



MYKOLAS ROMERIS
UNIVERSITY



MYKOLAS ROMERIS
UNIVERSITY

Proceedings of the 4th European Conference on Social Media

Mykolas Romeris University
Vilnius, Lithuania
3-4 July 2017



Edited by
Aelita Skaržauskienė and Nomeda Gudeliienė
Mykolas Romeris University (MRU)

A conference managed by ACPI, UK

acpi

**Proceedings of the
4th European Conference on Social Media
ECSM 2017**

**Hosted By
Mykolas Romeris University
Vilnius, Lithuania**

3rd-4th July 2017

**Edited by
Aelita Skaržauskienė and Nomedā Gudelienė**

Hate Speech on Facebook

Tuba Ciftci, Liridona Gashi, René Hoffmann, David Bahr, Aylin Ilhan and Kaja Fietkiewicz
Department of Information Science, Heinrich Heine University, Düsseldorf, Germany

tuba.ciftci@uni-duesseldorf.de

liridona.gashi@uni-duesseldorf.de

rene.hoffmann@uni-duesseldorf.de

david.bahr@uni-duesseldorf.de

aylin.ilhan@uni-duesseldorf.de

kaja.fietkiewicz@uni-duesseldorf.de

Abstract: Newspapers are increasingly posting news reports on social networking sites like Facebook. In this way, users can immediately read what happened and exchange their opinions in the comment section of the posts. In the wake of the refugee crisis in Europe and especially after the incidents on New Year's Eve 2015 in Cologne, the attitude towards refugees became more and more negative in Germany. Many users shared their opinions by commenting the news journals' posts on Facebook and many of these comments can be considered a hate speech, or since occurring on the web, cyber hate. The problem of cyber hate or hate speech spreading on Facebook is not new. Our paper considers the questions: How do higher educated German Facebook users classify cyber hate? How do they react to it? Which actions would they take to either support or oppose hateful comments? Are there gender dependent differences in users' attitude towards cyber hate? To answer these questions, we created an online questionnaire. It includes a news post reporting the incidents on New Year's Eve in Cologne with two corresponding comments to which the participants had to answer questions. The first comment was classified as hate speech. We formulated questions about these comments and what action would the users potentially take when seeing such posts online.

The outcomes of the study show that over 90% of the participants perceive the first comment as cyber hate; about 60% of all participants classified the second comment as cyber hate as well. Among the experienced emotions there are some significant differences between female and male respondents. The investigation revealed that the most occurring emotional responses when reading the comments are anger, disgust and frustration. The main reason for re-commenting the comment is because the comment contradicts the participant's opinion. Only a small share of the users would like the comments, whereas the second comment would get more likes than the first one. More than half of the participants would report the first comment, while the second comment would be reported by fewer users.

Keywords: Cyber Hate, Emotional Response, Facebook, Reaction on Hate Comments, Hate Speech

Cyber hate is a hot topic and is not enough researched. Especially the perception and reaction towards cyber hate among the higher educated Facebook users is completely unexplored, so we could add some new findings in this research area.

1. Introduction

The spread of hate comments in social media is a challenging issue that receives a lot of medial attention in Germany. This phenomenon has been known for a long time, however, now it is especially present in the context of the refugee crisis in Europe. Such hate comments are not allowed by the social network sites, such as Facebook or Twitter, and in some cases by law if they include defamation or call to hate crimes. Even though, a steady growth of them has been noted on Facebook. In addition, the number of criminal proceedings in Germany against people inciting hatred in social media is increasing, according to German Bundestag (2016). The social debate concentrates on the classification of hate speech and the accurate action against hate comments. The positions of the involved parties, such as politicians, social network operators and users are different and there is no optimal solution in sight. The German Federal Government considers the existing legislation as sufficient. According to the German Criminal Code (2013), it is prohibited to incite hatred against segments of the population or assault the human dignity of others by insulting, maliciously maligning or defaming segments of the population. Who disseminates a presentation of the content indicated above in the Internet can be sentenced to prison for up to three years or has to pay a fine.

For this study, we define the term of *cyber hate* as hate speech occurring on the web, since there exists no universal definition of cyber hate now. In this paper, we use the definition by Unger (2013, p. 259), also given to the participants of the questionnaire:

[Hate speech] should be understood utterances which degrade a person or a group based on a common characteristic. [...] [Hate speech] differs from an insult or defamation thereby that here is referred on characteristics of a particular group of people, such as white or black skin, Christian or Muslim faith or ethnicity.

