Football and Beer - a Social Media Analysis on Twitter in Context of the FIFA Football World Cup 2018

Roland Roller, Philippe Thomas, Sven Schmeier
Language Technology Lab, DFKI,
Berlin, Germany
{firstname.surname}@dfki.de

Abstract

In many societies alcohol is a legal and common recreational substance and socially accepted. Alcohol consumption often comes along with social events as it helps people to increase their sociability and to overcome their inhibitions. On the other hand we know that increased alcohol consumption can lead to serious health issues, such as cancer, cardiovascular diseases and diseases of the digestive system, to mention a few. This work examines alcohol consumption during the FIFA Football World Cup 2018, particularly the usage of alcohol related information on Twitter. For this we analyse the tweeting behaviour and show that the tournament strongly increases the interest in beer. Furthermore we show that countries who had to leave the tournament at early stage might have done something good to their fans as the interest in beer decreased again.

1   Introduction

Alcohol can lead to serious health issues. For instance, even though there is no apparent threshold, even one drink of alcohol per day on average can significantly increase the risk of cancer (Roerecke and Rehm, 2012).

Studies have shown, that the exposure to media and commercial communications on alcohol is associated with the likelihood that adolescents will start to drink alcohol, and with increased drinking amongst baseline drinkers (Anderson et al., 2009). In addition to that social events can have a influence on drinking behaviour. In course of this Curtis et al. (2018) apply a Twitter analysis and show that topics such as sporting events, art and food-related festivals are positively correlated to alcohol consumption on US county level.

Various other studies have also explored alcohol-related content on social media, particularly Twitter. Abbar et al. (2015) carry out a food

analysis on Twitter and identify weekly periodicities in context of daily volume of tweets mentioning food. Moreover, authors show a correlation between state obesity and caloric value of food (also alcoholic beverages). Instead Culotta (2013) analyse alcohol sales volume in context of Twitter messages. Kershaw et al. (2014) investigate regional alcohol consumption patterns in the UK, while Hossain et al. (2016) explore alcohol consumption patterns in various areas in the US. Curtis et al. (2018) target the prediction of excessive drinking rates and Huang et al. (2017) examine alcohol- and tobacco-related behavioural patterns. And finally, Moreno et al. (2010) carry out a content analysis of adolescents on social media.

This work examines alcohol consumption on Twitter in context of the FIFA Football World Cup 2018. We make use of the results of the above mentioned works, especially the observed correlation between people's behaviour on Twitter and in their real life in context of consumption. The study is carried out across all participating countries of the tournament and explores the influence of the event on the drinking behaviour of people.

2   Experimental Setup

The FIFA World Cup 2018 was taking place from 14th of June until 15th of July. Within the group stage 32 participating teams were playing in 8 groups and completed 3 matches each. After that the two best teams of each group went to the knockout stage. A match usually lasts 90 minutes plus 15 minutes of break ($\approx$2 hours). This work analyses the tweeting behaviours during a match. In the following a match is defined as a time period of one hour before kick-off and three hours after the kick-off. The hour before and after the game are included as supporters might express excitement for the game.

2.1 Data Collection

Tweets over a period between 05/31/2018 and 07/23/2018 were collected, which covers the period of the tournament, but also two weeks before and one week after. As this work analyses messages from all participating countries of the tournament, messages were crawled containing various emojis due to their language independence. The considered emojis are listed in Table 1. In the rest of the work we refer to them as *BEER*, *WINE*, *SAKE* and *BALL*.

Table 1: Emojis used for Twitter crawling

2.2 Country Assignment

In this work, Tweets are examined according to the different participating countries. As only a small number of Tweets contain country related information (8.63%), Tweets lacking this information had to be assigned automatically to the corresponding country of each user. A classifier was trained based on the approach of Thomas and Hennig (2018), which is able to detect the origin of a Tweet based on text and meta information.

As sanity check, Tweets which actually contain information about its origin were compared to the automatically assigned country. On those messages the model achieves an accuracy of above **91**%.

2.3 Preprocessing

Collected Tweets were then mapped to small time intervals of one hour, according to the target label (e.g. beer emoji). For instance a message sent on 06/27/2018 at 5:25 pm (UTC) and containing a beer emoji is assigned to its country, then assigned to the set of beer emoji Tweets, and finally mapped to the time interval 06/27/2018, 5:00 pm. All messages of a particular label and a particular time interval are summed up. In this way a list for each target label is generated containing *time intervals*, *country origin* and *number of relevant Tweets* for this interval. For the following analysis these lists are used as input. One line is considered as TCF (time-country-frequency) triple.

3 Analysis

In the following the different Tweets containing the target emojis are analysed in detail. All ex-

aminations which include significance testing use one-sided paired t-test.

