**Masterarbeit / Master's Thesis**

Titel der Masterarbeit / Title of the Master's Thesis

# "When They Break Up with You Online" – An Automated Content Analysis on People's Emotions When a Parasocial Relationship Ends


verfasst von / submitted by

**Kim Pamina Syed Ali**


angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

**Master of Science (MSc)**

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: UA 066 550

Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: Masterstudium Communication Science

Betreut von / Supervisor: Univ.-Prof. Hajo Boomgaarden, Ph.D.

**Abstract**

Due to the global pandemic, entirely new forms of media consumption have developed. By being forced to telecommute, people have to spend more time with their media options, and the fact that individuals are not allowed to spend their time outside with normal leisure activities. Therefore, statistics have shown that in 2020, television consumption increased (Broussard, 2020), which can be explained by the increase in social isolation and the need for people to entertain themselves with the resources they have. Therefore, people tend to form strong bonds with characters they interact with indirectly; they have a Parasocial Relationship with media characters. In particular, audiences of reality TV shows tend to have stronger Parasocial Relationships because these stars exist both in real life and in parallel online through social networks; thus, the relationship is more intense than that with a fictional character. The interesting question is how do individuals react when celebrities end their one-sided relationship by announcing the end of a show during a global pandemic with which they have built such a strong relationship? In this study, approximately 9000 Instagram comments were analyzed using an automated content analysis to determine the extent to which people react to such an event. In the end, it turned out that this kind of analysis is in no way applicable to the data collected, and thereupon this work is also based on the critical examination of computer-generated data.

***Keywords****: COVID-19, Social Distancing, Mood Management, Parasocial Interaction, Parasocial Relationship, Parasocial Break-Up, Audiovisual Media, Reality-TV, Online Commenting, Automated Content Analysis, Emotionality*

**Abstract**

Durch die globale Pandemie haben sich völlig neue Formen des Medienkonsums entwickelt. Nicht nur dadurch, dass die Menschen gezwungen sind von zuhause aus arbeiten zu müssen, müssen sie mehr Zeit mit ihren Medienoptionen verbringen, sondern auch durch die Tatsache, dass es den Individuen nicht erlaubt ist, ihre Zeit draußen mit normalen Freizeitaktivitäten zu verbringen. Daher haben Statistiken gezeigt, dass der Fernsehkonsum im Jahr 2020 zugenommen hat (Broussard, 2020), was durch die Zunahme der sozialen Isolation und dem Bedürfnis der Menschen erklärt werden kann, sich mit den Ressourcen, die sie haben, zu unterhalten. Aufgrund dessen neigen Menschen noch viel mehr dazu, starke Bindungen zu Figuren aufzubauen, mit denen sie indirekt interagieren; sie haben eine parasoziale Beziehung zu Medienfiguren. Insbesondere Zuschauer von Reality-TV Shows neigen dazu, stärkere parasoziale Beziehungen zu haben, weil diese Stars sowohl im realen Leben als auch parallel online über soziale Netzwerke existieren; die Beziehung ist also intensiver als die zu einer fiktiven Figur. Die interessante Frage ist, wie Individuen reagieren, wenn diese Mediencharaktere ihre einseitige Beziehung beenden, indem sie das Ende einer Show inmitten einer globalen Pandemie ankündigen, zu der sie eine so starke Beziehung aufgebaut haben? In dieser Studie wurden ca. 9000 Instagram-Kommentare automatisiert analysiert, um festzustellen, wie die Menschen auf ein solches Ereignis reagieren. Am Ende stellte sich heraus, dass diese Art der Analyse in keiner Weise auf die gesammelten Daten anwendbar ist und daraufhin fußt diese Arbeit auch auf der kritischen Auseinandersetzung mit computergenerierten Daten.

***Stichworte**: COVID-19, Soziale Distanzierung, Stimmungsmanagement, Parasoziale Interaktion, Parasoziale Beziehung, Parasoziale Trennung, Audiovisuelle Medien, Reality-TV, Online-Kommentarverhalten, Automatisierte Inhaltsanalyse, Emotionalität*

**Table of Contents**

## 1. Introduction

"I worked at a bar, and a girl came in & got REAL hammered. She started crying and trying to tell me about someone dying… it was someone from Grey's Anatomy. She had literally come to drink away the loss of her fictional friend. And I felt that. Like rlly felt it." (Whitneyyy, 2018). In 2018 this Tweet went viral and was retweeted over 40.000 times. The reason why this Tweet became so popular, the reason why that Tweet has nearly 300.000 likes, is the feeling of belonging to this topic and experiencing the same phenomenon (Dunlap & Lowenthal, 2009). Nowadays, nearly everyone has a favorite show, even the youngest or oldest individuals have an audiovisual form of art that they love to watch and maybe sympathize with one of the characters (Cohen, 1997; Cole & Leets, 1999). This social concept is nothing new and started not only because TV consumption is going to rise for the first time since 2011 (Broussard, 2020). Nevertheless, the way of bonding has changed, and the intensity of imaginary relationships has changed, too, especially in times where nearly everyone is forced to stay inside and occupy yourself with the resources you usually have in your home.

A crisis like the global pandemic starting in 2020 usually implements uncertainty and insecurities, which lead to higher emotional involvement of individuals (Hwang et al., 2020; Jarzyna, 2020). Since it is possible to consume audiovisual material not only through television, the consumption during the current COVID-19 crisis, increased (Broussard, 2020). Therefore, people tend to obsess over a show or film they recently ingested (Hwang et al., 2020; Jarzyna, 2020). Due to that, the relationships to characters in the medium are getting more robust, and human beings may develop a deeper relationship with the protagonist. This phenomenon is called *Parasocial Interaction* (Horton & Wohl, 1956).

## 2. Theoretical Framework

The following should clearly define why this work is highly relevant. Two basic media psychological approaches will be elaborated and placed in the appropriate theoretical and temporal context to highlight what this study is based on.

### 2.1 Media as Mood Management during Quarantine

As mentioned above, not only because of its immense general importance, Parasocial Relationships became even more essential due to current events because people need to stay at home for an extended period since March 2020 (Hwang et al., 2020; Jarzyna, 2020). Therefore, loneliness and social distancing correlate frequently (Hwang et al., 2020). However, intense relationships with individuals you adore are essential for well-being (Bowlby, 1969; 1973; 1980). In 1984, Caughey found that on average, when a child is 16, he or she will have spent more time watching TV than being at school. Applying this fact to the present time, one can assume that the time spent watching TV is even higher in relation to attending school since the daily TV consumption is frequently growing when one is older than 14 (Broussard, 2020). Thus, with the beginning of the global pandemic in 2020 because of COVID-19 - more commonly known as Coronavirus - it can be assumed that most people consume more television because of social isolation (Banerjee & Rai, 2020; Bryant & Zillmann, 1984; Bickham & Rich, 2006; Finn & Gorr, 1988).

One of the main mechanisms trying to explain media use in the context of psychological well-being is the *Mood Management Theory* (Zillmann & Bryant, 1985; Zillmann, 1988). Several studies demonstrated that entertaining media is a common way for individuals to impact their negative moods (e.g., Knobloch & Zillmann, 2002; Knobloch, 2003). According to Nabi and Krcmar (2004), positive emotions resulting from hedonic opportunities of media use have a briefly positive impact on individuals' stress levels. Therefore, Eden et al. (2020) assume that hedonically motivated media usage will rise during social isolation in the global pandemic of 2020. Due to the fact that adults often use television for bringing characters into their lives (Lim & Kim, 2017), Parasocial Interaction can be significant for those individuals, even more, when they are in an unstable psychological state because of quarantine (Bernhold & Metzger, 2018; Chory-Assad & Yanen, 2005). Therefore, a so-called *Parasocial Relationship* can fill the missing needs of security and affiliation (Cohen & Metzger, 1998; Rubin & Rubin, 1985). Hence, this study is of general interest, and a very urgent research gap is being studied. Thus, it is out of the question that this research is crucial.

### 2.2 Parasocial Relationships with Reality TV Stars

With the increasing importance of watching television or especially series and shows on different media services providers, where you are allowed to watch series endlessly, Parasocial Interaction's phenomenon become more present (Horton & Wohl, 1956; Rubin & McHugh, 1987). Horton and Wohl (1956) established the term of Parasocial Interaction first, where they called it an "illusion of a face-to-face relationship with the performer" (p. 1). In this case, the performer can be a media character, a celebrity, or a TV show character that can be fictional or non-fictional, but the relationship is one-sided and therefore an illusion. Moreover, interaction is higher when the character adjusts the performance to the audience's supposed response (Horton & Wohl, 1956).

Regarding this, a Parasocial Relationship describes a long-term relationship with a character by repeated Parasocial Interactions (Hartmann et al., 2008). Therefore, those two terms need to be distinguished. In this study, Parasocial Relationship will be examined due to the fact that the television show explored focuses on affective aspects with their audience. Nearly everyone is affected by having a Parasocial Relationship with someone. Everyone has a favorite TV show; everyone knows losing the most adorable character in a frequently watched show, or everyone has a celebrity they like the most. Applied at this time and given the increasing importance of audiovisual media, everyone has had a favorite show once that ended once. Thus, the research on that field is immensely relevant to society because almost everyone is affected, and with the increasing importance of working from home and therefore, spending more time watching TV, the concept is even more present. Since the establishment of the concept of Parasocial Relationship, several scholars researched the concept of a one-sided relationship with a preferable media character (e.g., Basil, 1996; Brown et al., 2003; Cohen, 2003; Cohen, 2004; Eyal & Cohen, 2006; Lather & Moyer-Guse, 2011; Papa et al., 2000). Furthermore, Eyal and Cohen (2006) argued that television characters could be seen as a significant part of one's social network, but Horton and Wohl (1956) pointed out that the intimacy that they offer is only at a distance and just one-way. It is important to emphasize that an imaginary friend's relationship feels similar to meeting a real friend but is not the same since the relationship is always and will always be one-sided (Paravati et al., 2019; Tian & Hoffner, 2010).

