Gender differences in language about Feminism: Results from Sentiment Analysis and Use of Emojis on Twitter

Peter Bußwolder, Independent Scholar, Germany

The European Conference on Arts and Humanities 2021
Official Conference Proceedings

Abstract
Social networks, such as Twitter with its around 192 million active users per day, are increasingly changing the way how people access information, communicate with each other, express opinions and discuss a wide range of topics. An example of a rather controversial topic is feminism. This study sheds light on the used language and emojis when discussing feminism on Twitter. Emojis are graphic symbols, representing inter alia facial expressions, but also objects, food or drinks, animals, or emotions, and feelings. For the analysis, 195,843 evaluable tweets were collected between the end of February until the beginning of March 2021, covering the International Women's Day and part of Women's History Month. A quantitative approach is employed to evaluate the sentiment value of tweets on a lexical level. Sentiment analysis enables the investigation of public emotions about events, opinions, persons etc. Together with the sentiment value of the emojis, it provides the basis to analyze the identified words and topics of the discussions on Twitter. Additionally, as Twitter does not provide the gender of a user, the gender is tried to be derived from unstructured data such as the screen or username as well as the description. Results indicate that female users send in average tweets with a more positive tone than male users, while negative tweets are not significantly different between genders. Emojis are only used in a part of all tweets. The emojis used are correlated to the sentiment value of the tweet.

Keywords: Feminism, Twitter, Sentiment Analysis, Public Opinion, Emojis

Introduction

On microblogging services, such as Twitter, Facebook, or Tumblr, users can write short text messages to express opinions on a variety of topics and to discuss current subjects. Caused by the accessibility and the potential audience for one's opinions, these platforms are seeing significant growth in users and exchanged messages. Due to the amount of created content and the expressed sentiment, opinions or experiences, such microblogging services become increasingly interesting for researchers of different fields to exploit this large database.

Feminism, as a controversially discussed topic of high societal importance, is a field of research regularly analyzed. Tweets are for example used to gain insights into trolling and how people cope with this behavior (Lopez et al., 2018), rape threats on social media (Hardaker & McGlashan, 2016) or the coping strategies and social reactions to #MeToo (Schneider & Carpenter, 2018).

A prevalent method of previous studies was the context analysis, which allows a deep analysis of the content and context of tweets but can only analyze a limited number of tweets. Furthermore, sentiment analysis has been used. Here the emotional content or text polarity is automatically processed, and the text is classified as positive, negative, or neutral (Jianqiang & Xiaolin, 2017). This allows for a higher number of analyzed tweets.

Despite the research, several research gaps remain. As Twitter does not provide the gender of its users, there is no analysis of the differences in revealed sentiments of female and male users about feminism. The gender needs first to be derived by clues given by the user (Vicente et al., 2018). Thus, this research tries to shed light on the sentiment of male and female users on the topic of feminism on Twitter and the language used, to express sentiments.

The linguistic study is in a second step extended by an analysis of emojis. Emojis are abstractions of facial expression, gestures, objects such as food, vehicles etc. and are used increasingly in computer mediated communication (Vidal et al., 2016; Walther & D'Addario, 2001). They provide additional emotional cues and therefore augment the statement of a message (Derks et al., 2007; Chen et al., 2017).

The alignment of the study at hand is explorative. It tries to describe various aspects of the language and emoji use differentiated by gender. This helps to outline further research directions.

The remainder of this paper is structured as follows. First, related studies and the research questions are discussed in the theory section. Second, the incorporated methods are descripted in detail. After that the results are presented. Finally, the paper concludes with a discussion of limitations and open questions for future research.

Related Work and Theory

Microblogging services such as Twitter allow users to share short messages and opinions about a variety of topics. In recent years microblogging services became increasingly more popular (Pak & Paroubek, 2010). Twitter alone has around 192 million active users per day (Twitter Inc., 2021), which generates a huge amount of data every minute. This data is of significant interest for researchers, focusing on user's opinion and sentiments (Vicente et al., 2018). Social and political activism happens on social media to initiate grass-root mobilization, providing interesting opportunities for social scientists (Tinati, et al., 2014). Studies have been conducted on, e.g., racism in elections (Stephens-Davidowitz, 2014), flow of information in protests (Tinati et al., 2014), gender differences in sports coverage (Sainz-de-Baranda et al., 2020) or rape threats (Hardaker & McGlashan, 2016).