This definition coincides with the one given by Facebook. For the social network hate speech is “content that attacks people based on their actual or perceived race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or disease” (Facebook, 2016).

In the course of the arrival of a large number of refugees, notably from Syria, Iraq, Northern Africa and Afghanistan, the predominant public mood in Germany towards the migrants changed from supportive to hostile within a few months. It came to tensions between Germans and refugees, both on- and offline. The refugee issue became a tremendous challenge in Germany and determines the public discourse. Hence, the sexual assaults of the New Year’s Eve 2015 in Cologne were selected as headline for the Facebook post. Other trigger events with large impact in Germany could have been the image of the drowned refugee boy, the closure of the Balkan route or Chancellor Merkel’s “We can do it” statement.

The intention of this study is not to answer the question what is considered hate or free speech. We do not want to define the circumstances under which hate comments on Facebook or other social media networks have to be deleted nor do we want to establish a criteria catalogue on how to act against such expressions of hate. Our contribution to the controversial issue of hate comments on social media is to analyze for the first time in an academic study how well higher educated German-speaking users of Facebook identify hate speech comments and how they react to it.

2. Related work

Burnap and Williams (2016) developed machine classification models to identify different types of cyber hate to be applicable to a range of categories such as race, disability and sexual orientation. These cyber hate classification models can help inform those responsible for managing content in social networks and allow them to verify and react, rather than have to search for offensive content in large data streams, in particular in the aftermath of a potential trigger event.

Erjavec and Poler Kovačič (2012) studied the characteristics of hate speech discourse in Slovenian news websites’ comments. The authors differentiated between organized and self-organized producers of hate speech comments. The organized producers were active members of political parties and other nongovernmental organizations and use hate speech according to the orders of superiors. The self-organized producers were divided in three groups. The first group faithfully followed their political and ideological role models, defended their political-ideological interests by attacking enemies. The second group wanted to have fun by humiliating others in the online community and had no clearly profiled values and beliefs, whereas the third group was motivated by social injustice and used hate speech to draw attention to social problems.

Awan (2014) analyzed tweets to examine how Muslims were viewed and targeted before and after a terrorist assassination of a British soldier by two Muslims in London, UK. He created a typology of online abusers with key characteristics. He found out that the most characteristic type of offender, the reactive type, will begin an online campaign following a major incident targeting that specific group. This type is followed by the accessory type, who is joining in with other people’s Twitter conversations to target vulnerable people.

Due to the lack of scientific work on this topic we also describe several recent surveys on perception of and reaction on cyber hate among German internet users. According to Bitkom (2015), among the readers who were not personally affected by cyber hate comments 77 % did not react at all, whereas 16% reported the comment to the operator. 7% answered the comment with a contrary posting. An Eco survey (2016) stated that about 45% ignore racist hate messages in social media, around 20% of users report such comments to the operator and 14 % would react with an own posting. LfM (2016) stated that private Internet users would ignore a hate comment on the web in 49% of the cases. 35% would report it to the operator or the police while 2% would answer the comment. 72% of the respondents state that hate comments make them angry, especially females. To deal with hate comments, is a waste of time according to 65% of the respondents. 58%

state that hate comments do not interest them whereas 34% are scared of such comments, in particular females. Among all these persons aged 60 or older agree slightly more than the rest.

The most relevant work to this study is that of Haque (2014), which examined cyber hate in Pakistan. Besides some more general aspects, part of his work deals with the perception of cyber hate. In order to determine how good respondents were able to identify online hate speech, a list of statements was presented, from which respondents had to check off all the statements they personally believed constituted hate speech. They found out that people over 25 years are slightly better at identifying hate speech compared to those aged 25 or younger and females identify hate speech less frequent than males. Our study goes beyond this approach in several ways as described in detail in the Methodology section.