Based on that specialty of interaction, it seems legitimate to assume that a Parasocial Relationship is even more urgent as nearly every celebrity intense this one-sided relationship via social media. This is caused by the fact that a Parasocial Relationship maintains a relationship not just through a show on a screen but also via a social networking site in everyday life of the receptions since a Parasocial Relationship is more realistic to occur when media characters engage with the audience (Stefanone et al., 2010). Moreover, it was found that a feeling of realism was important for individuals to build a relationship with a media character (Rubin et al., 1985). Furthermore, Chung and Cho argued that reality TV stars seem more approachable what is partly due to the fact that the media characters maintain the relationship via online presence; they seem more "personal, friendly, and understandable" (p. 6) and therefore the Parasocial Relationship with media characters that engage in social media are more intense than others (2020).

Consequently, Cohen (2003) also established the concept of *Parasocial Breakups*, which means that a Parasocial Relationship goes off the air. There are several possibilities for why this phenomenon comes up. One option is the ending of a show; another is that a character is taken off the show, or maybe something happens to the show's actor in real life. Moreover, the recipient may be willing to stop the relationship by itself because they become less interested in the character (Ellithorpe & Brookes, 2018; Eyal & Cohen, 2006; Gregg, 2018). Cohen emphasizes that the breakup of a Parasocial Relationship "is a painful experience, which elicits symptoms similar to those that follow the loss of a friend." (Cohen, 2003, p. 200).

In addition, Cole and Leeds (1999) found that the intensity of emotion is based on the recipients' different kinds of character traits. An intense attachment to people you look up to is essential in determining security and healthy functioning (Bowlby, 1969; 1973; 1980). Several mental models of attachment have been explored, and Hazan and Shaver (1987) described a very basic one. First, there are people with secure attachment styles who tend to feel safe in relationships; those have very positive beliefs about relationships in general and find it very simple to get into one. Second, individuals with avoidant attachment styles have more pessimistic beliefs about relationships, are unwilling to trust others, and often do not allow intimacy. Lastly, human beings are very anxious, and they tend to worship and seek intimacy and romance. Also, they often distrust their partners and tend to be dependent. Therefore, people have specific expectations from others on how they want to be treated (Hazan & Shaver, 1987). Thus, it can be assumed that if individuals are not treated like they expect it from the relationship, the reaction to that behavior will be very emotional.

Consequently, Garimella et al. worked on that concept and found that the breakup brings different reactions to it, for example, the process of unfollowing the television character (2017). Besides, Barbary and Dion (2000) mentioned that the answer to a close relationship's breakup differs, but even though they differ, all of them may include basic emotions 1. anger,

2. sadness, 3. regret, 4. rejection, 5. anxiety, 6. jealously, 7. loneliness, or 8. resentment. Thus, it may be assumed that those reactions may take place in Parasocial Breakups, too.

Since every TV show comes to an end or every character will not be present in the recipients' lives at some point, the Parasocial Relationship will break up someday. The breakup of an imaginary relationship has not yet been intensively investigated; most studies examine the loss of a media character because of their death (Cohen, 2004). In this regard, it is crucial to research Parasocial Breakup's emotions that cannot be justified with the media character's death. Since the ending of a social relationship has an immense impact on a human being (Bakermans-Kranenburg & Van IJzendoorn, 1997; Barbara & Dion, 2000), it can be assumed that an imaginary relationship may also influence the life of someone. In 2003, Cohen found out that the end of a Parasocial Relationship is a painful experience that triggers feelings similar to losing a friend. Therefore, it would be interesting to research what kind of emotions the end of a reality TV show triggers since Cohen (2003) always advise to explore different TV genres when it comes to Parasocial Relationships with media characters. It may be that the intensity of a relationship with a non-fictional character in a reality TV show is higher (Chung & Cho, 2014).

Consequently, the breakup may be worse for a recipient than the end of an interaction with a fictive character. Moreover, it can be assumed that the relationship's intensity is higher since most of the characters in a reality TV show strengthen their relationship with the consumer similarly via an online platform and keep them updated all the time (Shin, 2016). Recipients can actively communicate their PSI directly to celebrities (Sanderson & Cheong, 2010). Particularly today, many social relationships end online, and consequently, it seems interesting to examine the breakup of a Parasocial Relationship via a social media platform. Many streaming platforms announce the end of a series online to have many public reactions to the post. That kind of data can help research emotions when media characters finish a

Parasocial Relationship online since most of the research on virtual Parasocial Relationships focuses on traditional media like TV and radio (Hwang & Zhang, 2018; Jarzyna, 2020).

As mentioned in the chapter before, several studies researched the end of a Parasocial Relationship because of an actor or actress's death. Meyrowitz (1994) researched the extreme reactions regarding the death of Elvis Presley and John Lennon. Bingaman (2020) did so, too, regarding the death of Kobe Bryant. Those studies do show the potential of Parasocial Interactions. Moreover, previous research also found that the loss of an imaginary relationship to a media character follows a similar scheme like the end of social relationships (Cohen, 2003; 2004; Eyal & Cohen, 2006; Widiastuti, Mawarpury, Sulistyani, & Khairani, 2019). Besides, several studies have listed that although women mostly have more intense social relationships, they can handle the end of these relationships better (Cohen, 2006; Helgeson, 1994; Simpson, 1987; Sprecher et al., 1998). Furthermore, teenagers, that are more emotional in general tend to have a more emotional reaction to a Parasocial Breakup (Boon & Lomore, 2001; Cohen, 2003; 2006).

Since Kim Kardashian West stands out among many reality TV stars, she is an excellent example of celebrities who maintain a Parasocial Relationship extensively (Lueck, 2015). Therefore, after establishing her TV show *Keeping Up With The Kardashian* as a very successful TV format, she built her sole presence as a star on television using social media to build a valuable business and very intense Parasocial Relationships with her fans (Goudreau, 2010). Moreover, since Kardashian knows that the show is watched mainly by women from 18 to 34 (Ng, 2011), she actively seeks to strengthen the Parasocial Relationship with this same audience by promoting and releasing products that appeal to this type of fan (Goudreau, 2010). However, all of those strategies to deepen the connections with her fans and strength the Parasocial Relationships are based on the parallel relationship with her fans via social networking sites like Instagram.

This study will analyze the essential post of Kardashian in which she announces that her reality TV show will end after 20 seasons (see Kardashian West, 2020). At Instagram, because if you look at the social media channels of Kardashian, it is noticeable that she has by far the most followers on Instagram (232 million) and therefore, in contrast to Twitter (69.7 million) or Facebook (33 million), she reaches the most people there. Moreover, Bond (2016) also found that Instagram is the most popular medium for following a media personality, which is another argument for focusing on Instagram and neglecting the other two social media sites where she has announced the ending. Instagram provides an excellent database here, with far more than 200,000 comments under the post.

All in all, there is a research lack of Parasocial Breakups via social media and the phenomena that may arise with this imprinting experience. Furthermore, because of the current pandemic, the topic is more relevant than ever since the consumption of audiovisual material is increasing permanently because of the social isolation that is a worldwide status quo in 2020 and 2021 (Mikos, 2020). Like Netflix, streaming providers gained many users with the beginning of the pandemic; the worldwide subscriber increased from 25 million to 192.95 million in 2020 (Netflix, 2020).

All in all, this study will examine the reactions to having a hard breakup via Instagram, and for that reason, the post of Kim Kardashian West, where she announced the end of her family's reality TV show. Because of the research lack listed above, this research topic is significant and important. Parasocial Relationships are a social phenomenon and a helpful concept in the economy or even education. With the exploration of that concept and new findings that will arise with this study, Parasocial Relationship could be helpful in many ways. Because of its immense presence in daily life, Parasocial Relationships' concept has to be explored more intensely and adapted to the latest circumstances.

## 3. Methodological Approach

Based on the theoretical considerations mentioned above, the following research questions have been formulated:

1. *To What Extent Do Fans React When Celebrities End their Parasocial Relationship via an Instagram Post?*
2. *Who are the People that Express their Feelings online Visible for Everyone?*

Because of the argumentation sketched before, the following hypotheses has been derived. First of all, it is estimated that individuals that react sadly to the breakup also express their anger in an online comment (H1) (Bendelow, 1993). Moreover, comments that are positively denoted are given to express the love to the show's protagonists (H2) (Waterloo et al., 2017). Furthermore, women tend to express their emotionality with the help of emojis (H3) (Wolf, 2000). Nevertheless, the surprise score will be high in most of the comments (H4) (Bowlby, 1969; 1973; 1980). Furthermore, the most emotional comments will be given by real profiles (H5) (Waterloo et al., 2017). A mixed-method approach has been created to examine the hypothesis given, which will be explained in the next chapter.

### 3.1 Procedure

To answer the research questions, we applied a mixed-methods approach. First, a computer-aided quantitative content analysis of nearly 9000 comments on Instagram was performed. Subsequently a qualitative content analysis of several comments that came out as the most emotional ones should have been carried out.

Therefore, this study is a following a mixed-methods approach. Moreover, the necessary comments have been derived automated and examined with automated content analysis in R (see Appendix). Because over 200.000 comments have been published under the selected post until now, the sampling was limited to a random sampling of 8995 comments, since this number is statistically sufficient in relation to the approximately 200.000 comments under the post. Moreover, only comments in English will be included, whereby the number of relevant data reduces itself. Comments from the post of September 08 will be used because Kim Kardashian West announced the end of her reality TV show on this specific date via Instagram. Consequently, this post was chosen because of its immense online spread.

In order to test the validity of the automated analysis, the results of the analysis have been compared to the results of a subsample analyzed manually. Moreover, several pretests were made before completing the coding and running of the automated content analysis. The author conducted that study independently, and through automated content analysis with the help of R, reliability should be sufficiently high.

**3.2 Sample**

Data was scraped with a complex Python script that has run daily and stored in a local SQLite database (see Appendix). The following characteristics have been sampled: The type of text (comment or reply), the timestamp, the user's name, the counts of like, the counts of reply, and the text. As it appears in the paper before, for the scores of emotionality, only the text of the comment is vital; the other variables sampled are essential for the qualitative examination. The final sample is n = 8995. Even though the analysis was to be fully automated, it was not possible to scrape more than the n mentioned due to technical constraints. The computer was not able to collect all the comments under the post, because it was not technically capable of doing so.

