Sean McAteer

Student number: 11000040

Are there any differences in the ways in which male and female sportspersons represent themselves online?

Date of submission

I declare that this project is, excepting appropriately referenced and quoted material,

entirely my own work.

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date: \_

Acknowledgements.

I would like to thank my wife Sarah, for putting up with all the time I have spent stuck in front of a computer putting this dissertation together. She has been a great support through it all and love her more than she knows.

I would also like to thank my daughters Isabel and Sadhbh. They never fail to put a smile on my face every day. Most importantly they remind me of what is most valuable in life. This is only a piece of paper but they are what life is about.

Lastly, I would like to mention Donald Trump, the American President, whose impressive career advancement despite remedial language skills and a tenuous grasp on reality inspired me to believe that I was capable of authoring a research paper".

Table of Contents:

[Abstract 4](#_Toc511580500)

[Introduction. 5](#_Toc511580501)

[Literature review 6](#_Toc511580502)

[Sports and Twitter 7](#_Toc511580503)

[The role of gender in media coverage. 9](#_Toc511580504)

[Self-Presentation 10](#_Toc511580505)

[Methodology / Method: 12](#_Toc511580506)

[Selecting participants 13](#_Toc511580507)

[Harvesting the tweets. 14](#_Toc511580508)

[Cleaning the data. 15](#_Toc511580509)

[Analysing the Data 16](#_Toc511580510)

[Results: 19](#_Toc511580511)

[Twitter Activity 20](#_Toc511580512)

[Athlete Self-Presentation 21](#_Toc511580513)

[Clusters. 21](#_Toc511580514)

[Word-Cloud 25](#_Toc511580515)

[Discussion: 28](#_Toc511580516)

[Summary of results. 28](#_Toc511580517)

[Relationship between results and hypothesis or research question 29](#_Toc511580518)

[Relationship between results and literature 30](#_Toc511580519)

[Limitations and Future Research 31](#_Toc511580520)

[Conclusion 32](#_Toc511580521)

[References: 33](#_Toc511580522)

# Abstract

This research project is to explore gender differences in professional athletes' self-presentation on Twitter. To do this Cluster analysis was used on the tweets.

Athletes have engaged in tweeting at a fast pace, but there are major disparities in the number of followers male and female athletes get, which raises the question, are there any differences in the ways in which male and female sportspersons represent themselves online? This research investigated the tweets of the ten most followed female athletes and the ten most followed male athletes on Twitter over a one year period in an attempt to answer that question. The findings indicate that athletes are talking predominantly about their sport as well as their personal lives on Twitter. The results indicate that female athletes are also tweeting about issues related more specifically to women. So in relation to whether there are differences between how male and female sportspersons represent themselves online, the findings suggest that there are.

# Introduction.

**The ever-changing social-media landscape has revolutionized the ways in which sport is consumed. While traditional sports consumption occurs through means such as attending events, viewing games on television, listening to games on the radio, or reading about sports in newspapers and magazines, the infiltration of social media into the sports realm has led to changes in sports consumption. Consumers are using their electronic devices to participate in "second screen" behaviours instead of focusing solely on the sporting event through traditional methods of consumption. Second-screen behaviour entails using a smart phone, tablet, or laptop while watching television to access the Internet and other social-networking sites to obtain more information (de Zuniga, Garcia-Perdomo, & McGregor, 2015). In other words, consumers use multiple distribution systems to meet different needs that are not fulfilled by one system alone (Boehmer, 2015). Second-screen behaviours allow sports consumers to connect with other fans, feel the adrenaline rush of big moments in real time, be the first to broadcast content and have their voices heard (Yorke & Greenwood, 2014).**

**Web 2.0 encompasses the various segments of social media such as blogs, social networking sites (SNSs) (e.g. Facebook, Twitter), content communities (e.g. YouTube), team websites, forums and bulletin boards. While the underlying functionality of social media sites is similar, the way the sites function and the type of users they attract can be drastically different (Lipsman 2009). Some social networking sites focus on users and their life (e.g. Facebook), some focus on the message (e.g. Twitter), some focus on collecting and discovering ideas (e.g. Pinterest), some focus on knowledge creation (e.g. Quora), while some other focus on pictures (e.g. Instagram), or finding jobs (e.g. LinkedIn). Facebook and Twitter are the most utilized sites for both sharing and accessing news and updates on the unfolding events. Research shows that due to the ‘walled garden’ approach, Facebook has become less accessible than Twitter for public communication (Bruns and Stieglitz. 2012).**

Twitter is a unique medium where both traditional and more interpersonal forms of interaction are possible. Because social media platforms like Twitter provide multiple avenues for interaction, it is vital for sport communication scholars to understand the nature of these communication tools and how they impact fan–athlete relationships (Clavio & Kian, 2010; Hambrick, Simmons, Greenhalgh, & Greenwell, 2010)

Furthermore, Clavio and Kian (2010) stated that it would be worthwhile for sport communication scholars to begin formulating an understanding of the interactions taking place between sports personalities (i.e., athletes) and sports consumers (i.e., fans) on Twitter. This is what the following research will take for its focus.

# Literature review

Just like any other consumer-facing industry, the world of professional sports has seen a lot of disruption from the rise of social media. (Force 2016)

Quickly becoming a normal way to communicate, this has transformed the way sports are reported (Schultz & Sheffer, 2010) and consumed (Clavio & Kian, 2010), allowing athletes the freedom to interact directly with their fans and enabling them to become rulers of their own domain (Hambrick et al, 2010). It is argued that the rapid revolution social media has achieved in sport may be unrivalled regarding impact compared with any other industry (Sanderson, 2011)

## Sports and Twitter

Twitter has been particularly dominant in the digital-media-sport landscape and has been embraced by the sporting world at amazing speed. We have now reached the point where it is unusual to watch a sports event, attend a live match, or even read a sport-related article without a Twitter reference mentioned. Twitter defines an "impression", as how many times a tweet is seen online, both on Twitter and off. In 2015South America's Copa America soccer championship generated 14 billion impressions. The women's soccer world cup generated 9 billion impressions. Wimbledon generated 8 billion Impressions over two weeks, and the champions league final generated 834 million impressions in one day. (Laird 2015)

Findings from a study by Backman, Pertick and Yoon (2017) revealed that team attraction, team trust, and team involvement are positively related to team attachment and these are determinates of fan loyalty. While team attachment was found to positively influence fan loyalty, sports fans' Twitter use was found to significantly reinforce their loyalty. They also found that 'the more sports people strengthen the relationships with their fans through Twitter, the more likely those fans will be committed and attached.'

