A SENTIMENT ANALYSIS OF THE PHD EXPERIENCE EVIDENCED ON TWITTER

Panchali Guha*  
Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore  
panchali.guha@u.nus.edu

Devyani Pande  
Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore  
devyani@u.nus.edu

* Corresponding author

ABSTRACT

Aim/Purpose  
This article explores the content of PhD student tweets. It has three main aims: (a) to examine what is discussed regarding the PhD process, (b) to evaluate whether tweets express primarily positive or negative sentiments, and (c) to uncover the key themes discussed in both positive and negative tweets.

Background  
Recent surveys of PhD students have raised concerns about their wellbeing by pointing to high prevalence rates of adverse mental health conditions. However, our understanding of which factors pose the highest risks is still evolving. Self-selection into surveys also raises the possibility of discounting positive aspects of the PhD experience. We use a different data source (Twitter) to explore both these issues.

Methodology  
Using 16,928 tweets with the Twitter hashtags #phdlife and #phdchat, we first conduct dictionary-based sentiment analysis in R to determine whether tweets are dominated by positive or negative sentiment. We then hand-code the dominant sentiment of a randomly selected subset of 1,994 tweets and qualitatively analyse positive and negative tweets separately to uncover the key themes in each category.

Contribution  
This article contributes to the emerging literature on the wellbeing and mental health of PhD students by using a novel data source (tweets). It highlights both positive and negative aspects of the PhD student experience.

(C) BY-N C 4.0 This article is licensed to you under a Creative Commons Attribution-NonCommercial 4.0 International License. When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.
Findings
We find that most tweets express positive rather than negative sentiment, indicating that PhD students do enjoy many aspects of their experience. Negative tweets are dominated by mental health concerns. They also highlight problems with academic culture (especially the normalization of overwork) and the effects of the pandemic on students.

Recommendations for Practitioners
Our results indicate that there is a need to change the academic culture of normalizing overwork, ensure adequate institutional provision of mental health support and ability to spot signs of emotional distress, devise strategies to combat the imposter phenomenon, and respond to the particular challenges that the pandemic has created for PhD students.

Recommendations for Researchers
The authors recommend that future research explore the specific challenges and opportunities faced by PhD students in different disciplines and geographical locations. As the data used here were collected during the pandemic, it would be useful to track post-pandemic sentiments to observe changes.

Impact on Society
Improving the graduate experience of PhD students and providing them adequate mental health support will help to ensure their continued productivity and wellbeing.

Future Research
Future research in this area should focus on the efficacy of different interventions to address key problems, such as the imposter phenomenon, stress, anxiety, depression, and isolation.

Keywords
PhD, Twitter, sentiment analysis, mental health, #phdlife, #phdchat

INTRODUCTION

Concerns about the wellbeing of PhD students have been raised in recent years. A number of surveys have found that the prevalence of adverse mental health conditions such as stress, anxiety, and depression is much higher among PhD students relative to the general population, leading to a spate of sombre and alarming headlines such as “The mental health of PhD researchers demands urgent attention” (Nature, 2019) and “For some graduate students, the cost of doing science is their mental health” (Yu, 2020). While the increased research attention to this important issue is very welcome, there are gaps in the current state of knowledge. For one, there is imperfect understanding of the underlying factors causing high incidence of mental health issues among PhD students, especially of which factors pose the highest risks (Mackie & Bates, 2019). At the same time, some scholars suggest that inadequate attention is paid to the positive aspects of the PhD experience. Byrom et al. (2020) found that PhD students rated many aspects of their experience positively and emphasized that highlighting these positive experiences is important to counter the emerging cultural narrative that successful completion of doctoral research inevitably requires compromising on well-being.

