Estimating Determinants of Attrition in Eating Disorder Community on Social Media: An Instrumental Variables Approach

Tao Wang¹,²*, Emmanouil Mentzakis¹, Markus Brede³, and Antonella Ianni¹

¹Department of Economics, University of Southampton, UK
²The Alan Turing Institute, UK
³Department of Electronics and Computer Science, University of Southampton, UK
* t.wang@soton.ac.uk

ABSTRACT

High attrition is a major problem in online health interventions. However, little is known about risk factors of people dropping out online. Challenges exist due to the lack of baseline knowledge on how attrition naturally happens in online communities, and reliable methods that can identify causality between past traits and future dropout. Here, we examine characteristics of naturally occurring attrition in online health communities and use longitudinal statistical models to assess whether emotion and online social networks influence dropout behaviors in the future. From three comparable subpopulations sampled from Twitter, we find that individuals who self-identified as eating disordered have shorter active durations than general populations, with a half of cohort dropping out in 6 months after creating a Twitter profile. Applying instrumental variables estimation and survival analysis to longitudinal data on users’ activities spanning 1.5 year, we identify that negative emotions increase forthcoming dropout in general populations, while positive emotions instead increase dropout in disordered populations. Individuals who are surrounded by many active peers and those who are central in the network tend to stay longer in the future. We interpret our findings with clinical evidence and discuss their implications for designing network interventions that can promote organizational well-being in online communities.

1 Introduction

Eating disorders (ED), such as anorexia nervosa and bulimia, are a major public health concern due to a high mortality rate (highest of any mental illness)¹, intractable co-morbidities² and worldwide prevalence³. More than 2.7% of 13-17 year olds in the US³ and 725,000 people in the UK³ have been affected by ED, with a trend that is increasing over time. Although health interventions have been proposed, ED population are very hard to reach and as such study, plan and administer intervention to those in need⁴. Individuals often conceal their ED symptoms due to feelings of shame or fear of stigma⁵,⁶ and many never disclose their struggles with professionals⁷. Due to the secretive nature and the need to ensure anonymity, people suffering from ED often seek for social support and resources from peer-communities online⁸, particularly via social networking sites (SNS) such as Twitter and Facebook. Engagement in these online communities is common among individuals with ED⁹–¹⁴ and has recently been suggested as a screening factor for ED¹⁵. Given these facts, vast and growing research has focused on whether public health and policy can harness the power of such social networks for the benefit of ED sufferers and how to leverage online interventions over SNS to promote healthy behaviors and improve community-level well-being¹⁶–¹⁸.

Compared with traditional approaches, online interventions appear to be more accessible for broad audiences and more cost-effective in achieving short- or long-term goals in public health¹⁹. Existing interventions delivered via online health websites have shown a positive effect on healthy behavioral outcomes²⁰,²¹. However, using online communities to develop and deliver successful interventions requires stability and frequency of interactions within these communities themselves²². For communities with a very high dropout rate, it is unlikely that members will have adequate opportunity to promote a target behavior change. Attrition (i.e. participants stopping usage or are lost in follow-ups) has been identified as a crucial issue in the efficacy of online interventions²³–²⁵, since cost-effectiveness is largely reduced for population-level interventions as the number of people reaping their benefits goes down²⁶. A recent meta-analysis of 22 studies found that all studies suffered from decreased participation throughout the intervention period, with 12 studies reporting rates of more than 20%²⁷. Despite such high attrition rates, characteristics that differentiate dropouts from completers at various time points in an online intervention are still unknown in the literature²⁸–³⁰, even under-explored in the research on traditional face-to-face interventions on various behavior-related conditions, such as obesity, smoking and alcohol misuse³¹–³³.

Although understanding these characteristics is important to implement alternative strategies that can improve retentions
and ultimately health-related outcomes, several challenges exist in such investigation. First, and most importantly, lack of baseline knowledge on how attrition naturally take place within an online community makes it difficult to identify valid predictors that strongly effect dropout behaviors. Most previous studies have not found any differences in predictors of interest between dropouts and completers. Second, previous studies focus on the use of self-reported surveys and rely on participants’ estimates of their own cognitions (e.g., thoughts, emotions) and behaviors. This potentially introduces considerable measurement error and retrospective bias, while these issues can conceal the fact that risk factors influence dropout (i.e., endogenous behaviors). Third, engagement in an online community is inherently self-selected while members can drop out for many different reasons (e.g., effect of an online or offline event). It is difficult to include a comprehensive set of controls for all possible risk factors. An incomplete control set however leads to omitted-variable bias and further results in a biased estimation in causality analysis. Finally, in mental disorders, cognitive symptoms and traits are naturally the best predictors of sufferers’ behavior change like dropping-out of an intervention. While cognitions can cause behaviors, cognitions can also be the feedback of behaviors. It thus appears to exist a simultaneous causality between individuals’ traits and their dropout behaviors. Such simultaneity again causes endogeneity bias, resulting in causal estimators being inconsistent.

Here, we address these methodological challenges by studying the role of individual characteristics on dropout behaviors in health communities on SNS. To do so, we examine publicly shared longitudinal tweeting and activity data on individuals who self-identified suffering from ED on Twitter. As many SNS like Facebook and Instagram have taken moderation actions to counteract pro-ED content and user accounts, Twitter has not yet enforced actions to limit such content. This makes Twitter a unique platform to study the attrition process naturally happening in an ED online community and hence extend our baseline knowledge on attrition. Moreover, compared with other platforms like Facebook, Twitter provides a fairly anonymous environment of socialization, enabling people to be less concerned about self-presentation. This feature allows us to further examine individuals’ health-related characteristics by analyzing naturally occurring data in a non-reactive way.

In this study, we focus on measuring users’ dropout states based on their tweeting activities, since these activities can indicate the occurrence of dropout but also provide more details on when dropout occurs. In the light of well-recognized clinical evidence that emotional experiences and habits are the key trigger for abnormal behaviors among people with ED, our central research question is framed around the effects of individuals’ emotions on their dropout behaviors online. Emotions however often show a variety of physiological manifestations that can be characterized in different dimensions. Here, we focus on the Positive Activation - Negative Activation (PANA) model, since it provides overall better fits than other models (e.g., a circumplex model) and has been widely used in previous studies. In the PANA model, emotions are organized into two dimensions, i.e., positive and negative, based on their positive and negative valence ratings, and a neutral valence represents a medium level of arousal. Based on previous evidence that different linguistic constructs capture a wide range of psychological phenomena, such as happiness and sadness, we quantify individuals’ emotions through their language use in tweets. Meanwhile, psychological studies also show that emotional expressions are universal, regardless of language, culture or topic. In other words, the effects of emotions on dropout are comparable across different communities that are formed based on different interests. Thus, our second research question is whether emotions have varying effects on the dropout behaviors across disordered and general populations. Apart from cognitive factors, social dimension captured by social networks has also been shown to play an important role in influencing people’s concerns, behaviors and health conditions. Moreover, previous network interventions often identify community opinion leaders based on network structures, such as those with the greatest numbers of ties or central positions in local social networks, and train these opinion leaders as change agents to accelerate behavior change at community level. Hence, our last research question explores the effect of network attributes in terms of centrality on dropout in online communities. Specifically, this work makes the following contributions:

Using a hybrid methodology that integrates text mining and snowball sampling, we collect a group of users who have self-identified suffering from ED and their social networks on Twitter, leading to a connected community of individuals who are likely to exhibit ED on Twitter. We track tweeting activities for users from the ED community and two reference communities in two time periods spanning 1.5 year to obtain longitudinal data on attrition in online communities.

We use an instrumental variables (IV) econometric model to estimate the effects of users’ individual traits/characteristics (e.g., emotion and network centrality) manifested in the first period of observation on their dropout states in the second period. IV models consistently estimate causal relationships when controlled experiments are not feasible or endogeneity problems (e.g., measurement error, simultaneous causality and omitted variables) arise. We use the aggregated characteristics of a user’s followees on Twitter as instruments for potentially endogenous user’s individual characteristics.

We apply survival analysis methods to estimate active duration and time-to-dropout of our user populations on Twitter. Further, we incorporate IV techniques into survival models to achieve consistent estimation for the effects of users’ characteristics in early life (i.e., early stage of using a Twitter account) on their account longevity (i.e., how long the Twitter account keeps active in the future).

---

Followees are users whose Twitter activity the user in question follows; followers are users who follow the Twitter activity of the user in question.
Our findings show that ED users have higher dropout rates and shorter active durations on Twitter, compared with non-ED users. A half of our ED population is estimated to drop out in 6 months since the creations of their Twitter accounts, while those for non-ED populations are more than 79 months. We find that users’ emotions and online social networks in early life can affect dropout behaviors in the future, and these determinants function differently across disordered and general populations. In ED group, individuals who have more positive emotions are more likely to drop out and leave sooner, while non-ED people who have more positive emotions tend to stay active for a longer time on Twitter. Moreover, we find that individuals who are surrounded by many active peers and those who are located in the center of an ED community are less likely to drop out and will actively use Twitter for a longer time in the future. Our findings align with previous evidence in the psychology literature on ED and cognitive behavioral theory, and also provide new insights into the potential trajectories that ED communities develop online and can exacerbate ED risk factors through negative emotional content that promotes longevities of pro-ED users and enforces social ties among these users.