Haugh and Watkins (2016) found in their study that there is, 'a clear pattern for social-media use by sports fans'. Twitter is the leading destination for fans wishing to gather information, seek entertainment, show support for a team, express opinions on the team, and learn about rules and strategies pertaining to a sport.

This follows on from research by Clavio and Walsh (2013) indicating that the highest level of social-media use among their sample of college sports fans was to watch sports videos.

Further research by Brunette et al. (2014) found that there are 4 primary gratifications sought by Twitter users: interaction, promotion, live game updates, and news. Users wanted to hear about players and roster moves as they happen, learn information faster than other people, read tweets if unable to watch the game, learn about upcoming games and access special promotions.

Research surrounding Twitter use in sport has been conducted from a variety of perspectives. Clavio and Kian (2010) used an internet-based survey to ascertain the demographics, uses, and gratifications of a retired athlete's Twitter followers. Clavio and Walsh (2014) surveyed Division 1 college sports fans and found that social media should not be viewed holistically due to distinct variability between what attracts fans to Facebook and Twitter.

Aside from fans use of social media, attention has recently focused on the players/sports personalities themselves. Hull (2014) explored how professional golfers participating in the Master's tournament used Twitter during the week of the event. Basing the research on self-presentation theory the author conducted a content analysis of 895 tweets by 39 golfers. In tennis, Lebel and Danylchuk (2012) compared male and female athletes' tweets relayed by all professional tennis players. Furthermore, Hambrick et al. (2010) used content analysis to place 1,962 tweets by professional athletes into one of six categories: interactivity, diversion, information sharing, content, promotional, and fanship. Pegoraro (2010) also investigated athletes' use of Twitter and interestingly found that athletes are talking predominantly about their personal lives and responding to fans' queries through Twitter as opposed to solely discussing their respective sports.

The results indicate that Twitter is a powerful tool for increasing fan-athlete interaction. Professional cyclist Lance Armstrong once invited his fans to meet him for a ride around Dublin after completing the Tour of Ireland. 1,000 fans showed up hours later (Cromwell, 2009). Serena Williams even asked her Twitter followers for pregnancy advice, “Any tips on how to turn over at night? I'm having trouble from going from my left ....to my right.... to my left side,” She received over 1200 replies.

So it is clear what an important role social media, especially Twitter, now plays in sports from a fan perspective but also regarding the sports professionals usage of Twitter. What needs to be looked at now in more detail, is what the research states in relation to the differences, if any, between male and female.

## The role of gender in media coverage.

Researchers have found that female athletes are considerably underrepresented regarding the amount of media coverage they receive in comparison with their male counterparts (Kane & Maxwell, 2011). While Cooky (2015), found a decline in the amount of coverage of female athletes from 1989 to 2014. In 2014, only 3.2% of network television coverage focused on women's sports. The lack of representation of women in media's coverage of sport contributes to the perception of sport as a male domain. (Birrell, 2000)

The media maintain notions of sport as a male domain through displays of masculinity, strength, and power. On the other hand, when women athletes receive coverage, they are consistently portrayed in ways that emphasize their femininity through caregiver roles and sexuality versus their athletic competence (Cooky et al., 2015). When sports media do cover female athletes, they often describe their physical appearances and attire, delve into their personal lives and relationships, trivialise their accomplishments and athleticism, and focusing on perceived psychological weaknesses (Hardin et al. 2007).

Regarding online coverage of woman in sport, there seems to be a more level playing field. A study of gender-related descriptors employed within articles on men’s and women’s basketball produced surprising results, as it contradicted many of the findings previous sport media researchers cited to justify the presence of hegemonic masculinity. The researchers found that there were a significantly higher proportion of descriptors about the positive skill level/accomplishments and psychological/emotional strengths in women’s basketball articles than those on men’s basketball. (Kian et al. 2009). In another study, Cunningham (2003) found university websites provided more coverage of women's tennis than men's tennis. In further analysis of NCAA college websites, Cooper (2008) concluded that coverage of male and female athletes in the same sport was mostly equal.

An exploratory study analysing online media and print media for the 2007 US Open tennis tournament found that online media to be less likely to re-enforce the traditional stereotypes of male and female athletes than newspapers. (Kian and Clavio, 2011)

This suggests that there is a greater opportunity for women to expand their coverage in the online world rather than the traditional mainstream media. However, what exactly does expanding coverage look like or do?

## Self-Presentation

Self-presentation is behaviour that attempts to convey some information about oneself or some image of oneself to other people. It denotes a class of motivations in human behaviour. (Baumeister and Hutton. 1987)

This practice of self-presentation is a concept that was pioneered by Erving Goffman in his highly influential work, 'The Presentation of Self in Everyday Life' (1959).

Goffman’s theory of self-presentation suggests that individuals present themselves in manners in which they wish others to view them. Goffman alludes to life being a “drama,” and the two types of self-presentation depicted by individuals as daily life “performances” are front-stage performances and back-stage performances. In front-stage performances, individuals are especially concerned with the impression they create in the minds of others. On the other hand, in back-stage performances individuals are more candid, often sharing information as if no audience or an audience of familiar people were present (Goffman, 1959).

Research on the self-presentation of athletes suggests that athletes engage in back-stage performances on social-media platforms (Hambrick et al. 2010). Professional cyclists offered insights into the terrain and conditions of the event route, adding a level of personal knowledge for fans that journalists would not have necessarily focused on. (Kassing and Sanderson 2010).

When it comes to front-stage performances, Krane et al. (2010) investigated female college athletes’ self-presentation preferences. They found that female athletes wanted to be portrayed in ways that emphasized their power and strength. Support for the preferences of female athletes to be portrayed as powerful athletes was found in later research by Lebel and Danylchuk (2014).

Krane et al. (2010) found their qualitative study demonstrated how these women were driven toward masculine behaviours for their sport while at the same time they tried to stay in the feminine role off the field of play. As one athlete in the study said, “If you're an athlete, then you have to transform into entirely someone else when you come off the field.” This corresponds to Goffman's (1959) front-stage/back-stage performances theory.