3. Methodology

A research model (Figure 1) was created for this study that starts with an event which in the case of Facebook can be a public post by a user or an official site. The post can be just a statement or can relate to a real-life event. The study by Williams and Burnap (2016) further backs that cyber hate much like offline hate crime increases in the aftermath of so called trigger events. The model incorporates this by including the trigger event that leads to the Facebook post. News sites post an article about the event on Facebook and its comment section then becomes a place where users can express their opinions about the event or the topic it relates to. Some people use this opportunity to write hateful comments that are meant to hurt or damage the reputation of a specific group of people. The comment or, in the case of this study, the “hate comment” is therefore the following entity in this model. The Facebook user who reads the hate comment, since it is public, is the central point of this model. While the event and the comment are included for needed context the focus lies on the reaction of the user.

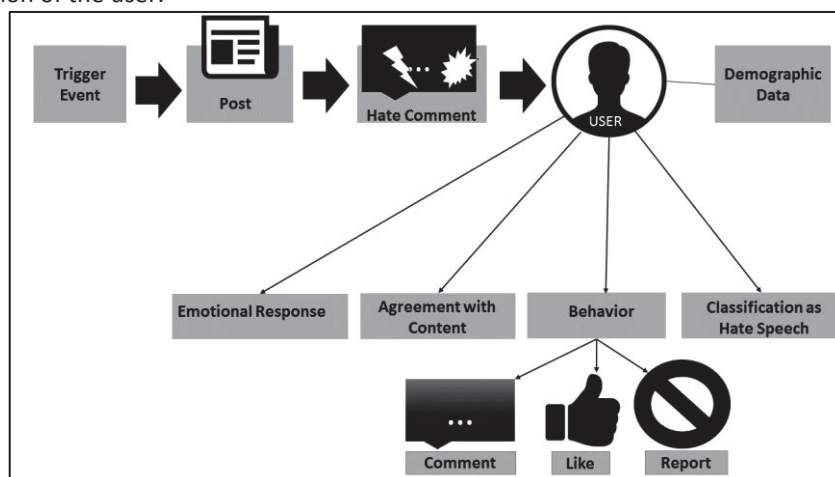


Figure 1: Research Model of the Reaction of Users to Hate Comments on Facebook

The demographic data like age, gender, immigration background or religion can be factors that influence the reaction and the decisions the user takes when confronted with cyber hate. As a first step, we focused on users with a higher education (*Abitur/Fachabitur, Hochschulabschluss*). Future research questions are defined in the limitations part of this study. The first important part of the reaction is the emotional response. Specific emotions may take part in the decision making of the user and can decide if the user likes, comments or reports the comment. Next, the agreement with the content of the comment is also included. There is a possibility that the user agrees with the opinion of the commenter but does not like how it was conveyed. The assessment whether the comment can be classified as cyber hate by the user is an important part because it sheds light on the motives behind the user’s actions. Finally, the behavior of the user shows the actions he will take to either ignore, support or combat the cyber hate. There are three main functions that are available for the user on Facebook: giving a like, writing a comment and reporting the comment.

An online survey in German was used to gather the necessary data. The questionnaire was shared in different German Facebook groups and completed by 771 participants. Many studies on cyber hate on social media use data mining (Burnap & Williams, 2016; Silva et al., 2016; Williams & Burnap, 2016). However, these studies investigate the cyber hate itself and not the reaction of the users. Although it would have been possible to count the likes and analyze the comments of hate comments, it would not have given much insight about the

motives of the users and the group of people that chose to ignore these comments. A post by the German news website Spiegel Online about the events on New Year's Eve in Cologne in 2015 was chosen because of the controversy of the topic of refugees in Germany and the present hate comments in the comment section. The article says that large groups of men that were supposed to be from North Africa or Arabia sexually assaulted and robbed a high number of women (Spiegel Online, 2016). Two comments found in the comment section of said post were picked as case examples to which the participants had to answer questions. The first comment was classified as hate speech:

"[...] it says in the text that the perpetrators come from North Africa and Arabia... If we take every filth in our country then no wonder it happens. Women are worth nothing in their cultures... We do not need anything from their cultures here. Out with these dirty people!!!" (semantic translation from German).

People from North Africa and Arabia are labeled as "filth" and "dirty people" which are roughly translated German ethnic slurs. Thus, the commenter attacks a specific group of people because of their nationality and ethnicity and also uses derogatory terms. The second comment on the other hand cannot be classified as clearly or even at all as hate speech because it does not contain any insults or obvious attacks against a group of people:

"It will get worse... because nothing is done against it, delinquent refugees will not even get deported and they know it, that is why they shit on our laws." (semantic translation from German).

