Research On Anorexia: Pro-recovery And Pro-anorexia Communities On Tumblr

Anorexia nervosa is an eating disorder that is characterised by a restricted-energy intake, a disturbed body image and an intense fear of gaining weight (National Eating Disorders Association, n.d.-a). Approximately 1% of women and 0.3% of men will develop anorexia during their lifetime (National Eating Disorders Association, n.d.-b) but on average only 33% (of women) achieve recovery with one third likely to relapse again (Herzog et al., 1999). On top of this, the standardized mortality rate for anorexia of 5.86 (deaths per 1000 person a year) is higher than for other psychiatric disorders (Arcelus, Mitchell, Wales, & Nielsen, 2011). These numbers emphasise the seriousness and immense impact of anorexia which is especially prevalent in young adults (National Eating Disorders Association, n.d.-a). One online platform that has received a lot of attention regarding anorexia is Tumblr.

Tumblr – a popular social networking and microblogging website founded in 2007 (Tumblr, n.d.-a) – allows users to create their own blogs, and post or consume a variety of content which includes anorexia-related content. Research identified two main communities on Tumblr: pro-recovery and pro-anorexia. The pro-recovery community aims to support individuals suffering from anorexia especially those on the way to recovery. Their interaction and self-disclosure highlight a constant conflict between the desire to recover and the persistent presence of anorexic thoughts and feelings (Branley & Covey, 2017). At the same time, the pro-anorexia community focuses on promoting and supporting anorexia, e.g. by providing weight loss tips and encouragement (Branley & Covey, 2017; Choudhury, 2015).

On one hand, pro-anorexic content has raised many concerns over the impact it might have on users. While the usage of pro-eating disorder sites is most likely a consequence of having an eating disorder (Peebles et al., 2012), evidence suggests that it can still have negative consequences. Participants of an experiment that involved viewing a pro-anorexia website (as opposed to control websites) reported significantly higher negative affect, lower social self-esteem, lower appearance self-efficacy, and higher perceived weight. They also felt less likely to overeat, and more likely to exercise (Bardone-Cone & Cass, 2007).

Another study found that the higher the usage level of pro-anorexic sites, the higher users scored on the Eating Disorder Examination Questionnaire (Peebles et al.). This correlation might indicate that pro-anorexic content reaffirms those with anorexia and potentially prevents them from recovering. On the other hand, interviews with users of pro-anorexia websites yielded quite opposite findings. These websites offer a safe, anonymous space (Mulveen & Hepsworth, 2006) where individuals find a supportive community and can freely express themselves without being stigmatized or judged; this helps them cope with their struggles (Yeshua-Katz & Nicole Martins, 2013). Many use it as a way to manage and maintain their disorder as no attempts of correction are made¬ (Mulveen & Hepsworth, 2006). When Tumblr’s policy changes in 2012 involved censoring content that actively promotes self-harm and the implementation of public service announcements (PSA’s) (Tumblr Staff, 2012), the same two views became evident. While those viewing the pro-anorexia community as harmful and triggering supported the policy, others contested it. These users emphasised Tumblr’s value as a social support system and worried that the changes would lead to increased isolation and stigmatisation (Schott & Langan,

2015).

This conflict highlights the need to understand the nature of these communities and evaluate their benefits and detriments in order to create appropriate interventions if necessary. Considering the popularity of tumblr and the severity of anorexia, it is essential to further investigate this as past research is scarce. We propose using a Big Data approach as Tumblr data already fulfils the criteria of Big Data: It is large in volume, comes in at a high velocity and includes a variety of content. By monitoring user behaviour over time, we plan to investigate how the anorexic-related content Tumblr users seek out affects their content output and behaviour. Based on past findings, we expect to find evidence for two distinct communities (pro-anorexia and pro-recovery).

Despite the positive view some users have of the pro-anorexia community, we hypothesise that a higher exposure to pro-anorexia content correlates with a higher level of pro-anorexia behaviour (e.g. creating, liking, searching or sharing pro-anorexia content) and a more negative attitude. 694 words Methods Participants The target population in this research contains all Tumblr users that consume and produce anorexia-related content. While this includes all ages and genders, data from underage users will be excluded for privacy reasons. To obtain reliable results only users active during that time frame will be included. Additionally, users that only engage with anorexic-content sporadically or are rarely active (less than once every two weeks) will also be excluded.