To our knowledge, this study is a first step to characterize dropout behaviors in online communities using naturally occurring longitudinal data on a large sample of individuals, laying out baseline knowledge for studying dropout behaviors on SNS. We incorporate psychological and social networking characteristics relevant to health behavior change, and demonstrate the need for research on dropout to move beyond socio-demographic and anthropometric factors as widely studied in previous work. Moreover, we present a practical approach to examine causal relationships in observational studies based on social media data, enabling the potential of user-generated data on SNS to understand underlying reasons why things happen and allow focused change activity. This work also raises important issues that need further research to enhance our understanding of attrition in online health communities, including consensus and clarity about the definition of dropout and its further refinement into different types, as well as many other factors related to dropout on SNS, such as individual personality, physical health states, perceptions and purposes of using a particular SNS platform. Our work has potential implications in using SNS like Twitter as a tool to conduct network interventions and promote organizational well-being.

2 Related Work

2.1 Clinical Studies on Eating Disorders

Eating disorders (ED) are often defined and diagnosed by certain physical criteria and identifiable behaviors around food, weight control, diet, and exercise. Growing evidence in psychological and clinical studies, however, has suggested that it is not really about food and weight at all — emotional issues lie at the core of ED. Difficulties in processing emotional states, particularly on negative emotions such as depression and anxiety, are implicated in the aetiology and maintenance of ED-related behaviors. A broad consensus in the field is that essential to recovery is learning how to recognize, regulate and healthfully express some of life’s most difficult and painful emotions. Given the role of emotions in developing healthy behaviors, psychologists have proposed an effective intervention called Cognitive Emotional Behavioral Therapy (CEBT) for treating ED. This therapy focuses on helping individuals to evaluate the basis of their emotional distress and hence reduce the need for associated dysfunctional coping behaviors, such as binging, purging, restriction of food intake, and substance misuse. Following this schema, we examine causality between past emotions and future dropout behaviors.

Although clinical studies are vital to learn about ED, most of them are carried out by surveys and interviews. These methods present two notable limitations. First, the denial of illness, ambivalence towards treatment and high dropout rates, as well as the social stigma of mental disorders make populations with ED hard to detect and reach. This may lead to the lack of quantifiable data to study ED. Even in cases where data can be collected, participants may conceal their condition and/or its severity, largely reducing the response accuracy and reliability of the data. Second, due to cost constraints, most surveys and interviews are conducted within small groups of individuals in a temporal granularity, which may not be representative of large populations (especially given the patients’ self-selection into such studies and trial) and miss longitudinal information.

Recently, people are increasingly using social networking sites (SNS) to express and exchange thoughts, develop social connections and relationships, or to document details of daily life. The use of SNS as a key health-information source and tool to manage chronic conditions has increased steadily among young people. Previous studies have shown that language use and socializing behaviors of people on SNS can indicate their feelings of worthlessness, depression, helplessness, anxiety and self-hatred that reflect mental disorders. Evidence has suggested that automatic analysis of data on SNS can offer an alternative and reliable tool to study a stigmatic and complex condition like ED. Compared with conventional data, analytics on SNS data has several advantages. First, the (semi-)anonymous nature of SNS encourages people to naturally socialize and self-disclose, enabling researchers to study individuals’ health problems by analyzing naturally occurring data in a non-reactive way. Second, the rich repository of SNS data can provide enormous and more fine-grained longitudinal features for identifying, tracking and predicting health risks at a large scale. Finally, data on SNS is often recorded in real time and preserved; analysis on such data alleviates the hindsight bias in retrospective analyses (e.g., survey-based methods), and may provide deeper insights into individuals’ physical and psychological states.
2.2 Online Intervention and Attrition

The rise of SNS has increasingly promoted people with health conditions to seek disease-related information and receive recovery-oriented support through online peer-communities \(^6\). These online communities play an important role in improving ailment recovery and coping in a variety of health challenges, ranging from diabetes to cancer \(^1\). Moreover, as online communities facilitate social connections and interactions among disordered peers \(^1\), recent studies on offline social networks show that people can be influenced by their social networks to adopt new conducts that affect their personal health \(^2\). Hence, community-oriented network interventions, which use online social networks among individuals to accelerate their behavior change and promote organizational well-being, have attracted increasing attention over recent years \(^3\). A widely used strategy in these sociometric network interventions is to identify community opinion leaders based on network attributes, such as those with the greatest numbers of ties or centrality, and train these opinion leaders as change agents to promote behavior change \(^4\).

However, attrition has been identified as a fundamental issue in the efficacy of online interventions \(^2\). As shown in a recent meta-analysis, all 22 studies examined suffer from decreased participation during intervention period, with 12 studies reporting rates of more than 20\% \(^5\). Yet, what determines individuals’ dropouts from an online community is still under-explored. Due to the strong presence of attrition across different systems \(^6\), identifying risks factors of individuals’ dropouts can help to target interventions correctly, pre-empt dropouts and improve overall health outcomes. Moreover, given the role of social networks on behavior change, it is also worth studying the effects of social network attributes (e.g., centrality) on individuals’ dropout behaviors. This work fills in this research gap by accessing the effect of network centrality on attrition in online communities.

2.3 Eating Disorders and Social Media

Studies on ED have been conducted on SNS data. Previous studies show that SNS harbor recovery-oriented communities, as well as pro-ED communities where people deny being disordered and instead promote ED as a lifestyle \(^7\). The pro-ED communities are easily accessible and can negatively impact the eating behaviors of people with and without ED, particularly through reinforcement of individuals’ identity on ED \(^8\), poor body image and thinness adoration \(^9\), or teaching harmful practices for weight loss \(^10\). Most prior studies focus on qualitative analyses and often involve intensive manual labor in data collection and analysis. However, the volume of user-generated content online is increasing explosively and this trend is likely to continue in the future. Thus, there is a need to devise more effective techniques to boost analyses on the large and rapidly increasing amount of data. This work contributes to this literature by developing computational techniques to automatically detect and quantitatively analyze ED-related communities on SNS.

On the other hand, most prior quantitative studies focus on identifying signs of ED or ED-recovery from user-generated content on SNS. Prior work examined the differences between pro-anorexia and pro-recovery content on Tumblr \(^11\), and the likelihood of a user recovering from ED measured from Tumblr posts \(^12\). Yom-Tov et al. \(^13\) studied social interactions between pro-anorexia and pro-recovery communities on Flickr \(^14\). Recently, ED-related tags have been intensively studied on Instagram, such as quantifying the severity of ED for users who posted ED-related tags \(^15\); examining the content moderation and lexical variation of ED-related tags \(^16\); and characterizing removed tags on ED \(^17\). Yet, limited studies have explored dropout behaviors in ED-related communities online.

2.4 Social Media Use and Mental Health

Another line related to our research is on the association between SNS use and mental health \(^18\). Some studies suggested an association between SNS use and the occurrence or exacerbation of mental health problems such as depression \(^19\), anxiety \(^20\), and ED \(^21\). However, others reported that there are no associations between mental illness and SNS use \(^22\); in some cases, significant improvements in social functioning are observed relating to engagement in SNS \(^23\). Such mixed findings in the field highlight the infancy of our understanding of the relationship between mental well-being and SNS use. Also, most prior studies focus on the use of self-report survey and rely on participants’ estimates of their behaviors on SNS, which may introduce considerable retrospective bias \(^24\). We extend this research by exploiting user-generated data on SNS in this work.

3 Data

We analyze a dataset collected from Twitter, a microblogging platform that allows millions of users to post and interact with short messages (“tweets”). Users can “follow” others to receive their updates, forward (“retweet” or “RT”) tweets to their own followers, or mention and reply (“@”) others in tweets. Tweets can be labeled with hashtags (“#”) to makes it easier for people to find tweets with a specific theme or content. All data used in our analysis is public information, available via the official Twitter APIs. We build our dataset as follows.

3.1 ED Data

To reach population with ED on Twitter, we adopt an approach used in previous work for detecting ED communities from SNS like Twitter \(^25\). We identify a user as ED-diagnosed if they self-report both ED-diagnosis information (e.g., “eating
disorder”, “edprob” and “proana”) and personal bio-information (e.g., body weight) in their profile descriptions (i.e., a sequence of user-generated text describing their accounts below profile images). We first 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 ED-diagnosed users as seed set and expand the set with 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 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. To verify the quality of our collected sample, we develop a labeling system by which the Twitter homepage of each user is automatically downloaded for inspectors. Inspectors annotate each user as to whether a user is suspected of having ED according to their posted tweets, images and friends’ profiles. Our annotation results on randomly selected 1,000 samples show that almost all of the checked samples are suspected of having ED and 95.2% of the samples are labeled as being highly likely to have ED.

