Title

Profiling commenters on mental health-related online forums: A methodological example focusing on eating disorder-related commenters

Running head

PROFILING MENTAL HEALTH-RELATED FORUM COMMENTERS

Authorship

Duncan McCaig* MSc, Mark T Elliott* PhD, Cynthia S Q Siewb,c PhD, Lukasz Walasekb PhD, and Caroline Meyera,c,d PhD

*Corresponding author: Duncan McCaig. Email address: d.mccaig@warwick.ac.uk. Address: WMG, University of Warwick, Coventry, UK, CV4 7AL

Authors’ institutional affiliations

aWMG, University of Warwick, Coventry, UK

bDepartment of Psychology, University of Warwick, Coventry, UK

cWarwick Medical School, University of Warwick, Coventry, UK

dCoventry and Warwickshire NHS Partnership Trust, Coventry, UK

eDepartment of Psychology, National University of Singapore, Singapore

Intended type of paper

Original paper
Abstract

Background: Understanding the characteristics of commenters on mental health-related online forums is vital for the development of effective psychological interventions in these communities. The way in which commenters interact can enhance our understanding of their characteristics. Objective: Using eating disorder-related (EDR) forums as an example, this study details a methodology that aimed to determine subtypes of mental health-related forums, and profile their commenters based on the other forums to which they contributed. Methods: The researchers identified all public EDR-forums (with ≥500 contributing commenters between March 2017 and February 2018) on a large online discussion platform (Reddit). A mixed-methods approach comprising network analysis with community-detection, text-mining and manual review identified subtypes of EDR-forums. For each subtype, another network analysis with community-detection was conducted using the EDR-forum commenter-overlap between 50 forums on which the commenters also commented. The topics of forums in each detected community were then manually reviewed to identify the shared interests of each subtype of EDR-forum commenters. Results: Six subtypes of EDR-forums were identified, to which 14024 commenters had contributed. The results focus on two subtypes – pro-eating disorder, and thinspiration – and communities of commenters within both subtypes. Within the pro-eating disorder subtype, three communities of commenters were detected that related to the body and eating, mental health, and women, appearance and mixed topics. Regarding the thinspiration group, 78% of commenters had also commented on pornographic forums, and 17% had contributed to pro-eating disorder forums. Conclusions: The article exemplifies a methodology that provides insight into subtypes of mental health-related forums, and the characteristics of their commenters. The findings have implications for future research, and online psychological interventions. With the publicly available data and code provided, researchers can easily reproduce the analyses, or utilise the methodology to investigate other mental health-related forums.

Keywords
mental health; eating disorders; social media; reddit; social networks; network analysis; text-mining
Introduction

Background

Compared to clinician-delivered psychological interventions, online interventions targeting mental health conditions offer several benefits, such as cost-effectiveness [e.g., 1], and potentially reaching a larger number of the target population [e.g., 2]. However, for these interventions to be effective, it is important that users adhere to them. A recent review indicated that several characteristics (e.g., gender, impersonality of intervention content) are important predictors of adherence to online psychological interventions [3]. Therefore, if adherence to online psychological interventions is to be improved, a clearer understanding of the characteristics of target populations (e.g., online mental health-related communities) is vital. The current article uses the example of eating disorder-related (EDR) forums to outline a reproducible methodology that can enhance our understanding of the characteristics of online mental health-related communities.

Eating disorder-related forums

EDR-forums are easily accessible online, and their content can be defined broadly as either ‘pro-eating disorder’, or ‘pro-recovery’ [cf. 4]. Pro-eating disorder content encourages the enactment of eating disordered behaviours (e.g., fasting, excessive exercise) without indicating a desire for recovery, and typically portrays eating disorders as a lifestyle choice rather than a mental health condition [4]. The term ‘thinspiration’ is often used to refer to pro-eating disorder material (e.g., photos) that encourages eating disordered behaviours [5]. In contrast, pro-recovery (or ‘anti-eating disorder’) content encourages recovery from eating disorders, and/or confronting eating disordered behaviours [4]. In addition to the pro-eating disorder/pro-recovery distinction, EDR-forums can also be characterised as relating to eating disorders in general, or specific diagnostic categories (e.g., anorexia nervosa, bulimia nervosa, binge eating disorder).

A deeper understanding of pro-eating disorder forums is particularly important [6]. Emphasising this importance, semi-structured interviews with current eating disorder patients found that they perceived pro-eating disorder websites as having reinforced and maintained their eating disorders [7]. Furthermore, a recent meta-analysis [8] showed that greater engagement with pro-eating disorder forums is associated with increased body dissatisfaction, dieting and negative affect, but not bulimic symptoms. Despite this issue’s importance, few studies have attempted to elucidate the characteristics of people who engage with EDR-forums.
Survey-based research

Most existing studies attempting to characterise users of EDR-forums have used survey-based methods, and samples recruited directly from EDR-forums or student samples. For example, Peebles et al. [9] recruited participants from pro-eating disorder websites and found that they were predominantly young women within the ‘healthy’ body mass index range. Over 70% of the respondents also indicated that they had purged, binged or used laxatives to control their weight. Exemplifying the use of student samples, Harper et al. [10] asked undergraduate women to complete self-report questions about viewing EDR-websites. Compared to controls who had not viewed EDR-websites, participants who had viewed pro-eating disorder websites reported higher appearance dissatisfaction, dietary restriction and bulimic symptoms, and more frequent viewing of cosmetic surgery websites.

While the research detailed above provides insight into commenters’ characteristics using standardised self-report scales, the studies’ samples are unlikely representative of everyone who engages with EDR-forums. For example, survey respondents on forums might differ from those who do not respond [e.g., 9], and relatively small student populations might not be representative of the actual forum users [e.g., 10].