Goffman's theory is now nearly 60 years old, is it still relevant? Arundale (2010) argues that Goffman’s work, is now outdated and should be modernised to include progress in research and technology, but Miller (2012) says that electronic interaction is a natural addition to what Goffman posited. He says that even though electronic communication is apparently limited in the depth of information it provides compared with face-to-face interaction, there is still enough room for information about the self to be given off in the way people use the medium, in what they say as well as what they don't say.

Oram (2009) says ' Goffman's approach certainly applies online, because our postings, even our instant messages, are more deliberate acts than our informal behaviours in real life. Although some participants play at being flippant and spontaneous on Facebook walls and microblogs, they must have a greater consciousness of their effects on the viewer than most dinner table guests or concert attendees. Our online personas, therefore, conform even more closely to Goffman's idea of everyday life than our everyday life does. '

While Goffman presented a framework for how people present themselves, Jones and Pittman (1982), put forward ideas for the motivations behind self-presentation.

Jones and Pittman believe that self-presentation motivations can extend beyond making a favourable impression. They put forward the idea of different self-presentation strategies such as self-promotion, ingratiation, exemplification, intimidation, and supplication.

Self-promotion happens when individuals bring attention to their own accomplishments, hoping to be viewed as capable by observers. Ingratiation occurs when individuals use flattery or favours to be seen as likeable. Exemplification results from individuals going above and beyond what is needed or expected to be seen as hardworking or committed. Intimidation occurs when individuals project their ability or power to punish to be perceived as dangerous and powerful. Lastly, supplication occurs when individuals present their weaknesses or deficiencies to receive assistance and compassion from others.

These strategies are used to elicit certain emotions from the audience depending on the attribution sought, including: likeable, competent, dangerous, worthy, and helpless.

While mainstream media still views sport as a male domain, the online world seems to afford a more equal footing for female athletes. Twitter is a powerful tool for increasing fan-athlete interaction (Pegoraro 2010) and allows the athlete themselves to control the output.

Athletes and teams have long used mainstream media to communicate with their fans, and extensive research has investigated the character of the communications. Social media platforms like Twitter have changed the communication between athletes and sports fans. As such, we are still learning about what is communicated and how it is communicated.

Further examination into how female athletes engage in strategic self-presentation compared to their male counterparts when they control the output themselves (i.e. their own social media account) is necessary.

Therefore, the purpose of this research is to explore the 'tweets' of top athletes and try to see, are there any differences in the ways in which male and female sportspersons represent themselves online?

# Methodology / Method:

This research project is to see if there are any differences in the way that top sportsmen and women present themselves on Twitter.

The previous section outlined the existing research and literature pertaining to the use of social media, especially Twitter, by athletes. It highlighted the gap in this literature, that is, the need to further explore, what differences, if any exists. This section will outline the methodology of this research project, that is, the steps that were taken to answer the research question.

## Selecting participants

The first step in the data-collection process was to identify the top ten Twitter accounts for athletes in various sports. This was done using the Web site fanpagelist.com. There you can select 'athletes' and then sort the results by the number of Twitter followers. Sports people who were wrestlers, i.e. WWE were omitted as they are 'performers' in a scripted setting and this was not deemed to be suitable for this research. Sportspeople were then omitted who tweeted in a non-English language as it was deemed that any translation would alter the words or meaning of some of the Tweets. It is also impossible to use Cluster analysis, the chosen method of analysis for this research, as will be outlined as we progress, on two languages at the same time. To do so, both languages' tweets would have to be analysed separately as well as results also being reported separately. As this was not the aim of this research, only English language tweets were used. From the resultant list, the top ten males and females athletes (by number of followers) were selected. This list is composed of athletes from tennis, NBA, soccer, track and field athletics, and cricket. The resulting list of athletes can be found in the tables below.

Table 1:

Top 10 Male Athletes Based on Followers.

|  |  |
| --- | --- |
| **Male Athletes** | **Followers (Million)** |
| Cristiano Ronaldo | 69.4 |
| LeBron James | 40.7 |
| Sachin Tendulkar | 24 |
| Kevin Durant | 17 |
| Wayne Rooney | 16.9 |
| Shaquille O'Neill | 15 |
| Kobe Bryant | 13.6 |
| Stephen Curry | 12.2 |
| Rodger Federer | 11.7 |
| Rio Ferdinand | 10 |
| Carmelo Anthony | 9.3 |

Table 2:

Top 10 Female Athletes Based on Followers.

|  |  |
| --- | --- |
| **Female Athletes** | **Followers (Million)** |
| Serena Williams | 10.6 |
| Maria Sharapova | 8.4 |
| Sania Mirza | 7.6 |
| Alex Morgan | 3.5 |
| Caroline Wozniacki | 3.1 |
| Ana Ivanovic | 2.7 |
| Venus Williams | 1.7 |
| Sydney LeRoux | 1.7 |
| Jessica Ennis | 1.6 |
| Victoria Azarenka | 1.3 |

## Harvesting the tweets.

Once the list was compiled, the tweets for each of these athletes were collected over a 12 month period covering October 1st, 2016 to September 30th, 2017.

A Python program was used along with Python's 'Tweepy library' to download the tweets from each user. Python is a high-level programming language for general-purpose programming. It is often used in projects involving data analysis, artificial intelligence, and scientific computing. The program uses a Twitter API (Application Programming Interface) to harvest the tweets. An API is a software go-between that allows two applications to talk to each other, in this case, the two applications are the Python program and the Twitter database. To use the Twitter API a Twitter account had to be set up. This is needed as the Twitter API requires you to enter an Access code, Access Secret, Consumer Key and Consumer Secret, which are only given to users with a Twitter account. Normally a limited number of Tweets are allowed to be downloaded at one time, so the script was repeatedly run until all the tweets were harvested for each athlete. This resulted in 20 data files, one for each of the athletes.

The Tweepy library has no way to select a range of dates from the Twitter API, so more Tweets than what were required were harvested along with the dates they were posted. Then the Tweets within the required dates were manually identified and extracted from each file.

## Cleaning the data.

Before text can be analysed, it needs to be cleaned. Data cleaning is the practice of detecting and correcting or removing words or characters from a data set that will cause problems when the text is being analysed. This is done because most text is created and stored so that humans can understand it, and it is not always easy for a computer to process that text.