3.2 Reference Data
We collect two sets of reference data for direct comparisons with ED users in behavioral patterns and characteristics. The first set of data serves as a baseline of the general population on Twitter. We build this dataset by a set of users at random, so called RD data. We first randomly sample 252,970 initial tweets via the Twitter APIs. To avoid sampling tweets on certain specific topics or from particular communities, these tweets are collected in three phases over two weeks. In each phase, only tweets written in English are collected. Second, from the unique authors of these initial tweets, 3,380 (the same number of core ED users) users are randomly selected. Third, to avoid preferentially sampling users who are very active on Twitter, we further crawl the friends of these users, resulting in 11,102,079 users. Finally, we retrieve historical tweets for these users. Due to the huge number of tweets, we stop this collection process after one-month running, obtaining 60,774,175 tweets of 30,684 users.

Since ED develop predominantly among young females3,45,84, we collect a set of young girls (YG) to further match well with ED users in age and gender. We first use the names of 14 popular artists2 as keywords to track a tweet stream and set the authors of these tweets as candidate users. This is motivated by the observation that popular music is always a hot topic of discussions among youngsters on Twitter. Candidates are then filtered by gender. To do so, we follow previous work85,86 by: (1) selecting candidate users that have given a full name in their profiles; (2) applying a lexicon-based method that identifies matches of the first name of each selected user to a dictionary of first names. Our name dictionary is built with the top 200 most popular first names for girls born in 2000s, as reported by the US Social Security Administration3. Further, we filter out accounts that have been verified so as to exclude celebrity friends of the listed artists. Next, we select 3,380 refined users as seed users and crawl the friends (only non-verified users with a female name are collected) for a random sample of YG users, resulting in 258,079 users. Finally, we collect 57,253,947 historical tweets for 37,983 unique users in one month.

3.3 Follow-Up Observation

The above data collection finished on Feb. 11, 2016 for ED users, Mar. 16, 2016 for RD users, and May. 16, 2016 for YG users respectively. We had a second observation for all three groups on Aug. 17, 2017. In the second observation, we only collect profile information that includes users’ last posted statuses for all users and their friends. Two observations afford us

---

3https://www.ssa.gov/oact/babynames/decades/names2000s.html
 flexibility in the subsequent analyses as discussed below. Figure 1 illustrates our two observations to collect longitudinal data on attrition in these three subpopulations.

4 Methods

We examine how user attributes exert causal influence on dropout behaviors on Twitter in two ways. First, we perform standard regression analysis on the whole sample to examine the effect of individual-level attributes (i.e., emotion) on the probability of dropout, i.e., using emotions that users exhibited in our first observation period to predict the occurrence of dropout in the second observation. Then, we perform survival analysis on sub-samples of users who have posting activities between our two observations to examine the effects of individual- and community-level attributes (i.e., emotion and network centrality in a social community) on the time to dropout, i.e., using users’ emotions and network centrality measured in the first observation to predict the duration of life in the second observation period6. Next, we present our methods in detail.

4.1 Effect of Emotion on Dropout

To study the effects of emotions on dropout behaviors, we need to eliminate the endogeneity problems (e.g., omitted variables, measurement error, and simultaneous causality) between the predictors and outcomes in regression analysis. An ideal approach to do this is a randomized controlled trial in which emotions are randomly assigned by investigators and emotional attribute is the only thing that differs among different individuals or groups87. Such an experiment, of course, is unfeasible due to ethical and practical issues. As an alternative, we use the instrumental variables (IV) approach which is widely used to achieve consistent estimates in a causal model containing endogenous variables88,89. The instrument is a variable that does not itself belong in the explanatory equation but is correlated with the endogenous explanatory variables, conditional on other covariates. Given a linear model, \( Y = \beta X + u \) for example, a valid instrument \( Z \) must satisfy two conditions:

- **Relevance Condition:** \( Z \) must be correlated with the endogenous explanatory variables \( X \), conditional on the other covariates, e.g., \( \text{cov}(Z, X) \neq 0 \). If this correlation is highly statistically significant, then \( Z \) is said to be a strong instrument.

- **Exclusion Condition:** \( Z \) cannot be correlated with the error \( U \) in the explanatory equation, conditional on the other covariates, i.e., \( \text{cov}(Z, U) = 0 \). In other words, \( Z \) cannot suffer from the same problem as the original predictors.

Given these properties, IV estimator captures only the effects on \( Y \) of shifts in \( X \) induced by \( Z \), but excludes the effects of endogeneity, enabling to consistently estimate a causal relationship90.

4.1.1 Variable Definition and Measurement

The following variables are defined and measured for the analysis:

- **Dependent Variable:** We code 1 (denoting dropout) for users who have no updated tweets, detect or set their account private in the second observation period; and 0 (denoting non-dropout) otherwise5.

- **Main Explanatory Variable:** The variable of interest is emotion. Given psychology evidence that different linguistic constructs capture a wide range of psychological phenomena, such as happiness and sadness38,39, we measure users’ emotions through their language use in tweets. To quantify emotions of tweets in a systematic, reproducible way, we adopt the widely used sentiment analysis tool for social media data — SentiStrength40. With removing mention marks, hashtags and URLs, each tweet is assigned with a scaled value in \([-4, 4]\) by SentiStrength, where negative/positive scores indicate the strength of negative/positive emotions respectively, and 0 indicates neutral emotions. Then, the averaged score over all tweets posted by a user is used as the user’s emotional variable. We exclude all retweets in this analysis, as retweets reflect more the emotions of their original authors than those of their retweeters. To obtain trustworthy results, we only consider users who have more than 10 tweets and post more than 50 words (widely used thresholds in computational linguistic analysis tools90) in our data.

- **Instruments:** As instruments for a user’s emotional attributes, we use the average emotions over all the user’s followees in our data. We use followees’ attributes for several reasons. First, the updates of followees act as information sources for a user, their original authors than those of their retweeters. To obtain trustworthy results, we only consider users who have more than 10 tweets and post more than 50 words (widely used thresholds in computational linguistic analysis tools90) in our data.

---

4. We exclude network attributes in the standard regression analysis. Some users might drop out a long time ago and their social networks might change a lot from the dates of their dropouts to the date of our first observation. For these users, network attributes are not applicable for a valid causality interpretation in the standard regression analysis, since the network attributes in the future cannot exert a causal influence on the dropouts in early life. Note that emotions are always measured from tweets before users dropped out, and hence the causality between emotions and dropout is interpretable in the standard analysis.

5 The activity of a user on Twitter can be indicated by other types of behaviors, e.g., following others. We focus on users’ tweeting activity because: 1. the timestamps when tweeting behaviors happened are accessible via Twitter APIs, allowing us to study not only the occurrence of dropout but also when dropout occurs; 2. the statistics on following behaviors (e.g., the numbers of followees and followers a user had at some stage) may not be able to accurately indicate the user’s activity. For example, even if a user has stopped using Twitter, the number of followees of the user can decrease because a followee deleted her/his own account; the number of followers can increase because this user has been followed by others.

6 http://liwc.wpengine.com/how-it-works/
and the followee’s behaviors as well as emotions manifested in their tweets can influence the user. Second, prior work\(^1\) has shown the presence of homophily among ED users on Twitter, i.e., users sharing similar emotional attributes tend to follow one another. These two facts suggest that the attributes of followees are correlated to a user’s attributes, satisfying the relevance condition of a valid IV. Finally, followees’ emotions are less likely to be directly correlated with a user’s outcome (i.e., decision to drop-out) other than through their effects on the user’s emotions, conditional on our controls as discussed below, which further satisfies the exclusion condition above.

**Control Variables:** As various factors (e.g., social capital and activity level) can relate to users’ dropouts, we condition our estimations on a number of variables, as listed in Table 1. To measure emotion change, we partition all tweets collected in the first observation for each user into two equal sets by chronological ordering. Then, we apply SentiStrength on these sets of tweets, resulting in two emotional states \(m_0\) and \(m_1\) for each user in an earlier and later stage respectively. Then, we define \(\Delta\text{emotion} = (m_1 - m_0) / \max(|m_0|, |m_1|)\), where \(\max(|m_0|, |m_1|)\) is reference value because it is unlikely to equal zero\(^1\). All variables on social capital, activity level and selection bias are measured from users’ profile information and tweets collected in our first observation. As discussed above, we hypothesize that followees’ attributes (e.g., emotions) are related to the dropout of a target user through their effects on the user’s emotions. However, followees’ emotions can also potentially effect followees’ own dropouts, and a growing feeling of loneliness due to friends’ leave then leads to the target user dropping out as well. To capture this alternative path, we further control the activity states of followees during our two observations.