Feminism is also being studied by using tweets to identify positive or negative sentiments or opinions. As feminism discusses issues of gender relations, a sentiment analysis should capture comparable dimensions of gender attitude (Scarborough, 2018). Lopez et al. (2018) made a content analysis of all tweets with the hashtag #feminism, which were send within a timespan of 24 hours. While many users engage with the topic of feminism, promote it, and learn about it, some also expressed disagreement, misogyny or even violence. Schneider and Carpenter (2018) conducted a study on the hashtag #MeToo to focus on the social reactions. Most tweets indicated a belief and offered emotional support. They used sentiment analysis along with content and context analysis to get their results.

Several studies contribute to methodical aspects of studying social media. Waseem and Hovey (2016) focused on hate speech and provided a list of criteria and a dictionary of indicative words for identifying offensive tweets. Dilai and Levchenko (2018) transferred the SentiStrength algorithm to Ukrainian to analyze the discourse about feminism in the Ukraine.

Sentiment analysis, which has been used in several previous studies on social networks and is also used in this study, is the automatic classification of a text as positive, negative, or neutral (Jianqiang & Xiaolin, 2017). It is part of Natural Language Processing, which can be applied from document classification down to determining the polarity of single words (Kouloumpis et al., 2011). As Twitter constitute a special environment for communication, especially because of its limitation to a maximum of 280 symbols per tweet (Rosen, 2017), the relevant features for analysis are also influenced by it. The baseline is constituted by n-grams, supplemented by sentiment lexica, such as the AFINN lexicon (Nielson, 2011). Lastly, particular micro-blogging features are included, like abbreviations or intensifiers (all-caps or character repetition) (Kouloumpis et al., 2011).

The goal is to try to gain insights into how male and female Twitter users differ in their language used when discussing feminism. This is done based on a sentiment analysis.

The micro-blogging features are complemented in this study by an analysis of emojis. Emojis are abstractions of facial expressions, bodily gestures etc. and help to communicate emotions or moods in computer mediated communications (Vidal et al., 2016; Walther & D'Addario, 2001). Thus, emojis can help to transport additional social information beyond the text and augment the meaning of the entire message (Derks et al., 2007). The usage of emojis in social media has increased significantly over the previous years (Huang et al., 2008; Huffaker & Calvert, 2005). The popularity is also reflected in the increasing number of available emojis (Ljubešić & Fišer, 2016). In case of Twitter, the user can select out of a set of 3.245 emojis (Twitter Open Source, n.d.).

By using geo-located tweets, Ljubešić & Fišer (2016) came to the result that about 20% of all tweets contained emojis and around 38% of all users used them. Users from the U.S. appeared to use relatively less emojis. Analyzing the demographics of U.S. Twitter users Mislove et al. (2011) showed a bias towards overrepresentation of urban regions and males. A sentiment lexicon for emojis was developed by Novak et al. (2015) by analyzing annotated tweets. Most of the 751 classified emojis were positive, which is especially the case for the most popular ones.

The present work aims to provide an analysis of how emojis are used by male and female users and how their sentiment value corresponds to the context in which they were used.

Methods

In this section the data collection via the Twitter API are described and the steps taken to pre-process the data. The evaluation of the sentiment value of each tweet is depicted in detail and the process of gender identification is explained.

Data Collection

The tweets were collected using the R package rtweet (Kearney, 2019). The search period was between February 24th until March 9th, 2021, covering the International Women's Day and part of Women's History Month. The search terms were related to feminism and taken from Scarborough (2018). The search terms were: feminist, feminism, women's rights, womens rights, women's rightist, womens rightist, women's liberationist, womens liberationist, women's libber, womens libber, women's liberation, womens liberation, women's lib, and womens lib. Search terms were used with and without apostrophes, as Twitter users often do not use punctuation. The initial sample consisted of 603,381 tweets originating from 402,054 different users.

Pre-Processing

The first step of pre-processing was the exclusion of all retweets and non-English tweets. Retweets constitute a form of indirect interaction (Hardaker & McGlashan, 2016), while tweets in other languages than English obscure the analysis. This resulted in a dataset of 195,843 tweets by 133,765 unique users.