Textual analyses of online content

Textual analyses of online content offer an alternative methodology to surveys, and complement findings obtained through self-report measures. Such analyses can be approached in different ways, including manual qualitative methods (e.g., thematic analysis, content analysis), and computerised methods (e.g., word-counts, topic modelling). However, these approaches all address a similar aim in identifying themes discussed on EDR-forums.

Through using manual qualitative methods, researchers have found that most EDR-forum commenters are women [4], and that eating and shape concerns are the most commonly expressed symptoms in a pro-eating disorder forum [11]. Despite variability between studies, applications of computerised word-count methods have indicated that pro-recovery and pro-eating disorder commenters differ in terms of factors such as affect and self-directed attention [12, 13], and mentions of fitness tracking technology [14]. A combination of both manual and computerised textual analyses has also indicated that recovery is less frequently mentioned by commenters in the least recovery-focused eating disorder stages of change [i.e., precontemplation, relapse; 15]. Recently, Moessner et al. [16] employed sophisticated computerised topic modelling to characterise a pro-eating disorder community in terms of the themes that its commenters discuss, including feedback and social support, and weight gain/loss.
Interaction between forum users

A complementary approach to textual analyses is to explore the way in which users on a forum interact. In the same study detailed above, Moessner et al. [16] investigated how commenters within a pro-eating disorder forum interacted, and were able to identify particularly influential commenters. In contrast, two studies have explored the interaction between communities of commenters on Twitter [13, 17]. Tiggemann et al. [17] theoretically selected and compared thinspiration and ‘fitspiration’ (i.e., content encouraging health and fitness) posts on Twitter. During a two-week period, the researchers identified users that included thinspiration or fitspiration tags in their posts, and found minimal overlap (i.e., interaction) between the commenters in the two communities. Using a more data-driven approach, Wang et al. [13] detected two communities of commenters who posted comments with EDR-tags. The two communities were found to reflect pro-eating disorder and pro-recovery stances, with network analyses indicating minimal overlap between the communities.

While Moessner et al. [16] explored the user-interaction within a forum, Tiggemann et al. [17] and Wang et al. [13] investigated the user-interaction between two communities. However, these approaches could be extended to explore the overlap between, theoretically, an infinite number of communities. While researchers have previously used network analyses to investigate traffic between over 500 EDR-websites [18], this has not been done with online discussion forums, nor at the level of the individual user.

Reddit

In February 2018, Reddit was the sixth most visited website in the world, with 234 million unique visitors [19]. Reddit is a large online discussion platform comprising thousands of forums (i.e., ‘subreddits’), and is perfectly suited to investigate the overlap between online communities on a large scale. Each subreddit relates to a specific topic (e.g., politics, films), and can be conceptualised as a community of people with a shared interest. A member of a subreddit can start a conversation (i.e., ‘thread’), or join a conversation by responding to an existing comment. Reddit can be viewed as a microcosm of the internet, as users are likely to engage only with a relatively small number of subreddits that interest them. Reddit therefore provides a unique opportunity to explore how online forums are related in terms of whether the same people comment on them. Furthermore, by identifying a group of commenters and identifying the other forums to which they contribute, the communities to include in analyses can be determined in a largely data-driven way, in contrast to the theoretical approach used in previous studies [17, 18]. Lastly, as eating disorders are the...
explicit focus of several subreddits, and have been the focus of previous studies [e.g., 11, 12, 16], Reddit presents a particularly beneficial opportunity to investigate the characteristics of EDR-forums in greater detail.

Objectives
While it is common for studies to consider EDR-online discussion forums in isolation [e.g., 11, 14, 16], recent studies have investigated how similar forums overlap in terms of commenters [e.g., 17]. The present study built on the latter approach, and used publicly available data from Reddit [20] to achieve two separate objectives, as detailed below.

Objective 1: Determining subtypes of eating disorder-related subreddits
The first objective of this study was to identify subtypes of EDR-forums on Reddit. In order to achieve this, large EDR-subreddits were identified, and the way in which they overlapped with regard to commenters was calculated. A mixed-methods approach was then applied, comprising network analysis with community-detection, text-mining and manual review of the EDR-subreddits’ focuses. This enabled the grouping of EDR-subreddits into subtypes, and, as a result, facilitated the elucidation of thematic heterogeneity in EDR-subreddits (e.g., pro-eating disorder, pro-recovery). The output regarding Objective 1 also determined the grouping of subreddits for the analyses relating to Objective 2.

Objective 2: Profiling eating disorder-related commenters based on contributions to ancillary subreddits
The second objective of the current study was to profile each subtype of EDR-forums (see Objective 1) in terms of their main interests, as represented by the topics of subreddits to which the commenters also contributed. In order to achieve Objective 2, all the other (public) subreddits to which EDR-commenters had contributed (‘ancillary subreddits’) were identified. A mixed-methods approach was then conducted consisting of network analysis with community-detection, and manual review of the ancillary subreddits’ topics. The overall topics of the subreddits comprising each detected community then enabled the profiling of groups of EDR-commenters in terms of the other topics in which they were interested.

Methods
Corpus selection and data analysis
All public Reddit comments, excluding the initial post to which commenters respond, are regularly archived and freely available [20]. While the archive includes all comments since December 2005, this study’s corpus comprises all comments posted between March 2017 and February 2018 (inclusive). This represents the one-year period preceding the most recent month’s data that were available when beginning the study. As publicly available data were used, this study was outside the remit of the University of Warwick’s Biomedical and Scientific Research Ethics Committee, from
whom an exemption from ethical review was obtained. All data were extracted, pre-processed and analysed with Python programming language [21], except where otherwise stated. The code used in this study is available as Supplementary Material and can be used to replicate the analyses.

