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

Tao Wang\textsuperscript{1,2}, MSc*; Emmanouil Mentzakis\textsuperscript{1}, PhD; Markus Brede\textsuperscript{3}, PhD; Antonella Ianni\textsuperscript{1}, PhD;

\textsuperscript{1}Department of Economics, University of Southampton, UK
\textsuperscript{2}The Alan Turing Institute, UK
\textsuperscript{3}Department 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: twang@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.
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 platforms like Twitter and Facebook. Participation in online communities is common in ED populations [8] and has recently been suggested as a screening factor for ED [3]. As such, a growing amount of research has focused on characterizing individuals’ behavioral patterns in online ED communities with an aim to better understand ED and promote population-level well-being [9–11].

One notable characteristic of online ED communities is their participants having widely different stances on ED [8, 12, 13]. 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 [12–14]. 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, 15]. These pro-ED communities have been shown to negatively impact the health and quality of life of people with and without ED, through reinforcement of individuals’ identity on ED [10], poor body image and thinness adoration [16], or teaching harmful practices for weight loss [6]. Recent studies further show that individuals’ language use online strongly indicate their pro-ED or pro-recovery stances [12, 17, 18], as well as emotions of depression, helplessness and anxiety that reflect their mental disorders [19]. Other studies have examined interactions between pro-ED and pro-recovery communities on Flickr [13], anorexia-related misinformation [9], sentiments of comments on ED-related videos on YouTube [20], characteristics of removed pro-ED content [21] and lexical variation of pro-ED tags on Instagram [11, 22]. Yet, prior studies have largely focused on examining how people engage in and maintain an online health 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 what determines and accelerates the dropouts of individuals can enhance our knowledge of the dynamics in online communities. In fact, studying the attrition process of a harmful or healthy community also has practical implications for disease prevention and online interventions [23, 24].

The decision of participation or dropout in online communities can be driven by a variety of factors [25], including personality traits [26], interests [27], and recognition in a community [28–30]. However, mixed results have been shown in prior empirical studies, one example being a positive association between individuals’ expertise and online participation found in [29] while a negative association was found in [28]. Such inconsistent results could be caused by several issues. First, prior studies often focus on the use of self-reported surveys and rely on participants’ estimates of their own personality, concerns and behaviors [26, 29], which can introduce considerable retrospective bias and measurement errors. Second, participation in an online community is inherently self-selected (e.g., sharing common interests), while members can drop out for many different reasons (e.g., effect of an online or offline event). Thus, unobservable factors (i.e., omitted variables) are likely correlated with both the main predictors and participation outcomes. Any of these issues can cause endogeneity and lead to biased results [31].

In this study, we estimate the effects of personal emotions and network centrality on dropout in an online ED community, while accounting for the endogeneity. Instrumental variables (IV) estimators both for the decision to drop out and for the time to the event are implemented on a unique dataset with longitudinal tweeting-activity information on a large set of individuals who self-identified with ED on Twitter. To our knowledge, this study is the first to systematically characterize the determinants of dropout behaviors in online health communities. Three research questions are examined: (i) what are the general characteristics of the attrition process in an online ED community, (ii) how do the intrinsic and extrinsic factors affect the decision of an individual to drop out of the community, (iii) how do these factors affect the duration of time until the occurrence of dropout?


**Methods**

**Data Collection**

Our data is collected from Twitter, a microblogging platform that allows millions of users to self-disclose and socialize. As many social media platforms like Facebook and Instagram have taken moderation actions to counteract pro-ED content and user accounts [22], Twitter has not yet enforced actions to limit such content [32]. This makes Twitter a unique platform to study the attrition process naturally happening in an online ED community and allows us to examine individuals’ behaviors in a non-reactive way. Our study protocol was approved by the Ethics Committee at the University of Southampton. All data used in our study is public information on Twitter and available through the Twitter APIs. No personally identifiable information is used in this study.

![Diagram of data collection and analysis procedures.](image)

- First, we collect a set of individuals who self-identified with ED on Twitter using a snowball sampling approach [33]. Specifically, we track the public tweet stream using “eating disorder”, “anorexia”, “bulimia” and “EDNOS” from Jan. 8 to 15, 2016. This results in 1,169 tweets that mention common ED. From the authors of these tweets, we identify 33 users who self-reported both ED-diagnosis information (e.g., “eating disorder”, “anorexia” and “bulimia”) and personal bio-information (e.g., body weight and height) in her profile descriptions (i.e., a sequence of user-generated text describing their accounts below profile images). Starting from these seed users, we expand the user set using snowball sampling through their social networks of followees/followers. At each sampling stage, we filter out non-English speaking accounts and finally obtain 3,380 unique ED users who self-identify with ED in their profile descriptions.