All links were removed from each of the data files and then characters not in the list: 0123456789abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ!"#$%&'()\*+,-./:;<=>?@[\]^\_`{|}~ were removed. This step resulted in the removal of emoji's.

Emoji's were removed as these are quite ambiguous, for example, the pizza slice can be used to represent pizza, food in general, being hungry, and even a party.

At this point another Python program was run to count the number of tweets, the number of links in tweets, the number of retweets and the number of hashtags for each of the athletes.

'Stop words' were then removed. Stop words are words which are usually filtered out before processing of natural language data. These are usually the most common words in a language, for example: and, the, my, etc. The stop word list used in this research is the one that comes with the Sklearn python data analysis package.  When the text was analysed using Cluster analysis more stop words were identified and added to the list.

One last step was to deal with 'stem' words. Stemming words normally means reducing inflected words to their root form. e.g. reducing weakness to weak. This accounts for words such as weakened, weaken, weaknesses, etc. In this data further steps were taken as there was also the issue of American versus British spelling. e.g. favourite and favorite. This required a manual examination of the data and using the 'replace' function of 'Notepad++' to change the data to British spelling. The issue of abbreviated words also needed to be dealt with. As tweets were only allowed to be 140 characters long (this has since been changed to 280), tweeters used abbreviations to get their message across within the character limit. This resulted in 'favourite' being abbreviated to 'fav'. Again this required a manual examination of the data and replacement of the abbreviations with the full word. When this was done a script was run to join all the data files for male athletes together and all the data file for female athletes together. This resulted in two large data files which were used for analysis. The file containing the tweets from male athletes contained 3,935 tweets and the file for female athletes contained 3983 tweets.

## Analysing the Data

Cluster analysis divides data into groups or clusters that are meaningful, useful or both. When looking for meaningful groups the clusters should capture the natural structure of the data.

Cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Cluster analysis (first used by Tryon, 1939) encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories.

The decision to use cluster analysis was made as up until now research into Twitter use has predominantly used frames into which each tweet is coded. Cluster analysis examines the data and groups the words into clusters. Some of the words will be in more than one cluster as the same word can have different meanings depending on the context. In a lot of qualitative text analysis on tweets, when the researchers use frames each tweet is placed into a specific frame even if the tweet could be applicable to another frame as well. This can lead to some subtleties within tweets being missed. If a tweet could be in two frames, the researcher has to decide on which frame to place it in and if there are enough of these decisions to be made then the researcher's choices can affect the final results. With this cluster analysis method more of these subtleties should be found and hopefully help us discover the underlying data structures.

The algorithm used in this research is 'K-Means' clustering. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity (Trevino. 2016).

K-means clusters the data into *k* clusters, even if *k* is not the right number of clusters to use. Therefore, when using k-means clustering, some way is needed to determine what is the best number of clusters.

The method used in this research is the elbow method. The elbow method runs k-means clustering on the dataset for a range of values of *k*, e.g. from 1 to 10, and for each value of *k* calculate the sum of squared errors. This produces a line chart which looks like an arm. The "elbow" on the arm is the value of *k* that is the best. In our analysis the best value was 5, so we looked for 5 clusters.

Finally, the text was analysed using K-means Cluster analysis. This analysis was done using a Python script and the Sklearn data analysis package. To perform this analysis, the data has to be converted to a numeric representation. This is done with a ' TFIDFVectorizer ' command which transforms text to feature vectors that can be used as input to our K-Means algorithm.

To run a k-means algorithm, you have to randomly initialize K points called the cluster centroids (A centroid is a data point at the centre of a cluster.) These initial points are referred to as seed points. There is much debate as to how to find the best starting seed point, and in this research, the value of the most common word was used as the initial seed value.

K-means is an iterative algorithm, and it performs two steps:

1. Cluster assignment step: the algorithm goes through each of the data points and depending on which cluster is closer, it assigns the data points to one of the K cluster centroids.

2. Move centroid step: in this step, the algorithm moves the centroids to the average of the points in a cluster. In other words, the algorithm calculates the average of all the points in a cluster and moves the centroid to that average location.

Once this is done the clusters are output on the screen.

After the cluster analysis was done, an analysis of the word frequency was run. To do this another Python program was written which removes 'stop words', and then counts the number of times each remaining word is used in the data file. From this output, a word cloud is generated. A word cloud is a graphical representation of word frequency that gives greater prominence to words that appear more frequently in the source data file.

In this section, I have outlined how I went about researching if there are any differences in the ways in which male and female sportspersons represent themselves online? I explained how I selected the participants, harvested the tweets, cleaned, and finally analysed the data. In the next section, I will present the results of the analysis.

# Results:

The main focus of this research was to determine how male and female professional athletes present themselves to their audience via Twitter. As pointed out in the literature review, research has shown that fans are looking for sports-related information when they follow athletes on Twitter (Haugh and Watkins. 2016). It was also shown in previous research that athletes who give fans the content they desire would both, 'reinforce their loyalty' and 'strengthen the relationships with their fans'.

This section will present the results of this research. The Twitter activity of both genders will be shown, followed by the results of the cluster analysis, which will be displayed in two tables, one for each gender. Finally, the word frequencies will be shown with the aid of a word-cloud.

## Twitter Activity

Anyone who wished to can follow professional athletes on Twitter. At present, there is no trustworthy method to determine the demographics of followers. At the time of analysis, the mean number of individuals following the sample of male athletes on Twitter was 22.79million (SD = 18.72). Two extreme cases were detected in Cristiano Ronaldo and LeBron James. These athletes among the most followed people on Twitter. Cristiano Ronaldo had accumulated 69.3 million followers at the time of study and was the 8th most followed person on Twitter, while LeBron James had accumulated 40.6 million followers and was the 24th most followed person. They were thus identified as outliers that skewed the collective data (Barnett & Lewis, 1984). With the outliers removed from calculations, the average number of followers of male athletes decreased to 13.41 (SD = 4.99). By contrast, the sample of female athletes had a mean of 4.22 million followers (SD = 3.36), with no influential cases detected.