Objective 1: Determining subtypes of eating disorder-related subreddits
A list of search terms was created to identify EDR-subreddits, which is provided as Supplementary Material (Supplementary table 1). This list was generated through consultation of EDR-sections of two clinical references, DSM-V and ICD-10 [22, 23], and previous research concerning EDR-online communities [14, 24]. Search terms were developed that related to: 1) eating disorders in general (e.g., eating disorder); 2) specific eating disorder diagnostic categories (e.g., anorexia nervosa, bulimia nervosa, binge eating disorder); or 3) online content associated with eating disorders (e.g., thinspiration). Subreddits were included that contained at least one search term in their name or brief description, but excluded if they 1) were unrelated to eating disorders (e.g., Anorexiclizardpeople), or 2) were private.

For each subreddit identified through this search and inclusion strategy, each comment and commenter name was then extracted from the Reddit data [20]. In order to focus the analyses on the largest EDR-subreddits, any subreddits with fewer than 500 commenters contributing within the one-year period were excluded. The Reddit commenters ‘AutoModerator’ and ‘[deleted]’ were not included in this count, and were excluded from all subsequent analyses.

A mixed-methods approach was then used in order to identify subtypes of the EDR-subreddits. This approach comprised three techniques, which are described below and were conducted in the order presented: 1) network analyses with community-detection; 2) text-mining; and 3) manual review of EDR-subreddits’ focuses.

Network analyses with community-detection
In order to conduct the network analyses with community-detection, a list of commenters was compiled separately for each of the included EDR-subreddits. For each pairwise comparison of the subreddits (e.g., subreddit A compared to subreddit B), the proportion of each subreddit’s commenters who had posted on the other subreddit was calculated, with the result ranging from 0 (no commenters overlap) to 1 (all commenters overlap). For example, 40/100 (0.4) of subreddit A’s commenters might post on subreddit B, while 40/50 (0.8) of subreddit B’s commenters post on subreddit A. The mean of these two proportions (i.e., 0.6) was calculated to account for differences in the number of commenters on each subreddit. A matrix was then created using all these pairwise comparisons, where each cell within the matrix represented the mean commenter-overlap between each pairing of the EDR-subreddits.
Using this matrix, a weighted and undirected (i.e., associative) network analysis was conducted using the \texttt{qgraph} package [25] for \textit{R} statistical software [26]. Many techniques exist for the purpose of detecting communities (i.e., subreddits with similar commenters) within networks. The \textit{walktrap} algorithm [27] was used for this purpose, as it is recommended for networks with fewer than 1000 nodes [i.e., subreddits; 28], and was observed to detect communities more reliably than other suitable algorithms [cf. 28].

As with individual subreddits, the detected communities might also overlap in terms of commenters (i.e., a commenter might contribute to subreddits from more than one community). As such, for each detected community, commenter lists were compiled and used to calculate the mean commenter-overlap between communities in the same way as detailed above for the pairwise comparisons of subreddits. In the event of more than two communities being detected, the \textit{VennDiagram} \textit{R} package [29] was used to visualise these commenter-overlaps using unscaled Venn diagrams.

\textit{Text-mining}

An existing text-mining approach was used to establish the degree to which each EDR-subreddit was recovery-focused [14]. For each EDR-subreddit, the percentage of its comment-threads that contained at least one recovery term (i.e., ‘recovery’, ‘recover’, ‘recovers’, ‘recovered’, ‘recovering’) was calculated. A higher percentage of threads containing at least one reference to recovery was interpreted as representing a greater recovery-focus, in line with findings that recovery is less frequently mentioned by people in precontemplation or relapse eating disorder stages of change [15]. For example, between May 2015 and January 2018 (inclusive), 10\% of the subreddit \textit{proED}'s threads contained a reference to recovery, compared to 50\% of the subreddit \textit{EatingDisorders}'s threads [14]. As this represented a 40\% difference between the two, \textit{EatingDisorders} was interpreted as having a greater recovery-focus than \textit{proED}.

\textit{Manual review of EDR-subreddits' focuses}

A manual review of each EDR-subreddit’s name and brief description was undertaken to define whether each EDR-subreddit related to: 1) eating disorders in general; 2) specific eating disorder diagnostic categories (e.g., anorexia nervosa, bulimia nervosa, binge eating disorder); or 3) online content associated with eating disorders (e.g., thinspiration). This step reflects the previously detailed distinctions used in the generation of the initial search terms. The definitions were then considered together with the results from the previous two steps (i.e., network analysis with community-detection, and text-mining) to guide the categorisation of the EDR-subreddits. For example, two EDR-subreddits specific to anorexia nervosa would have been categorised differently if they were present in two distinct
communities representing different levels of recovery-focus. All of the researchers reached an agreement on the
categorisation of EDR-subreddits at this stage.

Objective 2: Profiling eating disorder-related commenters based on contributions to ancillary subreddits
For each of the subtypes of EDR-subreddits identified through the previously described methods (Objective 1), a list
was compiled of all the commenters who had contributed within the one-year period to at least one subreddit within
the subtype (excluding ‘AutoModerator’ and ‘[deleted]’ commenters). In order to focus this report on the largest
subtypes of EDR-subreddits, only subtypes with 1000 or more commenters were included in the analyses. The
following analyses were repeated separately for each included subtype of EDR-subreddits.

Using the respective list of commenters for the EDR-subtype, all of the other subreddits to which the commenters had
contributed (i.e., ancillary subreddits) in the one-year period were identified. Any ancillary subreddits to which fewer
than 1% of the EDR-subtype’s commenters had contributed were excluded at this stage. This exclusion was made as
ancillary subreddits with so few EDR-commenters would not have been included in the final steps of the analysis
(detailed below), and therefore represented unnecessary data to extract.