After that the text was first tokenized and normalized. Upper-case letters were changed into lower case letters and numbers were removed along with Twitter tokens such as usertags (*@user*) and URLs. Hashtags (*#hashtag*) were kept for the analysis. Finally, stop words were excluded such as *the*, *all*, *of*, *and* or *with*. They are not associated with sentiments (Scarborough, 2018; Jianqiang & Xiaolin, 2017).

Sentiment Analysis

The sentiment analysis for Twitter is based on word n-grams, primarily on unigrams, expanded by bigrams. Values were taken from the AFINN lexicon, which lists English terms with an integer between -5 (negative) and +5 (positive) (Nielsen, 2011). Basic negation detection was done by identifying bigrams beginning with a negation and ending with an AFINN evaluated word (Kouloumpis et al., 2011). This is relevant for sentences such as "I'm not happy.", where the word "happy" would indicate a positive sentiment, but the negation "not" does reverse the meaning. Thus, considering bigrams does improve the accuracy of sentiment evaluation (Pak & Paroubek, 2010).

About 72.8% of all tweets included at least one word out of the AFINN lexicon. The remainder of the tweets were very short, including single words, mentions of other users, links or emojis. For the further analysis regarding sentiment only the evaluable tweets will be considered. The statistical analysis is based on the sum of the values from the AFINN lexicon for each tweet.

Gender Identification

The process of gender identification, adopted in this study, orients oneself towards the work of Vicente et al. (2018). Deducing the gender of a Twitter user is necessary, as the information Twitter provides about its users, is limited. The only required field for a user profile is a user name. Besides this, further information can be provided by the user through the choice of a screen name, a short description, a profile picture, or the content of the tweets. Here, features are extracted from the user and screen names as well as from the description.

A list with first names associated with gender and number of occurrences was compiled with data from the United States Social Security Administration (Social Security Administration, n.d.). Names with less than 4 characters and less than 150 occurrences in 2019 were excluded to avoid false-positive allocations and to reduce the computational burden. Table 1 states examples of screen names and matched names with genders. As in the last example, it is possible that several names may be allocated.

Screen Name	**Found name**
Rebecca__Rouse	Rebecca (Female)
FloraKingi	Flora (Female)
Ahmadabt212	Ahmad (Male)
trevortjames	Trevor (Male), James (Male)

Table 1: Examples Twitter Screen Names with Matched Names

The description was searched for clues indicating the gender of the user, such as words like “mom”, “mother”, “wife” (female) or “dad”, “daddy”, “husband” (male). Table 2 shows random examples for descriptions including gender indications.

<table>
<tr><th>Gender</th><th>Identified clue</th><th>Description</th></tr>
<tr><td rowspan="2">Female</td><td>Mother, wife</td><td>Feminist, mother, wife, fan of good wine and good conversation</td></tr>
<tr><td>Mom, wife</td><td>Book lover, mom, wife, ex-Sailor with a sailor’s mouth, Dem, RN and rabble rouser</td></tr>
<tr><td rowspan="2">Male</td><td>Father, husband</td><td>Flawed father & husband. Believe in #EqualityforAll. 🇬🇧 🇨🇮 🇫🇷 🇱🇰 🇵🇪 🇳🇬 are always special places to me. He/Him/#HeforShe. domwilliamsl on Instagram.</td></tr>
<tr><td>Husband, daddy</td><td>Malaysian. LGBT. Activist. Leader. Community Organiser. Human Rights Defender. Husband. Cat's daddy. Cuddler. Listener. Advisor.</td></tr>
</table>

Table 2: Examples Twitter Description with Gender Clues

Sample three in Table 2 reveals an additional gender reference. “He/Him” is an example for a so-called pronoun introduction, where people state their preferred pronouns, thus recognizing that gender is complex and that non-binary people are not alienated (Mahdawi, 2019). The description was additionally searched for the terms “she/her”, “she/they”, “he/him” and “he/they”, the first two as female, the last two as male indications.

In total, 49.9% of all users were assigned a gender, of which 25.5% are female and 24.4% male. The balance may be caused by the topic at hand, as Twitter has a higher proportion of male users, even though the gender bias is getting smaller (Mislove et al., 2011).