Nevertheless, the commenter automatically assumes that the perpetrators were refugees even though it is not stated in the article that refugees had any part in it. Whether it is hate speech or not is debatable and it is interesting to see how the respondents classify the two different comments and if they react differently to them.

From this observation, the first research question (RQ) is formulated:

RQ1: Are there differences in how the participants classify these two comments?

It is also interesting to investigate whether these two comments create different emotional responses, since one is less offensive or controversial. In the questionnaire, the users were asked what emotions the comment triggers in them. Most answer options of the emotions were adopted from the study by Gashi and Kanutz (2016). Additionally, the emotional responses may also be different between genders. There are studies of differences for emotion-eliciting tasks such as hypothetical emotional vignettes, which can be compared to the method of this study (self-reports of emotion): "[...] women report experiencing emotions more frequently and more intensely than do men [...]" (Chentsova-Dutton & Tsai, 2007, p. 164). Studies on the differences between emotional responses of men and women to cyber hate have not been found.

From this observation, the second RQ is formulated as follows:

RQ2a: Do the two comments trigger different emotions?

RQ2b: Do men and women have different emotional responses to hate comments?

The questions for the behavior of the respondents are based on the questions of the study by Ziegele et al. (2013) which assesses the motives of commenters on online news sites. The four dimensions information interest, addition of missing views, correction and contradiction were incorporated in our questions for the reasons why the respondent would comment the hate comment. The statements are in the form of "I comment, because..." and the respondents have to score the statement with a Likert-Scale (from 1= completely agree to 5= completely disagree). The distances between the scores are equal. The statements for liking and reporting are in the same format but were adjusted to fit in with the characteristics of these actions. While commenting can be used for positive and negative feedback, liking can only express positive feedback, whereas reporting only negative one. The statements for the cases of not commenting, liking or reporting incorporate following dimensions: pointlessness, laziness, being afraid and being found out by relatives, friends or acquaintances.

From this observation, the third, fourth and fifth RQ are formulated:

RQ3: Do users comment on hate comments and what is their motivation?

RQ4: Do users like hate comments and what is their motivation?

RQ5: Do users report hate comments and what is their motivation?

4. Results

The online survey was available for a period of four weeks; it was launched on July, 1st and closed on July 31st, 2016 and generated a rich amount of useful data. The questionnaire was answered by Facebook users only. The survey was completed by 771 participants. Those who did not finish the survey were excluded from the analysis. The questionnaire comprises a series of basic demographic questions (e.g. age, gender, education), alongside some supplementary information (Facebook registration, religion affiliation, immigrant background). Of the participants, 64.3% (N=496) are women, 34.4% (N=265) are men, and 1.3 % (N=10) of participants did not specify their gender. Most of the participants (51.1%) are between 25 and 34 years old. Additionally, most of the participants (64.6%) use their real name on Facebook, whereas 6.1% have indicated using a fake name. Most participants (73.0%) do not have an immigration background. To determine religious affiliation, we asked the participants: “What religion do you belong to?” Most them (45.0%) have stated that they are Christians. 30.2% have indicated that they are Atheists, 9.7% are Muslims and 1.6% are Buddhists. Only a small proportion of the participants (0.1 %) have indicated that they are Jewish.

To assess the two-sided statistical significance between value series, we performed a chi-square test, based on Pearson’s. Our study distinguishes levels of statistical significance: the 95% level (*), the 99% level (**), and the 99.9% level (***). All other cases with significance under the 95% level are “not significant” and are labeled “ns”.

RQ1: Are there differences in how the participants classify these two comments?

The outcomes of the study show that about 92.5% of the participants perceive the first comment (C1) as cyber hate; 61.9% of all participants classified the second comment (C2) as cyber hate. This result backs our analysis that the second comment is not as clearly hate speech as the first one. In addition, we also asked participants if they agree with these statements. Most participants do not agree with these comments (C1= 5.3% vs. C2= 13.5%).

RQ2a: Do the two comments trigger different emotions?

RQ2b: Do men and women have different emotional responses to hate comments?

