

Estimating Determinants of Attrition in Online Eating Disorder Community: An Instrumental Variables Approach

Tao Wang, Emmanouil Mentzakis, Markus Brede, Antonella Ianni

Submitted to: Journal of Medical Internet Research
on: May 06, 2018

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript 5



Estimating Determinants of Attrition in Online Eating Disorder Community: An Instrumental Variables Approach

Tao Wang^{1,2}, MSc; Emmanouil Mentzakis³, PhD; Markus Brede⁴, PhD; Antonella Ianni³, PhD

¹ University of Southampton, Department of Economics, University of Southampton, Southampton, United Kingdom.

² The Alan Turing Institute, , London, United Kingdom.

³ University of Southampton, Department of Economics, University of Southampton, Southampton, United Kingdom.

⁴ University of Southampton, Department of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom.

Corresponding Author:

Tao Wang, MSc

University of Southampton

Department of Economics

University of Southampton

University of Southampton

University Road

Southampton

United Kingdom

Phone: 44 7864373080

Email: t.wang@soton.ac.uk

Abstract

Background: The use of social media as key health-information source has increased steadily among people affected by eating disorders. Intensive research has examined characteristics of individuals engaging in online communities, while little is known about discontinuation of engagement and the phenomenon of participants dropping out of these communities.

Objective: This study aims to investigate characteristics of dropout behaviors among eating disordered individuals on Twitter and to estimate the causal effects of personal emotions and social networks on dropout behaviors.

Methods: Using a snowball sampling method, we collected a set of individuals who self-identified with eating disorders in their Twitter profile descriptions, as well as their tweets and social networks, leading to 241,243,043 tweets from 208,063 users. Individuals' emotions are measured from their language use in tweets using an automatic sentiment analysis tool, and network centralities are measured from users' following networks. Dropout statuses of users are observed in a follow-up period 1.5 years later (from Feb. 11, 2016 to Aug. 17, 2017). Linear and survival regression instrumental variables models are used to estimate the effects of emotions and network centrality on dropout behaviors. An individual's attributes are instrumented with the attributes of the individual's followees (i.e., people who are followed by the individual).

Results: Eating disordered users have relatively short periods of activity on Twitter, with one half of our sample dropping out at 6 months after account creation. Active users show more negative emotions and higher network centralities than dropped-out users. Active users tend to connect to other active users, while dropped-out users tend to cluster together. Estimation results suggest that users' emotions and network centralities have causal effects on their dropout behaviors on Twitter. More specifically, users with positive emotions are more likely to drop out and have shorter-lasting periods of activity online than users with negative emotions, while central users in a social network have longer-lasting participation than peripheral users. Findings on users' tweeting interests further show that users who attempt to recover from eating disorders are more likely to drop out than those who promote eating disorders as a lifestyle choice.

Conclusions: Presence in online communities is strongly determined by individual's emotions and social networks, suggesting that studies analyzing and trying to draw condition and population characteristics through online health communities are likely to be biased. Future research needs to examine in more detail the links between individual characteristics and participation patterns if better understanding of the entire population is to be achieved. At the same time, such attrition dynamics need to be acknowledged and controlled for when designing online interventions so as to accurately capture their intended populations.

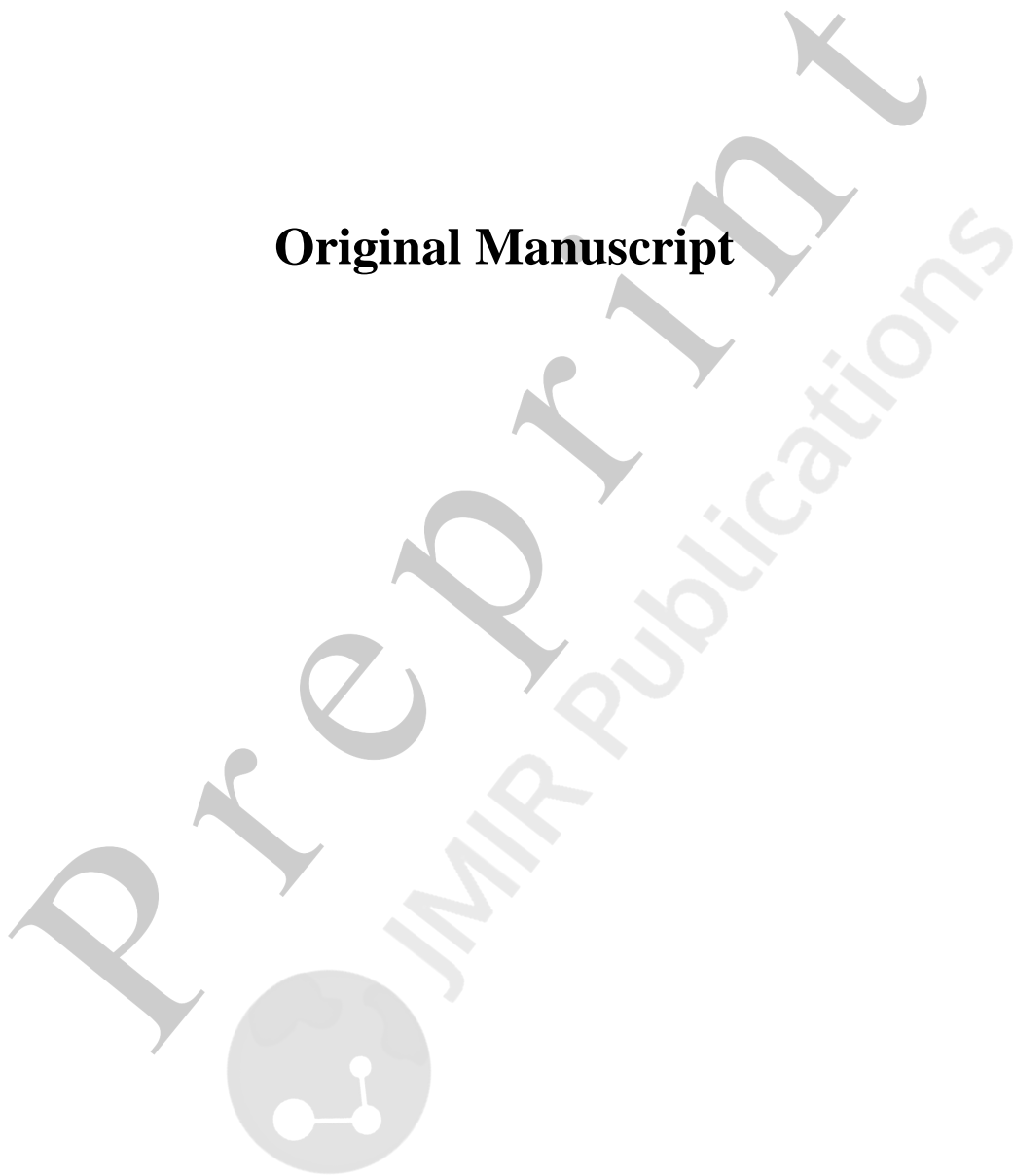
(JMIR Preprints 06/05/2018:10942) DOI: <https://doi.org/10.2196/preprints.10942>

Preprint Settings

- 1) Allow access to Preprints when not published (under review or accepted) for
 - (a) Open peer-review purposes.
 - (b) Logged-in Users only.
 - ✓ (c) **Anybody, anytime.**
 - (d) Nobody.
- 3) When a final version is published in a JMIR journal
 - (a) Allow download.
 - ✓ (b) **Show abstract only.**
- 4) Allow access to Preprints when no longer under consideration
 - (a) Logged-in Users only.
 - ✓ (b) **Anybody, anytime.**
 - (c) Nobody.
- 2) Accepted Setting
 - (a) Logged-in Users only.
 - ✓ (b) **Anybody, anytime.**
 - (c) Same as b), plus submit as ahead-of-print version to PubMed.
 - (d) Nobody.



Original Manuscript



1 Original Paper

2 **Estimating Determinants of Attrition in Eating Disorder** 3 **Communities on Twitter: An Instrumental Variables** 4 **Approach**

Tao Wang^{1,2}, MSc*; Emmanouil Mentzakis¹, PhD; Markus Brede³, PhD; Antonella Ianni¹, PhD;

¹Department of Economics, University of Southampton, UK

²The Alan Turing Institute, UK

³Department of Electronics and Computer Science, University of Southampton, UK

Corresponding Author:

Tao Wang,
MSc
Department of
Economics
University of
Southampton
Southampton,
SO17 1BJ
United
Kingdom
Phone: +44 7864373080
Email: t.wang@soton.ac.uk

Abstract

5 **Background:** The use of social media as key health-information source has increased steadily among people
6 affected by eating
7 disorders. Research has examined characteristics of individuals engaging in online communities, while little is
8 known about
9 discontinuation of engagement and the phenomenon of participants dropping out of these communities.
10 **Objective:** This study aims to investigate characteristics of dropout behaviors among eating disordered individuals
11 on Twitter
12 and to estimate the causal effects of personal emotions and social networks on dropout behaviors.
13 **Methods:** Using a snowball sampling method, we collected a set of individuals who self-identified with eating
14 disorders in
15 their Twitter profile descriptions, as well as their tweets and social networks, leading to 241,243,043 tweets from
16 208,063 users.
17 Individuals' emotions are measured from their language use in tweets using an automatic sentiment analysis tool,
18 and network
19 centralities are measured from users' following networks. Dropout statuses of users are observed in a follow-up
20 period 1.5 years
21 later (from Feb. 11, 2016 to Aug. 17, 2017). Linear and survival regression instrumental variables models are used

to estimate

17 the effects of emotions and network centrality on dropout behaviors. The average levels of attributes among an individual's

18 followees (i.e., people who are followed by the individual) are used as instruments for the individual's attributes.