There are two significant limitations to the described method for inferring gender. First, from a viewpoint of accuracy the method does not always provide a clear distinction. User and screen name may be freely chosen, thus not be related to any name in the list of the Social Security Administration. As previously mentioned, several names may be matched to one name, resulting in ambiguity. A description is not mandatory (92% of the users in the sample have a description) and may not contain information usable for gender identification.

Second, the binary division does not consider the fluid relationship between the biological sex and the gender identity (Blevins & Mullen, 2015). Especially in the dataset at hand, users may not locate themselves within a binary division of male and female, but see themselves as trans person, queer, or bisexual. Therefore, the results need to be interpreted with caution in the context of social and cultural practices.

Results

The discussion of the results is split firstly, into the presentation of the sentiment analysis. Secondly, the gender assignment is integrated and finally the emojis are analyzed in relation to the context of the sentiment and gender mapping.

Sentiment

The average sentiment value of a tweet is positive with a mean of 0.1477 and a median of 1. The spread is significant between -51 and +36. 50.1% of the tweets in the sample are positive, 44.4% negative, and 5.5% neutral. In the latter case the evaluation scores of positive and negative words cancel each other out. Figure 1 shows the histogram of the sentiment distribution. It has a long tail on both sides. For clarity it is truncated, ignoring values above +20 and below -20. Less than 0.1% of the cases are excluded.

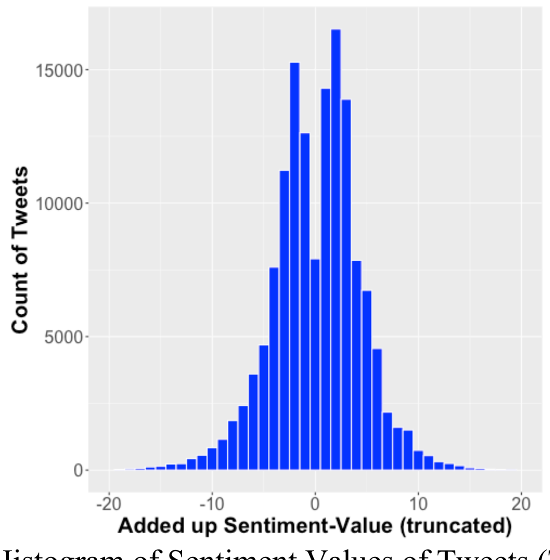

Figure 1: Histogram of Sentiment Values of Tweets (Truncated)

The use of hashtags allows Twitter users to link their contribution to one or several topics. This makes it easier to participate in online discussions. Table 3 shows the top 25 hashtags in the dataset, along with several statistics. In total 20,262 unique hashtags are used, with the 25 most used hashtags making up 40.4% of all mentioned hashtags, indicating a skewed distribution.

No.	Hashtag	Group	N	Average sentiment	No.	Hashtag	Group	N	Average sentiment
1	#internationalwomensday	IWD	8,358	3.3309	14	#womenpower		860	4.1265
2	#feminism	Gen	6,929	1.6193	15	#womensupportingwomen		794	4.1541
3	#iwd2021	IWD	5,047	2.8010	16	#happywomensday	IWD	765	4.8190
4	#womensday	IWD	2,929	3.9190	17	#generationequality		692	2.5943
5	#feminist	Gen	2,764	1.5724	18	#equality	Gen	634	1.7060
6	#twibbon		1,876	5.0011	19	#8m2021		535	1.4737
7	#women		1,778	3.0940	20	#womeninbusiness		513	4.2887
8	#womenshistorymonth		1,744	2.9092	21	#metoo		494	-1.4762
9	#choosetochallenge		1,464	3.4129	22	#inspiration		437	4.1696
10	#womenempowerment		1,396	4.1382	23	#womensday2021	IWD	411	4.8873
11	#iwd	IWD	1,373	2.7854	24	#love		408	3.4200
12	#girlpower		1,093	4.3127	25	#auratmarch2021		392	0.6335
13	#internationalwomensday2021	IWD	891	3.7205					

Note: IWD: Internatinal Women's Day; Gen: General; 8M: Aurat March

Table 3: Most Used Hashtags with Statistics (Sorted by Frequency)