The number of tweets that male and female athletes posted was very similar. Male athletes tweeted 3935 times during the period under review. Female athletes tweeted only 48 times more over the one year period with 3983 tweets. Even though the number of tweets is similar between the two groups, not all tweets are the same. Tweets can be direct messages, retweets, link to other social media sites, etc. The data gathered illustrates that the largest percentage of tweets were direct tweets. For male athletes, this accounted for 3367 tweets or 85.56% and for female athletes there were 3167 direct tweets or 79.51%. A large number of tweets contained a link, in fact, 2996 (76.13%) of male athletes tweets contained links and 3110 (78.08%) of female athletes tweets contained a link. Retweets made up 17.48% of the total tweets collected.

It is recognized that the number of followers a person accumulates is not necessarily indicative of the influence or reach of a Twitter account (Leonhardt, 2011). The analytical tool Klout was therefore used to calculate each athlete's impact score. As opposed to measuring influence solely based on the number of followers a Twitter user can collect, an impact score takes into account factors such as the number of times a Twitter name is mentioned by others and the frequency with which a person is communicating (Leonhardt, 2011). The mean impact score of female athletes was 80.25 (SD = 7.7), with the highest impact calculated for Serena Williams with a score of 90.2. The mean impact score for male athletes was 87.6 (SD = 3.5), with the highest impact achieved by LeBron James with a score of 92.4. A t-test was performed on the Klout impact scores with the result being a significance of 0.017 which is under 0.05 and is, therefore, a significant difference.

The athletes in the sample employed their Twitter accounts not only to be followed but also to follow other Twitter accounts. The mean number of Twitter accounts followed by female athletes was 227 (SD = 115), and they posted an average of 398 tweets per athlete (SD - 29.7) during the period under review.

By contrast, male athletes followed an average of 469 accounts (SD = 424). An outlier was again detected in Kevin Durant, who followed 1,453 accounts. With this outlier removed, the mean became 359 followed accounts (SD - 261). Male athletes posted a mean of 394 tweets per athlete (SD = 40.5) during the time of the study.

## Athlete Self-Presentation

### Clusters.

When the cluster analysis was performed the clusters for female athletes returned were:

Table 3:

Clusters and cluster words for female athletes.

|  |  |
| --- | --- |
| Cluster labels | Words in cluster |
| Cluster 1: feeling happy | Love, thank, happy, great, good, today |
| Cluster 2: Interview / photo-shoot | Cover, month, excited, pics, story, interview |
| Cluster 3: Friends | Best, proud, time, happy, beautiful, friends |
| Cluster4: Celebration | Baby, Santa, dearest, happy, love, hugs |
| Cluster 5: Gameday | Wait, play, league, love, tonight, favourite |

The labels for each cluster were created based on the content of the cluster.

Cluster 1, 'feeling happy', captured general 'good morning' comments and upbeat tweets like Ana Ivanovic's, 'Love walking up here every day so beautiful', as well as Sania Mirza's tweet, 'Me and all my friends having a great time ', Alex Morgan was delighted to be back living in Orlando and tweeted' I'm so in love with Orlando and so happy to be back living here. Today, and every day, we are #OrlandoUnited'.

Cluster 2, found words that related to photo shoots and interviews. This provided athletes with the opportunity to extend the coverage that they got from their mainstream media engagements into the social media world and is something that is used a lot by athletes. This cluster came from tweets like this form Maria Sharapova, 'Grazia Russia cover and interview with @Porsche'. Another tweet from Ana Ivanovic said, ' Thank you for the cover story. It was great talking to you guys' and Caroline Wozniacki was sharing her inclusion in Sports Illustrated photo shoot with the tweet, ' The new @si\_swimsuit is on the stands today!!! So excited and honored to be included in the issue for the third straight year! '.

Cluster 3: contained words in connection to 'friends'. Alex Morgan tweeted that, 'my beautiful and amazingly talented friend does it for me'. Serena Williams tweeted about the actor Scott Shilstone, 'So proud of my friend @shilstonescott yeaaa' when he got a part in the American daytime soap, Days of our lives.

Cluster 4, collected words that concerned the topic of celebration, Caroline Wozniacki posted, 'When one of your dearest and oldest friends gets married it's time to celebrate'. Sydney Leroux posted a picture of her baby in a Santa costume and tweeted 'My Santa baby' and Maria Sharapova celebrated the launch of her book with the tweet, ' This was incredible! The line along the NY streets!! The amount of love, hugs, kind words. THANK YOU'.

Cluster 5 was about a game that was being played soon. Comments leading to this cluster were, 'Landed in Manchester 24 hours until Champions League Semi Final! #uwcl ' from Alex Morgan. Venus Williams who was going to play for the Washington Kastles (a franchise competing in the World Team Tennis) tweeted, 'Can't wait to play for @WashKastles tonight in Philly' and Maria Sharapova commented that she was 'So thrilled to be playing at one of the biggest indoor arenas I've ever played at today in San Juan'.

The clusters for returned for male athletes were:

Table 4:

Clusters and cluster words for male athletes.

|  |  |
| --- | --- |
| Cluster Label | Words in Cluster |
| Cluster 1: Team | Good, great, happy, congrats, game, team |
| Cluster 2: Hard work | Work, hard, dedication, top, gym, guys |
| Cluster 3: Interview / photo-shoot / video | Story, full, cover, Instagram, dream, success |
| Cluster 4: friends / family support | Thank, wishes, support, kind, best, words |
| Cluster 5: Game Day | Ready, game, tonight, evening, tomorrow,  strong |

Cluster 1 were words related to the team, this resulted from comments like, 'Congrats to the team for bringing home that State Championship', from Carmelo Anthony. Kevin Durant commented on the Washington Redskins win by tweeting, 'Good win Skins, Cowboys pass the sticks', and Wayne Rooney said, 'Good win today and a solid performance from the team Time to get ready for Thursday'.

Cluster 2 contained words relating to hard work. These words came from Tweets like, 'Love it deserve it. Rio Ferdinand tweeted, ' @Chrisgunter16 @ReadingFC Love it... deserve it! Role model to the next gen... hard work and dedication!! Whilst Cristiano Ronaldo encouraged fans to,'Maximize your time in the gym. Work smart, focus on your goals and prepare for any challenge that comes your way #TrueToYourStory #sponsored'.