The ancillary subreddits were then ranked separately in descending order of: 1) the number of the EDR-subtype’s
commenters who had contributed to each ancillary subreddit; and 2) the proportion of each ancillary subreddit’s total
commenters (i.e., not only the EDR-subtype’s commenters) that had also commented on at least one subreddit within
the EDR-subtype. For each ancillary subreddit, the mean of these two ranks was then calculated. Due to the large
number (i.e., tens of thousands) of ancillary subreddits associated with each EDR-subtype, and to improve the
interpretability of the results, the mean rank was used to identify the most representative ancillary subreddits for
inclusion in the following analyses. Specifically, the 50 ancillary subreddits with the highest mean rank were included,
resulting in the inclusion of ancillary subreddits that were both large in size (i.e., comprised many commenters), and
included a large proportion of the EDR-subtype’s commenters. This avoided the inclusion of ancillary subreddits that
were very large in terms of the number of commenters, but of which the EDR-subtype’s commenters comprised a very
small proportion (e.g., general subreddits, such as AskReddit). At the same time, this step also avoided the inclusion
of very small subreddits that had very high proportions of the EDR-subtype’s commenters (e.g., subreddits comprising
a few commenters who had all contributed to at least one subreddit within the EDR-subtype). As this threshold (i.e.,
the top 50 subreddits) was used solely to facilitate a clear interpretation of the results, extensions of this study could set different thresholds to explore the communities at varying levels of detail.

A mixed-methods approach was then used to profile the EDR-subtype’s commenters in terms of their main interests, as represented by the thematic focuses of the ancillary subreddits. This approach comprised two techniques, which are described below and were conducted in the order presented: 1) network analyses with community-detection; and 2) manual review of the ancillary subreddits’ topics.

*Network analyses with community-detection*

The network and community-detection analyses were conducted in the same way as detailed for Objective 1. The only difference was that the commenter-overlaps (relating to both the ancillary subreddits and detected communities) were calculated using only the EDR-subtype’s commenters (i.e., excluding commenters on each ancillary subreddit who had not contributed to at least one of the EDR-subtype’s subreddits).

*Manual review of ancillary subreddits’ topics*

A manual review of each ancillary subreddit’s name and brief description was undertaken to describe each subreddit’s general topic. For example, the general topic of the subreddit *loseit* was described as ‘weight-loss’. All ancillary subreddits comprising the detected communities were then reviewed, and a label was produced to represent the general content of each community. For example, a community containing ancillary subreddits that related to eating behaviours and weight-loss was labelled ‘Eating/Body’. In order to ensure transparency at every stage of this process, the name and labels of all included ancillary subreddits are presented in tables in the Results section, along with a summary of how each label for the detected communities was generated.

Results

*Objective 1: Determining subtypes of eating disorder-related subreddits*

The search and inclusion strategy led to the identification of 50 EDR-subreddits, a list of which is provided as Supplementary Material (Supplementary table 2). Following the exclusion of any subreddits with fewer than 500 commenters, nine EDR-subreddits were identified: *BingeEatingDisorder, bulimia, EatingDisorders, eating_disorders, fuckeatingdisorders, MyProAna, proED, ProEDmemes, and thinspo*. In total, 14024 commenters posted on these nine EDR-subreddits. Of these commenters, 97 (0.7%) included the term ‘bot’ within their account name, with these
commenters contributing a mean of 9 comments (SD=16, median=3, minimum=1, maximum=94) to the EDR-subreddits. The network analysis with community-detection corresponding to the nine EDR-subreddits is presented in Figure 1.

Figure 1. Eating disorder-related subreddits network. Letters correspond to eating disorder-related subreddits (A=proED, B=BingeEatingDisorder, C=ProEDmemes, D=thinspo, E=MyProAna, F=fuckeatingdisorders, G=EatingDisorders, H=eating_disorders, I=bulimia). Light-grey circles represent community 1 (“low recovery-focus”), dark-grey circles represent community 2 (“high recovery-focus”). Thickness of lines represents the mean degree of commenter-overlap between each pair of subreddits (thicker line = larger overlap).

Figure 1 shows that the community-detection algorithm identified two communities in the EDR-subreddits network. The text-mining analyses found that community 1 (light-grey circles) comprised the four EDR-subreddits with the lowest percentage of threads mentioning recovery: MyProAna (9.98%), proED (11.72%), ProEDmemes (2.38%), and thinspo (0.23%). In contrast, community 2 (dark-grey circles) comprised the five EDR-subreddits with the highest percentage of threads mentioning recovery: BingeEatingDisorder (19.30%), bulimia (40.05%), eating_disorders (32.14%), EatingDisorders (46.38%), and fuckeatingdisorders (45.58%). These findings supported a conceptualisation
of community 1 comprising ‘low recovery-focus’ EDR-subreddits, and community 2 comprising ‘high recovery-focus’ EDR-subreddits.

Of the 14024 commenters, 9252 (65.97%) only posted on subreddits within the low recovery-focus community, while 4023 (28.69%) only commented on subreddits within the high recovery-focus community. However, 749 commenters (5.34%) posted on subreddits within both communities, indicating relatively little commenter-overlap between the communities.

In addition to the degree of recovery-focus, the EDR-subreddits also differed in terms of whether they concerned eating disorders in general (e.g., EatingDisorders), a specific eating disorder diagnostic category (e.g., BingeEatingDisorder), or online content associated with eating disorders (e.g., thinspo). Each subreddit’s focus was therefore used to categorise the subreddits within each detected community. Consequently, the low recovery-focus community comprised three subtypes of EDR-subreddits: ‘pro-eating disorder’, consisting of proED and ProEDmemes (8166 commenters); ‘thinspiration’, consisting of thinspo (1580 commenters); and ‘pro-anorexia nervosa’, consisting of MyProAna (731 commenters). As with the low recovery-focus community, the high recovery-focus community comprised three subtypes: ‘pro-recovery eating disorder’, consisting of EatingDisorders, eating_disorders and fuckeatingdisorders (1986 commenters); ‘pro-recovery binge eating disorder’, consisting of BingeEatingDisorder (2520 commenters); and ‘pro-recovery bulimia nervosa’, consisting of bulimia (524 commenters).