Moreover, it has to be mentioned that emojis are a common way to emphasize reactions, and thus the inclusion of those emojis in the analysis is mandatory (Cappallo et al., 2019; Barbieri et al., 2017; Na'aman et al., 2017; Raji & de Melo, 2020). There is one unit of analysis; the comments under the posts mentioned above. The measures are based on the literature review before.

However, emotionality in the comments has been analyzed using the *NRC Word-Emotion Association Lexicon* (Mohammad & Turney, 2012), and is therefore a compatible measure to measure the eight basic emotions anger, anticipation, disgust, fear, joy, sadness, surprise, and trust and two sentiments (negative and positive) in comments. This lexicon is based on the assumption that certain emotions are expressed by certain words (Mohammad & Turney, 2012), a so-called emotion lexicon, i.e., a list of emotions and words that can be indicated to certain emotions in texts. For example, words like *yummy* or *delightful* can be assigned to the emotion *joy* or *cry* to the emotion *sadness*. The lexicon by Mohammad and Turney (2012), also called *EmoLex*, is by far the most detailed lexicon of its kind. This program is of course subject to various restrictions. For example, it is difficult, if not impossible, for the program to understand the context of each word, unlike humans can. A word like *shout* evokes different emotions. Thus, for a human coder, it would be no problem to put this word in the context of the complete text, but for the program, analyzing and including this word can cause outliers in the analysis (Mohammad & Turney, 2012). However, even though this program has restrictions, various studies have already shown that EmoLex is quite useful in text analysis (e.g., Gupta et al., 2020; Mathur et al., 2020). Therefore, this program scores different words within a text block based on a very detailed dictionary and then the program calculates an overall score of each emotion by relating these words to the length of the text block (Mohammad & Turney, 2012).

Consequently, the dataset was subjected to some pre-processing steps. Also, the text corpus was tokenized, lemmatized, and irrelevant words and signs were removed using *udpipe* and *corpustools* (Straka & Straková 2017; Welbers & van Atteveldt 2020; Wijffels, 2020). Moreover, emojis have been replaced with words with the help of *textclean* (Rinker, 2018). For calculating emotions, the NRC Lexicon has been used to create scores that present the emotionality separated in the eight basic emotions mentioned before (Mohammad & Turney, 2012).

In order to answer all of the hypotheses, it is necessary to mix methods approaches and sample qualitative data, as well. H3 and H5 are only possible to be tested when the online comments are examined qualitatively, too. Therefore, the top 20 comments of all expressions of feelings mentioned above have been sampled manually. Since, with the scraping process, the user's name has been sampled, it is possible to have an insight into the profile of the commenters. In order to check the validity of this automated and computer-generated data analysis, it is crucial to run a few tests on a subsample to see if the resulting data matches what the computer-generated.

## 4. Results

### 4.1 Quantitative Automated Meta Data

First of all, it has to be mentioned that after applying the packages on the dataset, there were several measurement errors. There are 244 invalid values above 1, but this should be not possible at all. Due to the fact that those values can distort the overall outcome, it was inevitable to delete those comment IDs from our sample. Due to the fact that there was a vast dataset, the absence of the 244 comments did not pose a problem for further analysis. However, even after excluding the invalid values, it became apparent that the basic statements of the analysis did not differ significantly. Nevertheless, it was essential to continue with an adjusted data set to achieve the best possible outcomes for accuracy.

After completing the analysis and the modified dataset and applying all the packages mentioned above, there are several outcomes. First, Figure 1 shows the distribution of the top 20 emotional comments as an example of all emotional comments. The selection of the top comments is based on the descending sort of all comments related to each emotion. As shown in Figure 1, all eight emotions now have 20 top comments, which are sorted in such a way that it is visible which value the respective comment assumes. As already mentioned, a comment can have a maximum value of 1 for each emotion. Thus, 0 is not emotional at all and 1 is completely emotional.

| ID | Trust | ID | Joy | ID | Anger | ID | Fear | ID | Disgust | ID | Anticipation | ID | Positive | ID | Negative | ID | Surprise | ID | Sadness |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1086 | 0.5 | 1194 | 0.6 | 7173 | 0.8 | 8837 | 0.8 | 268 | 0.5 | 79 | 0.666666666666667 | 5710 | 0.8 | 7173 | 0.8 | 354 | 0.5 | 2259 | 0.666666666666667 |
| 1114 | 0.5 | 353 | 0.5 | 59 | 0.5 | 3440 | 0.75 | 558 | 0.5 | 1873 | 0.666666666666667 | 8450 | 0.8 | 2259 | 0.666666666666667 | 367 | 0.5 | 2314 | 0.666666666666667 |
| 1271 | 0.5 | 354 | 0.5 | 106 | 0.5 | 3441 | 0.75 | 1086 | 0.5 | 353 | 0.5 | 5230 | 0.75 | 2314 | 0.666666666666667 | 400 | 0.5 | 6995 | 0.666666666666667 |
| 1289 | 0.5 | 367 | 0.5 | 144 | 0.5 | 5371 | 0.75 | 1271 | 0.5 | 354 | 0.5 | 1194 | 0.6 | 6995 | 0.666666666666667 | 812 | 0.5 | 4 | 0.5 |
| 1341 | 0.5 | 400 | 0.5 | 230 | 0.5 | 3541 | 0.714285714285714 | 1341 | 0.5 | 367 | 0.5 | 353 | 0.5 | 4 | 0.5 | 1086 | 0.5 | 6 | 0.5 |
| 1478 | 0.5 | 570 | 0.5 | 254 | 0.5 | 317 | 0.666666666666667 | 1488 | 0.5 | 400 | 0.5 | 354 | 0.5 | 6 | 0.5 | 1271 | 0.5 | 7 | 0.5 |
| 1630 | 0.5 | 812 | 0.5 | 268 | 0.5 | 1764 | 0.666666666666667 | 1537 | 0.5 | 515 | 0.5 | 367 | 0.5 | 7 | 0.5 | 1289 | 0.5 | 8 | 0.5 |
| 1789 | 0.5 | 819 | 0.5 | 271 | 0.5 | 7919 | 0.666666666666667 | 2607 | 0.5 | 570 | 0.5 | 400 | 0.5 | 8 | 0.5 | 1341 | 0.5 | 12 | 0.5 |
| 1793 | 0.5 | 874 | 0.5 | 286 | 0.5 | 8634 | 0.6 | 2701 | 0.5 | 812 | 0.5 | 570 | 0.5 | 12 | 0.5 | 1630 | 0.5 | 19 | 0.5 |
| 1900 | 0.5 | 891 | 0.5 | 291 | 0.5 | 6255 | 0.571428571428571 | 2747 | 0.5 | 852 | 0.5 | 812 | 0.5 | 19 | 0.5 | 1925 | 0.5 | 59 | 0.5 |
| 1908 | 0.5 | 892 | 0.5 | 323 | 0.5 | 5794 | 0.526315789473684 | 2824 | 0.5 | 874 | 0.5 | 819 | 0.5 | 59 | 0.5 | 2101 | 0.5 | 65 | 0.5 |
| 1925 | 0.5 | 893 | 0.5 | 371 | 0.5 | 59 | 0.5 | 2827 | 0.5 | 891 | 0.5 | 874 | 0.5 | 65 | 0.5 | 2244 | 0.5 | 106 | 0.5 |
| 1980 | 0.5 | 894 | 0.5 | 451 | 0.5 | 106 | 0.5 | 2870 | 0.5 | 892 | 0.5 | 891 | 0.5 | 106 | 0.5 | 2508 | 0.5 | 111 | 0.5 |
| 2244 | 0.5 | 895 | 0.5 | 558 | 0.5 | 144 | 0.5 | 2958 | 0.5 | 893 | 0.5 | 892 | 0.5 | 111 | 0.5 | 2747 | 0.5 | 115 | 0.5 |
| 2438 | 0.5 | 896 | 0.5 | 647 | 0.5 | 230 | 0.5 | 3040 | 0.5 | 894 | 0.5 | 893 | 0.5 | 115 | 0.5 | 2870 | 0.5 | 142 | 0.5 |
| 2508 | 0.5 | 897 | 0.5 | 672 | 0.5 | 254 | 0.5 | 3185 | 0.5 | 895 | 0.5 | 894 | 0.5 | 142 | 0.5 | 2948 | 0.5 | 144 | 0.5 |
| 2747 | 0.5 | 898 | 0.5 | 673 | 0.5 | 271 | 0.5 | 3264 | 0.5 | 896 | 0.5 | 895 | 0.5 | 144 | 0.5 | 2958 | 0.5 | 172 | 0.5 |
| 2798 | 0.5 | 938 | 0.5 | 712 | 0.5 | 286 | 0.5 | 3470 | 0.5 | 897 | 0.5 | 896 | 0.5 | 172 | 0.5 | 3040 | 0.5 | 174 | 0.5 |
| 2870 | 0.5 | 1069 | 0.5 | 744 | 0.5 | 291 | 0.5 | 3557 | 0.5 | 898 | 0.5 | 897 | 0.5 | 174 | 0.5 | 3188 | 0.5 | 192 | 0.5 |
| 2958 | 0.5 | 1086 | 0.5 | 758 | 0.5 | 323 | 0.5 | 3658 | 0.5 | 1086 | 0.5 | 938 | 0.5 | 192 | 0.5 | 3470 | 0.5 | 212 | 0.5 |

Figure 1: The Top 20 Comments of Every Emotion and Sentiment

This table shows that there is a very balanced distribution of all the emotions. For example, looking at the highest value of sadness, a value of 0.6666666666667 is shown for comment number 2259. The program of Mohammad and Turney (2013) thus identified comment 2259 as the saddest. Nevertheless, it remains to be mentioned that this comment may also tend to contain other of the seven emotions, which would become apparent in a more detailed qualitative analysis. However, getting deeper into the numbers, there are several differences.

Furthermore, it is striking that some of the IDs occur more than once. For example, the Comment ID *1086* is the most emotional comment based on the trust score and the fifth-highest regarding the surprise score; this correlation may be interesting related to H1, where it is necessary to determine if individuals react sadly also react angrily to this announcement.