- Then, we collect all friends (including followees and followers) of each ED user, leading to a large social network consisting of 208,063 users. For each user, we retrieve up to 3,200 (the limit returned from Twitter APIs) of their most recent tweets and obtain 241,243,043 tweets in total. The data collection process finished on Feb. 11, 2016.

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

To verify the quality of our collected sample, two members of the research team classified a random sample of 1,000 users on whether they were likely to be a true ED user based on their posted tweets, images and friends’ profiles. The process revealed a 95.2% match between the identified ED individuals in the data collection stage and those classified as ED during inspection.

**Estimation Framework**
The empirical analysis builds on well-established incentive theory [34], where *intrinsic* (e.g., personal emotions and interests), and *extrinsic* motivation factors (e.g., sociometric status or centrality in social networks) determine participation in online communities. First, we specify a linear probability model on the whole sample to estimate the effects of individuals’ characteristics
observed in the first-observation period on the probability of dropping-out in the second-observation period. Second, we estimate survival models to explore the effects of individuals’ characteristics observed in the first observation on the time to dropout in the second observation (i.e., the duration from our first observation to the dropout in our second observation).

Like all social media studies, only a limited number of individuals’ characteristics are available for the estimations and these are mostly observed through user-generated data online. This leads to omitted variable bias, since unobservable factors can be correlated with both the main explanatory variables (i.e., emotions and network centrality) and dropout outcomes. For example, undergoing hospital treatment can simultaneously affect a person’s emotional state and the usage of social media. Further, both emotions and network centrality are complex concepts that can be defined and characterized in a variety of ways [35, 36] making it difficult to obtain precise measurements of these variables, and hence concerns of measurement error problems also arise.

Both omitted variables and measurement errors result in biased and inconsistent estimates. We address such endogeneity issues through instrumental variables (IV). Consider a model \( Y = \beta_1 X_1 + \beta_2 X_2 + u \), where \( X_1 \) is endogenous, \( X_2 \) is exogenous, \( u \) is a random error term and \( \beta_i \) are effects to be estimated. IV methodology uses an instrument \( Z \) (which is (i) not contained in the explanatory equation, (ii) correlated with \( X_1 \), i.e., \( \text{cov}(Z, X_1) \neq 0 \)), and (iii) uncorrelated with \( u \), i.e., \( \text{cov}(Z, u) = 0 \), conditional on the other covariates such as \( X_2 \) ) and runs a first stage reduced-form 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 regression \( Y = \beta_1 \hat{X}_1 + \beta_2 X_2 + u \), where \( \hat{X}_1 \) is the predicted values of \( X_1 \) from the first stage. For more details please see [31].

**Measures**

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

**Dropout Outcomes as Dependent Variables**

In the linear probability models, we encode the dropout status of a user as 0 (denoting non-dropout) if the user has updated posts in our second observation, and 1 (denoting dropout) otherwise.

In the survival models, each user has a two-variable outcome: (i) a censoring variable denoting whether the event of dropout occurs, and (ii) a variable of survival time denoting the duration of time until the occurrence of dropout. We censor the occurrence 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 interval \( \pi \) before our second observation (so called identical-interval censoring). As people use social media platforms with different activity levels, e.g., some users post every several hours while other users only post once every couple of days, our second censoring method further accounts for personalized posting activities of individuals (called personalized-interval censoring). In this method, users are said to drop out if they have not posted tweets for more than a variate threshold interval \( \lambda \pi + \left| 1 - \lambda \right| I_i \) before our second observation, where \( \pi \) is a fixed threshold, \( I_i \) is the average posting interval of individual \( i \) in our first observation period, and \( \lambda \) is a tunable parameter to control the effects of individual activities. For users who are censored as dropped-out, we set their survival times as the durations from our first observation to their last postings in our second observation. For those who are censored as non-dropped-out, we set their survival times as the whole time period between our two observations.