Several hashtags can be grouped as they represent the same topic. Primarily the International Women's Day is mentioned with seven different hashtags, three of them in the top four. The general hashtags covering feminism (#feminism, #feminist, #equality) show significantly lower sentiment values, indicating how controversial the topic is discussed. Lower values are only received by the hashtag #metoo, with the only negative evaluation, which is comparable to Schneider and Carpenter (2017), and the hashtags #8M2021 and #auratmarch2021. #8M2021 combines the fight against patriarchy with the one against capitalism. The Aurat March is a political demonstration organized in Pakistan also annually on March 8th.

The hashtag #twibbon deserves special attention. It relates to a website for organizing campaigns, where users can support causes, brands, or organizations[1]. It allows to add short text, logos, or a colored ring to the Twitter profile image of a user, to show that he or she is sympathetic with a statement. It is also possible to send a preset tweet via one's own account. For the "Violeta feminista", a Spanish speaking campaign related to feminism, an identical tweet was sent out about 1,872 times within the period of data collection. This explains the highest sentiment score of the inspected hashtags.

Gender and Sentiment

For the following part of the analysis, only tweets are considered with a gender-assigned user and where a sentiment calculation was possible. These amount to 71,740 tweets, which are about 50.3% of all tweets. Table 4 gives an overview of the distribution by gender and sentiment category.

[1] https://twibbon.com/

Sentiment Category	Number of Tweets		Average Sentiment of the Tweets	
	Female	Male	Female	Male
Positive Sentiment	19,572 – 52.1%	16,589 – 48.6%	3.57	3.42
Neutral Sentiment	1,965 – 5.2%	2,077 – 6.1%	0	0
Negative Sentiment	16,057 – 42.7%	15,480 – 45.3%	-3.66	-3.61
Total	**37,594**	**34,146**	**0.29**	**0.03**

Table 4: Number of Tweets by Gender and Average Sentiment Value

By inspecting the distribution, two aspects become noticeable. Firstly, the proportion of positive tweets is higher when written by female users, while correspondingly male users have a higher share of negative tweets. Secondly, the results show that male users have a significantly more negative sentiment value (t(71.453) = -8.3131, p < .001). A closer inspection reveals that the average sentiment of the negative tweets is slightly lower for female users ($t(31.535) = 1.744$, $p = .08$), while the mean of the positive ones is significantly higher for females ($t(35.628) = -5.6146$, $p < .001$). This indicates that, while both genders write similar negative tweets, women write more positive ones when it comes to feminism.

It is difficult to identify the sources for the different sentiment values by female and male Twitter users. Several sources on the word and on the topic level are possible, resulting in an elusive mixture of interacting causes. Looking at the usage of the AFINN words by gender, a slight shift in the distribution can be observed. Female users have a higher proportion of positive AFINN words, while the distribution of male users is shifted to a marginal higher share of negative AFINN words. Examples are the words “love”, “happy”, “thank” or “support”, which were used more by female users, or, on the other hand, “anti”, “wrong” or “racism” used by male users.

A variety of topics are discussed, which are related to feminism, like toxic masculinity, identity politics, sex workers or the movie Moxie, to name but a few. Different evaluations between genders become apparent for a discussion about feminism and racism related to the pop star Taylor Swift and the behavior of her fans (called “Swifties”). The average sentiment value is significantly lower for female users compared to male users (-.51 vs. 3.3; $t(151.3) = -3.5935$, $p < .001$). The topic of Women’s History Month is also less positively assessed by female Twitter users (2.30 vs. 2.94; $t(1087.4) = -2.6438$, $p < .01$). This does not mean that women see for example the Women’s History Month in a less positive way, but that in some cases they use the issue to bring up controversial topics to promote a discussion. Further research is necessary to analyze this style of debate in a deeper way. Context analysis may be a more practical research method, compared to the empirical approach applied in this study.

Emojis

In 13.6% of all tweets at least one emoji is used. The average usage in these tweets is 1.97 emojis. The distribution shows a long tail, with a maximum of 67 emojis in one tweet. These are exceptions. When it comes to different emojis used, in average 1.42 different emojis are incorporated, with a maximum of 26 in one tweet. From a user perspective, around 17% use an emoji at least ones. This is below previous results (Ljubešić & Fišer, 2016). The variability is rather small, as in average 1.54 different emojis were adopted by a user.