19 **Results:** Eating disordered users have relatively short periods of activity on Twitter, with one half of our sample dropping out at 6

20 months after account creation. Active users show more negative emotions and higher network centralities than dropped-out users.

21 Active users tend to connect to other active users, while dropped-out users tend to cluster together. Estimation results suggest

22 that users' emotions and network centralities have causal effects on their dropout behaviors on Twitter. More specifically, users

23 with positive emotions are more likely to drop out and have shorter-lasting periods of activity online than users with negative

24 emotions, while central users in a social network have longer-lasting participation than peripheral users. Findings on users'

25 tweeting interests further show that users who attempt to recover from eating disorders are more likely to drop out than those

26 who promote eating disorders as a lifestyle choice.

27 **Conclusions:** Presence in online communities is strongly determined by individual's emotions and social networks, suggesting

28 that studies analyzing and trying to draw condition and population characteristics through online health communities are likely

29 to be biased. Future research needs to examine in more detail the links between individual characteristics and participation

30 patterns if better understanding of the entire population is to be achieved. At the same time, such attrition dynamics need to be

31 acknowledged and controlled for when designing online interventions so as to accurately capture their intended populations.

KEYWORDS

Medical Informatics; Eating Disorders; Social Media; Attrition; Emotions; Social Network; Causality; Instrumental Variables

Introduction

Eating disorders (ED), such as anorexia and bulimia, are complex mental disorders defined by extreme obsessions with body weight or shape, and unusual eating behaviors [1]. These diseases have the highest mortality rate of any mental illness [2], intractable co-morbidities [3] and worldwide prevalence [4], having become a major public health concern. Although a variety of treatment options have emerged over recent years [5], populations affected by ED are often hard to reach through traditional healthcare services. This is mainly due to fear of stigma or a feeling of shame; many sufferers conceal their ED symptoms and never seek professional treatment or support [6, 7]. To keep struggles with illnesses private, people often seek health-related information and support through online peer-to-peer communities, particularly via social media sites like Twitter and Facebook. Participation in online communities is common in ED populations [8] and has been suggested as a screening factor for ED [3]. This provides an opportunity for healthcare professionals to deliver health support to these hard-to-reach populations through online communities [9–13]. Moreover, as online communities present a relatively anonymous environment for individuals to naturally self-disclose and socialize [14], user-generated data online provides a large amount of records about individuals' concerns, thoughts, emotions and social interactions [15–17], which may complement traditional data sources (e.g., surveys and interviews) in understanding risk factors of ED. Hence, growing research has focused on characterizing individuals' behavioral patterns in online communities [15–19], so as to better understand ED and promote population-level well-being.

One notable characteristic of online ED communities is their participants having widely different stances on ED [8, 20, 21]. Some communities encourage members to discuss their struggles with ED, share treatment options and offer support towards recovery from ED, so called *pro-recovery* communities [20–22]. There are also many anti-recovery or pro-ED communities in which members often deny ED being a disorder and instead promote ED as a healthy lifestyle choice [8, 23]. These pro-ED communities can negatively affect health and quality of life among people with and without ED, through reinforcing an individual's identity around ED [24], promoting thin ideals [25], and disseminating harmful practices for weight loss [8]. Recent studies have shown that individuals' language use online strongly indicate their pro-ED or pro-recovery stances [15, 17, 20], as well as emotions of depression, helplessness and anxiety that reflect their mental disorders [16]. Other studies have also examined

interactions between pro-ED and pro-recovery communities on Flickr [21], anorexia-related misinformation [18], sentiments of comments on ED-related videos on YouTube [26], characteristics of removed pro-ED content [27] and lexical variation of pro-ED tags on Instagram [19, 28]. Yet, prior studies have largely focused on examining how people engage in and maintain an online ED community, while little is known about how people drop out of such a community. As a dynamic process, people who join and actively engage in a community at earlier stages can have less participation and leave the community at later stages. Understanding the attrition processes of online communities can enhance our knowledge of the dynamics in these communities.

Studying the attrition process of an online community can also have practical implications for disease prevention and health interventions. Given the ease of accessibility of social media for many individuals (e.g., via mobile devices), increasing attention has focused on using online communities to deliver health interventions [9–13, 29, 30]. One of the most popular approaches is to deliver health lessons and behavior-changing instructions via online communities [9–13, 29]. Although pilot studies based on small samples have demonstrated the effectiveness of these approaches in reducing body dissatisfaction and disordered eating [12, 13], evidence from interventions for a variety of health behaviors (e.g., smoking, diet, exercise and sexual health) suggests that attrition (i.e., participant loss) is one of the most common challenges in online interventions [10, 29]. This is known as the "law of attrition" of online interventions [31]. A recent study has shown a high attrition rate in an online intervention for ED [32],

though this intervention is delivered via a purposely designed website rather than a general social media site. Thus, an important goal in conducting successful interventions via online communities is to improve members' retention, as members who remain longer are more likely to receive these interventions and have more opportunities to promote a target behavior change. To achieve

this goal, a critical first step is to understand what factors influence members' retention in an online community.

Prior studies have shown that people's decisions of retention or dropout in online communities are associated with a variety of factors [33, 34], including personality traits (e.g., shyness and the Big-Five traits) [35, 36], interests [37], recognition in a community [38–41] and support from others [42, 43]. However, such an association is not adequate to conclude the presence of

79 a causal relationship [44, 45] between an individual's attributes and her/his online participation. This is because an association



80 can arise from non-causal relationships. For example, most prior studies focus on the use of self-reported surveys and rely on
81 participants' reports of their own personality, concerns and behaviors [35, 39, 46]. This can introduce considerable retrospective
82 bias and measurement errors, leading to a coincidental association between two unrelated variables, particularly in small
samples.

83 Even if variables are measured rather than self-reported [43, 47], participation in an online community is inherently self-selected
84 (e.g., sharing common interests) and members can drop out for many different reasons (e.g., effect of an online or offline event).
85 Thus, unobservable factors (i.e., confounding variables) may affect both a main predictor and participation outcomes, causing a
86 spurious association. Moreover, in some cases reverse causality can lead to an association, e.g., prior studies suggest that
feelings

87 of social isolation are linked to frequent social media use [35, 48] whereas recent studies indicate that social media use is linked
88 to increased feelings of social isolation [49]. Technically speaking, the issues of measurement errors, confounding variables and
89 reverse causality can cause endogeneity which refers to an explanatory variable of interest being correlated with the error term
90 in a regression model [44]. In these cases, traditional methods such as ordinary least squares (OLS) give biased and inconsistent
91 estimates of the effect of interest. It is therefore not surprising that mixed results exist in prior studies, e.g., a positive association
92 between individuals' expertise and online participation was found in [46] while a negative association was found in [38].

93 This study aims to estimate determinants of dropout in an online ED community, while addressing the endogeneity issues
94 by using an instrumental variables (IV) approach [44]. Specifically, we analyze tweeting activities for a large set of individuals
95 who self-identified with ED on Twitter over 1.5 years and identify the presence of dropout if a user ceased to post tweets in the
96 observation period. We explore determinants of a user's dropout based on incentive theory [34, 50] which argues that people's
97 engagement in an activity can be driven by (i) intrinsic motivation which refers to doing something because it is interesting
98 or enjoyable, and (ii) extrinsic motivation which refers to doing something because it earns an external reward. We focus on
99 intrinsic motivation captured by personal emotions and extrinsic motivation captured by sociometric status in an online peer-to-
100 peer community. Rather than using self-reports [35, 39, 46], we measure users' emotions based on their emotional expressions in
101 tweets using sentiment analysis techniques [51] and quantify users' sociometric statuses by network centrality [52] in the social
102 network of an ED community on Twitter. Based on these measured variables, IV estimators both for the decision to drop out and
103 for the time to dropout are implemented to achieve consistent estimates of the effects of personal emotions and network
centrality

104 on dropout in an online ED community. To better understand the estimation results, we further examine heterogeneity in
tweeting

105 interests (i.e., topics discussed in tweets) of users with differing levels of characteristics (e.g., emotions) and dropout outcomes,
106 respectively. To our knowledge, this study is the first to systematically characterize the determinants of dropout behaviors in
107 online ED communities. Three research questions are examined: (i) what are the general characteristics of the attrition process
108 in an online ED community, (ii) how do intrinsic and extrinsic factors affect the decision of an individual to drop out of the
109 community, (iii) how do these factors affect the duration of time until the occurrence of dropout?

110 *Methods*

112 **Data Collection**

113 Our data is collected from Twitter, a microblogging platform that allows millions of users to self-disclose and socialize. As
114 many social media platforms like Facebook and Instagram have taken moderation actions to counteract pro-ED content and
115 user accounts [28], Twitter has not yet enforced actions to limit such content [53]. This makes Twitter a unique platform to
116 study the attrition process naturally happening in an online ED community and allows us to examine individuals' behaviors in
117 a non-reactive way. Our study protocol was approved by the Ethics Committee at the University of Southampton. All data
118 used in our study is *public* information on Twitter and available through the Twitter APIs (application programming interfaces).
119 No personally identifiable information is used in this study. Our data collection process includes three phases.