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

Individuals’ emotions are measured through their language used in tweets. There is a variety of sentiment analysis algorithms to measure emotional expressions in texts [37, 38]. In this study, we use SentiStrength [38] as (i) it has been used to measure the emotional content in online ED communities and shown good inter-rater reliability [20]; (ii) it is designed for short informal texts with abbreviations and slang, and thus suitable to process tweets [38]. After removing mention marks, hashtags and URLs, each tweet is assigned a scaled value in \([-4, 4]\) by SentiStrength, where negative/positive scores indicate the strength of negative/positive emotions respectively, and 0 denotes neutral emotions. We quantify a user’s emotional state by the average score of all tweets posted by the user. All re-tweets are excluded, as re-tweets reflect more the
emotions of their original authors than those of their re-tweeters. For robust results from the language processing algorithms, we only consider users who have more than 10 tweets and post more than 50 words.

Network centrality measures the importance of a person in a social network; people well-recognized by their peers often have high centralities in a group [36]. To measure a user’s centrality in the ED community, we build a who-follows-whom network
among ED users and their friends, where a directed edge runs from node A representing user A to node B representing user B if A follows B on Twitter. While there are various measures of network centrality, we focus on coreness centrality [39] as it has been shown to outperform other measures such as degree and betweenness centrality [36] in detecting influential nodes in complex networks [40] and cascades of users leaving an online community [30, 41]. We measure the sociometric status of a user in the ED community by the in-coreness centrality [42] of a node in the generated network using the package igraph 0.7.0 [43].

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

As IV for a user’s attributes, we use average emotions and network centrality over all followees of the user, i.e., people who are followed by the user. The choice of these IV are based on the following considerations.

First, we consider the relevance assumption of our instruments requiring that the characteristics of followees are correlated to the user’s characteristics, i.e., \( \text{cov}(Z, X_1) \neq 0 \). We expect that followees’ updates act as information sources for a user, and followees’ behaviors as well as emotions manifested in their tweets can influence the user. Prior work [33] has shown the presence of homophily among ED users on Twitter suggesting that users who share similar emotional and network attributes tend to follow one another. Further, the empirical existence and strength of the relevance property are tested in a first stage regression and presented along with the structural estimates of the models.

Second, we examine the exogeneity requirement (i.e., \( \text{cov}(Z, u) = 0 \)), where followees’ emotions and centrality must not 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 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.](image-url)

<table>
<thead>
<tr>
<th>Control effect</th>
<th>Covariate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social capital</strong>*</td>
<td>#Followees</td>
<td>Number of total followees</td>
</tr>
<tr>
<td></td>
<td>#Posts</td>
<td>Number of total posts, including tweets and retweets</td>
</tr>
<tr>
<td></td>
<td>#Followers</td>
<td>Number of total followers</td>
</tr>
<tr>
<td><strong>Activity level</strong>*</td>
<td>Active days</td>
<td>Number of days from account creation to last posting</td>
</tr>
<tr>
<td></td>
<td>#Followee/day</td>
<td>Average number of followees per day</td>
</tr>
<tr>
<td></td>
<td>#Posts/day</td>
<td>Average number of posts per day</td>
</tr>
<tr>
<td></td>
<td>#Followers/day</td>
<td>Average number of followers per day</td>
</tr>
<tr>
<td><strong>Observational bias</strong>*</td>
<td>#Tweets in use</td>
<td>Number of tweets in use to measure emotions</td>
</tr>
<tr>
<td></td>
<td>#Followees in use</td>
<td>Number of followees whose attributes are used as instruments</td>
</tr>
<tr>
<td><strong>Alternative causal channel</strong>*</td>
<td>%Active followees</td>
<td>Proportion of followees being active between two observations</td>
</tr>
<tr>
<td></td>
<td>&lt;Followee durations&gt;</td>
<td>Average days of followees being active between two</td>
</tr>
</tbody>
</table>

*Measured from the first observation.

**Activity Factors as Covariates/Controls**

We condition our estimations on covariates that may affect our outcomes of interest, as listed in Table 1. All variables on social capital, activity level and observational bias are measured from users’ profile information and tweets collected in our first
observation, where the covariates on observational bias are used to control effects caused by incomplete observations, e.g., a limited number of tweets are retrieved and used to measure emotions for each user.

**Model Estimations**

**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 centralities) 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 [44] and robust standard errors are computed.

**IV Estimation in Survival Model**

We use a Kaplan-Meier estimator [45] to estimate the survival function from data. Aalen's additive hazards model [46] is used to estimate the effects of users' attributes on the time to dropout. Compared to the proportional hazards models in which the ratios of hazard functions (i.e., hazard ratios) for different strata are assumed to be constant over time [47], the additive model is more flexible and applies under less restrictive assumptions. To compute an IV estimator in an additive hazards model, we use a control-function based approach which is proposed by Tchetgen et al. [48]. The TIMEREG package [49] is used for the implementation of the estimation algorithm. Standard errors are obtained through non-parametric bootstrap.