Cluster 3 was about interviews, videos and photo-shoots. Rio Ferdinand advertised his charity foundation with the tweet, 'Filming at the @riofoundation great work the team are doing.. follow my Instagram story to see behind the scenes!'. Kevin Durant linked to a video of his play with the tweet; I let my play do the talking. #BEHEARD Watch the full @BeatsByDre vid here: https://t.co/05zA4hgRhd https://t.co/Xlq2L6PncE'.

Cluster 4 related to support received from friends and family. Sachin Tendulkar tweeted this comment when he saw a picture of one of his fans who was delighted to receive a souvenir from him 'Now this really brought a big smile on my face I am very grateful for all the love and support'. Kobe Bryant tweeted, ' Thank you all for the birthday shouts and wishes! 40 is right around the corner #virgo #timehaswings'

Cluster 5 is about a game just played or one coming up soon. Sachin Tendulkar said that it was a 'Tough game today well fought out'. Stephen Curry posted the comment, 'S/O to #DubNation showing up strong for the preseason game tonight! Crazy atmosphere. Always Good to see my guy @Mospeights16 on the court!'.

### Word-Cloud

After the cluster analysis, the word frequency was examined and this resulted in the generation of the following table:

Table 5:

Word frequencies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Female\* | Male | Difference | Percentage difference |
| Team | 50 | 156 | -93 | -42.47 |
| Player | 26 | 42 | -12 | -16.67 |
| Game | 118 | 164 | -52 | -18.84 |
| Women | 51 | 17 | 92 | 73.02 |
| Men | 11 | 9 | 4 | 18.18 |
| Family | 28 | 64 | -29 | -54.69 |
| love | 144 | 113 | 52 | 18.71 |
| Happy | 106 | 149 | -23 | -8.36 |
| Best | 66 | 111 | -41 | -22.65 |
| Year | 55 | 94 | -12 | -6.81 |
| good | 110 | 193 | -73 | -23.32 |
| First | 60 | 57 | 8 | 6.56 |
| Today | 121 | 131 | -115 | -78.23 |
| Win | 52 | 106 | -54 | -34.18 |
| congrats | 77 | 156 | -67 | -27.35 |

\* Upon further analysis of the data, it was found that Sania Mirza was an outlier in relation to referencing 'women'. She had accounted for 58 of the 109 references to 'women' in her tweets. This is probably because she is a 'UN Women's Goodwill Ambassador'. As she was affecting the word count, she was replaced with the next highest female athlete on the list, Hope Solo, an America, soccer player.

The total for 'Women' includes the totals for the words: women, woman.

The total for 'Men' includes the totals for the words: men, man.

The total for 'Family' includes the totals for the words: family, families.

The total for 'Year' includes the totals for the words: year, years

The total for 'Congrats' includes the totals for the words: congrats, congratulations, congratulate.

From the word cloud analysis the following two images were created:

Figure 1:

Female athletes Word Cloud.



The most prominent words are: love, happy, today, women and game.

Figure 2:

**Male Athletes Word Cloud.**



For the male athletes, the most prominent words are: game, today, team, good, win and love.

This section stated the results of the research. The next section will look at the results combined with previous research discussed in the literature review. I will examine the relationship between results and the research question, as well as, the relationship between results and literature.

# Discussion:

## Summary of results.

Twitter provides an opportunity for athletes to reach a large audience, as evidenced by Cristiano Ronaldo's 69.7 million followers and Serena Williams 10.6 million. Hambrick et al. (2010) suggest that Twitter communications may be quite different from mainstream sports communications. Rather than sanitized, impersonal communications about the latest game filtered through a team's public relations department, professional athletes' tweets tend to be more direct and address topics in and beyond sport.

This study sought to explore the self-presentation strategies used by professional sports people in the Twitter setting and discern possible differences between genders. Five clusters were developed to help view athlete behaviour on Twitter. A significant difference was found regarding the number of followers each gender were able to attract, as well as the influence athletes have established, with male athletes enjoying sizeable advantages in both cases. The clusters that emerged from this research showed that while there are obvious similarities in the content of tweets, there is also a difference in the content that falls along gender lines. Male athletes tweeted about sport, sports-related activities (e.g. photo-shoots and interviews) as well as addressing family themes. Female athletes also tweeted about these areas, but they also tweeted about other themes that were specific to them, these included: friends, a celebration of events such as birthdays, weddings and Christmas, as well as areas such as equal rights for women.

## Relationship between results and hypothesis or research question

When looking at the data after the cluster analysis had been performed, it was observed that both male and female athletes had clusters representing 'game days' and 'interview, photo-shoots'.

One thing that emerges from the cluster analysis is that the male athletes have three clusters that have a strong connection to their sport, i.e. the 'team', 'hard work' and game day' clusters, whereas the female athletes only have one cluster (Gameday) which is, based on the data, strongly connected to their sport. When the word cloud is then looked at, we see that the top words for female athletes are: love, happy, good, game and women. The top words for men are: good, game, team, congrats, happy. This again points toward male athletes focusing more on sports than female athletes.

Looking at the word count, we see that the biggest differences between the word count for female and male athletes is that female athletes referenced 'women' 51 times compared to the 17 times male athletes referenced women. This is 66.67% more than men. When it comes to references to 'men', women made a reference 13 times, and males made a reference only nine times. The words where men ranked significantly higher were, team, good, and today. These words were nearly always used in connection with a comment about a sports team.

From this, it seems that women have a stronger tendency to tweet about women and trying to forward women's sports and female equality. Alex Morgan commented that she was, 'Proud to fight side by side with these strong women to continue to move the ball forward for the next generation', she also tweeted that, 'The pay gap penalizes women who work hard every day. It hurts our families, businesses and communities. Let's #ChangeTheGame #20percentcounts"'. Ana Ivanovic tweeted that she was, 'Having fun on set with @intimissimi and @mariotestino for the new campaign that celebrates empowered women'. Serena Williams tweeted that, 'On average, women in the US are paid 20% less than men. Black and Hispanic women are paid even less. I support equal pay', this was then retweeted by Caroline Wozniacki.