![Table 1. Covariates used in regression models.](image)

<table>
<thead>
<tr>
<th>Control Objective</th>
<th>Covariate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotion Change</td>
<td>(\Delta\text{emotion})</td>
<td>Percentage change of emotions</td>
</tr>
<tr>
<td>Social Capital</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>Activity Level</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>Selection Bias</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>Alternative Causal Path</td>
<td>%Active followees</td>
<td>Proportion of followees being active between two observations</td>
</tr>
<tr>
<td></td>
<td>(&lt;\text{Followee durations}&gt;)</td>
<td>Average days of followees being active between two observations</td>
</tr>
</tbody>
</table>

4.1.2 IV Estimation

Suppose that one has observed independent and identically distributed data on \((Y, E, X)\) for \(n\) users, where \(Y\) denotes the decision to dropout, \(E\) denotes endogenous variables (i.e., users’ emotions here), and \(X\) denotes control variables. A linear model relating the decision to dropout to the user attributes on Twitter would be:

\[
Y = \beta_0 + \beta_1 E + \beta_2 X + U, \tag{1}
\]

where \(U\) is an unobserved error term and \(\beta_s\) are parameters to be estimated. Yet, the endogeneity problem discussed previously, i.e., \(\text{cov}(E, U) \neq 0\), biases such estimates. To produce consistent estimates, a two step approach called 2SLS is used, where in the first step an auxiliary linear regression of the instruments and exogenous variables on the endogenous variables runs.

\[
E = \gamma_0 + \gamma_1 Z + \gamma_2 X + V, \tag{2}
\]

where \(Z\) denotes the instruments, \(\gamma_s\) are estimable parameters and \(V\) is an error term. Following the estimation, predicted values for \(E\) are obtained (i.e. \(\hat{E} = \hat{\gamma}_0 + \hat{\gamma}_1 Z + \hat{\gamma}_2 X\)) and are used in the second step to replace the original endogeneous variables \(E\).

\[
Y = \beta_0 + \beta_1 \hat{E} + \beta_2 X + U, \tag{3}
\]

Intuitively, \(\hat{E}\) capturers the variation of \(Y\) due to the shifts of \(E\) only induced by \(Z\), not accounted for confounding bias. Since \(\text{cov}(Z, U) = 0\), we yield \(\text{cov}(\hat{E}, U) = \text{cov}(\hat{\gamma}_0 + \hat{\gamma}_1 Z + \hat{\gamma}_2 X, U) = 0\). In words, there is no endogeneity problems when estimating

\(^{1}\)Strictly speaking, followees’ updates posted after a user’s dropout should be excluded in measuring instrument for the user’s emotions, since these updates may never been seen by the user and could not effect the user’s emotions. However, it is difficult to determine the exact timestamp when a user dropped out from observational data, and hence select data that satisfy the criteria. Even if these timestamps are accessible, a smaller number of tweets may reduce the statistical power of text and lead to an untrustworthy measure for a followee’s emotions. On the other hand, we use the aggregated statistics of attributes for all followees, not only accounting for the statistics of a single followee, which can alleviate the influence of these issues and improve the robustness of results.
the effect of \( \hat{E} \) on \( Y \). Thus, \( \beta_1 \), for sufficiently large samples, delivers consistent estimates of the effects of interest. The 2SLS estimation with IV has been well studied and can be carried out by the R package AER\(^8\).

4.2 Effects of Emotion and Network Centrality on Time to Dropout

We next perform survival analysis to examine how users’ attributes exhibited in our first observation period effect time to dropout in our second observation period. Survival analysis methods are widely used to analyze data where the outcome variable is the time to an event of interest\(^4\). Applying such analysis can help to reveal the effect of emotion on dropout in more detail (e.g., duration to dropout), and further examine the effect of users’ network centrality in local communities on dropout.

4.2.1 Data Censoring

The key component in performing survival analysis is identifying duration in the data. In our research, if a user frequently posts tweets until the time our second observation ended, the survival time of this user is considered to be at least as long as the duration between our two observations. However, the user may experience dropout after the end of the second observation. That is, we know that the event of dropout did not happen when this user is under observation, while we do not know her/his exact survival time. The survival time of this user is said to be censored. We censor our data in the following way to obtain two-variable outcomes \((\delta_i, \tilde{t}_i)\) for each user, where \(\delta_i\) is the censoring variable denoting whether the event of interest occurs and \(\tilde{t}_i\) is the survival time denoting the length of time until the occurrence of an event.

Censoring results of user 1:

\[
\tilde{t}_1 = \Delta(t_{11}^{o1}, t_{11}^{p1}), \delta_1 = 1
\]

Censoring results of user 2 with identical-interval censoring:

\[
\tilde{t}_2 = \Delta(t_{12}^{o1}, t_{12}^{p1}), \delta_2 = 1
\]

Censoring results of user 2 with personalized-interval censoring:

\[
\tilde{t}_2 = \Delta(t_{12}^{o1}, t_{12}^{p1}), \delta_2 = 0
\]

Censoring results of user 3:

\[
\tilde{t}_3 = \Delta(t_{13}^{o1}, t_{13}^{p1}), \delta_3 = 0
\]

Figure 2. Illustration of data censoring. Each time axis starts from the timestamp of a user creating a Twitter account to the timestamp of our second observation, and blue bar denotes the survival time in the second observation. The middle two axises compare the results of identical-interval censoring and personalized-interval censoring methods for user 2. Since user 2 has no updated tweets during \([t_{22}^{p2} - \pi, t_{22}^{o2}]\), i.e., \(\Delta(t_{22}^{p2}, t_{22}^{o2}) > \pi\), \(\delta_2 = 1\) assigned by the identical-interval censoring method, denoting that user 2 drops out. In contrast, since the average posting interval of the user is longer than the inactive interval in the second observation, meaning that our observation period may not be longer enough, \(\delta_2 = 0\) assigned by the personalized-interval censoring method, denoting that user 2 is unobserved to drop out.

\(^8\)https://cran.r-project.org/web/packages/AER/index.html
Censoring Dropout Event: Our event of interest is a “dropout event”. A user is said to experience a “dropout event” if they have not posted any tweets for some time. Since people can reactivate their accounts after a long period of inactivity, it is difficult to certainly identify whether a user drops out from Twitter. Hence, we propose two censoring methods to estimate the occurrence of “dropout event”. Data obtained with both censoring methods are analyzed for cross validation. The first censoring method is: if a user has no any updated tweets in the latest several days before our second observation, we assume that the user drops out from Twitter or experiences the “dropout event”. Formally, for the $i$th user, the occurrence indicator of “dropout event” (i.e., the censoring variable $\delta_i$) in our first censoring method is:

$$
\delta_i = \begin{cases} 
1 & \text{if } \Delta(t_{i,2}^{p2}, t_{i,2}^{o2}) > \pi \\
0 & \text{otherwise}
\end{cases}
$$

where $t_{i,2}^{p2}$ is the timestamp of $i$’s last posting in our second observation and $t_{i,2}^{o2}$ is the timestamp of our second observation for $i$. $\Delta(t_i^1, t_i^2)$ counts the length of interval from time $t_i^1$ to time $t_i^2$. $\pi$ is the threshold of inactive interval. As each user is said to drop out if they have been inactive for the identical interval, we call this method identical-interval censoring.

However, people use the Twitter platform at different activity levels. Some users post a tweet at a relatively short intervals, e.g., every several hours or even every a few minutes. In contrast, some users post tweets at relatively long intervals (e.g., every a few days or several weeks). To capture this, we further take personalized posting intervals for different users into account and propose another method for censoring “dropout event” (called personalized-interval censoring). Unlike the identical-interval censoring, we use a linear combination of a constant interval and the average posting interval of each user as a threshold. For the $i$th user, the occurrence indicator of “dropout event” $\tilde{\delta}_i$ in our second censoring method is:

$$
\tilde{\delta}_i = \begin{cases} 
1 & \text{if } \Delta(t_{i,2}^{p2}, t_{i,2}^{o2}) > \lambda \pi + (1 - \lambda)\bar{I}_i \\
0 & \text{otherwise}
\end{cases}
$$

where $\bar{I}_i = \Delta(t_{i,2}^{p2}, t_{i,2}^{o2}) / N_i$ is user $i$’s average posting interval. $\Delta(t_i^{c}, t_i^{p2})$ counts the length of interval from $i$’s account creation to her/his last posting in our observations. $t_i^c$ is the time of $i$’s account creation, $t_i^{p2}$ is the time of $i$’s last posting, and $N_i$ is the number of all posts. Given $t_i^{c}$ and $t_i^{p2}$, the last posting time of $i$ in our first observation and the time of our first observation for $i$ respectively, the chronological order of these timestamps is $t_i^c < t_i^{o1} < t_i^{c} < t_i^{p2} < t_i^{o2}$. As discussed below, we estimate the parameters of $\pi$ and $\gamma$ based on the results of our first observation.

Estimating Survival Time: The survival time is the duration until a user experiences a “dropout event”. For users who experience a “dropout event”, survival time is the timespan from our first observation to the user’s last posting in the second observation. For users who never experience a “dropout event” until our second observation ended, survival time is equal to the entire timespan from the first observation to the second observation. Formally, the survival time of $i$th user $\tilde{\tau}_i$ is defined as:

$$
\tilde{\tau}_i = \begin{cases} 
\Delta(t_{i,1}^{o1}, t_{i,2}^{p2}) & \text{if } \tilde{\delta}_i = 1 \\
\Delta(t_{i,1}^{o1}, t_{i,2}^{o2}) & \text{if } \tilde{\delta}_i = 0
\end{cases}
$$

Figure 2 illustrates the possible cases where censoring occurs in this work.

4.2.2 Network Centrality

To measure users’ centrality in their local social networks, we build a who-follow-whom network among our collected users and their friends, where an edge in the network runs from a node representing user A to a node representing user B if A follows B on Twitter. We use the eigenvector centrality, since this metric (compared with degree centrality) produces more reliable centrality results by capturing the correlation of centrality between an ego and its friends. For validation, we also measure other widely used but different centrality metrics for directed networks — hubs and authorities using the HITS algorithm. A hub value of a user is the sum of the scaled authority values of the user’s followees; and an authority value of a user is computed as the sum of the scaled hub values of the user’s followers.