**Results**

**Descriptive Statistics**

We obtain 2,906 users who posted more than 10 tweets (excluding re-tweets) and 50 words in our data, where 2,459 (85%) users had no posting activities during our two observation periods. Figure 2 shows more details on dates when users joined and dropped out on Twitter. Most users were active during 2012 to 2014, during which 1,944 users (67%) joined Twitter. Two notable peaks in the curve of last posting time occur at the dates of our two observations. The first peak indicates that some users were lost to follow up (e.g., accounts were deleted), and the second peak indicates that many users were still actively posting tweets until our observations ended. Based on the timestamps of account creation and last posting, we use the Kaplan-Meier estimator to estimate the “lifetime” of a user on Twitter, i.e., the duration from account creation to the last posting. Each user is censored by 1 (denoting “death”) if the user had no posts after our first observation, and 0 otherwise. The estimated median lifetime of these users on Twitter is 6 months, i.e., one half of the entire cohort drops out at 6 months after creating an account.

Table 2 shows the descriptive statistics of users stratified by dropout states that are observed in our second observation period. The differences between dropouts and non-dropouts are measured with Mann-Whitney U-tests and the Bonferroni correction is used to counteract the problem of multiple comparisons. Compared to dropouts, non-dropouts show more negative emotions and higher network centralities, where the network centralities are measured based on a following network containing 208,063 nodes and 1,347,056 directed edges. All nodes are connected in a single weakly connected component and the average degree of the network is 6.5.

Figure 3 visualizes the social network between dropouts and non-dropouts. We note that users with the same dropout states tend to cluster together. To quantify the strength of this assortative mixing pattern, we compute Newman’s homophily coefficient [51] of this network by users’ dropout states. The resulting coefficient is 0.09, suggesting that users with the same dropout states tend to befriend one another. To test the statistical significance of the homophily pattern, we randomly shuffle users’ dropout states and re-measure homophily coefficients based on the shuffled states. These coefficients can be viewed as observed values of a random variable. Repeating this procedure 3,000 times, we yield the empirical distribution of homophily coefficients with a mean of $\mu = 0$ and a standard deviation of $\sigma = 0.005$. The $z$-score for the observed homophily under this baseline distribution is $z = 16.84$ and $P < .001$, suggesting the presence of homophily.
Table 2. Descriptive statistics of users by dropout and non-dropout states.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>All (n = 2,906)</th>
<th>Non-dropout (n = 447)</th>
<th>Dropout (n = 2,459)</th>
<th>U-test</th>
<th>( \lambda )</th>
<th>Pc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Emotions</td>
<td>0.09</td>
<td>0.13</td>
<td>0.08</td>
<td>-3.06</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Centrality</td>
<td>29.35</td>
<td>16.22</td>
<td>28.16</td>
<td>16.11</td>
<td>9.26</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>#Followees</td>
<td>309.34</td>
<td>533.19</td>
<td>268.64</td>
<td>360.69</td>
<td>10.32</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>#Followers</td>
<td>899.62</td>
<td>2225.44</td>
<td>645.42</td>
<td>1281.62</td>
<td>16.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Active days</td>
<td>308.15</td>
<td>752.74</td>
<td>244.87</td>
<td>525.09</td>
<td>13.64</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Followee/day</td>
<td>4.16</td>
<td>19.07</td>
<td>4.60</td>
<td>20.60</td>
<td>-8.69</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Posts/day</td>
<td>4.41</td>
<td>7.54</td>
<td>4.51</td>
<td>7.81</td>
<td>-1.13</td>
<td>1</td>
</tr>
<tr>
<td>Followers/day</td>
<td>2.66</td>
<td>8.23</td>
<td>2.85</td>
<td>8.84</td>
<td>-4.97</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tweets in use</td>
<td>614.43</td>
<td>1244.17</td>
<td>499.96</td>
<td>715.65</td>
<td>16.47</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>%Active followers</td>
<td>242.13</td>
<td>405.27</td>
<td>212.47</td>
<td>280.92</td>
<td>9.56</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;Followee durations&gt;</td>
<td>193.18</td>
<td>243.42</td>
<td>184.05</td>
<td>78.65</td>
<td>13.23</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;Followee emotions&gt;</td>
<td>0.06</td>
<td>0.16</td>
<td>0.41</td>
<td>0.15</td>
<td>13.70</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;Followee centralities&gt;</td>
<td>29.38</td>
<td>28.04</td>
<td>29.62</td>
<td>9.08</td>
<td>-3.65</td>
<td>0.004</td>
</tr>
</tbody>
</table>

\(^a\) <Followee x> denotes the average values of a user’s followees in terms of statistics x.