Table 5 shows the 35 most used emojis in the dataset. Shown is the total number, the number by gender, and the rank of usage by gender. Rank 1 indicates that this is the most used emoji of female or male Twitter users. Please note that the numbers for female and male do not add up for total usage. The latter uses the entire dataset, while for female and male numbers only the part of the dataset was processed, where a gender allocation was possible.

No.	Emoji	N	N Female	N Male	Rank Female	Rank Male	No.	Emoji	N	N Female	N Male	Rank Female	Rank Male
1	😂	4.532	947	1,167	1	1	**19**	💪	413	118	72	20	23
2	😭	2.773	677	438	3	3	**20**	💕	399	123	67	19	24
3	🤣	2.593	547	644	4	2	**21**	😳	394	111	80	22	20
4	❤️	1.654	501	296	5	4	**22**	🤡	384	76	100	34	16
5	♀️	1.169	775	93	2	17	**23**	😄	367	88	72	28	23
6	👇	1.047	377	229	6	7	**24**	💖	342	102	73	23	22
7	🙄	1.027	293	246	8	5	**25**	✊	339	82	52	30	30
8	💜	989	332	165	7	10	**26**	👍	337	68	110	38	15
9	✨	928	284	148	9	12	**27**	😌	327	116	48	21	33
10	👏	708	180	198	13	8	**28**	🥺	302	92	67	25	24
11	😍	657	204	128	10	14	**29**	😊	288	86	58	29	29
12	🤔	632	153	170	15	9	**30**	😩	288	81	27	31	50
13	💀	620	131	147	17	13	**31**	💯	286	100	66	24	25
14	🕊️	619	187	231	12	6	**32**	😐	284	86	30	29	47
15	🥰	530	194	76	11	21	**33**	🤮	278	73	49	36	32
16	🙏	529	125	90	18	18	**34**	😅	273	55	64	48	26
17	🔥	524	168	87	14	19	**35**	🙌	263	78	49	32	32
18	☮️	436	137	159	16	11							

Table 5: 35 Most Often Used Emojis, Split by Gender

The most used emojis correspond with the overall distribution of emojis on Twitter as recorded by the Emojitracker[2]. There are a few exceptions, which are related to the topic of feminism, like e.g., ♀️ or ☮️.

From a gender perspective more female than male users insert at least one emoji into their tweets (18.3% vs. 14.5%). When emojis are used, there is no difference in average emojis per tweet ($t(8273.6) = .71514$, $p = .4745$), but a difference in unique emojis applied ($t(10.684) = 2.1195$, $p = .0341$). Female users show a higher variety of emojis.

When grouping the emojis by gender several clusters become visible. Women use the female sign ♀️ much more often than men. They also use positive faces more often (😍 🥰 😌 😊), but also negative ones (😩 😐 🤮). Additionally, different heart symbols (💕 ❤️ 💖 💜) are used regularly. This indicates a bigger range of emotional expressions. On the male side it is of interest, that laughing faces (🤣 😂 😅) along with the thinking face emoji 🤔 and the skull 💀 are used frequently, even in absolute numbers more than women. This is especially relevant, as the total number of emojis used by men is lower. The laughing faces indicate that some men deride the topic of feminism and gender equality. This study focuses on a quantitative methodology and cannot provide an in-depth content analysis. Thus, further research is necessary to further analyze this finding.

Following the analysis of the numerical usage of emojis, the focus is now on the sentiment value of the emojis itself and their relationship to each other. To calculate the sentiment value of an emoji, all tweets including the considered emoji are taken and again filtered by tweets

[2] https://www.emojitracker.com/

having a sentiment value based on the unigrams. The emoji sentiment value is then the mean of all isolated tweets sentiment values. The calculated value is then correlated with the emoji sentiment lexicon from Novak et al. (2015). A strong and significant correlation was found with $r(105) = .70$, $p < .001$. This indicates that the emojis used in the dataset are used in a comparable way to their usage in general tweets and it supports the evaluation method. Only emojis are considered, which appeared in at least 25 evaluated tweets, resulting in 107 emojis.