- 120 • First, we collect a set of individuals who self-identified with ED on Twitter using a snowball sampling approach. Specif-
121 ically, we track the public tweet stream using “eating disorder”, “anorexia”, “bulimia” and “EDNOS” from Jan. 8 to 15,
122 2016. This results in 1,169 tweets that mention ED. From the authors of these tweets, we identify 33 users who self-
123 reported both ED-related keywords (e.g., “eating disorder”, “anorexia” and “bulimia”) and personal bio-information (e.g.,

124 body weight and height) in her profile descriptions (i.e., a sequence of user-generated text describing their accounts be-
125 low profile images). Starting from these seed users, we expand the user set using snowball sampling through their social
126 networks of followees/followers. At each sampling stage, we filter out non-English speaking accounts and finally obtain



127 3,380 unique ED users who self-report ED-related keywords and bio-information in their profile descriptions. Note that
128 our focus in this work is studying individuals who are affected by ED rather than those who are related to ED. The
inclusion

129 of bio-information in user sampling allows us to filter out ED-related therapists, institutes or organizations, as these users
130 often display ED-related keywords but do not show bio-information in their Twitter profile descriptions. Details about the
131 data collection of ED users can be found in our prior work [54].

132 • Then, we collect all friends (including followees and followers) of each ED user, leading to a large social network
consisting

133 of 208,063 users. For each user, we retrieve up to 3,200 (the limit returned from Twitter APIs) of their most recent tweets
134 and obtain 241,243,043 tweets in total. The data collection process finished on Feb. 11, 2016.

135 • Finally, we open a follow-up observation period for all users on Aug. 17, 2017 to obtain measurements on users' activities
136 online. In the second observation, we only collect users' profile information which includes users' last posted statuses.

137 To verify the quality of our collected sample, two members of the research team classified a random sample of 1,000 users
on

138 whether they were likely to be a true ED user based on their posted tweets, images and friends' profiles. Users are classified as
139 "disordered" if they frequently and intensively post their body weights, details of their dietary regimen (e.g., calories), struggles
140 with eating (e.g., "I want to eat but cannot"), pictures of themselves, self-reports of being disordered or in recovery in tweets
141 and follow ED-related friends (e.g., user profiles with ED-related keywords). The process revealed a 95.2% match between the
142 identified ED individuals in the data collection stage and those classified as ED during inspection. Although it is impossible to
143 diagnose individuals' disorders based on their online behaviors, this inspection provides a strong indication that the collected
144 users are likely to be affected by ED rather than those who merely talk about ED online. See [54] for details of data validation.

145 Estimation Framework

146 Two different models are specified to estimate the effects of emotions and network centrality on dropout. First, we specify a
linear

147 probability model on the whole sample to estimate the effects of individuals' characteristics observed in the first-observation
148 period on the probability of dropping-out in the second-observation period. Second, we estimate survival models to explore
149 the effects of individuals' characteristics observed in the first observation on the time to dropout in the second observation (i.e.,
150 the duration from our first observation to the dropout in our second observation). However, like all social media studies, only
151 a limited number of individuals' characteristics are available for our estimations and these are mostly observed through user-
152 generated data online. This leads to confounding variable bias, since unobservable factors can be correlated with both the main
153 explanatory variables (i.e., emotions and network centrality) and dropout outcomes. For example, undergoing hospital treatment
154 can simultaneously affect a person's emotional state and the use of social media. Further, prior studies have shown that social
155 media use is associated with increased depression [55], social anxiety [49] and body dissatisfaction [56, 57], implying an effect
156 of online participation on individuals' emotions (i.e., reverse causality). Both confounding variables and reverse causality result
157 in biased estimates of the effects of emotions and network centrality on dropout. This problem can be addressed by using a
158 randomized controlled trial in which emotions or network centralities are randomly assigned to users by researchers [58]. Such
159 a trial, however, is not feasible, due to ethical and practical limitations [59].

160 Here, we utilize an alternative approach for estimating the effects of interest that is based on instrumental variables (IV)
161 regression, an econometric technique to infer causal relations from observational data [44]. This technique has been applied
162 to a variety of contexts, from identifying the causal effect of education on earning [60], the effect of a health treatment [61],
163 to estimating social contagion effects on both online [59] and offline behaviors [62]. Formally, consider a model $Y = \beta_1 X_1 +$
164 $\beta_2 X_2 + u$, where X_1 is endogenous, X_2 is exogenous, u is a random error term and β_s are effects to be estimated. IV
methodology

165 uses an instrument Z (which is (i) not contained in the explanatory equation, (ii) correlated with X_1 , i.e., $cov(Z, X_1) \neq 0$, and
166 (iii) uncorrelated with u , i.e., $cov(Z, u) = 0$, conditional on the other covariates such as X_2) and runs a first stage reduced-form
167 regression $X_1 = \gamma_1 Z + \gamma_2 X_2 + v$, where v is a random error. The causal effect of X_1 on Y is then given in a second
stage

168 regression $Y = \beta_1 X_1^{-1} + \beta_2 X_2 + u$, where X_1^{-1} is the predicted values of X_1 from the first stage. See [44] for more

details.

¹⁶⁹ **Measures**

¹⁷⁰ A number of variables are needed for estimations. All independent variables and IV are measured in the first-observation period
¹⁷¹ (unless otherwise stated), while dependent variables are measured in the second-observation period.



172 **Dropout Outcomes as Dependent Variables**

173 Following previous studies [42, 43], we identify the presence of dropout if a user ceases to post tweets. Specifically, in the linear
174 probability models, we encode the dropout status of a user as 0 (denoting *non-dropout*) if the user has updated posts in our
second

175 observation, and 1 (denoting *dropout*) otherwise.

176 In the survival models, each user has a two-variable outcome: (i) a censoring variable denoting whether the event of dropout
177 occurs, and (ii) a variable of survival time denoting the duration of time until the occurrence of dropout. We censor the
occurrence

178 of a “dropout event” in two ways. First, users are said to drop out if they have not posted tweets for more than a fixed threshold
179 interval π before our second observation (so called *identical-interval censoring*). As people use social media platforms with
180 different activity levels, e.g., some users post every several hours while other users only post once every couple of days, our
second

181 censoring method further accounts for personalized posting activities of individuals (called *personalized-interval censoring*). In
182 this method, users are said to drop out if they have not posted tweets for more than a variate threshold interval $\lambda\pi +$
 $(1 - \lambda)I_i$

183 before our second observation, where π is a fixed threshold, I_i is the average posting interval of individual i in our first
observation

184 period, and λ is a tunable parameter to control the effects of individual activities. We tune the parameters by maximizing the
185 agreement between the estimated dropout states based on users’ activities in our first observation and the observed states in our
186 second observation. See *Supplementary Information (SI)* for details. For users who are censored as dropped-out, we set their
187 survival times as the durations from our first observation to their last postings in our second observation. For those who are
188 censored as non-dropped-out, we set their survival times as the whole time period between our two observations.

189 **Emotions and Network Centrality as Main Explanatory Variables**

190 Individuals’ emotions are measured through their language used in tweets. There is a variety of sentiment analysis algorithms
191 to measure emotional expressions in texts [51, 63]. In this study, we use SentiStrength [51] as (i) it has been used to measure
192 the emotional content in online ED communities and shown good inter-rater reliability [26]; (ii) it is designed for short informal
193 texts with abbreviations and slang, and thus suitable to process tweets [51]. After removing mention marks, hashtags and
194 URLs, each tweet is assigned a scaled value in $[-4, 4]$ by SentiStrength, where negative/positive scores indicate the strength of
195 negative/positive emotions respectively, and 0 denotes neutral emotions. We quantify a user’s emotional state by the average
196 score of all tweets posted by the user. All re-tweets are excluded, as re-tweets reflect more the emotions of their original authors
197 than those of their re-tweeters. For robust results from the language processing algorithms, we only consider users who have
198 more than 10 tweets and post more than 50 words.

199 Network centrality measures the importance of a person in a social network; people well-recognized by their peers often
have

200 high centralities in a group [52]. To measure a user’s centrality in the ED community, we build a who-follows-whom network
201 among ED users and their friends, where a directed edge runs from node A representing user A to node B representing user B
202 if A follows B on Twitter. While there are various measures of network centrality, we focus on coreness centrality [64] as it
203 has been shown to outperform other measures such as degree and betweenness centrality [52] in detecting influential nodes in
204 complex networks [65] and cascades of users leaving an online community [66, 67]. We measure the sociometric status of a user
205 in the ED community by the in-coreness centrality [68] of a node in the generated network using the package igraph 0.7.0 [69].

206 **Aggregated Emotions and Network Centrality of Friends as Instrumental Variables**

207 As IV for a user’s attributes, we use average emotions and network centrality over all followees of the user, i.e., people who
208 are followed by the user. The choice of these IV are based on the following considerations. First, we consider the relevance
209 assumption of our instruments requiring that the characteristics of followees are correlated to the user’s characteristics, i.e.,
210 $cov(Z, X_1) \neq 0$. We expect that followees’ updates act as information sources for a user, and followees’ behaviors as well as
211 emotions manifested in their tweets can influence the user. Prior work [54] has shown the presence of homophily among ED
212 users on Twitter suggesting that users who share similar emotional and network attributes tend to follow one another. Further,
213 the empirical existence and strength of the relevance property are tested in a first stage regression and presented along with the

214 structural estimates of the models.