\(^b\) A z-score measures the extent of a variable in non-dropout group being larger than that in dropout group.

\(^c\) P -values for two-tailed tests with Bonferroni correction.

Figure 2. Number of users along with time points when users created a Twitter account and posted the last tweet.

Specifications of Data Censoring Methods
We tune the parameters of our data censoring methods based on users’ activities before and after our first observation. We apply each censoring method with different parameters to data on users’ activities before our first observation to estimate users’ dropout states, and choose parameters that achieve the best agreement between the estimated dropout states and the observed states in our second observation. By setting \( \pi \in [1,300] \) days in the identical-interval censoring method, we find the optimal parameter being \( \pi = 101 \), with Cohen’s \( \kappa = .68 \) of the estimated dropout states and the observed dropout states of users. Such good agreement illustrates the effectiveness of the censoring approach. Similarly, by searching in a parameter space of \( \pi \in [1,200] \) days and \( \lambda \in [0,1] \), we find the optimal parameters in the personalized-interval censoring method being \( \pi = 161 \) and \( \lambda = 0.6 \), with Cohen’s \( \kappa = .68 \) as well. We use these parameters censor the dropout states of users who were active in the second observation. Figure 4 shows the Kaplan-Meier curves of users’ survival time from our first observation until our second observation using the two censoring methods. The median survival time of users is 13 months in both two methods, and no significant difference is found between the two types of censorships (\( P = .93 \) in a log-rank test).
Figure 3. The who-follows-whom network among ED users on Twitter, laid out by the Fruchterman-Reingold algorithm [50]. Node colors represent dropout statues, where the red color denotes dropout and the green color denotes non-dropout. Node size is proportional to the in-coreness centrality.

Figure 4. Kaplan-Meier estimations of survival time.

Estimation Results of Linear Probability Models

Table 3 shows estimated results in the linear models with two different IV specifications. In the first specification, we use all 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 $F = 10$ [52], indicating the validity of our IV.

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. As shown in Figure 2, many users had dropped out long before our first observation, 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 drop-out decision, confirming our argument above on the irrelevance of centrality on this binary decision to drop-out or not.

### Estimation Results of Survival Models

In the survival models, we only consider users who were active past our first observation period, so as to examine a causal relation between network centralities in our first-observation and users’ activities in the second-observation period. Table 4 shows mean coefficients of emotions and network centrality in the survival models. Following [48], the effects of all covariates are assumed to be time dependent in estimations. Both the standard and IV models on the identical-interval censored data show that (i) positive emotions lead to a shorter survival time ($P < .05$ in the IV model), and (ii) a core position in social networks is associated with a longer survival time ($P < .05$ in both models). Estimations on the personalized-interval censored data and using different IV specifications give similar results. The strong relevance of IV in the first stage regressions confirms the validity of IV across 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.
**Table 4.** Estimated effects of emotions and centrality on survival time using Aalen’s additive hazards models.

<table>
<thead>
<tr>
<th></th>
<th>All followers</th>
<th></th>
<th>Single-way followers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(95% CI)</td>
<td>IV (95% CI)</td>
<td>(95% CI)</td>
<td>IV (95% CI)</td>
</tr>
<tr>
<td><strong>Identical-interval censoring</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotions</td>
<td>-0.018 (-0.037, 0.0002)</td>
<td>0.001 (0.0008, 0.0011)</td>
<td>-0.043 (-0.083, -0.004)</td>
<td>0.001 (0.0007, 0.0011)</td>
</tr>
<tr>
<td>Centrality</td>
<td>0.001 (0.0008, 0.0011)</td>
<td>0.001 (0.0007, 0.0011)</td>
<td>0.001 (0.0008, 0.0011)</td>
<td>0.001 (0.0006, 0.0011)</td>
</tr>
<tr>
<td><strong>Personalized-interval censoring</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotions</td>
<td>-0.016 (-0.034, 0.0031)</td>
<td>0.001 (0.0008, 0.0011)</td>
<td>-0.038 (-0.08, 0.002)</td>
<td>0.001 (0.0008, 0.0012)</td>
</tr>
<tr>
<td>Centrality</td>
<td>0.001 (0.0008, 0.0011)</td>
<td>0.001 (0.0007, 0.0011)</td>
<td>0.001 (0.0006, 0.0011)</td>
<td>0.001 (0.0005, 0.0011)</td>
</tr>
<tr>
<td>Observations</td>
<td>447</td>
<td>447</td>
<td>445</td>
<td>445</td>
</tr>
<tr>
<td>First-stage $F$-statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>66.11 (P &lt; .001)</td>
<td>2.785 (P &lt; .001)</td>
<td></td>
</tr>
</tbody>
</table>

*All 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. Confidence 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.