Objective 2: Profiling eating disorder-related commenters based on contributions to ancillary subreddits

In order to focus this report on the largest subtypes of EDR-subreddits, subtypes with fewer than 1000 commenters were excluded at this stage (i.e., MyProAna and bulimia). Consequently, the analyses regarding Objective 2 were conducted for four subtypes of EDR-subreddits: two that are conceptualised as low recovery-focus (i.e., pro-eating disorder subreddits, thinspiration); and two that are conceptualised as high recovery-focus (i.e., pro-recovery eating disorder subreddits, pro-recovery binge eating disorder). To examine the networks in sufficient detail, and due to the particular importance of the pro-eating disorder communities [e.g., 6], only the analyses for the low recovery-focus subtypes are presented below. The analyses for the high recovery-focus subtypes are provided as Supplementary Material (Supplementary analyses).
Low recovery-focus: Pro-eating disorder

In total, 974 ancillary subreddits had been contributed to by at least 1% of the 8166 commenters associated with the 'pro-eating disorder' subtype (i.e., proED and ProEDmemes). 50 ancillary subreddits were identified on which 5059 (61.95%) of the 8166 pro-eating disorder subtype’s commenters also posted. The network analysis with community-detection is presented in Figure 2, with a summary of the 50 ancillary subreddits presented in Table 1.

Figure 2. Pro-eating disorder network. Letters correspond to ancillary subreddits (Table 1 contains the names and descriptions of the subreddit to which each letter corresponds). Black circles represent community 1 (“Eating/Body”), dark-grey circles represent community 2 (“Mental health”), and light-grey circles represent community 3 (“Women/Appearance/Mixed”). Size of circles represents ancillary subreddit mean rank (larger circle = higher rank), and thickness of lines represents the mean degree of commenter-overlap between each pair of subreddits (thicker line = larger overlap). No lines representing <0.25 mean commenter-overlap are displayed.
Table 1. Names and descriptions of ancillary subreddits on which pro-eating disorder commenters posted

<table>
<thead>
<tr>
<th>Community</th>
<th>Rank</th>
<th>Label</th>
<th>Subreddit name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>A</td>
<td>1200isplenty</td>
<td>1200kcal daily energy intake</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>B</td>
<td>fatlogic</td>
<td>Weight-loss</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>C</td>
<td>fasting</td>
<td>Fasting</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>D</td>
<td>progresspics</td>
<td>Photos of &quot;body transformations&quot;</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>F</td>
<td>BingeEatingDisorder</td>
<td>Binge eating disorder [19.08% threads mention recovery]</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>H</td>
<td>loseit</td>
<td>Weight-loss</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>O</td>
<td>EDFood</td>
<td>Food in the context of eating disorders</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>S</td>
<td>MyProAna</td>
<td>Anorexia nervosa [10.26% threads mention recovery]</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>T</td>
<td>thinspo</td>
<td>Thinspiration [0.24% threads mention recovery]</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>V</td>
<td>intermittentfasting</td>
<td>Fasting</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>W</td>
<td>1200isjerky</td>
<td>1200kcal daily energy intake</td>
</tr>
<tr>
<td>1</td>
<td>46</td>
<td>AT</td>
<td>proEDadults</td>
<td>Eating Disorders</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>I</td>
<td>selfharm</td>
<td>Self-harm</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>J</td>
<td>BPD</td>
<td>Borderline personality disorder</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>X</td>
<td>bipolar</td>
<td>Bipolar disorder</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>AC</td>
<td>morbidquestions</td>
<td>Ask &quot;dark questions&quot;</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>AG</td>
<td>depression</td>
<td>Depression</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>AI</td>
<td>SanctionedSuicide</td>
<td>Suicide</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>AL</td>
<td>Anxiety</td>
<td>Anxiety</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>AM</td>
<td>selfharmpics</td>
<td>Self-harm</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>AN</td>
<td>MadeOfStyrofoam</td>
<td>Self-harm</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>AP</td>
<td>SuicideWatch</td>
<td>Suicide</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>AU</td>
<td>mentalhealth</td>
<td>Mental health</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>E</td>
<td>xxfitness</td>
<td>Female fitness</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>G</td>
<td>femalefashionadvice</td>
<td>Advice on female fashion</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>K</td>
<td>MakeupAddiction</td>
<td>Make-up addiction</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>L</td>
<td>muacirclejerk</td>
<td>Make-up addiction</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>M</td>
<td>fatpeoplestories</td>
<td>Stories about &quot;fat people&quot;</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>N</td>
<td>SkincareAddiction</td>
<td>&quot;Everything skincare&quot;</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>P</td>
<td>TheGirlSurvivalGuide</td>
<td>&quot;A survival guide of &quot;life pro-tips&quot; for the everyday girl&quot;</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>Q</td>
<td>muaqdiscussion</td>
<td>Make-up addiction</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>R</td>
<td>vegan</td>
<td>Veganism</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>U</td>
<td>badwomensanatomy</td>
<td>Women's anatomy</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>Y</td>
<td>awfuleyebrows</td>
<td>Photos of eyebrows that are judged to be &quot;embarrassing, ugly, and downright weird&quot;</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>Z</td>
<td>TrollXChromosomes</td>
<td>Women</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>AA</td>
<td>raisedbynarcissists</td>
<td>&quot;Support group for people raised by (or being raised by) a narcissistic parent&quot;</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>AB</td>
<td>FancyFollicles</td>
<td>Hair</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>AD</td>
<td>AsianBeauty</td>
<td>Beauty brands, cosmetics and skincare from Asia</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>AE</td>
<td>AskWomen</td>
<td>Ask women about any subject</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>AF</td>
<td>childfree</td>
<td>People who do not have or want children</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>AH</td>
<td>antiMLM</td>
<td>Multi-level marketing schemes</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>AJ</td>
<td>piercing</td>
<td>Piercing</td>
</tr>
</tbody>
</table>
As shown in Figure 2, the algorithm detected three communities within the pro-eating disorder network. Community 1 (black circles) was labelled “Eating/Body” as the ancillary subreddits related to restrictive eating (e.g., 1200isplenty, fasting, intermittent fasting), weight loss and/or body transformations (e.g., loseit, fatlogic, progresspics), and/or eating disorders (i.e., BingeEatingDisorder, MyProAna, proEDadults). Community 2 (dark-grey circles) was labelled “Mental health” as the subreddits mainly related to mental health conditions (e.g., depression, Anxiety, bipolar) and/or related issues (e.g., selfharm, SanctionedSuicide, SuicideWatch). Community 3 (light-grey circles) was labelled “Women/Appearance/Mixed” as the subreddits related to women (e.g., xxfitness, TheGirlSurvivalGuide, AskWomen), appearance (e.g., MakeupAddiction, BeautyGuruChatter, amiugly), and/or mixed topics (e.g., vegan, childfree, raisedbynarcissists).