The analysis was repeated by gender. The correlation between female and male emoji sentiment values is moderately positive and significant, with $r(197) = .56$, $p < .001$. This shows that female and male Twitter users use the emojis in a slightly different context.

Next the correlation between emojis is analyzed. Two emojis are correlated if they appear regularly in the same tweet. The analysis was limited to emojis with at least 50 appearances. Table 6 shows the highest correlating emojis.

No.	Emoji 1	Emoji 2	Correlation
1	☮️	🕊️	0.97
2	⚧️	🏳️‍⚧️	0.94
3	🇺🇦	🎬	0.81
4	😢	💥	0.60
5	✊🏻	✊🏼	0.35 to 0.26
	✊🏽	✊🏾	
	✊🏿		
6	🏳️‍⚧️	🏳️‍🌈	0.29 to 0.27
	⚧️	🏳️‍🌈	
7	🧡	💛	0.27 to 0.26
	💜	💚	
8	🇮🇳	💯	0.27

Table 6: Highest Correlating Emojis

The two highest correlations are achieved by pairs of synonym symbols (peace symbol – peace dove; transgender symbol – transgender flag). The following two pairs are related to feminist topics. In both cases nearly identical tweets have been send by different users. The first is refering to Lesya Ukrainka, a former Ukrainian female writer and feminist activist and the development of an animation film based on one of her poems. The second pair is a generic tweet for the International Women's Day. The same applies to pair number eight. The fists with different skin tone indicate the diverse struggle for equality. The rainbow flag (or LGBT pride flag) is correlated to the transgender flag and symbol, referring to the proximity of both movements. Finally, the different colored hearts are related. Table 5 shows that red- or purple-colored hearts are used much more often, but these are used regularly in a stand-alone fashion. The use of the correlation of emojis can be helping to identify current topics.

Conclusions

This study focused on the language used by male and female Twitter users when discussing feminism. The analysis of the sentiment values of tweets and the incorporated emojis provide interesting findings. Women on average seem to be more supportive to other women (Bogen et al., 2019; Schneider & Carpenter, 2018), but on the same side also more critical to foster discussions about ongoing topics related to feminism. A part of male users seems to be trying to disdain feminism, while another part shows support. This support seems to be less constructive as it does less often try to assist the discussion. These findings are underlined by the exploration of the emojis. Furthermore, the sentiment of the tweet correlates to the sentiment of the emojis used in that tweet, supporting previous research (Novak et al., 2015). The correlation of emojis among each other showed several synonymous emojis but can also facilitate topic identification.

Speaking about the limitations of this study several aspects need to be mentioned. Firstly, the gender classification, as a crucial element of the study, is not entirely accurate (Vicente et al., 2018). Moreover, the classification into binary groups of male or female does not recognize fluid relationships of gender identity. The results need to be interpreted carefully in the context of social or cultural practices (Blevins & Mullen, 2015). In the dataset at hand, intersections to the LGBTIQ+ community can be found. Secondly, the dataset stems from Twitter. Other platforms, where similar discussions are taking place, are not covered. Furthermore, non-English tweets were excluded, limiting the cultures involved and thus the attitudes towards feminism. A third limitation, which is related to the previous one, is the demographics of Twitter users. Densely populated areas, male and Caucasian users are overrepresented (Mislove et al., 2011). These points confine the generalizability of the findings.

Future research can tackle these topics by focusing on other cultures and languages. In this analysis, greater differences in genders' communication were revealed when looking on the emojis than on the linguistic analysis. Here, a deeper insight into the assigned meaning of emojis for both genders can be of interest. Finally, the differentiation into more groups, like trolls, neutral or indifferent males, females or activists is of high relevance. This study was only able to identify glimpses of these groups.

References

Blevins, C., & Mullen, L. (2015). Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction. *DHQ: Digital Humanities Quarterly, 9*(3).

Bogen, K. W., Bleiweiss, K. K., Leach, N. R., & Orchowski, L. M. (2019). #MeToo: Disclosure and response to sexual victimization on Twitter. *Journal of Interpersonal Violence*, 1-32.