$F$-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.

$F$-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.

**Underlying Connection between Emotions and Dropout**

To provide further intuition for the relationships between emotions and dropout found above, we examine associations of interests among users with different dropout statuses and emotional states. This follows past evidence that community interest is the primary motivating factor for participation in online communities [53] and people’s concerns/interests reflect their emotional states [54]. Since hashtags are explicit topic signals on Twitter and have been shown to strongly indicate users’ interests [55], we characterize users’ interests based on hashtags used in their tweets.

![Figure 5](image_url) The co-occurrence network of the most popular hashtags used by all ED users. Each node is a hashtag, and node size is proportional to the number of users who posted the tag. Node color is assigned based on the frequency of a tag so that high frequency is darker and low frequency is lighter. Edge width is proportional to the number of co-occurrences of two attached tags in tweets.

We first examine the prevalent topics of interest for the entire ED community. To capture relationships between different topics, we build an undirected, weighted hashtag network based on the co-occurrences of hashtags in tweets posted by ED users, where an edge is weighted by the co-occurrence count of two attached tags. To filter out noise from accidental co-occurrences...
and spam, we only consider hashtags used by more than 50 distinct users and observed in more than 50 tweets, resulting in a network of 312 nodes and 7,906 edges. Figure 5 shows the co-occurrence network of the most popular hashtags of interest for ED users. We observe that topics on promoting a thin ideal (e.g., “thinspo” and “thinspiration”) are very prevalent in the community.

**Table 5.** Popular hashtags used by ED users, grouped by users’ dropout states.

<table>
<thead>
<tr>
<th>Dropout states</th>
<th>Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-dropout</td>
<td>legspo, mythinspo, skinny4xmas, bonespo, goals, edlogic, eatingdisorders, edthoughts, ribs, bones, depressed, depression, edprobs, collarbones, bulimia, promia, replytweet, beautiful, anorexia, thin, hipbones, legs, ed, thighbap, weightloss, skinny, proed, selfharm, perfection, mia, thinspiration, perfect, proguna, diet,</td>
</tr>
<tr>
<td>Dropout</td>
<td>goaway, stopbullying, worthless, selfharmprobz, ew, anasisters, yay, oneday, reasonstobefit, bulimicprobz, anorexicprobz, fact, disgusting, thankgod, willpower, tweetwhatyoueat, wow, toofat, jealous, thankyou, true, ana-sister, anafamily, starveon, gross, teamfollowback, fuck, icandothis, tired, edfamily, relapse, stayingstrong</td>
</tr>
</tbody>
</table>

**Table 6.** Popular hashtags used by ED users, grouped by users’ emotional states.

<table>
<thead>
<tr>
<th>Emotional states</th>
<th>Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>bonespo, mythinspo, edlogic, bulimia, depression, starve, eatingdisorders, anxiety, anorexia, skinny4xmas, depressed, edprobs, proed, ribs, bones, edproblems, thinspo, edthoughts, thinspiration, selfharm, fuck, goals, thin, edgirlprobs, fat, sad, skinny, ednos, realityproject, ana, eatingdisorder, fatass, hipbones</td>
</tr>
<tr>
<td>Neutral</td>
<td>awkward, anorexicprobz, fast, mylife, bulimicprobz, please, sorry, fuckyou, myfitnesspal, ew, skinny4xmas, legspo, edfamily, gross, anafamily, ugh, ednos, workout, goals, replytweet, tmi, fatass, reversethinspo, edprobz, anaproblems, failure, flatstomach, fact, binge, fatty, fasting, suicide, depressed</td>
</tr>
<tr>
<td>Positive</td>
<td>eatclean, fitfam, inspiration, reasonstobefit, ff, noexcuses, fitness, loveit, winning, anasister, tweetwhatyoueat, twye, keepgoing, success, jealous, want, fitspo, beforeandafter, retweet, excited, proud, reasonstoloseweight, abcdiet, fail, justsaying, rt, motivated, workout, stayingstrong, love, myfitnesspal</td>
</tr>
</tbody>
</table>