However, having a look at this table, there are some tendencies. For example, it is noticeable that the comment IDs are often in order, meaning that several consecutive comments are among the top 20 most emotional comments. This can be observed in Figure 1 with the anticipation score, where it can be seen that several IDs are in identical order. This could be an indicator that, on the one hand, individuals who commented on the picture added a comment in the same emotion. On the other hand, individuals who commented on the photo in a particularly anticipatory way engaged others in the same emotion to comment on this photo. Concerning anticipation in particular, it is likely that people who are already anticipatory motivate others who also tend to react to the photo in a particularly anticipatory way.

Nevertheless, this arrangement can also follow logical sorting. Because if several comments have the same value, the program sorts the comments based on the ID. Accordingly, it would make sense that this arrangement has come about because the second characteristic has come into play. In order to be able to go into this in more detail, a more detailed examination of individual comments is necessary.

Also, due to the massive number of comments that have been analyzed, the distribution of the scores is very balanced what should mean that there are no significant outliers when looking at Figure 1. However, even though anger, fear, and positive have the highest scores with 0.8, when looking at Figure 2 and therefore the means of all the scores, all comments are on average rather negatively loaded, which goes hand in hand with the theories of this work. (Negative Score Mean = 0.119228922***).

| Score | Trust | Joy | Anger | Fear | Disgust | Anticipation | Positive | Negative | Surprise | Sadness | Overall |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 0,020439187 | 0,02976579 | 0,036049872 | 0,043612456* | 0,012842566 | 0,025148497 | 0,033838203 | 0,119228922*** | 0,014047074 | 0,115033875** | 0,045000644 |

*** = Highest Score, ** = Second Highest Score, * = Third Highest Score

Figure 2: Means of different Emotions

Moreover, the following second most emotional score is the sadness score since many individuals seem to be very sad about the announcement in the Instagram post (Sadness Score Mean = 0,115033875**). Besides, the third-highest score of emotions is represented by fear (Fear Score Mean = 0,043612456*). Therefore, given these results, it is not possible to determine whether our H1 is thus accepted or rejected, even if the anger follows in the fourth most emotional position. For a more detailed investigation of this hypothesis, it is therefore vital to consider whether individuals frequently express sadness and anger at the same time. This can only be done by manually examining the comments.

In order to answer H2, it is also essential to look at the data qualitatively. It can also be seen here that the quantitative investigation was an excellent way to work out scores and thus getting a good overview of the data. However, it is crucial to deal with the data inductively to answer H1 and H2 and H3, H4, and H5.

Also, to get a better overview of how emotional several comments overall are, it was necessary to merge all scores into one; the Emotions Score. This Emotion Score was created by averaging all emotions. Therefore, on average, one comment has an emotions score of 0,045000644. Nevertheless, based on this value alone, it is impossible to answer which emotion the separate comment can be assigned to; it can only be assumed that contradictory emotions can be ruled out if one takes a closer look at the data. For example, it can be observed that when a comment is posted that has a relatively high negative value of 0.5, it shows a low positive value near 0, as one can logically conclude. However, to be able to conclude how emotional a comment is, no matter in which direction this value goes, it is a good starting point. Nevertheless, it has to be mentioned that this overall score is not significantly high in relation to the individual scores of the emotions separated by Mohammad and Turney (2012). All in all, this quantitative analysis seems a good starting point to venture a more detailed qualitative analysis. Based on these 20 comments, a more detailed qualitative analysis can be carried out, which in one go will directly serve as a validity test.

### 4.2 Trial and Error

In order to answer some of the hypotheses, it was essential to look at the profiles of the top comments in relation to the comments as such. Therefore, the top three comments of every emotion have been selected and subjected to a qualitative analysis. An overall emotions score has been developed but having a further look at the outcome, the consequences of an overall mean have not been meaningful. Thus, it was necessary to focus on several emotions independently.

Accordingly, having a detailed look at the top comments, it turned out that the first validity check, which should have gone hand in hand with the qualitative analysis, did not match the results of the top comments. According to the NRC lexicon, many of the comments are among the top comments in the respective emotion do not make much. For example, the comment “Finally” was classified by the computer as a top comment in trust even though for an individual, it has no connection with *trust* as a stand-alone word. This phenomenon can also be seen in many other of the top comments. Another example that illustrates the problem at hand is the comment “Thank God", which is also classified as a top trust comment due to the fact that the lexicon may denote the word *god* as highly trustworthy. Therefore, it remains to be mentioned that this comment cannot be assigned to the emotion trust in the context of the post because it is essential to mention here that the computer does not seem to understand any irony at all. This shows very well that a computer has problems recognizing stylistic devices.

However, it was therefore essential to find other approaches that could filter out these measurement errors. One approach was to clean the data again. One problem that could explain these errors is the incorrect or general interpretation of emojis. Therefore, it was essential to remove them to find out whether such measurement errors result from the fact that these emojis were misinterpreted by a translation using *textclean*. Thus, one approach was the complete elimination of these emojis. However, there was no significant change. Maybe the comments under this post are too short. The NRC Lexicon measures the denoted words from the dictionary in relation to the length of the comments. Thus, if a text consists of only one or two words and a word like *god* appears, this comment is classified as very trustworthy based on this very heavily weighted word. Since the length of the comments could also play a role in the automated analysis, it was necessary to eliminate the shorter comments in order to be able to exclude exclude possibly the considerable misinterpretation of the machine.

Nevertheless, even after structuring the comments according to length, no significant difference was found since even the most extended comments often contained no more than ten words. Of course, it must be mentioned that the machine evaluation made much more sense with longer comments, as can be glanced from Figure 3, than it did with the short comments, which were far too frequent. Nevertheless, automated evaluation makes little sense in this case. As this example shows once again the computer does not recognize stylistic devices such as irony in any case. This comment has a disgust score of 0.05, a positive score of 0.05, a negative score of 0.1 and a sadness score of 0.1 (Figure 3). Again, this shows that the machine's interpretation makes little sense even with longer or complete sentences.

Despite all that, it should be mentioned that long comments show that the program seems to make sense. However, the comment 285: "When people spend so unnecessarily and excessively, I always wonder how much they need. Give that money to people who need it, not just out of charity, but out of solidarity - strategically reduce your burdensome, consumerist, materialistic and unsustainable lifestyle for the benefit of others 😍" shows a very logical and much-distributed weighting of the various emotion scores (Figure 4). Therefore, promising approaches arise, which in principle, adapted to the proper means of investigation, would make more sense.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **22** | comment | 0 | 2021-01-23T11:39:56 | teenandh | 8 | 0 | No, Kim, make no mistake, your prostitution movie made you famous, not this program@kimkardashian |

Figure 3: Example for a longer Comment

| | trust_score | joy_score | anger_score | fear_score | disgust_score | anticipation_score | positive_score | negative_score | surprise_score | sadness_score |
|---|---|---|---|---|---|---|---|---|---|---|
| 285 | 0.0483870967741935 | 0.0645161290322581 | 0.032258064516129 | 0 | 0 | 0.032258064516129 | 0.0645161290322581 | 0.0161290322580645 | 0.032258064516129 | 0 |

Figure 4: Scores of Comment 285

However, related to the first hypothesis; indeed, it can be seen that when the output is contrasted, the sadness score and anger score go together very often. Therefore, comment *3909* with a sadness score of 0.5 and an anger score of 0.5 (Figure 5) proves on the one hand that sadness and anger go together, and on the other hand that the computer again has problems interpreting such short sentences correctly. However, it remains to be mentioned that it is again undoubtedly tricky for a computer to classify this comment correctly. Figure 3 illustrates how the NRC Lexicon works and that it is not optimal for analysis in this case, probably since the comments on online platforms like Instagram may be – as mentioned before - too short for the sole interpretation of a computer. All three comments seem sad and not necessarily angry to the researcher. Therefore, maybe it can be interpreted here that someone sad is also angry at the same time and that a word like *bad* is denoted in such a way that it is automatically evaluated as additionally angry in a computer-assisted evaluation. However, it cannot be deduced from the word alone that the user also wants to express anger at the same time.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **3909** | comment | 0 | 2020-09-10T19:56:29 | nceshidoo | 0 | 0 | That's bad |
| **533** | comment | 0 | 2020-10-08T00:25:27 | leathy_lou_ | 0 | 0 | .💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 💔💔 |
| **7717** | comment | 0 | 2020-09-10T00:08:39 | blakeleymcmurrey | 0 | 0 | WHYYYYYYYYYY 💔💔 💔 |

Figure 5: Sadness and Anger Examples

This discussion of the first hypothesis demonstrates extremely well that while the machine followed precisely what it was taught, this type of interpretation is not readily applicable to Instagram comments. Therefore, it is essential not to neglect the context of the development of such programs. Moreover, this program was developed based on longer texts and, consequently, as this work shows, it is not readily applicable to such short sentence blocks as Figure 6 also displays.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **2338** | comment | 0 | 2020-09-12T13:20:52 | blackpearl5710 | 0 | 0 | Thank God. The story lines were  getting boring. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **8450** | comment | 0 | 2020-09-09T22:09:16 | julovesjules | 0 | 0 | 🌸Lbcb🌸 |
| **5230** | comment | 0 | 2020-09-10T12:11:07 | carinagermanosouza | 0 | 0 | beautiful cute look love |

Figure 6: Positive Top Comments

Moreover, the linguistic rendering under Instagram posts like this one is very likely to differ immensely from the linguistic elaboration in other texts. Thus, looking at the top positive comments (Figure 6), which is inevitable to answer hypothesis 2, shows once again how little sense the application of a pure machine evaluation makes in this context. The comments are indeed denoted positively in relation to the individual words, but the interpretation of these as top positive comments makes little sense. Furthermore, it shows that it is a big problem that the comments are read independently of the context. Especially in this post, it is crucial to put the comments into context when interpreting them in order to be able to recognize stylistic devices such as irony, sarcasm, or even metaphors.

In addition, as already pointed out, it is essential to consider the context. On the one hand, the context in which this package was created and on the other hand, the context in which the individual words are weighted and the comments are placed in the context of the post. It turns out that this program is simply not suitable for analyzing such short text phrases because this program was most likely developed under entirely different conditions, such as texts with a different kind of language. In the comments analyzed, much colloquial language is used, which is not suitable for this program. In addition, as mentioned, there are the short text phrases and the lack of context of the text to the post. All this leads to the fact that this program does not answer any hypotheses of our research question. Another approach is needed.