Table 1: Emotional Response of User Concerning both Comments

Emotions	All		Female		Male		Sig. difference between genders	
	C1	C2	C1	C2	C1	C2	C1	C2
Not applicable	0.6%	3.2%	0.4%	2.4%	0.8%	3.8%	*	*
Frustration	33.9%	42.7%	35.7%	44.8%	30.2%	38.9%	*	**
Indifference	4.5%	7.5%	3.4%	7.3%	6.0%	7.5%	*	Ns
Pity	9.1%	11.7%	6.0%	8.7%	14.7%	17.7%	*	***
Disappointment	30.9%	32.8%	29.4%	29.0%	33.6%	40.4%	**	***
Amusement	3.6%	5.1%	1.8%	4.16%	6.8%	9.4%	***	*
Shame	32.3%	19.2%	34.5%	20.4%	28.3%	17.0%	**	Ns
Surprise	1.7%	2.1%	1.6%	2.0%	1.9%	2.3%	ns	Ns
Disgust	58.8%	37.2%	59.5%	35.9%	57.7%	40.4%	ns	*
Fear	19.6%	11.2%	22.6%	12.1%	14.0%	9.4%	**	Ns
Anger	61.0%	44.9%	62.3%	44.2%	59.2%	46.8%	ns	Ns
Sadness	38.5%	32.7%	41.5%	33.7%	33.6%	31.3%	**	Ns
Happiness	0.4%	0.3%	0.0%	0.2%	0.8%	0.4%	ns	Ns

N = 771; Female: N = 496, Male: N =265; C1: Comment One; C2: Comment Two

Table 1 shows that the first comment triggers *Anger* in the users who read it (61.0%). The second most occurring emotional response is *Disgust*; 58.8% users feel disgusted by reading the first comment. About 33% feel frustrated when they read the first comment. The most occurring emotion for the second comment is also *Anger* (44.9%). About 42% feel frustrated because of the content of the second comment. *Disgust* is the third most occurring emotional reaction for the second comment, nearly 37% (37.2%) feel disgusted by reading this comment. *Happiness*, *Surprise* and *Amusement* are the emotions which occur the least among the participants. There seem to be some significant differences between the female and male respondents among the first and the second comment. The first comment has more significant differences between the genders. Female respondents seem to feel significantly more sadness when they read the first comment than male respondents (Female= 41.5%, Male= 33.6%). Women also seem to feel more fear than men by reading the first comment (Women=22.6%, Men= 14.0%). The emotions which are more positive also tend to have a significant difference between the genders, but in this case the men appear to be more emotionally affected. More men than women feel *Amusement* when they read the first comment (Women= 4.1%, Men= 9.4%). Some significant differences are to be found between the genders for the second comment, e.g. men feel more *Pity* than women (Women= 8.7%, Men= 17.7%). Men also feel more *Disappointment* than women (Women= 29.0%, Men= 40.4%).

There are statistically significant differences between female and male participants. Male respondents seem to be more amused by reading the first comment than the female respondents (Male= 6.8% vs Female= 1.8%). Significant differences are also among *Disappointment*, *Shame*, *Sadness* and *Fear*. Female respondents seem to feel significantly more negative emotions than male ones. Regarding the second comment, there are two extremely significant differences, *Pity* and *Disappointment*, which occur more often for men than for women. See Table 1.

RQ3: Do users comment on hate comments and what is their motivation?

To find out why people would respond to the comment, we asked the respondents if they would comment the respective comment (re-comment). Those who answer with “No” just skipped to the sub-question that asks for the reasons for not commenting.

Table 2: Reasons for Commenting a Post

	C1 Mean (SD)	C2 Mean (SD)	Sig.
Not my Opinion	1.51 (1.07)	1.73 (1.17)	ns
Make a Difference	1.79 (0.83)	1.92 (0.93)	ns
Reference on Misinformation	1.83 (1.09)	1.65 (0.96)	ns
Contradict Facebook Guidelines	2.54 (1.41)	3.22 (1.42)	*
Stimulate Discussions	2.68 (1.21)	2.42 (1.15)	*
Attack the Commentator	2.96 (1.41)	3.50 (1.41)	***
To Get Approval	3.98 (1.16)	4.16 (1.13)	*
I Like the Comment	4.85 (0.60)	4.75 (0.79)	ns

C1: N= 248; C2: N= 211; Scale: 1 (completely agree) to 5 (completely disagree)

Table 3: Reasons for not Commenting a Post

	C1 Mean (SD)	C2 Mean (SD)	Sig.
Pointless	1.73 (1.00)	1.77 (1.05)	ns
Do not Want	2.89 (1.41)	2.64 (1.51)	*
Fear of Virtual Attack	3.79 (1.33)	4.18 (1.25)	***
Do not Dare	4.00 (1.20)	4.36 (1.06)	**
Important People can see it (E.g. Boss, Friends)	4.05 (1.33)	4.26 (1.27)	**

C1: N= 528; C2: N=560; Scale: 1 (completely agree) to 5 (completely disagree)

Erjavec and Poler Kovačič (2012) found out that people comment on hate speech because they find it funny to humiliate others in the online community and because they feel social injustice and use Hate Speech to draw attention to social problems.