**Table 7.** Similarities of hashtags posted by users with different dropout and emotional states.

<table>
<thead>
<tr>
<th></th>
<th>Negative</th>
<th>Neutral</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-dropout</td>
<td>0.842</td>
<td>0.641</td>
<td>0.491</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.481</td>
<td>0.677</td>
<td>0.682</td>
</tr>
</tbody>
</table>

We then examine interests of users with different dropout states. We split ED users into two sets based on their dropout states in our second observation, and extract hashtags from tweets posted by each set of users. Again, tags that are used by less than 50 users and occur in less than 50 tweets in each set are excluded. To adjust tags that are popular in general, we use TF-IDF [56] to rank the specificity of a tag in each set of users. Table 5 lists the most representative hashtags in each user set, in which we find that users with different dropout states display distinct interests online. Non-dropouts are interested in advocating a thin ideal (“mythinspo” and “skinny4xmas”) and reinforcing a pro-ED identity (“edlogic” and “beautiful”). In contrast, dropouts engage more in discussing their health problems (“selfharmprobz”, “bulimicprobz”, “anorexicprobz” and “relapse”) and offering emotional support for others (“anasisters” and “stayingstrong”), which indicates a tendency of these users to recover from disorders [12–14]. Together, these results imply that pro-recovery users are more likely to drop out than pro-ED users. A comparison of interests between each individual set and the entire community (see Figure 5) further shows that the non-dropouts have dominated the topics of discussions within the community. This is expected because the non-dropouts have prolonged participation, with 732.17 active days on average compared to 278.40 days of the dropouts (see Table 2).

Similarly, we split ED users into three equal-size sets based on their emotional scores and obtain the most representative hashtags among each set of users in Table 6. The results show that users with negative emotions more engage in promoting thin ideals (“bonespo” and “mythinspo”), showing largely overlapping interests with the non-dropouts. In contrast, users with neutral and positive emotions are more interested in discussing their health problems (“anorexicprobz” and “bulimicprobz”), opposing pro-ED promotions (“reversethinspo”) and encouraging healthier body image and behaviors (“fitfam” and “fitness”).
showing similar interests with the dropouts. We further compute the cosine similarities of hashtags posted by users with different dropout
and emotional states. The results, as shown in Table 7, confirm our claim that negative users share more similar interests to non-dropouts and have more dissimilar interests to dropouts, as compared to positive users. The results on users’ interests reveal a possible underlying connection between positive emotions and dropout. Compared to users with positive emotions, those with negative emotions have more similar interests to active members in the ED community. Finding similarities with other members in a community can enhance a sense of belonging to the community and positively increase intention to engage in community activities [25, 27]. Therefore, it is not surprising that negative users are less likely to drop out than positive users in our estimations.

Discussion

Principal Findings

This study provides the first estimates of the effects of individual emotions and interpersonal social networks on dropout in online ED communities. Applying IV estimations with linear probability and survival models to a sample of individuals who self-identified with ED on Twitter, we present causal effects of individuals’ emotional states and network centrality on their dropout behaviors. Overall, we find that positive emotions increase the likelihood of dropout in ED individuals and accelerate the dropout process on Twitter. In contrast, a central position in the social network of ED individuals increases online participation. These findings are verified across a variety of robustness checks.

Our results align with prior evidence in psychological and social media research [5, 25, 26]. First, we find that ED individuals with negative emotions have high levels of participation on Twitter. This confirms a study on Facebook use in which people with high levels of anxiety and stress (i.e., negative emotions) are found to spend more time online [26]. An explanation for this is that negative emotions (e.g., shyness and inhibition in face-to-face interactions) may lead to reliance of negative individuals on the anonymous interactions in online communities [57]. Second, we find that negative individuals tend to engage in sharing pro-ED content which can encourage thin ideals, reinforce a pro-ED identity, promote harmful weight loss/control practices and exacerbate ED risk factors [10]. These results are in line with clinical evidence in ED treatment, i.e., more emotional distress is associated with a higher risk to learn and develop dysfunctional coping behaviors among ED sufferers [5]. Third, consistent with positive associations between recognition and participation in online communities [30, 58], we find that central individuals in the social network of an ED community are likely to have a longer-lasting participation in the community. Together, our findings confirm that both intrinsic (e.g., self-emotions and personal interests) and extrinsic factors (e.g., respect from others) significantly influence participation in online communities [34].