## 5. Discussion

Several limitations and problems must be considered in this project. First of all, it must be emphasized that ridiculing reality TV in this context makes no sense at all. As already pointed out, in COVID-19, television and reality TV are essential factors in coping with social isolation and loneliness (Banerjee & Rai, 2020; Bryant & Zillmann, 1984; Bickham & Rich, 2006; Finn & Gorr, 1988).

However, it is crucial to conduct this study differently by adopting other approaches. As mentioned before, this purely automated approach makes no sense when analyzing Instagram comments. Thus, a different methodological approach is needed, and further work is needed to answer the essential research question.

Besides, the limitations above need a detailed discussion to determine what kind of further research is appropriate. Various studies have already dealt with the problem at hand here in the study, which disregarded what the study has long been dealing with: Automated analysis can be an excellent complementary tool to manual analysis, but it never replaces it entirely (Grimmer & Stewart, 2013; Hopkins & King, 2010; Song et al., 2020; van Atteveldt & Peng, 2018).

However, there is an upcoming issue with the validity of the automated generated data, which can only be overcome with the help of human manual coding (Song et al., 2020). The post-measure of the subsample of the 20 top comments coded manually seems like an ideal approach to validate the automated generated data (e.g., DiMaggio, 2015; Grimmer & Stewart, 2013; Lowe & Benoit, 2013; Song et al., 2020). Indeed, there is a general problem with automated analysis because the reliability is extremely high, but this does not mean that the validity is that high, too (Conway, 2006). Therefore, Conway (2006) mentioned that words could have multiple meanings in the context of a sentence and that false positive or false negatives are very common in computer-based analysis. "Validate, Validate, Validate" (Grimmer & Stewart, 2013, p. 271) thus, is a vital principle of a successful analysis.

Grimmer and Stewart (2013) are very critical when it comes to automated analysis saying that almost all of those tools are wrong and only some of them could be useful and therefore just additional to human-based work. Accordingly, manual coding cannot be translated 1:1 into an automated analysis (Hase et al., 2020); optimally, one will combine manual and automated coding (Wettstein et al., 2014) since the automated analysis could help to sample standardized Metadata of the texts first. And when it comes to detailed analysis, one should still use humans to identify different stylistic devices. In this case, it was helpful that the data was scraped automatically, and with this, several Metadata have been sampled automatically, too. However, when it comes to the actual analysis, the work should have been done by humans. Therefore, as several researchers mentioned, automatic work should still be combined with the analysis done by humans (e.g., Song et al., 2020).

Nevertheless, the automated analysis does have advantages that could help to improve communication research. Automated analysis can provide more detailed evidence of what journalism researchers already suspect and sketch a more extensive picture due to the immense amounts of data that can be analyzed, and in addition, automated analysis saves time and money (Boumans & Trilling, 2015). However, the inclusion of computer science into the research field of social science is still an ongoing process (Flaounas et a. 2013), which will continue to be worked on. Because, as mentioned earlier, automated analysis can be an excellent complementary tool in communication research.

When criticizing a method so much, it is essential to consider why this type of approach did not work in this study. Thus, as already pointed out, it is crucial to familiarize oneself with the tool's dictionary and adjust it if necessary. The package provides an excellent basis on which to build and perfect the program.

Moreover, it was also shown that the dictionary used for this purpose does not show any improvement when short text phrases are sorted out. Nevertheless, this has been an essential factor in the automated analysis and why this tool was not valid. Thus, it remains to be mentioned that the dictionary used for this purpose should instead be applied to longer phrases since, in the case of this study, it was proven that analysis makes little to no sense to short phrases. Another factor could have been using the package *textclean*, which replaced emojis in words to include them in the analysis. This inclusion certainly worked, but probably here, as in the rest of the analysis, the problem was that emojis often convey humor, irony, or similar stylistic devices, which a computer cannot yet classify at all; at least not in the content analysis of text files.

Nevertheless, it must be mentioned that even after emojis were illuminated entirely in the analysis, the results did not show much of a difference. This means that it would be crucial to adapt the package to our needs. This requires, however, an extensive study of computer science and probably complicated programming.

However, it is essential to emphasize that this study has not failed in any case. Instead, this paper is intended to encourage further critical engagement with automated analysis. Related to this and as already mentioned, it is difficult for a communication scientist to entirely rely on a program that was not explicitly developed for this study. Machine learning is in principle able to understand humor, but this is based on excessive exposure to the text to be analyzed and the inclusion of many social factors. A program designed for a specific unit of analysis is not automatically applied to every other research.

Therefore, there are many incentives for further research. On the one hand, and this is already the case, it is vital to work in a much more interdisciplinary way. Therefore, computer science needs to become a part of social science to enable a more accurate reappraisal of the tools. Only if the coder is also familiar with the socio-psychological factors of communication science can he or she develop a proper tool that considers all relevant factors. Moreover, one should consider that even if a developed tool under different conditions is used, one will always have to adapt it.

Related to this study, a manual analysis based on the failed validity test would be helpful. In this case, an investigation using an automated tool does not make sense due to the problems mentioned above. Indeed, other factors have further complicated the analysis and why almost all comments were immensely misclassified.

Thus, a manual analysis of the comments based on a positive reliability and validity check would make sense. It is necessary to perform this coding with several independent coders and generate the best possible codebook based on various pretests with many such comments. In addition, with this approach, it would also be possible to identify and classify applied stylistic devices and difficulties such as the colloquial language of the users. Also, human coders could put the comments in context to the post, which was also missing from the computer-based analysis. This is incredibly important to correctly classify the comments emotionally, as it is precisely through those stylistic devices such as irony that comments become recognizable.

Taking into account that the NRC Lexicon of Mohammad and Turney (2013) has already been applied with a reasonable outcome, it can be said that it has been helpful and valuable to apply this type of method to other data and to investigate whether this lexicon is applicable to even more diverse data (e.g., Gupta et al., 2020; Mathur et al., 2020). Therefore, it can be concluded that this work is a suitable qualification for further research. In order to achieve a valid output and answer the research questions, it would be necessary for further research to code the comments manually and perhaps think about a smaller sample.

Nevertheless, as argued in the paper, the phenomenon of Parasocial Relationships is more relevant than ever and essential for society, especially in the time of social isolation due to COVID-19. Parasocial Relationships are becoming more robust, and the end of a series, which due to various factors is even more intense for one, can be even worse for many, especially in social isolation.

Additionally, this approach could also be used in strategic communication. If further research or a manual analysis shows that women and men react differently, as claimed in the hypotheses, this would be a fundamental approach in the strategic communication of media providers when it comes to producing other formats. The comprehensive research of this thesis could map precisely who reacts under which conditions and how exactly. Furthermore, this work could point out how to design an institution's communication precisely on this, be it in corporate communication, the media industry, advertising, or else.

In conclusion, this work should be seen as an incentive for further. Besides, it is essential to disclose that the computer alone cannot replace the human being yet, and instead, an interaction of both is essential. Furthermore, it is crucial to deal critically with new scientific methods and its possible impacts on science and our knowledge on communication.

## 6. References


Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (1997). Adult attachment and the break-up of romantic relationships. *Journal of Divorce and Remarriage*, *27*(3/4), 121–139. DOI:10.1300/J087v27n03_08

Banerjee, D., & Rai, M. (2020). Social isolation in Covid-19: The impact of loneliness. *International Journal of Social Psychiatry,* *66*(6), 525-527. DOI:10.1177/0020764020922269

Barbara, A. M., & Dion, K. L. (2000). Breaking up is hard to do: Especially for the strongly 'preoccupied' lovers. *Journal of Personal and Interpersonal Loss*, *5*, 315–342. DOI:10.1080/10811440008407850

Barbieri, F., Ballesteros, M., & Saggion, H. (2017). Are Emojis Predictable? *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. DOI:10.18653/v1/e17-2017

Basil, M. D. (1996). Identification as mediator of celebrity effects. *Journal of Broadcasting & Electronic Media, 40*(4), 478-495. DOI:10.1080/08838159609364370

Bendelow, G. (1993). Pain perceptions, emotions and gender. *Sociology of Health and Illness*, *15*(3), 273–294. DOI:10.1111/1467-9566.ep10490526

Bernhold, Q. S., & Metzger, M. (2018). Older Adults' Parasocial Relationships with Favorite Television Characters and Depressive Symptoms. *Health Communication, 35*(2), 168–179. DOI:10.1080/10410236.2018.1548336

Bickham, D. S., & Rich, M. (2006). Is Television Viewing Associated With Social Isolation? *Archives of Pediatrics & Adolescent Medicine, 160*(4), 387. DOI:10.1001/archpedi.160.4.387

Bingaman, J. (2020). “Dude I’ve Never Felt This Way Towards a Celebrity Death”: Parasocial Grieving and the Collective Mourning of Kobe Bryant on Reddit. *OMEGA - Journal of Death and Dying,* 003022282097153. DOI:10.1177/0030222820971531

Bond, B. J. (2016). Following Your “Friend”: Social Media and the Strength of Adolescents' Parasocial Relationships with Media Personae. *Cyberpsychology, Behavior, and Social Networking*, *19*(11), 656–660. https://doi.org/10.1089/cyber.2016.0355

Boon, S. D., & Lomore, C. D. (2001). Admirer–celebrity relationships among young adults: Explaining perceptions of celebrity influences on identity. *Human Communication Research*, *27*, 432–465. DOI:10.1111/j.1468-2958.2001.tb00788.x

Boukes, M., van de Velde, B., Araujo, T., & Vliegenthart, R. (2019). What’s the Tone? Easy Doesn’t Do It: Analyzing Performance and Agreement Between Off-the-Shelf Sentiment Analysis Tools. *Communication Methods and Measures, 14*(2), 83–104. DOI:10.1080/19312458.2019.1671966

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism, 4,* 8–23. DOI:10.1080/21670811.2015.1096598

Bowlby, J. (1969). *Attachment and loss: Vol. I. Attachment.* New York: Basic Books.