Of the 5059 pro-eating disorder commenters, 67.56% (n=3418) also posted on ancillary subreddits within the Women/Appearance/Mixed community, compared to 61.24% (n=3098) and 35.90% (n=1816) in the Eating/Body and Mental health communities, respectively. Figure 3 presents the commenter overlaps between the three pro-eating disorder communities.

Note. Community 1 = “Eating/Body”; Community 2 = “Mental health”; Community 3 = “Women/Appearance/Mixed”.

<table>
<thead>
<tr>
<th>Community 1</th>
<th>Community 2</th>
<th>Community 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 41 AK amiugly</td>
<td>3 42 AO vegetarian</td>
<td>3 45 AS veggie</td>
</tr>
<tr>
<td>3 43 AQ Youniqueamua</td>
<td>3 44 AR Shoplifting</td>
<td>3 46 AV beautyGuruChatter</td>
</tr>
<tr>
<td>3 44 AR Shoplifting</td>
<td>3 45 AS veggie</td>
<td>3 47 AR thesims</td>
</tr>
<tr>
<td>3 45 AS veggie</td>
<td>3 46 AV beautyGuruChatter</td>
<td>3 48 AS veggie</td>
</tr>
<tr>
<td>3 47 AR thesims</td>
<td>3 48 AV beautyGuruChatter</td>
<td>3 49 AW bulletjournal</td>
</tr>
<tr>
<td>3 49 AW bulletjournal</td>
<td>3 48 AV beautyGuruChatter</td>
<td>3 50 AX actuallesbians</td>
</tr>
</tbody>
</table>

Figure 3. Commenter-overlap between pro-eating disorder communities. Values represent the number of commenters in the pro-eating disorder network (N=5059) who posted in the three communities (represented by the three circles). Values in overlapping areas indicate the number of commenters who posted in two or more communities. Areas of circles are unscaled and do not represent size of communities.

Low recovery-focus: Thinspiration

In total, 3932 ancillary subreddits had been contributed to by at least 1% of the 1580 commenters associated with the ‘thinspiration’ subtype (i.e., thinspo). 50 ancillary subreddits were identified on which 1086 (68.73%) of the 1580 thinspiration subtype’s commenters also posted. The network analysis with community-detection is presented in Figure 4, with a summary of the 50 ancillary subreddits presented in Table 2.
Figure 4. Thinspiration network. Letters correspond to ancillary subreddits (Table 2 contains the names and descriptions of the subreddit to which each letter corresponds). Black circles represent community 1 ("Pornography: 1"), dark-grey circles represent community 2 ("Pornography: young/small"), mid-grey circles represent community 3 ("Pornography: 2"), the light-grey circle represents community 4 ("ProEDmemes"), and the white circle represents community 5 ("proED"). Size of circles represents ancillary subreddit mean rank (larger circle = higher rank), and thickness of lines represents the mean degree of commenter-overlap between each pair of subreddits (thicker line = larger overlap). No lines representing <0.25 mean commenter-overlap are displayed.

Table 2. Names and descriptions of ancillary subreddits on which thinspiration commenters posted

<table>
<thead>
<tr>
<th>Community</th>
<th>Rank</th>
<th>Label</th>
<th>Subreddit name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>B</td>
<td>CuteLittleButts</td>
<td>Pornography</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>C</td>
<td>SexyTummies</td>
<td>Pornography</td>
</tr>
</tbody>
</table>
Notes. Community 1 = “Pornography: 1”; Community 2 = “Pornography: young/small”; Community 3 = “Pornography: 2”; Community 4 = “ProEDmemes”; Community 5 = “proED”. The term ‘Pornography’ is used generally to describe any subreddit featuring material for the ostensibly exclusive purpose of sexual arousal.
As shown in Figure 4, the algorithm detected five communities within the thinspiration network. Community 1 (black circles) was labelled “Pornography: 1” as the ancillary subreddits were all pornographic in nature. Community 2 (dark-grey circles) was labelled “Pornography: young/small” as it mainly comprised pornographic subreddits that explicitly referred to women being young (e.g., LegalTeens, 18_19) and/or small (e.g., xsallgirls, TinyTits, dirtysmall). Community 3 (mid-grey circles) was labelled “Pornography: 2” as all the subreddits were pornographic. Community 4 (light-grey circle) was labelled “ProEDmemes” as it comprised only one subreddit, ProEDmemes. Similarly, community 5 (white circle) was labelled “proED”, as it consisted of proED only.

Of the 1086 thinspiration commenters, 67.59% (n=734) also posted on ancillary subreddits within the Pornography: 1 community, compared to 59.85% (n=650), 46.13% (n=501), 20.99% (n=228), and 10.87% (n=118) in the Pornography: young/small, Pornography: 2, proED, and ProEDmemes communities, respectively. Figure 5 presents the commenter-overlaps between the five thinspiration communities.