About 32.0% of users would comment the first comment. The second one would be re-commented (answer to the comment) by 27.4% of the respondents. Both comments would be re-commented because they contradict the user’s opinion (Mean: C1= 1.51 vs. C2= 1.73) (Table 2). The second most occurring reason for the first comment is “Make a Difference” (C1= 1.79 vs. C2= 1.92). They want to belong to the decision-makers and want to make things happen. In addition, they would re-comment the first one to point out some misinformation

(1.83). That misinformation can lead people to have false opinions. This is also the main reason why the second comment would be re-commented (1.65). The answer option “I Like the Comment” ranked last, with a mean of 4.85 (C2= 4.75). Finally, it is to be said that the respondents would not re-comment both comments for the same reason.

There seem to be some significant differences between comment one and comment two. An extremely significant reason is “To Attack the Commentator”. More users comment the first comment to attack the author of the comment. Some other significant differences are that the respondents just re-comment the first comment because they want to stimulate discussions and because the comment contradicts the Facebook guidelines.

As we can see in Table 3, there are more people who would not re-comment the comment. 68% would not re-comment the first comment and 72.6% the second one. People mostly do not comment because they think it makes no sense (C1=1.73 vs. C2=1.77). The second reason for not re-commenting is that people simply do not want to react to the comment (C1=2.89 vs. C2=2.64). The fear to be virtually attacked by other users seems to not be the reason that prevents users from re-commenting (C1= 3.79 vs. C2=4.18). “Not Dare” is the highest significant difference regarding the reasons for not commenting a post.

RQ4: Do users like hate comments and what is their motivation?

Table 4: Reasons for Like

	C1 Mean (SD)	C2 Mean (SD)	Sig.
Like the Comment	2.22 (1.30)	2.08 (1.02)	ns
Attract Attention	3.33 (1.00)	3.25 (1.26)	ns
To get Notifications	3.67 (1.41)	3.33 (1.44)	*
Other People Liked it	4.33 (1.00)	4.46 (0.83)	ns

C1: N= 9; C2: N=25; Scale: 1 (completely agree) to 5 (completely disagree)

Table 5: Reasons for not Like

	C1 Mean (SD)	C2 Mean (SD)	Sig.
I do not Want to Support the Comment	1.14 (0.57)	1.32 (0.85)	*
Pointless	3.45 (1.64)	3.22 (1.65)	*
Do not Want	4.15 (1.30)	3.72 (1.48)	***
Important People can see it (E.g. Boss, Friends)	4.51 (1.10)	4.48 (1.10)	ns
Fear of Virtual Attack	4.68 (0.80)	4.58 (0.92)	ns
Do not Dare	4.77 (0.65)	4.63 (0.81)	ns

C1: N= 762; C2: N=746; Scale: 1 (completely agree) to 5 (completely disagree)

There are only a few users who would give a “like” to the two comments. The first comment would only be “liked” by 1.2% of the respondents and the second one by 3.2%. Hence, the majority does not agree with the content of the comments. Most of those who stated that they would click the “like” button, would do this because they like the comment, whereby the second comment would get more likes than the first one (C1=2.22 vs. C2=2.08). People would not “like” the comment just because other people “liked” it as well (C1= 4.33 vs. C2= 4.46). Getting notifications or attracting attention are also not the reasons for “liking” the comment. There is a statistically significant difference between comment one and comment two: People like the second comment to get notifications.

According to Table 5, those who stated that they would not “like” the comment (C1=98.8% vs. C2=96.8%), would not do this because they do not want to support the comment (C1=1.14 vs. C2=1.32). Thinking it would be pointless, the fear that important people can see the like or the fear of virtual attacks are not the reasons preventing the respondents from “liking” the comment. As stated in Table 4 the highest significant difference is for the reason “Do not Want”. Another significant difference is “Pointless”. More respondents see no point in liking the second comment compared to the first one.

