Komorowski, M., Huu, T. and Deligiannis, N. (2017). Twitter data analysis for studying communities of practice in the media industry. *Telematics and Informatics*, 35, 195-212

Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: Sage

Lamirán-Palomares, J. M., Baviera, T., & Baviera-Puig, A. (2020). Sports Influencers on Twitter. Analysis and Comparative Study of Track Cycling World Cups 2016 and 2018. Social Sciences, 9(10), 169.

Lamirán-Palomares J., Baviera T. and Baviera-Puig, A. (2019). Identifying Opinion Leaders on Twitter during sporting events: lessons from a case study. *Soc Sci* 8:141.
Lemay, D., Basnet, R., Doleck, T. and Bazelais, P. (2018). Social network analysis of Twitter use during the AERA 2017 annual conference. *Education and Information Technologies*, 24, 459-470
Li, M., Wang, X., Gao, K., Zhang, S. (2017). A Survey on Information Diffusion in Online Social Networks: Models and Methods. Information 2017, 8, 118.
Manoli, A. (2020). Brand capabilities in English Premier League clubs. *European Sport Management Quarterly*, 20:1, 30-46
McCarthy, J., Rowley, J., Ashworth and C., Pioch, E. (2014). Managing brand presence through social media: the case of UK football clubs. *Communication and Ethics in Society*, 24, 181–204
Melancon, G. (2006). Just how dense are dense graphs in the real world? A methodological note. P*aper presented at BE-LIV 2006: BEyond time and errors: novel evaLuation methods of Information Visualization*, Italy
Mclean, S., Salmon, P., Gorman, A., Stevens, N. and Solomon, C. (2018). A social network analysis of the goal scoring passing networks of the 2016 European Football Championships, Human Movement Science, Volume 57, 400-408,
Meng, M., Stavros, C. and Westberg, K. (2015). Engaging fans through social media: implications for team identification. *Sport, Business and Management: An International Journal,* Vol. 5 No. 3, 199-217
Naraine, M., Pegoraro, A. and Wear, H. (2019). #WeTheNorth: Examining an online brand community through a professional sport organization's hashtag marketing campaign. *Communication and Sport*
Ozanian, M. (2019). The World's Most Valuable Soccer Teams 2018. *Forbes*.
Ozer, M., Kim, N. and Davulcu, H. (2016). Community Detection in Political Twitter Networks using Nonnegative Matrix Factorization Methods. *IEEE/ACM International Conference on Advances in Social Network Analysis and Mining*, United States of America
Pacheco, D., Pinheiro, D., de Lima Neto, F., Ribeiro, E. and Menezes, R. (2016). Characterization of Football Supporters from Twitter Conversations. 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 169-176
Rowe, A. and Pitfield, D. (2018). The challenge facing existing airport campaign groups when incorporating social media into their campaign: A social network analysis of Airport Watch's social media utilisation. *Geoforum*, 96, 236-247
Sadri, A., Hasan, S., Ukkusuri, S. and Suarez, J. (2018). Analysis of social interaction network properties and growth on Twitter. *Social Network Analysis and Mining*, 8, 56-69
Scott, J. (2012). Social network analysis. London: SAGE Publications
Schubert, M. and Seyffert, J. (2017). Fan motives for interacting on social media – the example of the International Table Tennis Federation and Facebook. *Current Issues in Sport Science* 2:004
Stavros, C., Meng, M., Westberg, K. and Farrelly, F. (2014). Understanding fan motivation for interacting on social media Constantino Stavros. *Sport Management Review*, 17, 455–469
Surian, D., Nguyen, D., Kennedy, G., Johnson, M., Coeira, E. and Dunn, A. (2016). Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of Medical Internet Research*, 19, 1-12
Wasserman, S. and Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press.
Winard, M., Belot, M., Merten, S. and Kolyperas, D. (2018). International Sport Federations' Social Media Communication: A Content Analysis of FIFA's Twitter Account. *International Journal of Sport Communication*, 12 (2), 209-233

Yan, G., Watanabe, N., Shapira, S., Naraine, M. and Hull, K. (2019). Unfolding the Twitter scene of the 2017 UEFA Champions League Final: social media networks and power dynamics, *European Sport Management Quarterly*, 19:4: 419-436

Yan, G., Pegoraro, A. and Watanabe, N. (2018). Student-Athletes' Organization of Activism at the University of Missouri: Resource Mobilization on Twitter. *Journal of Sport Management*, 32, 24-37

List of Appendices

Appendix A. Examples of tweets belonging to different topic categories

Topic Category	Example
Goal!	“RT @ManU: Goooal! @AnderHerrera draws us level!”
Player’s Birthday	“RT @ManUtd: @AnthonyMartial feeling those birthday vibes”
Good Memories	“Arsenal … scoring goals at Old Trafford under the lights for @Ma-nUts”
Highlight Plays	“RT @rioferdy5: Get in there … sublime from @MarcusRashford”
Line-up	“RT @ManUtd: Here’s how #MUFC line up for #MUNARS”
Statistics	“RT @ManUtd: @AnthonyMartial has seven goals in his last eight #PL outings”
Advertisement	“RT @BwoyYorubad: If you’re a @ManUtd fan, please gather here, drop your handle, let’s follow”
Critique of Man. United	“RT @GaryLineker: If only @ManUtd had a decent keeper”
Criticism of referees	“The ref was HORRENDOUS, shouldn’t be allowed to ref at Sunday League level”
Mocking to Rivals	“Mike Tyson: I’ve been in Manchester for a long period of time and I’ve never heard of Manchester City”
Personal Memories	“Getting a lift from Milan to Turin. The driver is a happy Serbian. Happy because Red Star won last night”
Player’s Welcoming	“Welcome back, Marcos Rojo”
Players and Man-ager’s Trading	“Real Madrid fans want Mourinho back”
Politics	“Mueller: Was Trump aware of your contact w/the Russians?”
Supporting Messages	“I can’t stand anyone who can’t stand @ManUtd I am a @ManUtd fan”

Appendix B. Distribution of popular topics per match

Match / Topic category	vs Man. City	vs Crystal Palace	vs Southampton	vs Arsenal
Advertisement	2	1	1	2
Criticism of referees	-	1	-	-
Critique of Man. Utd	1	4	3	9
Goal!	-	-	1	6
Good Memories	6	-	2	10
Highlight Plays	5	-	1	4
Line-up	-	2	2	3
Mocking to rivals	1	-	-	-
Personal Memories	1	2	-	-
Player’s Birthday	-	-	-	10
Player’s Welcoming	-	-	-	1
Players and Manager’s Trading	1	-	-	1
Politics	2	-	2	2
Statistics	1	2	2	8
Supporting Message	1	3	2	6
Total	21	15	16	62