Men, on the other hand, tended to tweet about gender-neutral themes such as sport and family. Male athletes tweeted about a 'team' 93 times more than women. Kobe Bryant said that he was, 'Excited to be a part of the team'. Cristiano Ronaldo said, 'Great session this morning!!!! Winning team'. Sachin Tendulkar said after a cricket test match, 'What a fabulous way to end the test series! A big hug to the entire team for their outstanding performance. While Carmelo Anthony tweeted that, 'We Are Not A Team Because We Work Together We Are A Team Because We Respect, Trust And Care For Each Other'. When it came to referencing 'family' or 'families' male athletes tweeted 54% more than female athletes. LeBron James told us that he had, "Just saw ""Girls Trip"" with the wife, family and friends. Man I/we was dying laughing all movie!! Stomach still hurting. #MustSee". Rio Ferdinand had a night out and tweeted, 'Top night with spent with family & friends... good to catch up with @WestwoodLee and @anthonyfjoshua #dubai'. These texts let the fan see the athlete away from the competitive environment that they are used to seeing them in. It is one of the 'behind-the-scenes' glimpses that fans look for in twitter interaction.

## Relationship between results and literature

Findings from a study by Backman, Pertick and Yoon (2017) revealed that team attraction, team trust, and team involvement are positively related to team attachment, and these are determinates of fan loyalty. While team attachment was found to positively influence fan loyalty, sports fans' Twitter use was found to significantly reinforce their loyalty. They also found that 'the more sports people strengthen the relationships with their fans through Twitter, the more likely those fans will be committed and attached.'

Haugh and Watkins (2016) found in their study that there is, 'a clear pattern for social-media use by sports fans'. Twitter is the leading destination for fans wishing to gather information, seek entertainment, show support for a team, express opinions on the team, and learn about rules and strategies pertaining to a sport. This follows on from research by Clavio and Walsh (2013) indicating that the highest level of social-media use among their sample of college sports fans was to watch sports videos.

Further research by Brunette et al. (2014) found that there are four primary gratifications sought by Twitter users: interaction, promotion, live game updates, and news. Users wanted to hear about players and roster moves as they happen, learn information faster than other people, read tweets if unable to watch the game, learn about upcoming games and access special promotions.

If sports fans are accessing twitter to receive information related almost exclusively to the sport in which the athlete they are following is a participant it may be that male athlete's willingness to tweet to this very requirement is a factor in their greater number of followers.

The motivations behind the male athlete's tweets are primarily to elicit respect and to be seen as competent according to Jones and Pittman (1982) Their concentration on tweeting about their sport falls into the self-promotion strategy. Their tweets about family life and are most likely intended to elicit affection and be seen as likeable. This is Jones and Pittman's ingratiation strategy. The female athletes tweet about the same themes and are therefore trying to elicit the same responses from their audience. Their tweets about more female-specific topics fall into the 'Exemplification' strategy. A person can accomplish exemplification by presenting him or herself as honest, disciplined, self-sacrificing, generous, or principled. When successful, a person who exemplifies integrity and moral worthiness may be able to influence other people to follow his or her example. (Stone. 2007)

## Limitations and Future Research

While the results of this research are promising, there are limitations that should be addressed. As this is one of the first studies to focus on evaluating the content of tweets using a cluster analysis method, further research is required to validate the methodology and the results.

In addition, social media content and platforms change rapidly. MySpace was one of the top social media sites as recently as July 2007 but by May 2011 it had fallen out of use (Google Trends. 2018). So this research should be seen as an inspection of a point in time. Future research could be used to identify trends in Twitter use and content over time. Other potential research could also see if athletes self-presentation changes over time and if so, does this affect their number of followers.

Also, this research combined individual athletes and team athletes. Future research might wish to look at the differences that may exist between team and individual sports. It is possible that team athletes would be more active on social media as they have opportunities to congratulate teammates on their achievements, etc. As individual athletes do not have these teammates to engage with it may result in a difference in how they engage with social media.

While this research has identified a difference in the way that male and female athletes tweet, further research should be done to determine if indeed this has any correlation to why male athletes gain more twitter followers.

## Conclusion

In the literature review, it was shown that previous research has looked at fan-athlete interaction on social media. The uses and gratifications were identified as well as self-presentation strategies and their motivations.

In the method section, the details of how this piece of research was carried out and the reason for using cluster analysis was explained. The results section detailed the findings of the analysis, and the discussion section looked at these results in combination with the research references in the literature review.

In the very competitive worlds of sport and sport-marketing, professional athletes are increasingly looking for exposure to gain publicity, attract fans, and eventually attract sponsorship and lucrative contracts. As this research shows, both genders have embraced Twitter as a way to engage fans, create positive exposure, and increase their visibility. Both male and female athletes tweet about their sport and their experiences of day to day life as a professional athlete. Male athletes tend to stay in this area whereas female athletes also try to address women's rights inside and outside sport. This is done via the 'exemplification strategy' to influence other people to follow their example and therefore address issues like equal pay and empowering women.

Of the studies that have been done concerning tweets by sports people most focus on using frames to analyse the data, with most of these frames stemming from Goffman's self-presentation theory. All of these studies, including this one, analyse what sports people are tweeting about but they don't address whether it actually affects the number of followers or the amount of enjoyment that followers get from following an athlete. Does it matter to men that female athletes tweet more about women's topics than male athletes? If it does matter, is the effect similar or not in individual sports compared to team sports.

# References:

Arundale, R. 2010. Face as emergent in interpersonal communication: an alternative to Goffman. In: Bargiela-chiappini F. and Haugh M (eds.) Face, communication and social interaction. London: Equinox Publishing, pp.33-54.

Baumeister, R. and Hutton D. 1987. Self-Presentation Theory: Self-Construction and Audience Pleasing. In: Mullen B., Goethals G. (eds) Theories of Group Behaviour. Springer Series in Social Psychology. Springer, New York, NY

Backman, S., Petrick, J. and Yoon, S. 2017. Twitter Power and Sport-Fan Loyalty: The Moderating Effects of Twitter. *International Journal of Sport Communication*, 10(2), pp. 153-177

Brunette , M., Gibbs, C. and O’Reilly, N. 2014. Professional Team Sport and Twitter: Gratifications Sought and Obtained by Followers*. International Journal of Sport Communication*, 7(2), pp. 188 -213

Clavio, G. and Kian, T. 2010. Uses and Gratifications of a Retired Female Athlete’s Twitter Followers. *International Journal of Sport Communication*, 3(4), pp. 485-500

Clavio, G. and Walsh, P. 2014. “Dimensions of Social Media Utilization among College Sport Fans.” *Communication & Sport*, 2 (3), pp. 261–281.