Figure 5. Commenter-overlap between thinspiration communities. Values represent the number of commenters in the thinspiration network (N=1086) who posted in the five communities (represented by the five ovals). Values in overlapping areas indicate the number of commenters who posted in two or more communities. Areas of ovals are unscaled and do not represent size of communities.
As shown in Figure 4, a clear distinction was observed between the pornography communities (i.e., Pornography: 1, Pornography: 2, Pornography: young/small) and pro-eating disorder communities (i.e., ProEDmemes, proED). As such, the overlap between these two groups of communities was calculated. Of the 1086 commenters in the thinspiration commenter network, 849 (78.18%) only posted on ancillary subreddits within the pornography communities, while 181 (16.67%) only commented on subreddits within the pro-eating disorder communities. However, 56 commenters (5.16%) posted on subreddits within both groups of communities, indicating a small commenter-overlap between these groups.

Discussion
Using the example of EDR-subreddits, the current study demonstrated a methodology that addressed two objectives: 1) determine subtypes of forums related to a similar mental health issue; and 2) elucidate the characteristics (i.e., shared interests) of the subtypes’ commenters by identifying other forums to which they contribute (i.e., ancillary subreddits), and investigating the commenter-overlap between these subreddits. These two objectives were achieved using mixed-methods approaches, comprising techniques that included network analysis with community-detection, text-mining, and manual review of the forums’ topics. Following the identification of six subtypes of EDR-subreddits, the report focused on two specific subtypes – pro-eating disorder, and thinspiration. The pro-eating disorder commenters also contributed to subreddits relating to the body and eating, mental health, and women, appearance and mixed topics. Regarding the thinspiration subtype, 78% of the commenters also contributed to pornographic subreddits, while 17% also commented on pro-eating disorder subreddits.

Principal results
Objective 1: Determining subtypes of eating disorder-related subreddits
Concerning the first objective, through the use of network analyses with community-detection, and a previously detailed text-mining technique [14], two communities of EDR-subreddits were identified that differed in terms of their degree of recovery-focus (i.e., low recovery-focus and high recovery-focus). The detection of these two communities, and the relatively small (5%) commenter-overlap between them, is in line with similar findings relating to pro-eating disorder and pro-recovery communities on Twitter [13]. Furthermore, previous analyses of online data from an EDR-
online forum indicated that commenters in the least recovery-focused eating disorder stages of change (i.e., precontemplation, relapse) used recovery words less frequently than commenters in more recovery-focused stages of change [15]. The current findings offer support for this, as recovery was indeed mentioned less frequently in ostensibly pro-eating disorder subreddits (i.e., MyProAna, proED, ProEDmemes, thinspo) than more recovery-focused subreddits (i.e., BingeEatingDisorder, bulimia, eating_disorders, EatingDisorders, fuckeatingdisorders). While the text-mining approach used textual data to assess the frequency of words’ occurrence in comment-threads, the network analyses with community-detection utilised behavioural data (i.e., data about the subreddits to which commenters contributed). As the results of the text-mining approach (i.e., degree of recovery-focus) clearly align with the detected communities (i.e., the communities appear to differ on the basis of recovery-focus), a strength of this mixed-methods approach is that the two distinct techniques appear to provide a degree of convergent validity to each other.

Objective 2: Profiling eating disorder-related commenters based on contributions to ancillary subreddits

Regarding the second objective, the topics in which commenters on pro-eating disorder subreddits were interested are in line with other research [10, 11, 16]. Specifically, commenters on pro-eating disorder subreddits were also found, unsurprisingly, to be interested in the body, eating, mental health and appearance. As several identified subreddits were specific to women (e.g., femalefashionadvice, TheGirlSurvivalGuide), the results also support previous findings that suggest women are more likely to engage with pro-eating disorder-related online content [4, 9, 30].

In contrast to the pro-eating disorder results, the findings concerning thinspiration commenters were of great surprise. Namely, a clear majority of thinspiration commenters (78%) had also contributed to pornographic subreddits. Furthermore, a specific group of commenters contributed to pornographic subreddits that had names suggesting that women were young (including terms such as ‘legal’ or ‘18’) and/or small (including terms such as ‘tiny’, ‘petite’ and ‘small’). This finding is in line with previous research, which concluded that thinspiration images were typically sexually suggestive [30-32]. In fact, one study [32] actually identified pornographic images in their search for thinspiration (and fitspiration) content, although these were excluded from the subsequent analyses. As indicated in the current study, the distinction between thinspiration and pornographic material is not clear. As such, it is important that researchers do not exclude specific material (e.g., pornographic images) from future analyses, as this might lead to a sanitised understanding of thinspiration content. The current study therefore highlights an issue of potentially great concern.

Specifically, as thinspiration content typically comprises photos of extremely thin women [e.g., 30], the people
submitting this content might not be fully aware of how their content is subsequently used. Speculatively, this lack of a complete understanding might lead to people unintentionally entering into vulnerable situations, and would therefore clearly warrant further research.

While the current findings might suggest that people engage with thinspiration for pornographic reasons, it is also possible that engagement with pornography and eating disorder symptomatology are related. For example, pornographic content might be viewed for the purpose of body comparison. However, only two studies appear to have explicitly investigated the relationship between pornography use and eating disorder symptomatology, both of which recruited exclusively male samples [33, 34]. Given the apparent lack of research investigating this relationship in a female sample, and due to pro-eating disorder commenters typically being women [e.g., 9], this topic also represents an important avenue for future research.