215 Second, we examine the exogeneity requirement (i.e., $cov(Z, u) = 0$), where followees' emotions and centrality must not
216 be have a direct effect on the drop-out decision of the user other than through their effect on the user's emotions. While we
217 take such assumption to be reasonable, we identify a pathway through which direct links could arise. Followees' attributes (e.g.,



emotions) could affect a user's dropout through their effects on followees' own dropouts, e.g., followees' emotional states may affect their own dropouts, and a feeling of loneliness due to friends' leaving may then drive the target user to drop out. To control for this channel, we measure the proportion and durations of followees that remain active in our second observation (regardless of whether the target user drops out or not). Further, we change the definition of followees (that are used to create the instruments) to those who are followed by a user but do not follow the user back (called *single-way followees*). In this setting, the reverse causality of a user's dropout on followees' attributes is nullified, which strengthens the exogeneity assumption on IV and controls.

Table 1. Covariates used in estimations.

Control effect	Covariate	Description
Social capital		
	#Followees	Number of total followees
	#Posts	Number of total posts, including tweets and re-tweets
	#Followers	Number of total followers
Activity level		
	Active days	Number of days from account creation to last posting
	#Followee/day	Average number of followees per day
	#Posts/day	Average number of posts per day
	#Followers/day	Average number of followers per day
Observational bias		
	#Tweets in use	Number of tweets in use to measure emotions
	#Followees in use	Number of followees whose attributes are used as instruments
Alternative causal channel		
	%Active followees (Followee durations)	Proportion of followees being active between two observations Average days of followees being active between two

224 Estimation Covariates

Our estimates control for several covariates that may affect users' tweeting activities, as listed in Table 1. We first measure users' social capital on Twitter (e.g., the numbers of social connections and the levels of engagement in sharing content) to capture the fact that people with different levels of popularity may have different tendencies to share content online [70]. Note that, although the numbers of followees and followers can be regarded as the in- and out-degree centralities of a user in the whole social network on Twitter (i.e., the "global" social capital), we are interesting in the "local" network centrality in the ED-specific communities. Second, as prior studies show an association between social media use and depression [55], we measure historical activity levels of users (i.e., active days) to capture effects that previous engagement may relate to both users' emotions and their future engagement. We also measure users' activity frequencies (e.g., posting frequency) to capture their patterns of Twitter usage. Third, the covariates on observational bias are used to control for effects caused by incomplete observations, e.g., a limited number of tweets are retrieved and used to measure emotions for a user. All variables on social capital, activity level and observational bias are measured from users' profile information and tweets collected in our first observation. Moreover, as discussed above, we include the proportion and average durations of followees that are active in our second observation to capture the channel that followees' emotions affect a user's dropout through their effects on followees' own dropouts.

238 Model Estimations

239 IV Estimation in Linear Regression Model

We use standard two-stage least squares (2SLS) estimators for linear probability models. In the first stage, we run an auxiliary regression and predict the endogenous variables (i.e., an individual's emotional state and network centrality) based on IV and exogenous covariates. In the second stage regression, we substitute the endogenous variables of interest with their predicted values from the first stage. Estimation is conducted through the AER package [71] and robust standard errors are computed.

244 **IV Estimation in Survival Model**

245 We use a Kaplan-Meier estimator [72] to estimate the survival function from data. Aalen's additive hazards model [73] is used
246 to estimate the effects of users' attributes on the time to dropout. Compared to the proportional hazards models in which the
247 ratios of hazard functions (i.e., hazard ratios) for different strata are assumed to be constant over time [74], the additive
248 model is more flexible and applies under less restrictive assumptions. To compute an IV estimator in an additive hazards model, we
249 use a control-function based approach which is proposed by Tchetgen et al. [61]. The TIMEREG package [75] is used for the
250 implementation of the estimation algorithm. Standard errors are obtained through non-parametric bootstrap.

251 *Results*

253 **Descriptive Statistics**

254 We obtain 2,906 users who posted more than 10 tweets (excluding re-tweets) and 50 words in our data, where 2,459
255 (85%) users had no posting activities during our two observation periods. Among 357 users who self-reported gender
256 information in their Twitter profile descriptions, 84% of them ($n = 300$) are female. The mean age is 17.3 among ED
257 users who self-reported age ($n = 1,015$). Based on the timestamps of account creation and last posting, we use the
258 Kaplan-Meier estimator to estimate the "lifetime" of a user on Twitter, i.e., the duration from account creation to the last
259 posting. The estimated median lifetime of these users on Twitter is 6 months, i.e., one half of the entire cohort drops out
260 at 6 months after creating an account. Figure 1 visualizes the social network between dropouts and non-dropouts among
261 ED users. We note that users with the same dropout states tend to cluster together. Computing Newman's homophily
262 coefficient r [76] of this network by users' dropout states, we find $r = 0.09$ ($z = 16.84$ and $P < .001$ compared to a
263 null model, see *SI*), suggesting that users with the same dropout states tend to befriend one another. See *SI* for details of
data statistics.



Figure 1. The who-follows-whom network among ED users on Twitter, laid out by the Fruchterman-Reingold algorithm [77]. Node colors represent dropout statuses, where the red color denotes dropout and the green color denotes non-dropout. Node size is proportional to the in-coreness centrality.

264 **Estimation Results of Linear Probability Models**

265 Table 2 shows estimated results in the linear models with two different IV specifications. In the first specification, we use all
266 followees of a user to create IV for the user's attributes. The results are given in columns 2-3, in which both OLS and IV

estimators show that positive emotions are associated with a higher probability of dropout ($\beta = 0.044$, $P = .007$ and $\beta = 0.29$, $P < .001$, respectively), with largely comparable coefficients for covariates. Compared to the OLS estimator, the IV estimator of the effect of emotions on dropout is remarkably stronger. The Wu-Hausman test further shows a significant difference between the OLS and IV estimators ($P < .01$), suggesting the presence of endogeneity. These results indicate that ignoring endogeneity in the OLS estimation leads to an underestimation of the effect of interest. Moreover, the F -statistics in the first stage regressions show that the relevance of IV exceeds the conventional standard of $F = 10$ [78], indicating the validity of our IV.

Table 2. Estimated effects of emotions on dropout using OLS and IV models.

	All followees				Single-way followees			
	OLS		IV		OLS		IV	
	β	P	β	P	β	P	β	P
Emotions	0.044	.005	0.290	< .001	0.064	< .001	0.304	< .001
#Followees	-0.0004	.01	-0.0002	.18	-0.0001	.11	-0.0001	.22
#Posts	-0.00000	.18	-0.00000	.35	-0.00001	.06	-0.00001	.14
#Followers	0.00001	.53	0.00001	.58	0.00001	.7	0.00000	.82
Active days	-0.0003	< .001	-0.0003	< .001	-0.0003	< .001	-0.0003	< .001
#Follower/day	0.001	.01	0.002	< .001	0.001	.03	0.002	.003
#Posts/day	0.0002	.72	-0.001	.38	0.0001	.85	-0.001	.32
#Followers/day	-0.003	.01	-0.005	< .001	-0.003	.02	-0.004	.002
#Tweets in use	-0.00004	.002	-0.00004	.003	-0.00003	.03	-0.00003	.03
#Followees in use	0.0004	.03	0.0002	.38	0.00000	.96	-0.0001	.41
%Active followees (Follower durations)	-1.159	< .001	-0.812	< .001	-0.939	< .001	-0.655	< .001
Constant	1.270	< .001	1.273	< .001	1.246	< .001	1.251	< .001
Observations	2,906		2,906		2,898		2,898	
First-stage F -statistic ^a			440.26 ($P < .001$)				158.21 ($P < .001$)	
Wu-Hausman test ^b			42.24 ($P < .001$)				14.54 ($P < .001$)	

^a F -statistic tests the significance of the instrument from a first-stage regression of a user's emotions on followees' emotions (i.e. the instrument) and the rest of the covariates.

^bTest the difference in estimates between OLS and IV; rejecting the null hypothesis suggests the presence of endogeneity.

^c P -values are computed based on heteroscedasticity-consistent standard errors.

Columns 4-5 show results of the second IV specification in which only single-way followees are used to create IV. Users who have no any single-way followees are excluded as instruments for these users' attributes are not available. Thus, the number of observations decreases as compared to that in the first IV specification. Moreover, as data on a smaller number of friends is used in the second IV specification, the relevance of IV becomes weaker but still passes the conventional test in the first stage regression. Despite such changes, the two specifications produce largely similar results. Computing Wald tests of equality of coefficients between the two IV models, we find that the estimated effects of emotions on dropout are statistically the same across different IV specifications ($P = .8$), potentially suggesting robustness of the results.

Note that network centrality is excluded from the linear models. This is because many users had dropped out long before our first observation (see Figure 2, SI), and the social networks of such users might largely change from the dates of their dropouts to our first observation, e.g., a user might be followed by new followers when these followers were unaware of the dropout of this user. In these cases, network centralities in the future are used to explain dropouts in the past, which can produce misleading results in the linear models. Nevertheless, including network centrality and instrumenting for it return statistically insignificant effect of centrality on the dropout decision, confirming our argument above on the irrelevance of centrality on this binary decision

286 to drop out or not.