Further, our study also offers new insights into online ED communities. First, ED users have a high dropout rate (85% in our sample) and a short lifespan between an account creation to lost posting on Twitter (with 6 months of median time to drop out), highlighting the dynamic characteristics of online ED communities. Second, users who discuss their health problems and have a tendency to recover from ED (i.e., pro-recovery users) are more likely to drop out (or with lower levels of posting activities) than those who share pro-ED content (i.e., pro-ED users) on Twitter. A possible explanation for this is as follows. Due to common interests in ED, pro-recovery and pro-ED groups are likely to be connected in the same social networks, and content shared within a group is hence likely to be visible to the other group. However, exposure to the content from the antagonist group can have distinct effects in pro-ED and pro-recovery groups. Exposure to pro-ED content is harmful for pro-recovery users and can impede their recovery process [3, 10], while exposure to pro-recovery content can instead stimulate harmful behaviors in pro-ED users (e.g., actively sharing pro-ED content) [13]. Thus, pro-recovery users might tend to leave such an online community to avoid a risk of further deterioration or relapse. Our finding may also explain why pro-ED content is found being more pervasive than pro-recovery content across social media sites [13, 17, 18], e.g., almost five times in terms of unique publishers on Tumblr [17]. Finally, ED users tend to connect with others with the same dropout states on Twitter. This implies that whether an individual drops out from online communities depends on whether others in the individual’s social networks drop out. In other words, dropout in online ED communities is not only a function of individual experience or individual choice but also a property of group interactions, e.g., homophily [59] and social contagion effects [60].
Implications

Our findings are of practical relevance to the promotion of public health over social media. First, the decision to maintain active participation in an online community can be caused by intrinsic and extrinsic characteristics/traits of the participants, e.g., personal emotions, interests and social networks. Such self-selection bias can lead to the sample not being representative of the whole population, and hence researchers need to consider both active and dropped-out users for a well-rounded picture of online health communities. This is particularly important for public health officials to make special efforts to reach these dropouts and offer more intensive support when they are trying to recover. Second, while high attrition is a common issue in Internet-based interventions for challenging health problems like ED [23, 24, 61], our findings suggest that early dropout is a typical feature in individuals with ED on social media like Twitter. Hence, attrition in an online intervention can be caused by inappropriate settings of intervention programs (e.g., content, protocol or intensity) [23, 24], but also some inherent traits of individuals (e.g., frustration and fear of stigma [7]). Disentangling the risk factors between the two aspects is important for successful implementation of alternative strategies that can improve retention and ultimately health-related outcomes. Third, interventions that expose individuals to positive emotions and positive images about recovery may decrease their engagement in a harmful online community and increase well-being [62]. Finally, intervention strategies could be tailored for different individuals depending on their positions in the social network of an online community. Identifying central individuals as change agent might enhance the efficacy and cost effectiveness of an intervention, due to their greater influence potential through larger numbers of social ties [63], but also their longer-lasting effects through longer-term participation in the community.

Limitations

First, we recognize that self-diagnosis information on Twitter may be itself self-censored by users to align with their personality traits and perceptions of their audience on the platform. People may not use tags like “eatingdisorder” to self-report their experience of illness and would be excluded by our collection methods. Second, although over 208K users and over 241M tweets are studied in this work, a small sample of rich social media data is used to explore the attrition of ED communities on Twitter. Thus, our results cannot be generalized to all ED-related communities online. Third, our measures of dropout are based on posting activity, while some people primarily use Twitter to receive outside information but rarely post their own information. We have little activity data on these users and hence less understanding of the characteristics of their dropout. This thus raises important issues that need further research to enhance our understanding of attrition in online health communities, such as consensus and clarity about the definition of dropout. Fourth, our study focuses on the Twitter platform, without validation on other platforms. However, stopping using a platform can be related to the attractiveness of the platform. Hence, future research is also in need to examine many other factors that we did not explore but can effect dropout on social media, such as individual personality, physical health states, perceptions and purposes of using a particular social media platform. Finally, social media profiles are not identical and nonrenewable identities. Individuals who dropped out from an online community may have other profiles on the same platform or other platforms. We cannot be sure whether they will migrate to a similar or different community in the future.

Conclusions

This study presents a systematic characterization of attrition in online ED communities. Our analysis offers a first attempt towards the estimators of the causal effects of emotions and network centrality on dropout behaviors in individuals affected by ED on Twitter. Our results provide new insights into the trajectories that ED communities develop online, and can guide public health officials in designing online interventions for individuals at risk of ED.

References