Overall, a strength of the current findings is that they complement previous studies. While there are limitations (detailed below) to the current mixed-methods approach, its techniques can compensate for the methodological limitations of the previous studies. For example, survey-based studies concerning users of online forums are unlikely to have representative samples. As the current approach used data concerning all the commenters on public forums, it is not subject to this limitation. As a result, by consolidating the findings generated from these distinct methodological approaches, greater confidence can be had that results do not simply represent an artefact of one particular technique [cf. 35].

Limitations

These current findings must be considered in relation to the limitations of the data and methodology. First, it is important to note that the Reddit data are unsolicited. While this represents a strength of the data (e.g., the data are not liable to demand characteristics), this also results in a significant amount of noise in the data. Steps were taken to reduce this noise, such as only including 50 ancillary subreddits with a large number and proportion of EDR-commenters in the network analyses. This approach identified the most representative subreddits by excluding very small subreddits (many of which had a high proportion but small number of EDR-commenters) and very large subreddits (many of which had a large number but low proportion of EDR-commenters). However, the effects of other sources of noise in the data are more difficult to mitigate. For example, the same person might have more than one
Reddit username, which might lead to an underestimation of commenter-overlap. Additionally, ‘bots’ (i.e., automated software) exist that comment widely on subreddits, and which might contribute to an overestimation of commenter-overlap. While strategies exist to identify bots [e.g., 16, 36], these might also exclude actual users. For this reason, bots (except ‘AutoModerator’, a generic Reddit bot) were not excluded, in order to adopt a conservative approach to our analyses. To minimise the effects of these sources of noise, the conclusions are based on communities of subreddits, rather than individual subreddits. While not necessarily a limitation of the present study, caution should be exercised in generalising these findings to people who have read, but not commented on, the online content. Exploring communities based on the content that users read would be important but ethically problematic, as this would likely require access to data that is not publicly available. Despite not being able to generalise to readers of forums, Aardoom et al. [37] found that most survey respondents (87.2%) recruited from a pro-recovery EDR-forum posted content, while the remainder only read content. Furthermore, algorithms exist that recommend subreddits in which users might be interested [cf. 38]. Therefore, by focusing on commenters, users can be ensured to be actively engaged with the content, rather than being passively exposed to it. Similarly, private subreddits were not included in the present study for primarily ethical reasons. However, a private recovery-focused EDR-subreddit is advertised on a number of the public EDR-subreddits. As such, the results regarding the high recovery-focus subreddits (detailed in the Supplementary Material) might differ were the private subreddit’s commenters included. However, the private recovery-focused subreddit is less likely to influence the low recovery-focus analyses presented above.

Implications
Regarding the approach used for the first objective, the current findings have implications for how mental health-related online communities should be conceptualised and investigated in future research. Specifically, when comparing multiple communities, the degree of user-overlap between these should be acknowledged. For example, a previous study compared the frequency of fitness tracker mentions between three EDR-subreddits [i.e., proED, fuckeatingdisorders, EatingDisorders; 14]. By considering how these forums overlap, a clearer and more detailed interpretation of the characteristics of these forums’ commenters could be achieved.

Concerning the second objective’s methodological approach, the findings have implications for future research, and the design of online psychological interventions for mental health issues. Regarding future research, the approach presented here is entirely reproducible and can be used to explore similar questions in other groups of commenters.
of particular theoretical interest (e.g., relating to other mental health conditions). The methodology could also be easily extended to explore longitudinal – and, therefore, causal – patterns of commenting. The approach is also useful for hypothesis-generation and identifying new avenues of research. As detailed above, the current study generated a surprising finding, in that over three quarters of thinspiration commenters also commented on pornography. As this relationship had not been identified before, it is a clear indicator of how this approach can be used to identify areas that require greater research attention. Concerning implications for psychological interventions, the current approach can identify other topics that are of interest to people commenting on mental health-related online discussion forums. For example, pro-eating disorder commenters were observed to be also interested in topics such as body, eating, mental health and appearance. Consequently, these topics confirm the importance of existing EDR-intervention focuses [e.g., eating, and body shape and weight concerns; cf. 39]. These findings could also be used to more accurately tailor interventions to the target population’s characteristics (e.g., topics of interest), potentially increasing adherence to the programmes [cf. 3]. Another implication for psychological interventions is that this approach can identify other forums in which there is a high activity of mental health-related forums’ commenters. In the case of pro-eating disorder commenters, they were also observed to be active in subreddits including 1200isplenty, loseit, and progresspics. As some mental health-related communities (e.g., pro-eating disorder) might be unlikely to promote psychological interventions, the current approach could be utilised to identify the communities in which these users also tend to post. As a result, these communities could be approached to provide an alternative way in which to reach these people, and to target prevention-focused interventions.

Conclusions

In summary, this study presents a reproducible and primarily data-driven methodology that can be used to: 1) identify subtypes of mental health-related forums; and 2) identify the interests of the commenters who post on the forums comprising these subtypes. This offers a powerful technique for hypothesis-generation, and informing strategies for psychological intervention. Employing different methodologies to explore the same research question is vital to ensure that findings are not solely a result of a particular methodological design [35]. This approach therefore offers one way in which to triangulate methodologically the findings obtained through previous – e.g., – survey-based research, and consequently contributes to a more robust evidence base.
Acknowledgements

The authors would like to extend their gratitude to Miss Helen A. Fay and Dr Lara Ayala-Nunes for their feedback on aspects of a draft of this manuscript. Duncan McCaig is the recipient of a doctoral studentship from WMG, University of Warwick, and Coventry and Warwickshire NHS Partnership Trust. Cynthia S.Q. Siew is supported by an Overseas Postdoctoral Fellowship from the National University of Singapore. The authors also extend their gratitude to the peer-reviewers of this article, whose feedback was both considered and constructive in nature.

Conflicts of interest

None.

Abbreviations

EDR: eating disorder-related
References


5. Gerrard, Y. Beyond the hashtag: Circumventing content moderation on social media. New Media & Society; 2018. DOI: 10.1177/1461444818776611.