3.1 Outside the Tournament

Figure 1 shows the average number of Tweets per day containing *BEER* emojis before/after the tournament from the participating countries. Participants with less than 50 *BEER* Tweets per day are excluded, to show more meaningful results. We refer to this group as *avg50*. In average, more than half of the *BEER* Tweets per day come from Brazil and England together. Results show no direct correlation to the statistics *Harmful use of alcohol*[1] of the WHO. The reasons for this can only be possibly found out by a longterm analysis going further than pure statistics. In that list Brazil with 7.8 litres of pure alcohol per capita is actually ranked further to the end. Instead countries such as Germany (13.4), France (12.6), England (11.4) and Australia (10.6) would be ranked to the top.
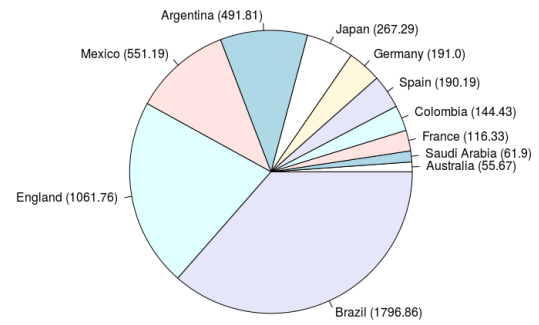

Figure 1: Average number of alcohol related (BEER) Tweets per day outside the tournament

In order to draw a fair comparison to *pure alcohol per capita* the number of active Twitter users must be taken into account. It turns out that dividing the avg. *BEER* Tweets by the number of active Twitter users does not change much. Argentina switches place with England, Saudi Arabia moves to the very end and Japan drops just in front of it. Columbia moves slightly up.

Next mean is calculated for all *BEER* Tweets from *avg50* for each day of the week. The resulting graph is presented in Figure 2 and visualises, similarly to Abbar et al. (2015), particular periodicities. Firstly single days can be recognised as small peaks. Moreover, towards the end of the week, peaks are slightly increased compared to the beginning of the week. Using this data it is for in-

---

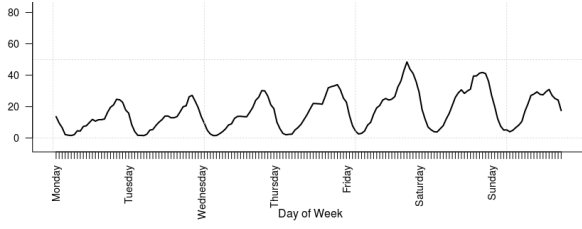[1] http://apps.who.int/gho/data/node.sdg.3-5-viz, accessed 19.07.2018

Figure 2: Mean of all alcohol related (*BEER*) Tweets before/after the tournament from *avg50*

stance possible to deduct, that **people tweet significantly more about alcohol on the evening** (from 4pm until 1am) (p<0.001). Moreover, the data also shows, that **people tweet significantly more about alcohol at the weekend** (Friday 4pm - Monday 6am) (p<0.001). in comparison to the rest of the week.

3.2 The Tournament

In this subsection alcohol related Tweets during the tournament are examined. The first question to address is whether supporters of their national team tweet more during the match in comparison to other periods. Reference periods are the days after each match during the same time slots.

The analysis shows that **people from 19 countries use *BEER* significantly more when their team is playing** (p<0.05, 10 of them with p<0.001). Among the 13 other countries, only Japan and Saudi Arabia are from *avg50*. Interestingly Croatia, which reached the final, does not show any significant increase, but the general usage of *BEER* is generally very low here. Considering *WINE* Tweets, only Brazil, Poland and Belgium and for *SAKE* only Mexico show a significant increase in Tweets during the matches of their team (p<0.05). However the number of Tweets are low in comparison to *BEER*.

Figure 3 presents an overview on how the tournament influences the average usage of *BEER* per day of *avg50*, while the team is in the tournament. France shows the largest increase of BEER Tweets per day in average of more than 107%, followed by Japan with 35%. The increase from Japan is surprising as Japanese people do not tweet significantly more during the matches of their team. Possible explanations might be that matches are broadcasted often late in the evening due to the time difference to Russia. For this reason people might meet up earlier, thus start drinking earlier. Another explanation can be just the fact that there is a high interest for the tournament in general in
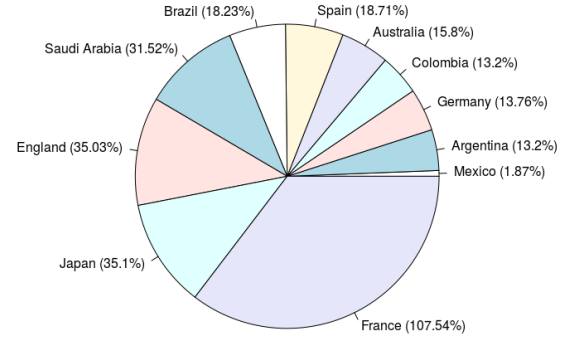
Japan.

Figure 3: Increase of beer related Tweets per day during World Cup, until leaving the tournament

Brazil instead shows 'only' an increase of 18%. Even with this small increase (in comparison to others) Brazil remains the country with the largest number of avg. *BEER* Tweets per day. Considering all countries Morocco has the strongest increase (407.5%) and also the strongest decrease after leaving the tournament (-68.97%). On the other hand, outside the tournament Morocco has a very low number of *BEER* Tweets per day (0.48), so the increase might be not too serious. Interestingly only Peru shows a decrease during the tournament (-16.4%). Considering the *avg50*, Colombia and Brazil showed the strongest decrease when their team left the tournament with -18.81% and -10.59% respectively.