4.2.3 Survival Analysis with IV

We define a survival function to characterize the probability that a user will survive (i.e., not experience a “dropout event”) beyond any given specified time, i.e., $S(t) = Pr(T > t) = 1 - F(t)$, where $T$ is a random variable on the survival time with cumulative distribution function $F(T)$ on the interval $[0, \infty)$. We use the nonparametric Kaplan-Meier estimator to estimate and graph the survival function. This estimator gives univariate descriptive statistics for survival data, including survival curves and the median survival time at which one half of the entire cohort experiences an event. We use the log-rank test to assess overall differences between the estimated survival curves for different groups of users.

To gain more insight into the survival mechanism, we further examine the effects of users’ attributes on the time to dropout. The most popular techniques for such purpose is the Cox proportional hazards model, whose validity, however, relies on
the proportional hazards assumption that the ratios of hazard functions (i.e., hazard ratios) for different strata are constant over time. The Aalen’s additive hazards model offers a flexible alternative for modeling associations on the hazard scale\textsuperscript{98}. Unlike the proportional hazards model that estimates the hazard ratio, the additive model estimates hazard differences, i.e., the change in hazards due to the change of values for an explanatory variable. The hazards difference is measured in an additive way, which leads to two notable advantages. First, covariates act in additive manner on an unknown baseline hazard function, which can recover a marginal additive hazards model under fairly reasonable assumptions and lend the regression results to a natural interpretation as the excess hazards due to the effect of a covariate. Second, the computational methods for fitting additive hazards regression make it relatively easy to model effects of covariates over time\textsuperscript{99}. The additive hazards model has gained increasing attention and has been widely used in survival analysis across different fields\textsuperscript{100–102}. For our research purpose, we build an additive hazards model as follow. Suppose that one has observed independent and identically distributed data on \((\tilde{T}, E, X)\) for \(n\) users, where \(E\) is the endogenous variables (i.e., emotion and centrality), \(X\) is the control variables, and \(\tilde{T}\) is the time to event outcome. Let \(U\) denote the unobserved error terms. Then, an additive hazards model that estimates the effect of \(E\) on \(\tilde{T}\) is:

\[
h(\tilde{t}|E, X, U) = \beta_0(\tilde{t}) + \beta_1(\tilde{t})E + \beta_2(\tilde{t})X + \beta_3(U|E, X, \tilde{t}),
\]

where \(h(\tilde{t}|E, X, U)\) is the hazard function of \(\tilde{T}\) evaluated at \(\tilde{t}\), conditional on \(E, X\) and \(U\). \(\beta_0(\tilde{t})\) is the unknown baseline hazard function, while \(\beta_1(\tilde{t}), \beta_2(\tilde{t})\) and \(\beta_3(U|E, X, \tilde{t})\) are regression functions that measure the effects of their corresponding covariates. All of these functions are allowed to vary freely over time. The model posits that conditional on \(X\) and \(U\), the effect of \(E\) on \(\tilde{T}\) is linear in \(E\) for each \(\tilde{t}\), although, the effect size \(\beta_1(\tilde{t})\) may vary with \(\tilde{t}\). A sub-model is the partially-constant hazards model which can be obtained by setting \(\beta_1(\tilde{t}) = \beta_1\), where \(\beta_1\) is an unknown constant. We use this sub-model to measure and compare the effects of \(E\) on \(\tilde{T}\) in different groups of users.

However, as previously, the endogeneity problems can also bias the estimates of an additive hazards model\textsuperscript{101, 103, 104}. To obtain consistent estimations, we again use an IV method where user’s emotions and network centralities are instrumented by these attributes of user’s followees. For centralities measured by HITS, based on their definitions, we use the averaged authority of followers as instrument for a user’s hub value. Similarly, we condition our estimations on the covariates in Table 1, where the variable of <followee durations> is used to control the alternative causal path as discussed above.

To compute an IV estimator in the survival context, we use a method developed by Tchetgen et al.\textsuperscript{101}. This method is based on the control-function approach. Like 2SLS, it also has two separated steps, but adds the residual from a first-stage regression of the exposure on the IV to the additive hazards model. Specifically, similar to 2SLS, the first-step regression model is:

\[
E = \gamma_0 + \gamma_1 Z + \gamma_2 X + V,
\]

where \(Z\) is the IV, \(\gamma_1\)s are model parameters and \(V\) is the error term. Once we estimate the model parameter \(\hat{\gamma}\), we compute the residual errors as \(\hat{V} = E - \hat{\gamma}_0 + \hat{\gamma}_1 Z + \hat{\gamma}_2 X\). Then, we specify the linear projection of the error \(U\) in Eq. 7 on \(V\) as:

\[
\beta_3(U|E, X, \tilde{t}) = \rho(t) V + \epsilon(t)
\]

where \(\rho(t)\) is the regression coefficient, and \(\epsilon(t)\) is a random error independent of \((V, Z)\). The model makes explicit the dependence between \(V\) and \(U\), encoded in a non-null value of \(\rho(t) \neq 0\), and induces confounding bias. The residual error \(\epsilon(t)\) introduces additional variability to ensure that the relation between \(U\) and \(V\) is not assumed deterministic. Apart from independence with \((V, Z)\), the distribution of \(\epsilon(t)\) is unrestricted. Let \(h(\tilde{t}|E, X, Z, U)\) denote the observed hazard function of \(\tilde{T}\) given \((E, X, Z)\), evaluated at \(\tilde{t}\). Then, we plug Eq. 9 into Eq. 7 and have the following result:

\[
h(\tilde{t}|E, X, Z, U) = \beta_0(\tilde{t}) + \beta_1(\tilde{t})E + \beta_2(\tilde{t})X + \rho(t)\hat{V} + \epsilon(t),
\]

where \(\beta_0(\tilde{t})\) is a baseline hazard function, while \(\beta_1(\tilde{t})\) and \(\beta_2(\tilde{t})\) are regression functions. Intuitively, the residual \(V\) captures any variation in the hazard function due to unobserved correlates of \(E\), not accounted for in \(\gamma_0 + \gamma_1 Z + \gamma_2 X\). These unobserved correlates must include any confounders of the association between \(E\) and \(\tilde{T}\), and so \(V\) can be used as a proxy measure of unobserved confounders. For this reason, \(\rho(t)\hat{V}\) is referred to as a control function. For estimation, we use \(\hat{V}\) as an estimate of the unobserved residual \(V\) that we use to fit an additive hazards model, with regressors \((E, X, \hat{V})\) under Eq. 10. Thus, in the second stage, we use Aalen’s least-squares to estimate the following hazard model:

\[
h(\tilde{t}|E, X, U) = \beta_0(\tilde{t}) + \beta_1(\tilde{t})E + \beta_2(\tilde{t})X + \rho(t)\hat{V} + \epsilon(t)
\]

Inference about \(B(t) = (\beta_1(\tilde{t}), \beta_2(\tilde{t}), \beta_3(\tilde{t}), \rho(t))_T\) for such a model has been well studied and can be obtained using the R package TIMEREQ\textsuperscript{7}. Correct standard errors are obtained through non-parametric bootstrap.

\footnote{https://cran.r-project.org/web/packages/timereg/index.html}
5 Results

5.1 Descriptive Statistics
Table 2 lists the descriptive statistics for selected variables among three groups of users. The samples are limited to users who have followees with enough posts (i.e., more than 10 tweets and 50 words) in our data and these followees count for more than 1% of all followees of a user on Twitter, since they are the samples for whom we can estimate an IV model. We measure network centralities for ED users based on the following network of that contains 208,063 nodes and 1,347,056 directed edges, with all nodes connected in a single weakly connected component. Due to the random collection processes, many disconnected components exist in the who-follow-whom networks among RD \((n = 10)\) and YG \((n = 481)\) users, as well as their friends, respectively, suggesting the presence of multiple different communities in each of these groups. However, centrality values in a network with multiple isolated components may not be comparable since these values are related to the sizes of components\(^{10}\). Hence, we only study the centrality variable for ED users in the following analysis. The differences between dropouts and non-dropouts in each group of users are measured with Mann-Whitney U-tests and the Bonferroni correction is used to counteract the problem of multiple comparisons. Among 2,906 ED users, 85% of them have no tweeting activity during our two observations. This dropout rate is remarkably higher than those in RD (i.e., 20%) and YG (i.e. 42%) groups.

![Figure 3](image-url)

**Figure 3.** Number of users along with time points when users created an Twitter account and posted the last tweet in YG users (a) in ED users (b). (c) plots Kaplan-Meier estimates of lifetimes for ED and YG users on Twitter.

5.2 Lifespan Statistics
To examine when dropouts happened, we count the numbers of users along with time points when they created a Twitter account and posted the last time in our two observations. Figure 3(a)-3(c) shows the number of users who created an account and posted
Table 2. Baseline comparisons of dropouts and non-dropouts in each group along measures. <Followee x> denotes the average values of a user’s followees in terms of statistics x.