Table 3. Estimated effects of emotions and centrality on survival time using Aalen's additive hazards models^a.

	All followees		Single-way followees	
	Standard (95% CI) ^b	IV (95% CI)	Standard (95% CI)	IV (95% CI)
Identical-interval censoring				
Emotions	-0.018 (-0.037, 0.0002)	-0.043 (-0.083, -0.004)	-0.018 (-0.036, 0.0006)	-0.061 (-0.116, -0.011)
Centrality	0.001 (0.0008, 0.0011)	0.001 (0.0007, 0.0011)	0.001 (0.0008, 0.0011)	0.001 (0.0006, 0.0011)
Personalized-interval censoring				
Emotions	-0.016 (-0.034, 0.0031)	-0.038 (-0.08, 0.002)	-0.015(-0.034, 0.0026)	-0.056 (-0.115, -0.007)
Centrality	0.001 (0.0008, 0.0011)	0.001 (0.0008, 0.0012)	0.001 (0.0008, 0.0011)	0.001 (0.0007, 0.0011)
Observations	447	447	445	445
First-stage F - statistic ^c		66.11 ($P < .001$)		34.99 ($P < .001$)
First-stage F		27.85 ($P < .001$)		12.62 ($P < .001$)

^aAll models are estimated controlling for the full list of covariates but are omitted from the tables due to space concerns. Results are available from the authors.

^bConfidence intervals (CI) for coefficients are obtained from 1,000 bootstrap replicates. A coefficient is significant at $P < .05$ if 0 is not in 95% CI.

^c F -statistic tests the joint significance of the two instruments from a first-stage regression of a user's emotions on followees' emotions and followees' centralities (i.e. the instruments) plus the rest of the covariates.

^d F -statistic tests the joint significance of the two excluded instruments from a first-stage regression of a user's centrality on followees' emotions and followees' centralities (i.e. the instruments) plus the rest of the covariates.

287 Estimation Results of Survival Models

288 In the survival models, we only consider users who were active past our first observation period, so as to examine the effect of
 289 network centralities in our first-observation on users' activities in the second-observation period. Table 3 shows mean coefficients
 290 of emotions and network centrality in the survival models. Following [61], the effects of all covariates are assumed to be
 291 time dependent in estimations. Both the standard and IV models on the identical-interval censored data show that (i) positive
 292 emotions lead to a shorter survival time ($P < .05$ in the IV model), and (ii) a core position in social networks is associated
 293 with
 294 a longer survival time ($P < .05$ in both models). Estimations on the personalized-interval censored data and using different
 295 IV specifications give similar results. The strong relevance of IV in the first stage regressions confirms the validity of IV across
 296 different models. A comparison of results between the linear and survival models further shows that these models have consistent
 estimators for the effect of emotions on dropout, i.e., positive emotions increase the likelihood to drop out.

297 Underlying Connection between Emotions and Dropout

298 To better understand the relationships between emotions and dropout, we examine posting interests among users with different
 299 dropout statuses and emotional states based on hashtags used in users' tweets (see SI for details). We find non-dropouts are in-
 300 terested in advocating a thin ideal (e.g., using hashtags "mythinspo" and "skinny4xmas") and promoting a pro-ED identity (e.g.,
 301 "edlogic" and "beautiful"). In contrast, dropouts engage in discussing their health problems (e.g., "selfharmprobz", "bulim-
 302 icprobz" and "anorexicprobz") and offering emotional support for others (e.g., "anasisters" and "stayingstrong"), which implies
 303 a tendency of these users to recover from disorders [20–22]. Similarly, we split all ED users into three equal-size sets based
 304 on their emotional scores and examine hashtags used by each set of users. We find that users with negative emotions often
 305 engage in promoting thin ideals (e.g., "bonespo" and "mythinspo"), showing largely overlapping interests with the non-dropouts. In
 306 contrast, users with neutral and positive emotions are more interested in discussing their health problems (e.g., "anorexicprobz"
 307 and "bulimicprobz"), opposing pro-ED promotions (e.g., "reversethinspo") and encouraging healthier body image and behaviors
 308 (e.g., "fitfam" and "fitness"), showing similar interests with the dropouts. See SI for more detailed lists of hashtags.

309 Measuring the Spearman rank correlation ρ between pairwise lists of hashtags posted by users with a given state (e.g.,
 310 dropped-out or not, and positive or negative), we find a positive correlation between negative users and non-dropouts in hashtag
 311 usage ($\rho = 0.36, P = .003$ in Table 4), indicating similar posting interests among these users. A similar pattern occurs between
 312 positive users and dropouts. In contrast, users with other pairs of states show a negative correlation or non-correlation in hashtag
 313 usage, indicating their discrepancies in posting interests. These results reveal a possible underlying connection between positive
 314 emotions and dropout. Compared to users with positive emotions, those with negative emotions have more similar interests to
 315 active members (i.e., non-dropouts) in the ED community. Finding similarities with other members in a community can enhance

Table 4. Spearman rank correlations between pairwise lists of hashtags posted by users with a given dropout state and by users with a given emotional state respectively^a.

	Negative ($n =$	Neutral ($n =$	Positive ($n =$
Non-dropout ($n =$	0.36 ($P = .003$) ^c	-0.21 ($P = .03$)	-0.66 ($P < .001$)
Dropout ($n = 227$)	-0.33 ($P < .001$)	-0.04 ($P = .57$)	0.12 ($P = .07$)

^aAll tags in two lists I_i and I_j are considered in computing the correlation $\rho(I_i, I_j)$. Tags in each list are ranked by TF-IDF scores [79] and the TF-IDF score of tag t in list I_i is 0 if I_i does not contain t .

^bThe number of hashtags posted by users with a given state.

^cThe Spearman correlations r of hashtags posted by users with different dropout and emotional states, where $\rho \in [-1, 1]$ with 0 indicating no correlation. P -values testing for non-correlation are reported in parentheses.

316 a sense of belonging to the community and positively increase intention to engage in community activities [33, 37]. Therefore, it
317 is not surprising that negative users are less likely to drop out than positive users in our estimations.

318 Discussion

320 Principal Findings

322 This study provides the first estimates of the effects of personal emotions and interpersonal social networks on dropout in
323 online ED communities. The present work has several strengths. First, we base our analysis on incentive theory to explore
324 determinants of users' online behaviors (i.e., dropout), allowing us to study users' behaviors in a more systematic way than
325 most prior studies that often focus on a single type of determinant (e.g., individual attributes [35, 36] or social attributes [42,
326 43, 67]). Second, we use automated sentiment analysis techniques to measure users' emotions and network analysis
327 methods to quantify users' sociometric statuses in an online community, leading to higher efficiency than traditional
328 research methods such as surveys [35, 37, 39, 41, 48]. Third, we apply an IV approach to both linear probability and
329 survival models, which enables us to achieve a more consistent estimate of human behavior in online settings than traditional
330 methods (e.g., OLS) used in prior studies [39, 41, 47]. Overall, we find that positive emotions increase the likelihood of
331 dropout in ED individuals and accelerate the dropout process on Twitter. In contrast, a central position in the social network of
332 ED individuals at an earlier stage is associated with prolonged participation of an individual at a later stage. These findings are
333 verified across a variety of robustness checks.

334 Despite differences in methodology, our findings align with prior studies in psychological and social media research [5, 33,
335 35]. Our results suggest that ED users with negative emotions have high levels of participation on Twitter. This aligns with
336 prior survey studies on social media use (e.g., Facebook use), where people with social anxiety and shyness (i.e., personality
337 traits that are often correlated with multiple negative emotions such as feeling lonely, isolated and unhappy [80]) are found to
338 spend more time online [35, 48, 81]. An explanation for this is the online disinhibition effect [82], i.e., because of anonymity
339 in online interactions, people with social inhibitions (e.g., those who are socially anxious or shy and those with a stigmatized
340 health problem [83]) might be more willing share personal feelings and reveal themselves in online interactions than offline
341 interactions, in order to meet their social and intimacy needs [48]. Additional analyses on users' posting interests reveal that
342 users with negative emotions share similar interests with active users. This allows us to confirm the validity of our results via
343 the social capital theory [39, 40], i.e., sharing common attributes (e.g., interests and vision) with other members can enhance a
344 sense of belonging and positive feeling toward a community, which drives people to actively engage in the community.

345 Consistent with positive associations between network centrality and active participation in other online communities [67,
346 70], we find that central users in the social network of an ED community tend to have a longer-lasting participation in the
347 community. This result is to be expected for several reasons. First, users who are centrally embedded in a group have a
348 relatively high number of social ties to other members, which can lead these users to feel being socially accepted and
349 approved, as well as a strong sense of belonging to the group. Prior studies have consistently shown that recognition from other
350 members and identification within an online community increase an individual's commitment to the community [34, 39–41].
Second, information shared by central users is likely to spread to the majority of a community through social ties, and their
central positions in the community may promote other members to trust such information [70]. This implies that central

u
s
e
r
s

h
a
v
e

a

g
r
e
a
t
e
r

p
o
t
e
n
t
i
a
l

t
h
a
n

p
e
r
i
p
h
e
r
a
l



351 users in influencing members' opinions, emotions and behaviors in online communities [84]. Thus, compared to peripheral
352 users,

353 feeling influential may provide an additional incentive for central users to continue participating.