Bowlby, J. (1973). *Attachment and loss: Vol. II. Separation*. New York: Basic Books.

Bowlby, J. (1980). *Attachment and loss: Vol. III. Loss.* New York: Basic Books.

Broussard, G. (2020). *Television Update Fall 2020*. eMarketer. https://www.emarketer.com/content/television-update-fall-2020#page-report

Brown, W. J., Basil, M. D., & Bocarnea, M. C. (2003). The influence of famous athletes on health beliefs and practices: Mark McGwire, child abuse prevention, and androstenedione. *Journal of Health Communication, 8,* 41-57. DOI:10.1080/10810730305733

Bryant, J., & Zillmann, D. (1984). Using television to alleviate boredom and stress: Selective exposure as a function of induced excitation states. *Journal of Broadcasting*, *28*(1), 1–20. DOI:10.1080/08838158409386511

Cappallo, S., Svetlichnaya, S., Garrigues, P., Mensink, T., & Snoek, C. G. (2019). New Modality: Emoji Challenges in Prediction, Anticipation, and Retrieval. *IEEE Transactions on Multimedia, 21*(2), 402–415. DOI:10.1109/tmm.2018.2862363

Caughey, J. (1984). *Imaginary social worlds: A cultural approach*. Lincoln: University of Nebraska Press.

Chory-Assad, R. M., & Yanen, A. (2005). Hopelessness and Loneliness as Predictors of Older Adults' Involvement With Favorite Television Performers. *Journal of Broadcasting & Electronic Media*, *49*(2), 182–201. DOI:10.1207/s15506878jobem4902_3

Chung, S., & Cho, H. (2014). Parasocial relationship via reality TV and social media. *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '14*. DOI:10.1145/2602299.2602306

Cohen, J. (1997). Parasocial relations and romantic attraction: Gender and dating status differences. *Journal of Broadcasting and Electronic Media, 41*, 516–529. DOI:10.1080/08838159709364424

Cohen, J. (2003). Parasocial Breakups: Measuring Individual Differences in Responses to the Dissolution of Parasocial Relationships. *Mass Communication and Society, 6*(2), 191-202. DOI:10.1207/s15327825mcs0602_5

Cohen, J. (2004). Parasocial Break-Up from Favorite Television Characters: The Role of Attachment Styles and Relationship Intensity. *Journal of Social and Personal Relationships, 21*(2), 187-202. DOI:10.1177/0265407504041374

Cohen, J., & Metzger, M. (1998). Social affiliation and the achievement of ontological security through interpersonal and mass communication. *Critical Studies in Mass Communication, 15*, 41–60. DOI:10.1080/ 15295039809367032

Cole, T., & Leets, L. (1999). Attachment styles and intimate television viewing: Insecurely forming relationships in a parasocial way. *Journal of Social and Personal Relationships*, *16*, 495–511. DOI:10.1177/0265407599164005

Conway, M. (2006). The Subjective Precision of Computers: A Methodological Comparison with Human Coding in Content Analysis. *Journalism & Mass Communication Quarterly, 83*(1), 186-200.

DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society, 2*(2). DOI:10.1177/2053951715602908

Dunlap, J. C., & Lowenthal, R. R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education, 20*, 129–135.

Eden, A. L., Johnson, B. K., Reinecke, L., & Grady, S. M. (2020). Media for Coping During COVID-19 Social Distancing: Stress, Anxiety, and Psychological Well-Being. *Frontiers in Psychology*, *11*. DOI:10.3389/fpsyg.2020.577639

Ellithorpe, M. E., & Brookes, S. E. (2018). I didn't see that coming: Spoilers, fan theories, and their influence on enjoyment and parasocial breakup distress during a series finale. *Psychology of Popular Media Culture, 7*(3), 250-263. DOI:10.1037/ppm0000134

Eyal, K., & Cohen, J. (2006). When good friends say goodbye: parasocial breakup study. *Journal of Broadcasting & Electronic Media, 50*(3), 502-523. DOI:10.1207/s15506878jobem5003_9

Finn, S., & Gorr, M. B. (1988). Social isolation and social support as correlates of television viewing motivations. *Communication Research*, *15*, 135–158. DOI: 10.1177/009365088015002002

Flaounas, I., Ali, O., Lansdall-Welfare, T., De Bie, T., Mosdell, N., Lewis, J., & Cristianini, N. (2013). RESEARCH METHODS IN THE AGE OF DIGITAL JOURNALISM. *Digital Journalism*, *1*(1), 102–116. DOI:10.1080/21670811.2012.714928

Garimella, K., Cohen, J., & Weber, I. (2017). Understanding Para Social Breakups on Twitter. *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17*. DOI:10.1145/3091478.3098877

Goudreau, J. (2010). *Reality TV stars turned entrepreneurs*. Forbes. http://www.forbes.com/2010/04/ 13/reality-tv-celebrity-fame-forbes-woman-entrepreneurs-business.html.

Gregg, P. B. (2018). Parasocial Breakup and Twitter: The Firing of Barb Abney. *Journal of Broadcasting & Electronic Media, 62*(1), 38-50. DOI:10.1080/08838151.2017.1402900

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*, 267–297. DOI:10.1093/pan/mps028

Gupta, V., Jain, N., Katariya, P., Kumar, A., Mohan, S., Ahmadian, A., & Ferrara, M. (2021). An Emotion Care Model using Multimodal Textual Analysis on COVID-19. *Chaos, Solitons & Fractals*, *144*, 110708. DOI:10.1016/j.chaos.2021.110708

Hartmann, T., Stuke, D., & Daschmann, G. (2008). Positive parasocial relationships with drivers affect suspense in racing sport spectators. *Journal of Media Psychology: Theories, Methods, and Applications, 20*, 24–34. DOI:10.1027/1864-1105.20.1.24

Hase, V., Engelke, K. M., & Kieslich, K. (2020). The Things We Fear. Combining Automated and Manual Content Analysis to Uncover Themes, Topics and Threats in Fear-Related News. *Journalism Studies (London, England), 21*(10), 1384-1402.

Hazan, C., & Shaver, P. R. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52*, 511–524. DOI: 10.1037/0022-3514.52.3.511

Hopkins, D. J., & King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science, 54*, 229–247. DOI:10.1111/j.1540- 5907.2009.00428.x

Horton, D., & Wohl, R. R. (1956). Mass Communication and Para-Social Interaction. *Psychiatry, 19*(3), 215-229. DOI:10.1080/00332747.1956.11023049

Hwang, T.-J., Rabheru, K., Peisah, C., Reichman, W., & Ikeda, M. (2020). Loneliness and social isolation during the COVID-19 pandemic. *International Psychogeriatrics*, *32*(10), 1217–1220. DOI:10.1017/s1041610220000988

Hwang, K., & Zhang, Q. (2018). Influence of parasocial relationship between digital celebrities and their followers on followers' purchase and electronic word-of-mouth intentions, and persuasion knowledge. *Computers in Human Behavior, 87*, 155-173. DOI:10.1016/j.chb.2018.05.029

Jarzyna, C. L. (2020). Parasocial Interaction, the COVID-19 Quarantine, and Digital Age Media. *Human Arenas*. DOI:10.1007/s42087-020-00156-0

Kardashian West, K. [kimkardashian]. (2020, September 08). To our amazing fans - It is with heavy hearts that we've made the difficult decision as a family to say goodbye to Keeping Up with the Kardashians. … [Instagram Photo]. Instagram. https://www.instagram.com/p/CE5AwirA4Cx/

Knobloch, S. (2003). Mood Adjustment via Mass Communication. Journal of Communication, 53(2), 233–250. DOI:10.1111/j.1460-2466.2003.tb02588.x

Knobloch, S., & Zillmann, D. (2002). Mood Management via the Digital Jukebox. *Journal of Communication, 52*(2), 351–366. DOI:10.1111/j.1460-2466.2002.tb02549.x

Lather, J., & Moyer-Guse, E. (2011). How Do We React When Our Favorite Characters Are Taken Away? An Examination of a Temporary Parasocial Breakup. *Mass Communication and Society, 14*(2), 196-215. DOI:10.1080/15205431003668603

Lim, C. M., & Kim, Y.-K. (2011). Older consumers' TV home shopping: Loneliness, parasocial interaction, and perceived convenience. *Psychology and Marketing, 28*, 763–780. DOI:10.1002/mar.20411

Lowe, W., & Benoit, K. (2013). Validating estimates of latent traits from textual data using human judgment as a benchmark. *Political Analysis, 21*, 298–313. DOI:10.1093/pan/mpt002

Lueck, J. A. (2015). Friend-zone with benefits: The parasocial advertising of Kim Kardashian. *Journal of Marketing Communications, 21*(2), 91-109.

Mathur, A. Kubde, P., & Vaidya, S. (2020). Emotional analysis using Twitter Data during Pandemic Situation: COVID-19. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, p. 845-848, DOI: 10.1109/ICCES48766.2020.9138079.

Mayring, P. (2010). *Qualitative Inhaltsanalyse*. Weinheim, Germany: Beltz Verlagsgruppe.

Meyrowitz, J. (1994). The life and death of media friends: New genres of intimacy and mourning. In Cathcart, R. & Drucker, S. (Eds.), *American heroes in a media age* (p. 52-81). Cresskill, NJ: Hampton.

Mikos, L. (2020). Film and Television Production and Consumption in Times of the COVID-19 Pandemic – The Case of Germany. *Baltic Screen Media Review, 8*(1), 30-34. DOI: 10.2478/bsmr-2020-0004

Mohammad, S. M., & Turney, P. D. (2012). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, *29*(3), 436–465. DOI:10.1111/j.1467-8640.2012.00460.x

Na'aman, N., Provenza, H., & Montoya, O. (2017). Varying Linguistic Purposes of Emoji in (Twitter) Context. *Proceedings of ACL 2017, Student Research Workshop.* DOI:10.18653/v1/p17-3022

Nabi, R. L., & Krcmar, M. (2004). Conceptualizing Media Enjoyment as Attitude: Implications for Mass Media Effects Research. *Communication Theory*, *14*(4), 288–310. DOI:10.1111/j.1468-2885.2004.tb00316.x

Netflix (2020). *Letter to Shareholders Quarter 2 2020*. https://s22.q4cdn.com/959853165/files/doc_finan- cials/2020/q2/FINAL-Q2-20-Shareholder-Letter-V3- with-Tables.pdf; (February 15 2021)

Ng, P. (2011). *Kardashian sisters continue to draw high TV ratings*. Reuters. http://www.reuters.com/article/2011/01/24/us-televisionidUSTRE70N6S620110124.