 Cooper, C. 2008. NCAA Website Coverage: An Analysis of Similar Sport Team Gender Coverage on Athletic Department's Home Web Pages. *Journal of Intercollegiate Sports*, 1(2), pp. 227-241

Cunningham, G. 2003. Media coverage of women’s sport: A new look at an old problem. *Physical Educator*, 60(2), p.43

Cromwell, G. (2009). Armstrong's tweet turns out more than 1,000 riders for a jaunt around Dublin. VeloNews. [Online]. Available from: http://www.velonews.com/2009/08/news/armstrongs-tweet-turns-out-more-than-1000-riders-for-a-jaunt-around-dublin\_97144 [Accessed 1 November 2017]

Force, J. 2016. Social Media Trends in Sports Marketing [Online]. Available from: https://sysomos.com/2016/08/26/social-media-trends-sports-marketing/ [Accessed 9 November 2017]

Google trends. 2018. MySpace [Online]. Available from: https://trends.google.com/trends/explore?date=all&q=myspace [Accessed 23 February 2018]

Hambriek, M., Simmons, J., Greenhalgh, G. and Greenwell, T. 2010. *Understanding  Professional Athletes' Use of Twitter: A Content Analysis of Athlete Tweets.* International Journal of Sport Communication, 2010, 3, pp. 454-47

Haferkamp, N., Eimler, S., Papadakis, A. and Kruck, J. 2012. Men are from Mars, women are from Venus? Examining gender differences in self- presentation on social networking sites.  *Cyberpsychology, Behavior and Social Networking*, 15(2), pp.91-98.

Haugh, B. and Watkins, B. 2016. Tag Me, Tweet Me if You Want to Reach Me: An investigation Into How Sports Fans Use Social Media. International Journal of Sport Communication, 2016, 9,278-293

*Gender stereotypes in Facebook profiles: Are women more female online? (PDF Download Available)*. Available from: https://www.researchgate.net/publication/295402737\_Gender\_stereotypes\_in\_Facebook\_profiles\_Are\_women\_more\_female\_online#pf6 [accessed Nov 23 2017].

Hardin, M., Simpson, S., Whiteside, E and Garris, K. 2007. The Gender War in U.S. Sport: Winners and Losers in News Coverage of Title IX. *Mass Communication and Society*, 2(10), pp. 211-233

Hull, K. 2014. “A Hole in One (Hundred Forty Characters): A Case Study Examining PGA Tour Golfers’ Twitter Use during the Masters.” *International Journal of Sport Communication*, 7 (2), pp. 245–260.

Jones, E. and Pittman, T. 1982. Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self,* 1(1), pp. 231–262). Hillsdale, NJ: Erlbaum.

Kian, M. and Clavio, G. 2011. A Comparison of Online Media and Traditional Newspaper Coverage of  the Men’s and Women’s U.S. Open Tennis Tournaments. *Journal of Sports Media*, 6(1), pp. 55-84

 Kian, M., Mondello, M. and Vincent, J. 2009. ESPN—The women’s sports network? A content analysis of internet coverage of march madness. *Journal of Broadcasting and Electronic Media*, 53(3), pp.477–495.

Kane, M. and Maxwell, H. 2011. Expanding the boundaries of sport media research: Using critical theory to explore consumer responses to representations of women’s sports. *Journal of Sport Management*, 25(3), pp. 202-216.

Kassing, J. W., & Sanderson, J. (2010). Fan-athlete interaction and Twitter: Tweeting through the Giro: A case study. *International Journal of Sport Communication*, 3(1), pp. 113–128.

 Laird, S. 2015.The top 15 sporting events that blew up Twitter in 2015 [Online]. Available from: http://mashable.com/2015/12/07/2015-top-sports-events-twitter/#jsSMyHwK\_Sqz [Accessed 9 November 2017]

Lebel, K. and Danylchuk, K. 2012. HowTweet It Is: A Gendered Analysis of Professional  Tennis Players' Self-Presentation on Twitter. *International Journal of Sport Communication*, 5(4), pp. 461 -480

Manago, A., Graham, M., Greenfield, P. and Salimkhan, G. 2008. Self-presentation and gender on MySpace. *Journal of Applied Developmental Psychology*, 29(6), pp.446–458

Mehdizadeh, S. 2010. Self-Presentation 2.0: Narcissism and Self-Esteem on Facebook. *Cyberpsychology, Behavior and Social Networking,* 13(4), pp.357-364

Miller, H. 2012. The presentation of self in electronic life: Goffman on the Internet [Online] Available from:  http://www.dourish.com/classes/ics234cw04/miller2.pdf

Newman, T., Peck, J., Harris, C. and Wilhide, B. (2017). Social media in sport marketing. Scottsdale, AZ: Holcomb Hathaway.

Oram, A. 2009 What sociologist Erving Goffman could tell us about social networking and Internet identity. [Online] Available from: http://radar.oreilly.com/2009/10/what-sociologist-erving-goffma.html

Pegoraro, A. 2010. “Look Who’s Talking – Athletes on Twitter: A Case Study.” International Journal of Sport Communication 3 (4): 501–514.

R-bloggers 2017. Clustering Search Keywords Using K-Means Clustering [Online] Available from: http://tagteam.harvard.edu/hub\_feeds/1981/feed\_items/274117 [Accessed 20 November 2017]

Sanderson, J. and Kassing, J. 2012, Tweets and blogs: Transformative, adversarial, and integrative developments in sports media. *Sports Media: Transformation, Integration, Consumption.* Taylor and Francis, pp. 114-127.

Serena Williams 2017. https://twitter.com/serenawilliams/status/887491335980298240?lang=en

Stone, J. 2007. Exemplification. In: Baumeister, R. and Vohs, K.(eds) Encyclopedia of Social Psychology. SAGE Publications.

Shultz, B. and Sheffer, M. 2010. Paradigm Shift or Passing Fad? Twitter and Sports Journalism. *International Journal of Sport Communication*, 3(4), pp. 472-484

Trevino, A. 2016. Introduction to K-means Clustering [Online]. Available from: https://www.datascience.com/blog/k-means-clustering [Accessed 27th February 2018]