354 In line with prior studies on online ED communities [15, 17, 21], we find that ED users on Twitter have different stances on
355 ED, where users with negative emotions often share pro-ED content and those with positive emotions often share pro-recovery
356 content. As pro-ED content often contains thin-ideal images and harmful tips for weight loss/control [8, 24, 25], this result
357 aligns

358 with clinical evidence on ED treatment showing that more emotional distress is associated with a higher risk to learn and
359 develop

360 dysfunctional coping behaviors among ED sufferers [5]. Thus, as suggested by prior studies [85], engaging in pro-ED content
361 may serve as a coping mechanism to deal with emotional pressures and stress of ED. A possible explanation for the association
362 between engaging in harmful online content and coping with stress is sensation seeking [86], a basic personality trait defined as
363 the seeking of varied, novel, complex and intense sensations and experiences, and the willingness to take risks. Several studies
364 have shown that sensation seeking is prominent in adolescence (i.e., the age that disordered eating often develops [1]) and
365 closely

366 related to pathological Internet use, such as use of violent sites [87] and Internet dependence [88].

367 Our study also offers new insights into online ED communities. First, ED users have a high dropout rate (85% in our
368 sample)

369 and a short lifespan between an account creation to lost posting on Twitter (with 6 months of median time to drop out). This
370 aligns

371 with views of online ED communities as hidden, secretive groups [30], but also indicates the dynamic characteristics of these
372 communities. Second, users who discuss their health problems and share pro-recovery content (i.e., pro-recovery users) have
373 lower levels of posting activities (i.e., a higher dropout rate) than those who share pro-ED content (i.e., pro-ED users) on Twitter.
374 This can be explained as follows. Due to common interests in ED, pro-recovery and pro-ED groups are likely to be connected in
375 the same social networks, and content shared within a group is hence likely to be visible to the other group. However, exposure
376 to content from the antagonist group can have distinct effects in pro-ED and pro-recovery groups. Exposure to pro-ED content
377 is harmful for pro-recovery users and can impede their recovery process [3, 24], while exposure to pro-recovery content can
378 instead stimulate harmful behaviors in pro-ED users (e.g., actively sharing pro-ED content) [21]. Thus, pro-recovery users might
379 tend to leave such an online community to avoid a risk of further deterioration or relapse. Our finding may also explain why
380 pro-ED content is found being more pervasive than pro-recovery content across social media sites [15, 17, 21], e.g., almost five
381 times in terms of unique publishers on Tumblr [15]. Third, ED users tend to connect with others with the same dropout states on
382 Twitter. This implies that whether an individual drops out from online communities depends on whether others in the
383 individual's

384 social networks drop out. In other words, dropout in online ED communities is not only a function of individual experience or
385 individual choice but also a property of group interactions, e.g., homophily [89] and social contagion effects [59].

379 Implications

380 Our findings are of practical relevance to the promotion of public health over social media. First, the decision to maintain
381 active participation in an online community can be caused by intrinsic and extrinsic characteristics/traits of the participants, e.g.,
382 personal emotions, interests and social networks. Such self-selection bias can lead to the sample not being representative of the
383 whole population, and hence researchers need to consider both active and dropped-out users for a well-rounded picture of online
384 health communities. This is particularly important for public health officials to make special efforts to reach these dropouts and
385 offer more intensive support when they are trying to recover. Second, high attrition rates are often regarded as negative outcomes
386 in online interventions, particularly in those delivered over a purposely designed website [11, 31, 32]. However, this may or
387 may not be the case in interventions over general social media sites (e.g., Twitter) depending on how targeted populations use
388 these sites. For example, when an intervention is delivered in an online community in which members often shared harmful
389 content, a high attrition rate (i.e., members dropping out of the harmful community) may not be a negative outcome. Using
390 automated data-mining techniques to track users' behaviors (e.g., emotions and posting interests), as used in this work, can
391 provide more detailed information about people's use of online health communities and improve our understanding of attrition
392 in online interventions. Third, interventions that recommend content containing positive emotions to ED users (not limited to
393 ED-related content but more general content containing happiness and inspiration) may reduce their engagement in a harmful

394 online community. This aligns with Fredrickson's broaden-and-build model which argues that cultivating positive emotions is
395 useful to prevent and treat mental health problems [90]. Finally, intervention strategies could be tailored for different individuals
396 depending on their positions in the social network of an online community. For example, identifying central individuals as
change
397 agent might enhance the efficacy and cost effectiveness of an intervention, due to their greater influence potential through larger
398 numbers of social ties [91], but also their longer-lasting effects through longer-term participation in the community.



399 **Limitations**

400 First, we recognize that self-diagnosis information on Twitter may be itself self-censored by users to align with their personality
401 traits and perceptions of their audience on the platform. People may not use tags like “eatingdisorder” to self-report their
402 experience of illness and would be excluded by our collection methods. Also, although over 208K users and over 241M tweets
403 are studied in this work, a small sample of rich social media data is used to explore the attrition of ED communities on Twitter.
404 Thus, generalization of our results to all ED-related online communities should be cautious. Second, our measures of dropout
405 are based on posting activity, while some people primarily use Twitter to receive outside information but rarely post their own
406 information. We have little activity data on these users and hence less understanding of the characteristics of their dropout. This
407 thus raises important issues that need further research to enhance our understanding of attrition in online health communities,
408 such as consensus and clarity about the definition of dropout. Third, our study focuses on the Twitter platform, without
validation

409 on other platforms. However, stopping using a platform can be related to the attractiveness of the platform. Hence, future
research

410 is also in need to examine many other factors that we did not explore but can effect dropout on social media, such as individual
411 personality, physical health states, perceptions and purposes of using a particular social media platform. Fourthly, user accounts
412 on a social media site are often not unique—an individual may have multiple accounts on the same site. Thus, we cannot be
413 certain whether individuals who stopped using an account will engage in the same or similar online communities through other
414 accounts. That is, the dropout of a user account may not necessarily imply that an individual abandons a specific identity (e.g.,
415 pro-ED) shared within a community. Finally, although our methodology allows us to establish a causal link from emotions to
416 dropout behavior, it offers limited insights into the pathways through which this link exactly operates and future work is needed
417 to explore such issues in detail.

418 **Conclusions**

419 This study presents a systematic characterization of attrition in an ED community on Twitter. Our analysis offers the first attempt
420 towards the estimation of the effects of personal emotions and network centrality on dropout behaviors in individuals affected by
421 ED on Twitter. Our results provide new insights into the trajectories that ED communities develop online, which can help public
422 health officials to better understand individual needs in using online ED communities and provide tailored support for
individuals

423 with different needs.

424 **References**

- 426 1. Association AP, et al. Diagnostic and statistical manual of mental disorders (DSM-5®); 2013. [DOI:
427 10.1176/appi.books.9780890425596.529303].
- 428 2. Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders: a
429 meta-analysis of 36 studies. Archives of general psychiatry. 2011;68(7):724–731. [PMID: 21727255] [DOI: 10.1001/arch-
430 genpsychiatry.2011.74].
- 431 3. Campbell K, Peebles R. Eating disorders in children and adolescents: state of the art review. Pediatrics. 2014;134(3):582–
432 592. [PMID: 25157017] [DOI: 10.1542/peds.2014-0194].
- 433 4. Merikangas KR, He Jp, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorder
434 in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A).
435 Journal of the American Academy of Child & Adolescent Psychiatry. 2010;49(10):980–989. [PMID: 20855043] [DOI:
436 10.1016/j.jaac.2010.05.017].
- 437 5. Corstorphine E. Cognitive–emotional–behavioural therapy for the eating disorders: Working with beliefs about emotions.
438 European Eating Disorders Review. 2006;14(6):448–461. [DOI: 10.1002/erv.747].
- 439 6. Swanson SA, Crow SJ, Le Grange D, Swendsen J, Merikangas KR. Prevalence and correlates of eating disorders in ado-
440 lescents: Results from the national comorbidity survey replication adolescent supplement. Archives of general psychiatry.

441 2011;68(7):714–723. [PMID: 21383252] [DOI: 10.1001/archgenpsychiatry.2011.22].