Generally the results show, that from almost all countries more alcohol related Tweets can be found during the tournament. Moreover, in most cases the avg. number of *BEER* Tweets decreases when the team leaves the tournament. However, in some cases an increase in Tweets can be detected. Senegal for instance increases the number of *BEER* Tweets up to 54%, followed by Uruguay (28%) and Australia (10%).

3.3 Top-5 Matches

This subsection analyses the different matches of the tournament for popularity in terms of *BEER* and *BALL*. In order to have a fair comparison data is normalized first. The average number of Tweets of each country outside the tournament is subtracted from the number of Tweets during the tournament at a given time and day.

Table 2 presents the Top-5 matches involving *BEER* and *BALL*. The table shows that more people use the football emoji than the beer emoji. In terms of *BALL*, the Top-5 list contains the final

3

(France-Croatia), the opening (Russia-Saudi Arabia) and some other games involving recent European and World Champions.

🍺	#	⚽	#
Mexico-Sweden German-S. Korea	1313	Portugal-Spain	5100
Brazil-Belgium	1305	France-Croatia	4637
Serbia-Brazil Switzerland-C. Rica	1250	Russia-S. Arabia	4253
Nigeria-Iceland	1104	Germany-Mexico	3798
Brazil-Costa Rica	1092	France-Argentina	3616

Table 2: Top-5 matches of the tournament in terms of beer and football emoji (normalized)

From *BEER* perspective we find on the first and third palce matches which took place in parallel. Considering that, the most popular single match was Brazil-Belgium in the Quarter Final. Ranked 4th is Nigeria-Iceland, which is surprising, as both countries are not tweeting much about beer. Analysing the results in more detail reveals, that all 7 matches took place on a day Brazil played. Even though Tweets were normalized, the influence of Brazilian *BEER* Tweets before and after a match of their team is enormous, so that even the Nigeria-Iceland match achieved a high rank.

4 Results

This work presented a short analysis of alcohol related emojis in context of the FIFA football World Cup 2018. With the start of the tournament we showed, that most countries strongly increase the number of Tweets containing beer emojis. As many people tweet less after their national team left the tournament, we draw the conclusion that leaving the tournament early, as Germany did, is the healthiest solution - unless you are Peru. We also showed that people of many participating teams of the tournament tweet significantly more about alcohol during a match of their team. Furthermore we presented the increase of alcohol related Tweets during the tournament and the most popular games in terms of beer and football emojis. Finally we showed, that Brazil tweets by far the most about beer. Cheers!

Acknowledgements

References

Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3197–3206, New York, NY, USA. ACM.

Peter Anderson, Avalon de Bruijn, Kathryn Angus, Ross Gordon, and Gerard Hastings. 2009. Impact of alcohol advertising and media exposure on adolescent alcohol use: A systematic review of longitudinal studies. *Alcohol and Alcoholism*, 44(3):229–243.

Aron Culotta. 2013. Lightweight methods to estimate influenza rates and alcohol sales volume from twitter messages. *Language resources and evaluation*, 47(1):217–238.

Brenda Curtis, Salvatore Giorgi, Anneke E. K. Buffone, Lyle H. Ungar, Robert D. Ashford, Jessie Hemmons, Dan Summers, Casey Hamilton, and H. Andrew Schwartz. 2018. Can twitter be used to predict county excessive alcohol consumption rates? *PLOS ONE*, 13(4):1–16.

Nabil Hossain, Tianran Hu, Roghayeh Feizi, Ann Marie White, Jiebo Luo, and Henry A. Kautz. 2016. Precise localization of homes and activities: Detecting drinking-while-tweeting patterns in communities. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016.*, pages 587–590.

Tom Huang, Anas Elghafari, Kunal Relia, and Rumi Chunara. 2017. High-resolution temporal representations of alcohol and tobacco behaviors from social media data. *Proceedings of the ACM on human-computer interaction*, 1(CSCW).

Daniel Kershaw, Matthew Rowe, and Patrick Stacey. 2014. Towards tracking and analysing regional alcohol consumption patterns in the uk through the use of social media. In *Proceedings of the 2014 ACM conference on Web science*, pages 220–228. ACM.

Megan A. Moreno, Leslie R. Briner, Amanda Williams, Libby Brockman, Leslie Walker, and Dimitri A. Christakis. 2010. A content analysis of displayed alcohol references on a social networking web site. *Journal of Adolescent Health*, 47(2):168 – 175.

Michael Roerecke and Jürgen Rehm. 2012. Alcohol intake revisited: Risks and benefits. *Current Atherosclerosis Reports*, 14(6):556–562.

Philippe Thomas and Leonhard Hennig. 2018. Twitter Geolocation Prediction Using Neural Networks. In *Language Technologies for the Challenges of the Digital Age*, pages 248–255, Cham. Springer International Publishing.