<table>
<thead>
<tr>
<th></th>
<th>Non-dropout</th>
<th>Dropout</th>
<th>U-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>RD (n = 2,980)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.24</td>
<td>0.37</td>
<td>0.22</td>
</tr>
<tr>
<td>ΔEmotion</td>
<td>-0.02</td>
<td>0.66</td>
<td>0.00</td>
</tr>
<tr>
<td>#Followees</td>
<td>991.94</td>
<td>3538.59</td>
<td>690.53</td>
</tr>
<tr>
<td>#Posts</td>
<td>25861.79</td>
<td>59797.51</td>
<td>19693.62</td>
</tr>
<tr>
<td>#Followers</td>
<td>19936.71</td>
<td>244306.24</td>
<td>1052.65</td>
</tr>
<tr>
<td>Active days</td>
<td>1181.95</td>
<td>766.74</td>
<td>758.20</td>
</tr>
<tr>
<td>#Followee/day</td>
<td>1.67</td>
<td>7.55</td>
<td>3.08</td>
</tr>
<tr>
<td>#Posts/day</td>
<td>31.68</td>
<td>98.17</td>
<td>35.50</td>
</tr>
<tr>
<td>#Followers/day</td>
<td>15.68</td>
<td>136.48</td>
<td>2.64</td>
</tr>
<tr>
<td>#Tweets in use</td>
<td>1195.52</td>
<td>980.98</td>
<td>1009.68</td>
</tr>
<tr>
<td>%Active followees</td>
<td>0.25</td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td>&lt;Followee durations&gt;</td>
<td>135.51</td>
<td>194.70</td>
<td>262.83</td>
</tr>
<tr>
<td>&lt;Followee emotions&gt;</td>
<td>0.36</td>
<td>0.27</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>YG (n = 1,984)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.42</td>
<td>0.41</td>
<td>0.49</td>
</tr>
<tr>
<td>ΔEmotion</td>
<td>-0.00</td>
<td>0.64</td>
<td>-0.06</td>
</tr>
<tr>
<td>#Followees</td>
<td>500.05</td>
<td>2922.55</td>
<td>111.83</td>
</tr>
<tr>
<td>#Posts</td>
<td>8589.86</td>
<td>50252.40</td>
<td>1078.07</td>
</tr>
<tr>
<td>#Followers</td>
<td>1426.54</td>
<td>14807.84</td>
<td>96.64</td>
</tr>
<tr>
<td>Active days</td>
<td>1257.88</td>
<td>836.84</td>
<td>731.83</td>
</tr>
<tr>
<td>#Followee/day</td>
<td>0.80</td>
<td>4.47</td>
<td>1.87</td>
</tr>
<tr>
<td>#Posts/day</td>
<td>7.80</td>
<td>43.26</td>
<td>2.26</td>
</tr>
<tr>
<td>#Followers/day</td>
<td>1.20</td>
<td>10.85</td>
<td>0.96</td>
</tr>
<tr>
<td>#Tweets in use</td>
<td>877.39</td>
<td>988.17</td>
<td>267.14</td>
</tr>
<tr>
<td>%Active followees</td>
<td>0.75</td>
<td>0.30</td>
<td>0.90</td>
</tr>
<tr>
<td>&lt;Followee durations&gt;</td>
<td>324.04</td>
<td>143.14</td>
<td>403.29</td>
</tr>
<tr>
<td>&lt;Followee emotions&gt;</td>
<td>0.42</td>
<td>0.30</td>
<td>0.32</td>
</tr>
<tr>
<td><strong>ED (n = 2,906)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td>-0.13</td>
<td>0.33</td>
<td>-0.08</td>
</tr>
<tr>
<td>Eigenvector</td>
<td>0.07</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>Hub</td>
<td>0.10</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>ΔEigenvector</td>
<td>-0.10</td>
<td>0.76</td>
<td>-0.12</td>
</tr>
<tr>
<td>#Followees</td>
<td>533.19</td>
<td>862.29</td>
<td>268.64</td>
</tr>
<tr>
<td>#Posts</td>
<td>2298.01</td>
<td>4566.24</td>
<td>645.42</td>
</tr>
<tr>
<td>#Followers</td>
<td>656.25</td>
<td>1422.54</td>
<td>244.87</td>
</tr>
<tr>
<td>Active days</td>
<td>732.17</td>
<td>524.86</td>
<td>278.40</td>
</tr>
<tr>
<td>#Followee/day</td>
<td>1.75</td>
<td>4.87</td>
<td>4.60</td>
</tr>
<tr>
<td>#Posts/day</td>
<td>3.85</td>
<td>5.82</td>
<td>4.51</td>
</tr>
<tr>
<td>#Followers/day</td>
<td>1.57</td>
<td>3.03</td>
<td>2.85</td>
</tr>
<tr>
<td>#Tweets in use</td>
<td>858.87</td>
<td>844.10</td>
<td>361.69</td>
</tr>
<tr>
<td>%Active followees</td>
<td>0.53</td>
<td>0.17</td>
<td>0.41</td>
</tr>
<tr>
<td>&lt;Followee durations&gt;</td>
<td>243.42</td>
<td>89.64</td>
<td>184.05</td>
</tr>
<tr>
<td>&lt;Followee emotions&gt;</td>
<td>0.01</td>
<td>0.16</td>
<td>0.06</td>
</tr>
<tr>
<td>&lt;Followee eigenvectors&gt;</td>
<td>0.07</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>&lt;Followee authorities&gt;</td>
<td>0.08</td>
<td>0.05</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Significance levels with Bonferroni correction: * p < 0.1; ** p < 0.05; *** p < 0.01.

the last post in our observations per year among RD, YG and ED groups. Two peaks in the curve of last-posting-time occur at
the dates of two observations for each group of users. One peak occurring at the first observation indicates that some users have been lost in the follow-up (e.g., accounts are deleted), and the other peak occurring at the second observation indicating that many users are still actively posting tweets until our observation ended. From the account creation times, we see that most ED users ($n = 1,944$ counting for 67% of all ED users) joined Twitter from 2011 to 2013. In contrast, there is no such intensive time slot for account creation among RD and YG users. Moreover, since we have not collected new users after the first observation, the number of users who created an account after the first observation is, by construction, zero.

Based on these creation and last-posting timestamps, we further count the number of months of duration from account creation to last-posting to examine users’ lifetimes on Twitter. We use the Kaplan-Meier estimator to estimate such lifetime, and censor each user by measuring whether a user has a post after our first observation. The censor is 1 (indicating “death”), if a user has no such tweet, and 0 otherwise. Figure 3(d) shows the Kaplan-Meier estimates for lifetimes of three groups of users.

We see that the lifetimes of ED users on Twitter are much shorter than those of RD and YG users. The median lifetimes for ED and YG users are 6 and 79 months respectively, while about 75% of RD users are still active in 125 month after they created an account. The lifetimes of three groups are statistically different ($p < 0.0001$) one another in a log-rank test.

5.3 Network Structure and Dropout

In combination with the above statistics, we present the who-follow-whom network among ED users in Figure 4. One notable feature is that users with similar dropout states are likely to cluster together. To examine the strength of such assortative mixing, we compute the assortativity coefficient $r$ for this network in terms of users’ dropout states. The resulting assortativity is 0.09, indicating that users with the same dropout states tend to befriend one another. To further test the statistical significance of the assortativity outcome within this particular network, we randomly shuffle users’ dropout states and re-measure assortativity 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 assortativity coefficients with a mean of $\mu = 0$ and a standard deviation of $\sigma = 0.005$. The $z$-score for the observed assortativity under this baseline distribution is $z = 16.84$ and $p < 0.001$, suggesting that the following network is significantly assortative mixed by dropout states.
5.4 Validity of Instruments and Censorships

To test the strength of our instruments, we assess the F-statistic against the null that the instruments are irrelevant in the first-stage regression of 2SLS on users’ emotions and centralities. The results are given in Table 3. In each group, the first column shows the results by using only followees’ emotions as instruments. The right two columns for ED users are obtained by setting both followees’ emotions and centralities (eigenvector and HITS centralities respectively) as instruments. In each model, the power of instruments exceeds the conventional standard of $F = 10^{.07}$, indicating strong relevance for our instruments.

![Table 3. F-statistics test on the instruments in the first-stage regression.](image)

To validate our methods on data censoring, we apply these censoring methods on users’ activity statuses before our first observation to estimate users’ dropout states and then check the agreement between these estimated dropout states and the observed dropout states in our second observation. Specifically, for the identical-interval censoring, if the period from a user’s last posting before the first observation to our first observation (i.e., $\Delta(t_i^u, t_i^o)$) is longer than $\pi$ days, the estimated dropout state for this user is 1, and 0 otherwise. Figure 5(a) shows the Cohen’s $\kappa$ between censored and observed dropout states for three groups of users, along with values of $\pi$ in identical-interval censoring methods. The largest $\kappa$ value is 0.73, achieved at $\pi = 82$ days. For the personalized-interval censoring, if the period from a user’s last posting before the first observation to our first observation is longer than $\lambda \pi + (1 - \lambda) \bar{t}_i^u$ (where $\bar{t}_i^u = \Delta(t_i^u, t_i^o)$/ is the average posting interval (counted in days), and $N_i^u$ is total number of tweets before the first observation), the estimated dropout state for this user is 1, and 0 otherwise. Figure 5(b) shows the Cohen’s $\kappa$ between censored and observed dropout states for three groups of users along with values of $\pi$ and $\lambda$ in the personalized-interval censoring method. Searching parameter space of $\pi \in [0,202]$ days and $\lambda \in [0,1]$, the largest $\kappa$ between observed dropout states and estimated dropout states by personalized-interval censoring methods is 0.73 as well. Such good agreement suggests validity of the censoring approaches.