Type of data The data that will be collected can be sorted into two categories: Blog posts and user data. While we only plan to analyse text posts, other formats like videos or photos will be included for secondary analysis via their tags. Regarding user data, this will include a user’s feed - which is composed of posts of all followed blogs - , followers, likes, comments, posted and reblogged content, frequency of personal messages and search terms. This data will give an insight into a user’s behaviour, by showing what community they are in, who they interact with and how frequently, and what content they consume, like and create. Data collection Before data collection, relevant data and users must be identified by determining frequently used tags in relation with anorexic content. For data collection itself, the Firehose Tumblr provides will be employed in order to receive a real-time feed of relevant data (Tumblr Engineering, 2012). At the same time, Tumblr also offers an API which after authentication can aid in collecting user, blog, and blog post data (Tumblr, n.d.-b).

Data collection would start on a randomly selected day and continue for six months. Monitoring data as it is created ensures a higher representativeness as opposed to collecting past data which might have undergone changes or even been deleted. Data analysis The aim is to analyse the connections between user networks, consumed and produced content and search terms over a certain time span to identify patterns or changes in behaviour. Before the actual data analysis, relevant tags must be identified for which a website like www.hashtagify.me could be used. Additionally, frequently co-occurring tags will also be added manually. Regarding the actual data analysis, this will be separated in several parts. Following the approach of distributed processing via MapReduce, incoming data will first be sorted by user, which over time allows us to identify both frequently active and infrequently active users for further selection or exclusion. Different aspects of the data will be analysed directly, and only the results are saved.

Data analysis attempts to provide answers to the following three questions: First, what content

does a user consume and does this change over time? This would be identified by looking at blogs a user follows and their content (text, tags, attitudes). Secondly, how does a user behave and does this change over time? This would be identified by looking at a user's posted content (texts, tags, attitude), likes, search terms, comments, activity patterns, and communication (number of messages send and received). Lastly, are there any relations or common patterns between consumed content and user behaviour? This would be identified by comparing the main topics, attitudes and activity changes.

Analysing the data will rely primarily on natural language processing (NLP) which will include word-frequency and sentiment analysis. An example of frequently used words when searching Tumblr with the tag 'ana' can be seen in Figure 1 (Appendix). By including different functions themes and topics of posts can be identified. Past research by Choudhury (2015) yielded that there are differences in content and language between pro-anorexia and pro-recovery posts. Choudhury was also able to develop an algorithm that could predict whether a post involved anorexia (83% accuracy) and differentiate between pro-anorexia and pro-recovery (74%). This supports the idea that algorithms and applications can be successfully used to categorize posts and users. Since many posts fall into one of the two categories (pro-anorexia or pro-recovery), k-clustering will be used to help identify the main communities and which community user fall. This can help in classifying posts. 663 words Methodological, ethical, and legal issues

The main methodological problem concerns identifying relevant content. For one, the PSA implementations led users to adapt new tags to hide their content (Lauren Rae Orsini, 2012). Moreover, the pro-recovery community tends to 'infiltrate' the pro-anorexia community by using their tags (Choudhury, 2015). Since multiple methods are used to identify topics and the number of posts will be high, the significance of occasional mistakes might be small. Secondly, there are a few ethical considerations. Users have reported a fear of disclosure in the past (Yeshua-Katz & Nicole Martins, 2013) which further highlights their vulnerability. Paired with the difficulty of obtaining consent from every user, we propose a collaboration with Tumblr and user notifications of ongoing research with an option to withdraw their data. Transparency, an emphasis on the non-intrusiveness of the research and guaranteed anonymity (no personally identifiable data is saved or published) are all crucial here. Concerning the legality of data and text mining some uncertainty still exists. However, there will be a plenary vote regarding changes to the European copyright in March or April of 2019 (Julia Reda, n.d.). These changes will completely legalise data and text mining for research purposes and institutes (European Commission, 2016). 198 words