Papa, M. J., Singhal, A., Law, S., Pant, S., Sood, S., Rogers, E.M., et al. (2000). Entertainment-education and social change: An analysis of parasocial interaction, social learning, collective efficacy, and paradoxical communication. *Journal of Communication,* 50(4), 31-55. DOI: 10.1111/j.1460-2466.2000.tb02862.x

Paravati, E., Naidu, E., Gabriel, S., & Wiedemann, C. (2019). More than just a tweet: The unconscious impact of forming parasocial relationships through social media. *Psychology of Consciousness: Theory, Research, and Practice*. DOI:10.1037/cns0000214

Raji, S., & de Melo, G. (2020). What Sparks Joy: The AffectVec Emotion Database. *Proceedings of The Web Conference 2020*.DOI:10.1145/3366423.3380068

Rinker, T. W. (2018). *textclean: Text Cleaning Tools version 0.9.3*. Buffalo, New York. https://github.com/trinker/textclean

Rubin, R. B., & McHugh, M. P. (1987). Development of parasocial interaction relationship. *Journal of Broadcasting & Electronic Media, 31*(3), 279-292. DOI: 10.1080/08838158709386664

Rubin, A.M., Perse, E.M. & Powell, R.A. (1985). Loneliness, parasocial interaction, and local television news viewing. *Human Communication Research, 12*, 155–80.

Rubin, A. M., & Rubin, R. C. (1985). Interface of personal and mediated communication: A research agenda. *Critical Studies in Mass Communication, 2*, 36–53. DOI:10.1080/15295038509360060

Sanderson, J., & Cheong, P. H. (2010). Tweeting Prayers and Communicating Grief Over Michael Jackson Online. *Bulletin of Science, Technology & Society, 30*(5), 328-340. DOI:10.1177/0270467610380010

Shin, D. (2016). Do Users Experience Real Sociability Through Social TV? Analyzing Parasocial Behavior in Relation to Social TV. *Journal of Broadcasting & Electronic Media, 60*(1), 140-159. DOI:10.1080/08838151.2015.1127247

Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgaarden, H. G. (2020). In Validations We Trust? The Impact of Imperfect Human Annotations as a Gold Standard on the Quality of Validation of Automated Content Analysis. *Political Communication*, *37*(4), 550–572. DOI:10.1080/10584609.2020.1723752

Stefanone, M., Lackaff, D., & Rosen, D. (2010). The relationship between traditional mass media and “social media”: Reality television as a model for social network site behavior. *Journal of Broadcasting & Electronic Media. 54*, 3, 508-525.

Straka, M., & Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88–99. DOI:10.18653/v1/k17-3009

Tian, Q., & Hoffner, C. A. (2010). Parasocial Interaction With Liked, Neutral, and Disliked Characters on a Popular TV Series. *Mass Communication and Society, 13*(3), 250-269. DOI:10.1080/15205430903296051

van Atteveldt, W., & Peng, T. Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods & Measures, 12*, 81–92. DOI:10.1080/19312458.2018.1458084

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2017). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, *20*(5), 1813–1831. DOI:10.1177/1461444817707349

Welbers, K., & van Atteveldt, W. (2020). *corpustools: Managing, Querying and Analyzing Tokenized Text. R package version 0.4.2*. https://CRAN.R-project.org/package=corpustools

Wettstein, M. (2014). "Best of Both Worlds": Die halbautomatische Inhaltsanalyse. In Sommer, K., Wettstein, M., Wirth, W., & Matthes, J. (Eds), *Automatisierung in der Inhaltsanalyse* (p. 16-39). Cologne, Germany: Von Harlem.

Whitneyyy. [@feelingwhitneyd]. (2018, 01. July). *I worked at a Bar and a girl came in …* [link attached] [Tweet]. Twitter. https://twitter.com/feelingwhitneyd/status/1013517815897378816

Widiastuti, R., Mawarpury, M., Sulistyani, A., & Khairani, M. (2019). The Relationship between Celebrity Worship and Parasocial Interaction on Emerging Adult. *Proceedings of the 1st International Conference on Psychology*. DOI:10.5220/0009438000900094

Wijffels, J. (2020). *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit*. R package version 0.8.5. https://CRAN.R-project.org/package=udpipe

Wolf, A. (2000). Emotional Expression Online: Gender Differences in Emoticon Use. *Cyber Psychology& Behavior, 3*(5), 827–833. DOI:10.1089/10949310050191809

Zillmann, D. (1988). Mood Management Through Communication Choices. *American Behavioral Scientist*, *31*(3), 327–340. DOI:10.1177/000276488031003005

Zillmann, D., & Bryant, J. (Eds.). (1985). *Selective Exposure To Communication* (1st ed.). New York, United States: Routledge. DOI:10.4324/9780203056721

## Appendix

## R Script

```
# Load Packages
install.packages("corpustools")
library(corpustools)
install.packages("readr")
library(readr)
install.packages("stringr")
library(stringr)

# Load the Data
> kimk_insta <- read_csv("kimk_insta.csv")
> View(kimk_insta)

# Resolve
quote_pattern <- "“|”|„"
url_pattern <- "https?:.*(?=\\s|\\))"
email_pattern <- "\\w+@\\w+\\.\\w{2,3}"
# replace quotation marks „ “
kimk_insta$text <- str_replace_all(kimk_insta$text, quote_pattern, '"')
# replace urls
kimk_insta$text <- str_replace_all(kimk_insta$text, url_pattern, "URL")
# replace email
kimk_insta$text <- str_replace_all(kimk_insta$text, email_pattern, "EMAIL")

# Mithilfe von TextClean
> install.packages("textclean")
library(textclean)
kimk_insta$text <- replace_emoji(kimk_insta$text)
View(lexicon::hash_emojis)

# Corpus(without Lemmatisierung)
tc = create_tcorpus(kimk_insta, doc_column = "mID",
                    text_columns = "text")

# Corpus (with Lemmatisierung)
corpustools::show_udpipe_models()
tc = create_tcorpus(kimk_insta, doc_column = "mID",
                    text_columns = "text",
                    udpipe_model = "ewt",
                    udpipe_cores = 4L

# Lexicon URL
# http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

nrc <- read_excel("NRC-Emotion-Lexicon/NRC-Emotion-Lexicon-v0.92/NRC-Emotion-Lexicon-v0.92-
In105Languages-Nov2017Translations.csv.xlsx")
nrc$string <- nrc$`English (en)`
nrc <- nrc[,c(13, 3:12)]

tc$code_dictionary(nrc, token_col = "token")

install.packages("tidyverse")
library(tidyverse)

# Scores
anger_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Anger, na.rm = T)/n())
View(anger_score)
> write.csv(anger_score, file = "anger_score.csv")

sadness_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Sadness , na.rm = T)/n())
View(sadness_score)
write.csv(sadness_score, file = "sadness_score.csv")

surprise_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Surprise , na.rm = T)/n())
View(surprise_score)
write.csv(surprise_score, file = "surprise_score.csv")
```

```
negative_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Negative , na.rm = T)/n())
View(negative_score)
write.csv(negative_score, file = "negative_score.csv")

positive_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Positive , na.rm = T)/n())
View(positive_score)
write.csv(positive_score, file = "positive_score.csv")

anticipation_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Anticipation , na.rm = T)/n())
View(anticipation_score)
write.csv(anticipation_score, file = "anticipation_score.csv")

disgust_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Disgust , na.rm = T)/n())
View(disgust_score)
write.csv(disgust_score, file = "disgust_score.csv")

fear_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Fear , na.rm = T)/n())
View(fear_score)
write.csv(fear_score, file = "fear_score.csv")

joy_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Joy , na.rm = T)/n())
View(joy_score)
write.csv(joy_score, file = "joy_score.csv")

trust_score <- tc$tokens %>% group_by(doc_id) %>%
  summarise(score = sum(Trust , na.rm = T)/n())
View(trust_score)
write.csv(trust_score, file = "trust_score.csv")

#Plot Scores
> plot(trust_score)
> plot(joy_score)
> plot(fear_score)
> plot(disgust_score)
> plot(anticipation_score)
> plot(positive_score)
> plot(negative_score)
> plot(surprise_score)
> plot(sadness_score)
> plot(anger_score)

#Mean
mean(sadness_score$score)

#Sentiment in Comments


#Counting hits and plotting
install.packages("ggplot2")
library(ggplot2)

queries = data.frame(label = c('sad','crying','end','era'),
                     query = c('sad* OR crying OR end*', 'era*', 'cry* OR dying*', 'bad* OR
conflict*'))
hits = search_features(tc, query=queries$query, code=queries$label)

count_tcorpus(tc, hits=hits)

count_tcorpus(tc, hits=hits, meta_cols = 'user')

plot(hits)

# Weightening?
# In information retrieval, tf–idf, TF*IDF, or TFIDF, short for term frequency–inverse
# document frequency, is a numerical statistic that is intended to reflect how important a word is to
a document in a collection or corpus.
tf_idf <- feature_stats(tc, "token", context_level = c("document", "sentence"))
```