- 442 7. Rich E. Anorexic dis (connection): managing anorexia as an illness and an identity. *Sociology of health & illness*.
443 2006;28(3):284–305. [PMID: 16573717] [DOI: 10.1111/j.1467-9566.2006.00493.x].
- 444 8. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: A pilot study of pro-eating disorder web site usage in
445 adolescents with eating disorders. *Pediatrics*. 2006;118(6):e1635–e1643. [PMID: 17142493] [DOI: 10.1542/peds.2006-
446 1133].
- 447 9. Latkin CA, Knowlton AR. Social network assessments and interventions for health behavior change: a critical review.
448 *Behavioral Medicine*. 2015;41(3):90–97. [PMID: 26332926] [DOI: 10.1080/08964289.2015.1034645].
- 449 10. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on
450 health behavior change: a systematic review and meta-analysis. *Journal of the American Medical Informatics Association*.
451 2014;22(1):243–256. [PMID: 25005606] [DOI: 10.1136/amiajnl-2014-002841].
- 452 11. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions
453 that use online social networks effective? A systematic review. *Journal of medical Internet research*. 2014;16(2). [PMID:
454 24550083] [DOI: 10.2196/jmir.2952].
- 455 12. McLean SA, Wertheim EH, Masters J, Paxton SJ. A pilot evaluation of a social media literacy intervention to reduce risk
456 factors for eating disorders. *International Journal of Eating Disorders*. 2017;50(7):847–851. [PMID: 28370321] [DOI:
457 10.1002/eat.22708].
- 458 13. Stapleton JL, Manne SL, Day AK, Levonyan-Radloff K, Pagoto SL. Healthy Body Image Intervention Delivered to Young
459 Women via Facebook Groups: Formative Study of Engagement and Acceptability. *JMIR research protocols*. 2018;7(2).
460 [PMID: 29463495] [DOI: 10.2196/resprot.9429].
- 461 14. Bazarova NN, Choi YH. Self-disclosure in social media: Extending the functional approach to disclosure motivations and
462 characteristics on social network sites. *Journal of Communication*. 2014;64(4):635–657. [DOI: 10.1111/jcom.12106].
- 463 15. De Choudhury M. Anorexia on tumblr: A characterization study. In: *Proceedings of the 5th International Conference on*
464 *Digital Health 2015*; 2015. p. 43–50. [DOI: 10.1145/2750511.2750515].
- 465 16. Chancellor S, Lin Z, Goodman EL, Zerwas S, De Choudhury M. Quantifying and Predicting Mental Illness Severity in On-
466 line Pro-Eating Disorder Communities. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative*
467 *Work and Social Computing (CSCW)*. ACM; 2016. p. 1171–1184. [DOI: 10.1145/2818048.2819973].
- 468 17. Chancellor S, Mitra T, De Choudhury M. Recovery Amid Pro-Anorexia: Analysis of Recovery in Social Media. In:
469 *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*; 2016. p. 2111–2123. [DOI:
470 10.1145/2858036.2858246].
- 471 18. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information
472 promoted
473 through video-based social media: anorexia on YouTube. *Journal of medical Internet research*. 2013;15(2):e30. [PMID:
474 23406655] [DOI: 10.2196/jmir.2237].
- 475 19. Chancellor S, Kalantidis Y, Pater JA, De Choudhury M, Shamma DA. Multimodal Classification of Moderated Online Pro-
476 Eating Disorder Content. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM;
477 2017. p. 3213–3226. [DOI: 10.1145/3025453.3025985].
- 478 20. Lyons EJ, Mehl MR, Pennebaker JW. Pro-anorexics and recovering anorexics differ in their linguis-
479 tic Internet self-presentation. *Journal of psychosomatic research*. 2006;60(3):253–256. [PMID: 16516656]
[DOI:10.1016/j.jpsychores.2005.07.017].
- 480 21. Yom-Tov E, Fernandez-Luque L, Weber I, Crain SP. Pro-anorexia and pro-recovery photo sharing: a tale of two warring
481 tribes. *Journal of medical Internet research*. 2012;14(6):e151. [PMID: 23134671] [DOI: 10.2196/jmir.2239].
- 482 22. Wolf M, Theis F, Kordy H. Language use in eating disorder blogs: Psychological implications of social online activity.
483 *Journal of Language and Social Psychology*. 2013; [DOI: 10.1177/0261927X12474278].

- 484 23. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: A content analysis of pro-eating disorder web sites.
485 American journal of public health. 2010;100(8):1526–1534. [PMID: 20558807] [DOI: 10.2105/AJPH.2009.172700].
- 486 24. Maloney P. Online networks and emotional energy: how pro-anorexic websites use interaction ritual chains to (re) form
487 identity. *Information, Communication & Society*. 2013;16(1):105–124. [DOI: 10.1080/1369118X.2012.659197].
- 488 25. Bardone-Cone AM, Cass KM. Investigating the impact of pro-anorexia websites: A pilot study. *European Eating Disorders
489 Review*. 2006;14(4):256–262. [DOI: 10.1002/erv.714].
- 490 26. Oksanen A, Garcia D, Sirola A, Näsi M, Kaakinen M, Keipi T, et al. Pro-anorexia and anti-pro-anorexia videos on
491 YouTube: Sentiment analysis of user responses. *Journal of medical Internet research*. 2015;17(11). [PMID: 26563678]
492 [DOI: 10.2196/jmir.5007].
- 493 27. Chancellor S, Lin ZJ, De Choudhury M. “This Post Will Just Get Taken Down”: Characterizing Removed Pro-Eating
494 Disorder Social Media Content. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.
495 ACM; 2016. p. 1157–1162. [DOI: 10.1145/2858036.2858248].
- 496 28. Chancellor S, Pater JA, Clear T, Gilbert E, De Choudhury M. #thyghgapp: Instagram Content Moderation and Lexical
497 Variation in Pro-Eating Disorder Communities. In: *Proceedings of the 19th ACM Conference on Computer-Supported
498 Cooperative Work and Social Computing (CSCW)*. ACM; 2016. p. 1201–1213. [DOI: 10.1145/2818048.2819963].
- 499 29. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours:
500 a systematic review and meta-analysis of randomised controlled trials. *BMJ open*. 2014;4(2):e003926. [PMID: 24525388]
501 [DOI: 10.1136/bmjopen-2013-003926].
- 502 30. Casilli AA, Pailler F, Tubaro P, et al. Online networks of eating-disorder websites: why censoring pro-ana might be a bad
503 idea. *Perspectives in public health*. 2013;133(2):94–95. [PMID: 23467530] [DOI: 10.1177/1757913913475756].
- 504 31. Eysenbach G. The law of attrition. *Journal of medical Internet research*. 2005;7(1). [PMID: 15829473] [DOI:
505 10.2196/jmir.7.1.e11].
- 506 32. ter Huurne ED, Postel MG, de Haan HA, van der Palen J, DeJong CA. Treatment dropout in web-based cognitive be-
507 havioral therapy for patients with eating disorders. *Psychiatry research*. 2017;247:182–193. [PMID: 27918968] [DOI:
508 10.1016/j.psychres.2016.11.034].
- 509 33. Malinen S. Understanding user participation in online communities: A systematic literature review of empirical studies.
510 *Computers in human behavior*. 2015;46:228–238. [DOI: 10.1016/j.chb.2015.01.004].
- 511 34. Kollock P. The economies of online cooperation. *Communities in cyberspace*. 1999;220. [DOI: 10.4324/9780203194959].
- 512 35. Orr ES, Sisic M, Ross C, Simmering MG, Arseneault JM, Orr RR. The influence of shyness on the use of Face-
513 book in an undergraduate sample. *CyberPsychology & Behavior*. 2009;12(3):337–340. [PMID: 19250019] [DOI:
514 10.1089/cpb.2008.0214].
- 515 36. Hughes DJ, Rowe M, Batey M, Lee A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social
516 media usage. *Computers in Human Behavior*. 2012;28(2):561–569. [DOI: 10.1016/j.chb.2011.11.001].
- 517 37. Casaló LV, Flavián C, Guinalíu M. New members’ integration: Key factor of success in online travel communities. *Journal
518 of Business Research*. 2013;66(6):706–710. [DOI: 10.1016/j.jbusres.2011.09.007].
- 519 38. Cook E, Teasley SD, Ackerman MS. Contribution, commercialization & audience: understanding participation in an online
520 creative community. In: *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM; 2009. p.
521 41–50. [DOI: 10.1145/1531674.1531681].
- 522 39. Chiu CM, Hsu MH, Wang ET. Understanding knowledge sharing in virtual communities: An integration of social capital
523 and social cognitive theories. *Decision support systems*. 2006;42(3):1872–1888. [DOI: 10.1016/j.dss.2006.04.001].