Chen, Z., Lu, X., Shen, S., Ai, W., Liu, X., & Mei, Q. (2017). Through a gender lens: an empirical study of emoji usage over large-scale android users. arXiv preprint arXiv:1705.05546.

Derks, D., Bos, A. E., & Von Grumbkow, J. (2007). Emoticons and social interaction on the Internet: the importance of social context. *Computers in human behavior, 23*(1), 842-849.

Dilai, M., & Levchenko, O. (2018). Discourses Surrounding Feminism in Ukraine: A Sentiment Analysis of Twitter Data. *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 2*, 47-50.

Hardaker, C., & McGlashan, M. (2016). ‘‘Real men don’t hate women’’: Twitter rape threats and group identity. *Journal of Pragmatics, 91*, 80-93.

Huang, A. H., Yen, D. C., & Zhang, X. (2008). Exploring the potential effects of emoticons. *Information & Management, 45*(7), 466-473.

Huffaker, D. A., & Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. *Journal of computer-mediated communication, 10*(2), JCMC10211.

Jianqiang, Z., & Xiaolin, G. (2017). Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access, 5*, 2870-2879.

Kearney, M. W. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software, 4*(42), 1829.

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the OMG!. *Fifth International AAAI conference on weblogs and social media (ICWSM-2011).*

Ljubešić, N., & Fišer, D. (2016). A global analysis of emoji usage. *Proceedings of the 10th Web as Corpus Workshop*, 82-89.

Lopez, K. J., Muldoon, M. L., & McKeown J. K. L. (2018). One Day of #Feminism: Twitter as a Complex Digital Arena for Wielding, Shielding, and Trolling talk on Feminism. *Leisure Sciences, 41*(3), 203-220.

Mahdawi, A. (2019, September 13). He, she, they … should we now clarify our preferred pronouns when we say hello? Pronoun introductions are seen as a positive step towards a better understanding of gender – but it's complicated. The Guardian. https://www.theguardian.com/lifeandstyle/2019/sep/13/pronouns-gender-he-she-they-natalie-wynn-contrapoints

Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. (2011). Understanding the demographics of Twitter users. *Proceedings of the International AAAI Conference on Web and Social Media, 5*(1), 554-557.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts' in CEUR Workshop Proceedings*, 93-98.

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PloS one, 10*(12).

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *LREc, 10*(2010), 1320-1326.

Rosen, A. (2017, November 7). Tweeting Made Easier. Twitter Company Blog. https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier

Sainz-de-Baranda, C., Adá-Lameiras, A., & Blanco-Ruiz, M. (2020). Gender Differences in Sports News Coverage on Twitter. *International Journal of Environmental Research and Public Health, 17*(14), 5199.

Scarborough, W. J. (2018). Feminist Twitter and Gender Attitudes: Opportunities and Limitations to Using Twitter in the Study of Public Opinion. *Socius, 4*, 1-16.

Schneider, K. T., & Carpenter, N. J. (2019). Sharing #MeToo on Twitter: incidents, coping responses, and social reactions. *Equality, Diversity and Inclusion: An International Journal, 39*(1), 87-100.

Social Security Administration (n.d.). Popular Baby Names - Beyond the Top 1000 Names.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics, 118*(1), 26-40.

Tinati, R., Halford, S., Carr, L., & Pope, C. (2014). Big Data: Methodological Challenges and Approaches for Sociological Analysis. *Sociology, 48*(4), 663-681.

Twitter Inc. (2021). Q4' 2020 Investor Fact Sheet [Fact Sheet]. https://s22.q4cdn.com/826641620/files/doc_financials/2020/q4/Q4_20__InvestorFactSheet.pdf

Twitter Open Source (n.d.). Emojis for everyone. https://twemoji.twitter.com

Vicente, M., Batista, F., & Carvalho, J. P. (2019). Gender detection of Twitter users based on multiple information sources. In L. T., Kóczy, J., Medina-Moreno, & E., Ramírez-Poussa (Eds.), *Interactions Between Computational Intelligence and Mathematics Part 2* (pp. 39-54). Springer.

Vidal, L., Ares, G., & Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference, 49*, 119-128.

Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review, 19*(3), 324-347.

Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? predictive features for hate speech detection on twitter. Proceedings of the NAACL student research workshop, 88-93.