To censor the dropout states of users in the second observation, we specify parameters that produce good and robust performance in the above evaluation, i.e., $\pi = 82$ days for the identical-interval censoring method and $\pi = 131$ days and $\lambda = 0.7$ for the personalized-interval censoring method. The Cohen’s $\kappa$ between the dropout states censored by two methods is 0.986, and there are 53 users to whom two censoring methods give different censorships. Figure 6 shows the Kaplan-Meier estimates for survival times of ED and YG users in the second observation period, where the samples are limited to users who have a post after our first observation. In both data censoring methods, the median survival time for ED users is 13 months. In contrast, by the end of our second observation, more than 75% of non-ED users are still active. The log-rank tests on both types of censorship show there is a statistical difference ($p < 0.0001$) between each pairwise group of users in survival times.
5.5 Effects of Emotion on Dropout

Table 4 shows the estimated effects of users’ emotions measured in our first observation and their dropout states in our second observation. In all models, only users’ emotions are studied as main explanatory variable and only average emotions of a user’s followees are used as instruments in IV models. From columns 1-2 for RD group, both ordinary least squares (OLS) and IV models indicate that emotions have no significant influence on the dropout outcomes. From columns 3-4 for YG group, the OLS results indicate that emotions have no significant influence on the dropout outcomes, while the IV results show that more positive emotions are significantly associated with lower probability of dropout. Columns 5-6 show results of two models for ED group. Both two models show that more positive emotions are significantly related to higher probability of dropouts. Moreover, the Wu-Hausman tests for both YG and ED groups indicate the existence of endogeneity.

Table 4. Estimated effects of emotion on dropout using OLS and IV models. Parentheses refer to standard errors.

<table>
<thead>
<tr>
<th></th>
<th>RD</th>
<th>YG</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS IV</td>
<td>OLS IV</td>
<td>OLS IV</td>
</tr>
<tr>
<td>Emotion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) (2)</td>
<td>(3) (4)</td>
<td>(5) (6)</td>
<td></td>
</tr>
<tr>
<td>(−0.024, 0.019)</td>
<td>(−0.035, 0.073)</td>
<td>(0.032, 0.023)</td>
<td>(−0.381, 0.098)</td>
</tr>
<tr>
<td>Observations</td>
<td>2.980</td>
<td>2.980</td>
<td>1.984</td>
</tr>
<tr>
<td>Wu-Hausman</td>
<td>0.021</td>
<td>22.46***</td>
<td>45.42***</td>
</tr>
</tbody>
</table>

* p < 0.1; ** p < 0.05; *** p < 0.01

5.6 Effects of Emotion and Centrality on Survival Time

Based on the observations for users who have posting activity past our first observation, we estimate the effects of emotion and network centrality in early life on survival time in the future. Table 5 lists the estimated coefficients and 95% confidence intervals. The upper part of the table compares standard and IV survival analyses using identical-interval censoring method. Columns 1-3 show results for RD, YG and ED groups respectively, given only emotional variable in question. We see that more positive emotions in early life are related to longer survival times in the future among both RD and YG groups, in both standard and IV survival models. In contrast, as shown in column 3, more positive emotions are related to shorter survival times among ED users. These results are compatible with our previous comparisons in Table 4. While there is no a significant association between emotion and dropout for RD users in the standard regression analysis, these survival analyses show that emotion influence time to dropout in RD users. Such difference can be explained by two reasons. First, unlike selecting YG and ED users based on their names and body-related information, our collection methods for RD users have no filtering processes. This may have included some users who use Twitter purely for business purposes (e.g., company and organization accounts).
Table 5. Survival analysis to estimate the effects of emotion and centrality in the first observation on survival times in the second observation under Aalen additive hazards models. Parentheses refer to 95% confidence intervals obtained from 1,000 bootstrap replicates. All results significant at \( p < 0.05 \) are labeled with an asterisk.

<table>
<thead>
<tr>
<th></th>
<th>RD (1)</th>
<th>YG (2)</th>
<th>ED (3)</th>
<th>ED (Eigen) (4)</th>
<th>ED (HITS) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identical-interval</strong></td>
<td><strong>Censoring</strong></td>
<td><strong>Standard Survival Analysis</strong></td>
<td><strong>Survival Analysis with IV</strong></td>
<td><strong>Standard Survival Analysis</strong></td>
<td><strong>Survival Analysis with IV</strong></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.015*</td>
<td>0.028*</td>
<td>-0.054*</td>
<td>-0.046*</td>
<td>-0.041*</td>
</tr>
<tr>
<td></td>
<td>(0.012, 0.017)</td>
<td>(0.023, 0.033)</td>
<td>(-0.075, -0.036)</td>
<td>(-0.065, -0.029)</td>
<td>(-0.059, -0.024)</td>
</tr>
<tr>
<td>Centrality</td>
<td></td>
<td>0.154*</td>
<td>0.149*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.110, 0.219)</td>
<td>(0.111, 0.205)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.038*</td>
<td>0.038*</td>
<td>-0.133*</td>
<td>-0.114*</td>
<td>-0.104*</td>
</tr>
<tr>
<td></td>
<td>(0.032, 0.043)</td>
<td>(0.033, 0.044)</td>
<td>(-0.179, -0.091)</td>
<td>(-0.158, -0.076)</td>
<td>(-0.151, -0.071)</td>
</tr>
<tr>
<td>Centrality</td>
<td>0.134*</td>
<td>0.138*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.083, 0.201)</td>
<td>(0.093, 0.193)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Personalized-interval</strong></td>
<td><strong>Censoring</strong></td>
<td><strong>Standard Survival Analysis</strong></td>
<td><strong>Survival Analysis with IV</strong></td>
<td><strong>Standard Survival Analysis</strong></td>
<td><strong>Survival Analysis with IV</strong></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.014*</td>
<td>0.025*</td>
<td>-0.053*</td>
<td>-0.045*</td>
<td>-0.039*</td>
</tr>
<tr>
<td></td>
<td>(0.012, 0.017)</td>
<td>(0.021, 0.030)</td>
<td>(-0.071, -0.035)</td>
<td>(-0.064, -0.027)</td>
<td>(-0.058, -0.023)</td>
</tr>
<tr>
<td>Centrality</td>
<td></td>
<td>0.153*</td>
<td>0.149*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.110, 0.211)</td>
<td>(0.109, 0.202)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.035*</td>
<td>0.035*</td>
<td>-0.130*</td>
<td>-0.109*</td>
<td>-0.100*</td>
</tr>
<tr>
<td></td>
<td>(0.031, 0.041)</td>
<td>(0.030, 0.040)</td>
<td>(-0.171, -0.091)</td>
<td>(-0.151, -0.069)</td>
<td>(-0.143, -0.058)</td>
</tr>
<tr>
<td>Centrality</td>
<td>0.136*</td>
<td>0.139*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.085, 0.196)</td>
<td>(0.094, 0.189)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2,376</td>
<td>1,160</td>
<td>447</td>
<td>447</td>
<td>447</td>
</tr>
</tbody>
</table>

and hence undermine the statistical significance of the results on RD users. Second, survival analyses model the duration to dropout more than solely the occurrence of dropout, which can capture finer-grained characteristics of different users.

Columns 4-5 show co-effects of emotion and network centrality (eigenvector and HITS centralities respectively) on survival times of ED users. The effects of emotion estimated in these models are in line with the results in the previous model-3 in which only emotional variable is studied. In terms of centrality, both models using two measures of centrality show higher centrality in early life is related to longer survival time in the future for ED users. In addition, analyses on personalized-interval censored data show similar results.

5.7 Link between Emotion and Dropout

To understand how emotion effects dropout in ED users through a channel that is different from those in non-ED users, we investigate the difference of interests for ED users with different dropout and emotional states. Our investigation approach builds on previous findings that community interest is the primary motivating factor for participation in online communities\(^{108}\) and people’s interests connect their emotional states\(^{109}\). We examine users’ interests based on their posted hashtags, since hashtags are explicit topic signals on Twitter and have been shown to present a strong indication of users’ interests\(^{110}\).

First, we give a glimpse of prevalent topics in our collected ED community. We construct undirected, weighted hashtag networks 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 nodes. To filter out noise from accidental co-occurrence 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 7 shows the co-occurrence network of the most popular hashtags in ED users. We see that topics on promoting thin ideals are prevalent in the community. Subsequently, we examine the difference of hashtag usage between users with different dropout states. We split ED users into two sets based on their dropout states in our second observation period, and extract hashtags from tweets posted by each set of users. Again, we exclude tags that are used by less than 50 users and occur in less than 50 tweets in each set. To adjust tags that are more popular in general, we use the ratio between the number of users using a tag in a set...
Figure 7. Co-occurrence network of hashtags from tweets of ED users. Each node is a hashtag, and node size is proportional to the number of users who used 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.