- 524 40. Nahapiet J, Ghoshal S. Social capital, intellectual capital, and the organizational advantage. In: Knowledge and social
525 capital. Elsevier; 2000. p. 119–157. [DOI: 10.2307/259373].
- 526 41. Lai HM, Chen TT. Knowledge sharing in interest online communities: A comparison of posters and lurkers. *Computers in*
527 *Human Behavior*. 2014;35:295–306. [DOI: 10.1016/j.chb.2014.02.004].
- 528 42. Wang YC, Kraut R, Levine JM. To stay or leave?: the relationship of emotional and informational support to commitment
529 in online health support groups. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work.
530 ACM; 2012. p. 833–842. [DOI: 10.1145/2145204.2145329].
- 531 43. Xing W, Goggins S, Introne J. Quantifying the Effect of Informational Support on Membership Retention in On-
532 line Communities through Large-Scale Data Analytics. *Computers in Human Behavior*. 2018;86:227–234. [DOI:
533 10.1016/j.chb.2018.04.042].
- 534 44. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American*
535 *statistical Association*. 1996;91(434):444–455. [DOI: 10.2307/2291629].
- 536 45. Stovitz SD, Verhagen E, Shrier I. Distinguishing between causal and non-causal associations: implications for sports
537 medicine clinicians. *BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine*; 2017. [PMID:
538 29162620] [DOI: 10.1136/bjsports-2017-098520].
- 539 46. Tausczik YR, Pennebaker JW. Participation in an online mathematics community: differentiating motivations to add. In:
540 Proceedings of the ACM 2012 conference on computer supported cooperative work. ACM; 2012. p. 207–216. [DOI:
541 10.1145/2145204.2145237].
- 542 47. Wu B. Patient Continued Use of Online Health Care Communities: Web Mining of Patient-Doctor Communication. *Journal*
543 *of medical Internet research*. 2018;20(4). [PMID: 29661747] [DOI: 10.2196/jmir.9127].
- 544 48. Sheeks MS, Birchmeier ZP. Shyness, sociability, and the use of computer-mediated communication in relationship develop-
545 ment. *CyberPsychology & Behavior*. 2007;10(1):64–70. [PMID: 17305450] [DOI: 10.1089/cpb.2006.9991].
- 546 49. Primack BA, Shensa A, Sidani JE, Whaitte EO, Yi Lin L, Rosen D, et al. Social media use and perceived social isolation
547 among young adults in the US. *American journal of preventive medicine*. 2017;53(1):1–8. [PMID: 28279545] [DOI:
548 10.1016/j.amepre.2017.01.010].
- 549 50. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational*
550 *psychology*. 2000;25(1):54–67. [DOI: 10.1006/ceps.1999.1020].
- 551 51. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. *Journal of the*
552 *Association for Information Science and Technology*. 2010;61(12):2544–2558. [DOI: 10.1002/asi.21416].
- 553 52. Friedkin NE. Theoretical foundations for centrality measures. *American journal of Sociology*. 1991;96(6):1478–1504. [DOI:
554 10.1086/229694].
- 555 53. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. # Proana: Pro-Eating Disorder Socialization on Twitter. *Journal*
556 *of Adolescent Health*. 2016;58(6):659–664. [PMID: 27080731] [DOI: 10.1016/j.jadohealth.2016.02.012].
- 557 54. Wang T, Brede M, Ianni A, Mentzakis E. Detecting and characterizing eating-disorder communities on social media. In:
558 Proceedings of the Tenth International Conference on Web Search and Data Mining (WSDM) 2017. ACM; 2017. p. 91–100.
559 [DOI: 10.1145/3018661.3018706].
- 560 55. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression
561 among US young adults. *Depression and anxiety*. 2016;33(4):323–331. [PMID: 26783723] [DOI: 10.1002/da.22466].
- 562 56. Mabe AG, Forney KJ, Keel PK. Do you “like” my photo? Facebook use maintains eating disorder risk. *International Journal*
563 *of Eating Disorders*. 2014;47(5):516–523. [PMID: 25035882] [DOI: 10.1002/eat.22254].

- 564 57. de Vries DA, Peter J, de Graaf H, Nikken P. Adolescents' social network site use, peer appearance-related feedback, and
565 body dissatisfaction: Testing a mediation model. *Journal of youth and adolescence*. 2016;45(1):211–224. [PMID: 25788122]
566 [DOI: 10.1007/s10964-015-0266-4].
- 567 58. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through so-
568 cial networks. *Proceedings of the National Academy of Sciences*. 2014;p. 201320040. [PMID: 24889601] [DOI:
569 10.1073/pnas.1320040111].
- 570 59. Coviello L, Sohn Y, Kramer AD, Marlow C, Franceschetti M, Christakis NA, et al. Detecting emotional contagion in
571 massive
social networks. *PLoS one*. 2014;9(3):e90315. [PMID: 24621792] [DOI: 10.1371/journal.pone.0090315].
- 572 60. Card D. The causal effect of education on earnings. In: *Handbook of labor economics*. vol. 3. Elsevier; 1999. p. 1801–1863.
573 [DOI: 10.1016/S1573-4463(99)03011-4].
- 574 61. Tchetgen EJT, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context.
575 *Epidemiology (Cambridge, Mass)*. 2015;26(3):402. [PMID: 25692223] [DOI: 10.1097/EDE.0000000000000262].
- 576 62. Aral S, Nicolaides C. Exercise contagion in a global social network. *Nature communications*. 2017;8:14753. [PMID:
577 28418379] [DOI: 10.1038/ncomms14753].
- 578 63. Gonçalves P, Araújo M, Benevenuto F, Cha M. Comparing and combining sentiment analysis methods. In: *Proceedings of*
579 *the first ACM conference on Online social networks*. ACM; 2013. p. 27–38. [DOI: 10.1145/2512938.2512951].
- 580 64. Seidman SB. Network structure and minimum degree. *Social networks*. 1983;5(3):269–287. [DOI: 10.1016/0378-
581 8733(83)90028-x].
- 582 65. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex
583 networks. *Nature physics*. 2010;6(11):888. [DOI: 10.3410/f.8218956.8658056].
- 584 66. Garcia D, Mavrodiev P, Schweitzer F. Social resilience in online communities: The autopsy of friendster. In: *Proceedings*
585 *of the first ACM conference on Online social networks*. ACM; 2013. p. 39–50. [DOI: 10.1145/2512938.2512946].
- 586 67. Garcia D, Mavrodiev P, Casati D, Schweitzer F. Understanding popularity, reputation, and social influence in the twitter
587 society. *Policy & Internet*. 2017;9(3):343–364. [DOI: 10.1002/poi3.151].
- 588 68. Giatsidis C, Thilikos DM, Vazirgiannis M. D-cores: measuring collaboration of directed graphs based on degeneracy.
589 *Knowledge and information systems*. 2013;35(2):311–343. [DOI: 10.1109/icdm.2011.46].
- 590 69. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems*.
591 2006;1695(5):1–9.
- 592 70. Wasko MM, Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of
593 practice. *MIS quarterly*. 2005;p. 35–57. [DOI: 10.2307/25148667].
- 594 71. Kleiber C, Zeileis A. *Applied econometrics with R*. Springer Science & Business Media; 2008. [ISBN: 978-0-387-77318-
6].
- 595 72. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American statistical*
596 *association*.
1958;53(282):457–481. [DOI: 10.1007/978-1-4612-4380-9_25].
- 597 73. Aalen O, Borgan O, Gjessing H. *Survival and event history analysis: a process point of view*. Springer Science & Business
598 *Media*; 2008. [ISBN: 038768560X, 9780387685601].
- 599 74. Cox DR. Regression models and life-tables. In: *Breakthroughs in statistics*. Springer; 1992. p. 527–541. [DOI: 10.1007/978-
600 1-4612-4380-9_37].
- 601 75. Scheike TH, Zhang MJ. Analyzing competing risk data using the R timereg package. *Journal of statistical software*.
602 2011;38(2). [DOI: 10.18637/jss.v038.i02].

603 76. Newman ME. Mixing patterns in networks. Physical Review E. 2003;67(2):026126. [DOI: 10.1103/physreve.67.026126].



- 604 77. Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. *Software: Practice and experience*.
605 1991;21(11):1129–1164. [DOI: 10.1002/spe.4380211102].
- 606 78. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments.
607 *Journal of Business & Economic Statistics*. 2002;20(4):518–529. [DOI: 10.1198/073500102288618658].
- 608 79. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.
609 1972;28(1):11–21. [DOI: 10.1108/eb026526].
- 610 80. Farmer AS, Kashdan TB. Social anxiety and emotion regulation in daily life: Spillover effects on positive and negative
social
611 events. *Cognitive behaviour therapy*. 2012;41(2):152–162. [PMID: 22428662] [DOI: 10.1080/16506073.2012.666561].
- 612 81. Amichai-Hamburger Y, Gazit T, Bar-Ilan J, Perez O, Aharony N, Bronstein J, et al. Psychological factors behind the lack of
613 participation in online discussions. *Computers in Human Behavior*. 2016;55:268–277. [DOI: 10.1016/j.chb.2015.09.009].
- 614 82. Suler J. The online disinhibition effect. *Cyberpsychology & behavior*. 2004;7(3):321–326. [PMID: 15257832] [DOI:
615 10.1089/1094931041291295].
- 616 83. Lapidot-Lefler N, Barak A. The benign online disinhibition effect: Could situational factors induce self-disclosure and
proso- 617 cial behaviors? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*. 2015;9(2). [DOI:
10.5817/CP2015-2- 618 3].
- 619 84. Tang J, Li J. Semantic mining of social networks. *Synthesis Lectures on the Semantic Web: Theory and Technology*.
620 2015;5(2):1–205. [DOI: 10.2200/S00629ED1V01Y201502WBE011].
- 621 85. Mulveen R, Hepworth J. An interpretative phenomenological analysis of participation in a pro-anorexia internet site and
622 its relationship with disordered eating. *Journal of health psychology*. 2006;11(2):283–296. [PMID: 16464925] [DOI:
623 10.1177/1359105306061187].
- 624 86. Zuckerman M. Sensation seeking. *The International Encyclopedia of Communication*. 2008; [DOI:
625 10.1002/9781405186407.wbiecs029].
- 626 87. Slater MD. Alienation, aggression, and sensation seeking as predictors of adolescent use of violent film, computer, and
627 website content. *Journal of communication*. 2003;53(1):105–121. [DOI: 10.1111/j.1460-2466.2003.tb03008.x].
- 628 88. Lin SS, Tsai CC. Sensation seeking and internet dependence of Taiwanese high school adolescents. *Computers in human
629 behavior*. 2002;18(4):411–426. [DOI: 10.1016/S0747-5632(01)00056-5].
- 630 89. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. *Annual review of sociology*.
631 2001;p. 415–444. [DOI: 10.1146/annurev.soc.27.1.415].
- 632 90. Fredrickson BL. Cultivating positive emotions to optimize health and well-being. *Prevention & Treatment*. 2000;3(1):1a.
633 [DOI: 10.1037//1522-3736.3.0001a].
- 634 91. Valente TW. Network interventions. *Science*. 2012;337(6090):49–53. [DOI: 10.1093/acprof:oso/9780195301014.003.0011].