RQ5: Do users report hate comments and what is their motivation?

Table 6: Reasons for Reporting a Comment

	C1 Mean (SD)	C2 Mean (SD)	Sig.
It is Hate Speech	1.12 (0.37)	1.28 (0.49)	ns
It Contradicts Facebook Guidelines	1.74 (1.10)	1.88 (1.15)	ns
Harming the Author	3.09 (1.42)	3.12 (1.53)	ns
Feeling Attacked	3.28 (1.40)	3.33 (1.45)	ns
To Protect the Image of the Page	3.55 (1.25)	3.40 (1.37)	ns

C1: N = 441; C2: N = 201; Scale: 1 (completely agree) to 5 (completely disagree)

Table 7: Reasons for not Reporting a Comment

	C1 Mean (SD)	C2 Mean (SD)	Sig.
It is Pointless	2.09 (1.23)	2.22 (1.39)	ns
It Does not Contradict Facebook Guidelines	3.06 (1.37)	2.60 (1.39)	***
Do not Want	3.40 (1.45)	3.52 (1.47)	ns
Do not Want to be Deleted	4.12 (1.31)	4.15 (1.24)	ns
Not Knowing How to Report it	4.21 (1.29)	4.52 (1.08)	**

C1: N = 330; C2: N = 570; Scale: 1 (completely agree) to 5 (completely disagree)

What prompts Facebook users to report comments? There were five answer options to this question. Over half of respondents (57.2%) would report the first comment. The second one would be reported by 26.1% of respondents (see Table 6). The reasons for reporting both comments are nearly equal. The most important reason for reporting is “It is Hate Speech” (C1= 1.12 vs. C2= 1.28). In addition, our study shows that the Facebook Guidelines could be a reason why users report comments on Facebook. They would report both comments because they believe that these contradict the Facebook Guidelines (C1= 1.74 vs. C2= 1.88). The answer option “I Would Like to Harm the Author” is not a main reason for reporting a comment. Furthermore, we can clearly observe that the statement “To Protect the Image of the Page” is not a trigger for reporting these comments (C1= 3.55 vs. C2= 3.40). There are no significant differences between comment one and comment two (see Table 6).

Participants who stated that they would not report the two comments were asked about their motivation. Nearly three thirds of the respondents (73.9%) would not report the second comment. In contrast to the second comment, the first comment would not be reported by 42.8%. According to the users’ specifications, they do not report comment one and comment two because it is pointless (C1= 2.09 vs. C2= 2.22). It seems that many users would not report it because they believe that these comments do not contradict Facebook Guidelines. In this respect, the second comment (2.60) would be rather reported than the first comment (3.06). Another reason is that users do not feel like reporting comments (C1= 3.40 vs. 3.52). Users have mostly not agreed with the statements: “Not Wanting it to be Deleted”, “Not Knowing How to Report it” and “Not Daring to”. Our study shows that there were no big differences between comment one and comment two. “It Does not Contradict Facebook Guidelines” is the highest significant difference (see Table 7).

5. Conclusion

The purpose of this study was to investigate higher educated Facebook users’ reactions to cyber hate. We defined a research model which focuses on the reaction and behavior of users to cyber hate. We created an online questionnaire including a news post reporting the incidents on New Year’s Eve in Cologne with two corresponding comments, from which one could be classified as cyber hate.

To summarize, both comments trigger Anger in the users who read it. Comment one can clearly be classified as cyber hate. Therefore, it is not surprising that the relative frequency for the first comment is higher than for the second one. The second most occurring emotional response for the first comment is Disgust and for the second comment is Frustration. There are some significant differences between female and male respondents for the first and the second comment. Female respondents seem to feel significantly more sadness when they read the first comment than male respondents. The emotions which are more positive tend to also have a significant difference between the genders but in this case men show more tendencies to be emotionally affected. Regarding the second comment, there are also some significant differences among the genders, e.g. men feel more Pity than women. Men feel also more Disappointment than women. Users tend to reply a comment, if they consider the comment as hate speech. Our study shows that users would be recommend a comment if this comment do not represent their opinion. In this way, they want to make a difference with

their comment and want also reference on misinformation. People mostly do not comment because they think it makes no sense. The second reason for not recommending is that people simply do not want to react to the comment. Both of comments would be “liked” only by a few numbers of respondents. The main reason for liking is “I Like the Comment”. People would not “like” the comment just because other people “liked” it as well. More than the half of participants would report a comment if they think that they think the comment is Hate Speech. The reasons for not reporting are “Pointless” and “Not Contradicting Facebook Guidelines”.