Table 6. 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, ednos, ed, thighgap, weightloss, skinny, proed, selfharm, perfection, mia, thinspiration, perfect, proana, diet, eatingdisorder</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, anasister, anafamily, starveon, gross, teamfollowback, fuck, icandothis, tired, edfamily, relapse, stayingstrong</td>
</tr>
</tbody>
</table>

Table 7. 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, anaprobz, failure, flatstomach, fact, binge, fatty, fasting, suicide, depressed</td>
</tr>
<tr>
<td>Positive</td>
<td>eatclean, litfam, inspiration, reasonstobefit, ff, noexcuses, fitness, loveit, winning, anasister, tweetwhatyoueat, twye, keepgoing, success, jealous, want, fitspo, beforehandafter, retweet, excited, proud, reasonstoloseweight, abcdiet, fail, justsaying, rt, motivated, workout, stayingstrong, love, myfitnesspal, fact, tmi</td>
</tr>
</tbody>
</table>

Table 8. Similarities between hashtags from tweets 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></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropout</td>
<td>0.797</td>
<td>0.486</td>
<td>0.451</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.269</td>
<td>0.476</td>
<td>0.485</td>
</tr>
</tbody>
</table>

and that in all sets to rank the specificity of a tag in each set of users. Table 6 lists the most representative hashtags for users with different dropout states. We see that non-dropout users are more interested in advocating pro-ED lifestyle choice, e.g., promoting a thin ideal and reinforcing a pro-ED identity. In contrast, dropped-out users are more interested in discussing their health problems and offer emotional support to other peers. Similarly, we split ED users into three equal-size sets based on their emotional states, and obtain the most representative hashtags for users with different emotional states in Table 7. We see that users with negative emotions are also more interested in promoting thin ideals, while users with more positive emotions are more interested in promotion of healthier body image and behaviors. To quantify the association between users’ dropout and emotions in terms of hashtag usage, we calculate the Jaccard similarity of hashtags used by users with different dropout and
emotional states in Table 8. The results suggest that users with more positive emotions are more similar to users who dropped out than those who did not drop out in terms of their interests in hashtags.

Taken together, individuals with more positive emotions have more healthier body image and concerns, and hence tend to leave a harmful online community in which members deny ED being illnesses and instead promote ED as a lifestyle choice. Such evidence may explain why more positive emotions lead to higher dropout among ED users in our regression analyses.

6 Discussion

We examined characteristics of attrition in online health communities, particularly ED-related communities on Twitter. Compared to the general populations, people who self-identified with ED exhibit higher dropout rates and shorter active durations on Twitter. Users’ emotions and social networks in early life can effect dropout behaviors in the future, and these determinants function differently across disordered and general populations. Disordered people who have more positive emotions tend to drop out sooner, while general populations who have more positive emotions tend to stay active for a longer time on Twitter. Moreover, individuals surrounded by many active peers and those located in the center of ED communities are less likely to drop out. Our results provide new insights into the potential trajectories that ED communities develop online and can exacerbate ED risk factors through negative emotional content that promotes longevities of pro-ED users and enforces social ties among these users. The observations from our approach have potential implications of using a SNS platform like Twitter as a tool to conduct network interventions and promote organizational well-being.

Although high attrition is common in Internet-based interventions for challenging health problems like ED and has become a crucial issue for these programs to achieve positive outcomes. Clinical studies on attrition focus on comparative analyses of risk factors (e.g., socio-demographic and anthropometric attributes) between dropouts and completers of participants in an intervention. This may thus produce biased results in favor of those who actively seek treatment. Our findings suggest that early dropout is also a considerably prevalent phenomenon in individuals with ED on general SNS like Twitter. In other words, attrition in an online intervention can be related to a faulty of intervention programs (e.g., content, protocol or intensity), but is also related to some inherent characteristics of mental disorders (e.g., feelings of shame or fear of stigma). Distinguishing risk factors of the two aspects is important to implement alternative strategies that can improve retention and ultimately health-related outcomes. Considering individuals who have the same set of symptoms but have not experienced a wide range of online interventions in online health communities as a control group for these treatment groups, our findings provide insights into understanding attrition in a natural setting.

Despite significant past work attempting to identity predictors to dropout, most prior studies rely on retrospective analyses of risk factors obtained from self-identified clinical in-patients and may therefore introduce retrospective bias, self-selection bias and measurement errors that can invalidate causal inference. Our findings suggest that social media research can complement clinical and psychological studies of dropout behavior by providing new information that is hard-to-access otherwise. Our work opens up a promising opportunity of using an unobtrusive data source like social media to not only understand how and what characteristics of individuals may promote or hinder dropout behaviors online, but also quantify prospective risks based on naturally occurring behavioral data for a large sample of individuals in a non-reactive way.

We find that positive emotional content in early tweet written is strongly predictive of the presence of attrition at a later stage, which is consistent with evidence in the cognitive behavioral theory and highlights the impact of emotions on behaviors. We provide further insights into how the underlying mechanisms can vary across different populations. In the same way that positive emotions effect people’s offline longevity, more positive emotions in early lifetime are found to increase longevity online among non-ED samples. Conversely, more positive emotions increase likelihood of dropping out from an online community full of pro-ED content and decrease online longevity for individuals with ED. Those with more positive emotions tend to use online communities to discuss their illness and to support peers, showing positive impact of online social communities for promoting health outcomes. In contrast, those with more negative emotions tend to promote ED as a lifestyle choice rather than an illness, suggesting a negative impact of these communities by exacerbating ED risk factors, such as thin ideals, reinforcing a pro-ED identity, or exposing and adopting harmful weight loss/control practices. These results also imply that less emotional distress is linked to a lower risk to learn and develop dysfunctional coping behaviors among individuals suffering from ED, in line with practical evidence in CBT for treating ED.

In addition to internal and psychological determinants, whether an individual drops out from pro-ED communities also depends on whether others in the individual’s social networks drop out. Active users who have not dropped out tend to be located in the center of their local social networks and in large clusters of other active users. These results are in line with previous evidence on social contagion effects on human behavior, and suggest that dropout is not only a function of individual experience or individual choice but also a property of group interaction. Also, core group members tend to have further long-term engagement in online pro-ED communities at a later stage than peripheral members, which can increase the chance of core members being exposed to pro-ED content and exacerbate their unhealthy behaviors. This result is even more
remarkable considering that unhealthy information shared by these core users can reach broad audiences and raise long-standing public health concerns given these are users with large numbers of followers and long lifespans online.

Our findings have practical relevance for promoting public health over SNS. First, by measuring emotions from user-generated content online, health professionals and researchers might be able to identify and direct efforts towards disordered individuals most likely in need of support and treatment. Second, to the extent that clinical or policy maneuvers raise positive emotions of individuals, they might have a lower level of engagement in an unhealthy community and decreased exposure to harmful content online. In other words, apart from education programs (e.g., raising health awareness or instructing behavioral restructuring)\textsuperscript{111} and moderation practices (e.g., deleting content or banning users)\textsuperscript{112}, increasing exposure to positive emotions for people with ED, such as interjecting with pop-up messages containing inspirational information in their SNS homepages, might be useful to improve their health statuses. Such an understated way is not only good for cultivating positive emotions to optimize health and well-being\textsuperscript{113}, but also prevent users from feeling shame, stigma or offense to trigger an avoidant or rebellious response to the intervention (e.g., early dropout, relapse and even deterioration)\textsuperscript{114,115} and counterproductive outcomes. Third, given that a member of online pro-ED communities dropping out might have cascade effects on others, online intervention should be delicately crafted to avoid a large scale of attrition, thereby facilitating positive health outcomes. Finally, intervention strategies might be tailored for different individuals depending on their network role. Identifying key players for training might enhance the efficacy and cost effectiveness of an intervention due to their greater influence potential on larger numbers of social ties\textsuperscript{16,17}, but also their longer-lasting effect through longer-term involvement in the community.

We acknowledge that our study has some 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 “edprob” 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 is used to explore rich social media data on the attrition of ED communities on Twitter rather than produce generalizable results 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 understand the characteristics of their dropout behaviors. 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 and its further refinement into different types. Fourth, we have limited external validity on other SNS platforms, while stopping usage of a platform can be related to the general features and attraction 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 SNS, such as individual personality, physical health states, perceptions and purposes of using a particular SNS platform. Finally, SNS profiles are not identical and nonrenewable identities. Individuals who dropped out from an online community in our observations may have other profiles on Twitter or other cyberspaces. We cannot be sure whether they will migrate to a similar or different community in the future.

References

27. Gow, R. W., Trace, S. E. & Mazzeo, S. E. Preventing weight gain in first year college students: an online intervention to prevent the “freshman fifteen”. *Eat. behaviors* 11, 33–39 (2010).


**Acknowledgements**

This work is supported by ESRC Doctoral Training Centre (NO. ES/J500161/1), Institute for Life Sciences, WSI-RCSF and SocSCI-SIRF, University of Southampton, UK.

**Author contributions**

T.W., E.M., A.I. and M.B. designed research; T.W. and E.M. analysed data and wrote the manuscript; all authors contributed to revisions.