## Python Script (Insta-Scraper)

```
import os
import csv
import time
import math
import random
import traceback
import pandas as pd
from bs4 import BeautifulSoup
from datetime import datetime
from selenium import webdriver
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.support.ui import WebDriverWait
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.action_chains import ActionChains

def check_cookies(browser):
    try:
        # accept cookies
        browser.find_element_by_xpath('/html/body/div[2]/div/div/div/div[2]/button[1]').click()
    except Exception as e:
        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'No cookies.')
        print(e)

def check_login(browser, credentials):
    try:

browser.find_element_by_xpath('/html/body/div[1]/section/nav/div[2]/div/div/div[3]/div/span/a[1]/butto
n').click()
        time.sleep(2)
        # username
        browser.find_element_by_xpath('//*[@id="loginForm"]/div/div[1]/div/label/input').clear()

browser.find_element_by_xpath('//*[@id="loginForm"]/div/div[1]/div/label/input').send_keys(credentials
['user'])
        # password
        browser.find_element_by_xpath('//*[@id="loginForm"]/div/div[2]/div/label/input').clear()

browser.find_element_by_xpath('//*[@id="loginForm"]/div/div[2]/div/label/input').send_keys(credentials
['password'])
        time.sleep(1)
        browser.find_element_by_xpath('//*[@id="loginForm"]/div/div[3]/button/div').click()
        time.sleep(2)
        browser.find_element_by_xpath('/html/body/div[1]/section/main/div/div/div/div/button').click()
    except Exception as e:
        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Already logged in.')
        print(e)

def load_more_comments(browser):
    time.sleep(1)
    browser.execute_script('window.scrollTo(0, document.body.scrollHeight);')
    time.sleep(1)
    browser.execute_script('window.scroll(0, 0);')
    time.sleep(1)
    post = browser.find_element_by_xpath('/html/body/div[1]/section/main/div/div[1]/article/div[3]')

post.find_element_by_xpath('/html/body/div[1]/section/main/div/div[1]/article/div[3]/div[1]/ul/li/div/
button/span').click()

def find_replies(browser):
    post1 = browser.find_element_by_xpath('/html/body/div[1]/section/main/div/div[1]/article/div[3]')
    all_comments1 = post1.find_elements_by_tag_name('ul')[-12:]
    replies = 0
    for element in reversed(all_comments1):
        try:
            #temp_repl = element.find_element_by_xpath('//*[contains(text(), "View replies")]')
            #temp_repl = element.find_element_by_xpath('//*[contains(text(), "Antworten ansehen")]')
            temp_repl = element.find_element_by_class_name('EizgU')
            # print(temp_repl.text)
```

```
            # print('###')
            # print(element.get_attribute('innerHTML'))
            # print('########################')
            # print('########################')
            if 'ansehen' in temp_repl.text:
                temp_repl.click()
                replies += 1
        except:
            pass
    return replies

def scrape_comments(browser, existing_data, id_distr, last_id, writer):
    start_time = datetime.now()#.replace(microsecond=0)
    browser.implicitly_wait(1)
    # open all replies
    st_scrape = datetime.now()#.replace(microsecond=0)
    replies = find_replies(browser)
    check_c = 0
    print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'{replies} replies found.')
    while True:
        check_c += 1
        print('NOW',check_c,':',replies)
        if check_c % 50 == 0: time.sleep(20)
        if check_c == 100: break
        replies = find_replies(browser)
        if replies == 0: break
    et_open_replies = datetime.now()#.replace(microsecond=0)
    # narrow down
    post = browser.find_element_by_xpath('/html/body/div[1]/section/main/div/div[1]/article/div[3]')
    all_comments = post.find_elements_by_tag_name('ul')[-12:]
    for comment in reversed(all_comments):
        already_seen = False
        if comment.get_attribute('class').strip() == 'Mr508':
            # COMMENT
            # unpack in BeautifulSoup
            soup = BeautifulSoup(comment.get_attribute('innerHTML'), 'lxml')
            # username
            user = soup.find('h3').find('a').text
            # timestamp
            created_at = soup.find('time')['datetime'].split('.')[0]
            # check if already exists
            if user in existing_data:
                if created_at in existing_data[user]:
                    already_seen = True
                    parent_id = id_distr[f'{user}_{created_at}']
                    break
            # text
            text = soup.find('h3').next_sibling.text
            # likes
            like_count = 0
            for button in soup.find_all('button'):
                #if 'like' in button.text:
                if 'Gefällt' in button.text:
                    like_count = button.text.split()[1]
                    break

            # replies
            reply_count = 0
            if len(soup.find_all('ul')) > 0:
            #if len(soup.find_all('div', {'class':'ZyFrc'})) > 0:
                p_replies = soup.find('ul')
                #p_replies = soup.find_all('div', {'class':'ZyFrc'})
                all_replies = p_replies.find_all('h3')
                reply_count = len(all_replies)


            if not already_seen:
                last_id += 1
                row = [last_id, 'comment', '0', created_at, user, like_count, reply_count, text]
                parent_id = last_id
                writer.writerow(row)
                id_distr[f'{user}_{created_at}'] = last_id
                if user in existing_data:
                    existing_data[user].append(created_at)
```

```
                else:
                    existing_data[user] = [created_at]

            # replies
            if reply_count > 0:
                p_replies = soup.find('ul')
                #p_replies = soup.find_all('div', {'class':'ZyFrc'})
                for ireply in p_replies.find_all('div'):
                    if ireply['class'] == ['P9YgZ']:
                        # username
                        user = ireply.find('h3').find('a').text
                        # timestamp
                        created_at = ireply.find('time')['datetime'].split('.')[0]
                        # text
                        text = ireply.find('h3').next_sibling.text
                        # likes
                        like_count = 0
                        for button in ireply.find_all('button'):
                            if 'like' in button.text:
                                like_count = button.text.split()[0]
                                break
                        # check if already exists
                        if user in existing_data:
                            if created_at in existing_data[user]:
                                already_seen = True
                        if not already_seen:
                            last_id += 1
                            row = [last_id, 'reply', parent_id, created_at, user, like_count, '0',
text]
                            writer.writerow(row)
                            id_distr[f'{user}_{created_at}'] = last_id
                            if user in existing_data:
                                existing_data[user].append(created_at)
                            else:
                                existing_data[user] = [created_at]
    end_time = datetime.now()#.replace(microsecond=0)
    elapsed_time = end_time - start_time
    elapsed_time_replies = et_open_replies - start_time
    elapsed_time_scrape = end_time - st_scrape
    ret_data = [existing_data, id_distr, last_id, elapsed_time, elapsed_time_replies,
elapsed_time_scrape]
    return ret_data

# main function
def get_comments(browser, existing_data, id_distr, last_id, out):
    c = 0
    clicked = 0
    with open(out, 'a') as o_f:
        writer = csv.writer(o_f)
        # press '+' as long as it's possible
        while True:
            scthro = False
            # debug
            #if c == 5: break
            c += 1
            # random breaks to prevent bot detections
            time.sleep(random.randint(1,2))
            try:
                # scrape comments
                print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Scraping comments page {c}.')
                if math.floor(int(last_id) / 12) <= clicked:
                    ret_data = scrape_comments(browser, existing_data, id_distr, last_id, writer)
                    existing_data = ret_data[0]
                    id_distr = ret_data[1]
                    last_id = ret_data[2]
                    scthro = True
                browser.implicitly_wait(60)
                load_more_comments(browser)
                clicked += 1
                print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Already scraped {last_id}
comments.')
                if scthro:
                    print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'TIMING. Total:
{ret_data[3]} | Open replies: {ret_data[4]} | Scrape: {ret_data[5]}')
```

```
            except Exception as e:
                no_success = True
                print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Failed to load more comments.
Try again...')
                print(e)
                print(traceback.format_exc())
                for x in range(10):
                    try:
                        if x == 5:
                            time.sleep(55)
                        if x > 9:
                            time.sleep(595)
                        time.sleep(5)
                        load_more_comments(browser)
                        clicked += 1
                        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'...successful.')
                        no_success = False
                        break
                    except Exception as e2:
                        print(e2)
                        print(traceback.format_exc())
                        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'...no success (try nr
{x+1}).')
                if no_success:
                    print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f"Can't load any more
comments:")
                    print(e)
                    print(traceback.format_exc())
                    print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Exiting main loop.')
                    break

print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Start script.')
too_many_errors = 0
while too_many_errors < 101:
    try:
        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Read existing data.')
        #usr_path = '/Users/heidenreit89/'
        usr_path = '/Users/reminder/'

        header = ['mID', 'type', 'parentID', 'timestamp', 'user', 'like_count', 'reply_count', 'text']

        existing_data = {}
        id_distr = {}
        last_id = 0
        out = f'{usr_path}ucloud/Shared/Instagram_Pamina/data/kimk_insta.csv'
        # check if output file exists
        if not os.path.isfile(out):
            # if not: create file, write header
            with open(out, 'w') as w:
                writer = csv.DictWriter(w, header)
                writer.writeheader()
        # if exists: read in rows
        else:
            data = pd.read_csv(out)
            for i,row in data.iterrows():
                usr = row['user']
                tmstmp = row['timestamp']
                mid = row['mID']
                id_distr[f'{usr}_{tmstmp}'] = mid
                if int(mid) > last_id: last_id = int(mid)
                if usr in existing_data:
                    existing_data[usr].append(tmstmp)
                else:
                    existing_data[usr] = [tmstmp]
        # set parameters
        kimk = 'https://www.instagram.com/p/CE5AwirA4Cx/?utm_source=ig_web_copy_link'
        credentials = {'user':'hrgygax', 'password':'universitaet_wien'}
        # start google chrome browser
        options = Options()
        options.add_argument('--disable-images')
        options.add_argument('disable-infobars')
        options.add_argument('--disable-extensions')
        if usr_path == '/Users/heidenreit89/':
            browser = webdriver.Chrome(ChromeDriverManager().install(), chrome_options=options)
```

```
        elif usr_path == '/Users/reminder/':
            #browser = webdriver.Chrome(executable_path = f'{usr_path}Desktop/chromedriver',
chrome_options=options)
            browser = webdriver.Firefox()
        browser.implicitly_wait(10)
        # open url
        browser.get(kimk)
        # accept cookies
        check_cookies(browser)
        # login
        check_login(browser, credentials)
        # start scraping comments
        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Start scraping.')
        get_comments(browser, existing_data, id_distr, last_id, out)
    except Exception as e:
        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Error (Nr {too_many_errors}). Error
type:')
        print(e)
        too_many_errors += 1
        print(datetime.now().strftime('%d.%m.%Y, %H:%M:%S:'),f'Restart script...')
        browser.quit()
```

**Declaration of Authorship**

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources, I agree that it shall be entered in a database where it shall also remain after examination to enable comparison with future theses submitted. Other rights of reproduction and usage, however, are not granted here. This paper was not previously presented to another examination board and has not been published.

Vienna, July 10, 2021, Kim Pamina Syed Ali