This study has some limitations. Given the high number of respondents, we believe that our results are reliable. However, investigation of a controlled group (for example, left- versus right-wing), or an experiment, would enrich these findings. Another critical point could be that the survey was conducted in German-speaking countries. Therefore, the results only reflect cyber hate behavior of German-speaking people and we cannot draw a conclusion for other countries. Furthermore, this study considers only Facebook users with at least a higher education entrance qualification. For future work it would be interesting to study different levels of education and compare them to each other. We believe the field is wide open for future research, for example, what roles do a person’s age? Are there cultural differences in posting hate speeches? All of these questions will provide fertile soil in which to cultivate further study.

References

- Awan, I. (2014). Islamophobia and Twitter: A typology of online hate against muslims on social media. *Policy & Internet*, 6(2), 133–150.
- Bitkom. (2015). Hasskommentare: Jeder neunte Internetnutzer war selbst schon Opfer [Press release]. Retrieved from <https://www.bitkom.org/Presse/Presseinformation/Hasskommentare-Jeder-neunte-Internetnutzer-war-selbst-schon-Opfer.html>.
- Burnap, P., & Williams, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science*, 5(1), 1.
- Chentsova-Dutton, Y. E., & Tsai, J. L. (2007). Gender differences in emotional response among European Americans and among Americans. *Cognition and Emotion*, 21(1), 162–181.
- Eco. (2016). Eco-Umfrage: Jeder Dritte ist schon rassistischen Hassbotschaften im Internet begegnet [Press release]. Retrieved from https://www.eco.de/wp-content/blogs.dir/20160210_pm_-umfrage-hate-speech.pdf.
- Erjavec, K., & Poler Kovačič, M. (2012). "You don't understand, this is a new war!": Analysis of hate speech in news web sites' comments. *Mass Communication & Society*, 15(6), 899–920.
- Facebook. (2016). Retrieved from <https://www.facebook.com/help/135402139904490/>.
- Gashi, L., & Knautz, K. (2016). Unfriending and becoming unfriended on Facebook. In K. Knautz & K. S. Baran (Eds.), *Facets of Facebook* (pp. 1–44). Berlin, Boston: De Gruyter Saur.
- German Bundestag. (2016). Drucksache 18/7941.
- German Criminal Code. (2013). § 130.
- Haque, J. (2014). *Hate Speech: A study of Pakistan’s Cyberspace*. Islamabad, Pakistan: Bytes for All, Pakistan.
- LfM. (2016). Ethik im Netz. [PDF file]. Retrieved from http://www.lfm-nrw.de/fileadamin/user_upload/lfm-nrw/Service/Veranstaltungen_und_Preise/Medienversammlung/2016/EthikimNetz_Hate_Speech-PP.pdf.
- Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the targets of hate in online social media. In Tenth International AAAI Conference on Web and Social Media (pp. 687-690). Palo Alto, CA: AAAI Press.
- Spiegel Online (2016). Retrieved from <http://www.spiegel.de/panorama/justiz/koeln-60-frauen-erstatten-anzeige-nach-sexuellen-uebergreifen-an-silvester-a-1070418.html>.
- Unger, D. (2013). Kriterien zur Einschränkung von Hate Speech: Inhalt, Kosten oder Wertigkeit von Äußerungen? In J. Meibauer (Ed.), *Hassrede/Hate Speech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion* (pp. 257–285). Gießen: Gießener Elektronische Bibliothek.
- Williams, M. L., & Burnap, P. (2016). CyberHate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, 56(2), 211–238.
- Ziegele, M., Johnen, M., Bickler, A., Jakobs, I., Setzer, T., & Schnauber, A. (2013). Männlich, rüstig, kommentiert? Einflussfaktoren auf die Aktivität kommentierender Nutzer von Online-Nachrichtenseiten. *SC| M Studies in Communication| Media*, 2(1), 67114.