In this article, we aim to address both these issues by examining the content of PhD related tweets. With about 313 million active users (Malik et al., 2019), Twitter is a popular microblogging site where users can communicate using a maximum of 280 characters (this was increased from 140 characters in 2017). Tweets can include hashtags, which consist of the symbol # followed by a keyword or unbroken phrase, to link them to all other tweets on the same theme or topic. PhD students have used Twitter hashtags such as #phdchat, #phdlife and #phdadvice to form online communities that provide information, advice, and support, aid professional socialization, reduce feelings of isolation, and enable interaction without time and space constraints (Bennett & Folley, 2014; Ferguson & Wheat, 2015; Zhu & Procter, 2012). Over the last decade, the analysis of Twitter data has emerged as a promising way to explore public opinion on a variety of issues ranging from politics (Lopez et al.,
Twitter users often tweet in a naturalistic way about their day-to-day experiences and feelings (Lachmar et al., 2017), making Twitter a good source of data for examining sentiments and attitudes. Research has shown that people experiencing mental health issues such as depression often find it useful to talk about them on Twitter as it helps them to connect with others experiencing similar issues and provides a sense of community (Berry et al., 2017). This suggests that it is also a promising avenue for information about the mental health challenges faced by PhD students.

With this backdrop, our research question for this study is: What sentiments are expressed through tweets regarding the PhD journey? Using a sample of tweets collected from the Twitter hashtags #phdlife and #phdchat, we (a) examine what is discussed regarding the PhD process, (b) evaluate whether tweets express primarily positive or negative sentiments, and (c) explore the key themes discussed in positive and negative tweets. Most research studies on the wellbeing of PhD students are survey-based. Here, we use a different data source (tweets) to investigate these issues and complement the results of PhD student surveys.

The article is organized as follows. We provide a brief review of the literature on the use of Twitter by scholars and on the sentiments and wellbeing of PhD students in the next section. This is followed by a methodological description of how we collected the Twitter data and conducted the sentiment and thematic analyses. Next, we present our results. The sentiment analysis shows that the majority of tweets express positive rather than negative sentiments. Our thematic analysis of positive tweets reveals that PhD students use Twitter to share updates about work and life, give and receive advice, appreciate others, make connections with others, share their feelings and achievements, and express their interest in research and analysis. Our thematic analysis of negative tweets, on the other hand, suggests that students struggle with mental health concerns, overwork and fatigue, difficulty of the work, problematic aspects of the academic culture, and – currently – COVID-related issues. The final two sections conclude by discussing the implications of the findings, potential limitations of the research methodology, and recommendations for practice and future research.

**LITERATURE REVIEW**

Social media platforms enable individuals to create a public or semi-public profile within a system, communicate with other users and further connect with their connections within the system (Boyd & Ellison, 2007). Twitter has been a popular medium of communication amongst academics. Most of them use it for professional purposes that range from tweeting during conferences to disseminating research to discussions on books and scientific papers (Sugimoto et al., 2017). We have considered the social media platform of Twitter for our study because majority of the users identify professionally on the platform (Bowman, 2015) and due to the ease of data availability through the application programming interface (API) (Veletsianos et al., 2018).

In order to review relevant literature, we used the keywords ‘PhD’, ‘doctorate’, ‘students’, ‘Twitter’, ‘mental health’ with Boolean operators like AND and OR in SCOPUS database. There have been no studies that specifically examine sentiments expressed about the PhD process on Twitter. The closest study of this kind was conducted by Gomez-Vasquez and Romero-Hall (2020), examining the demographics and sentiments of users on #AcademicTwitter, which is a hashtag used by graduate students, educators, researchers, professionals, and others. Their sentiment analysis concluded that the majority of tweets expressed a positive sentiment, and that conversations were dominated by a “call-to-action language”, using active words such as give, make, check, learn, and find. There is however a rapidly expanding corpus of studies that attempt to gauge the sentiments and wellbeing of PhD students using survey-based methods. The prestigious journal Nature, for example, administers biennial surveys to graduate students worldwide. Overall, these surveys find that positive sentiments generally outweigh negative ones: 70 percent of respondents were ‘satisfied’ or ‘very satisfied’ with their graduate school experience in 2015 (Woolston, 2015), and 75 percent or more were at least somewhat satisfied with their decision to pursue a PhD degree in 2017 and 2019 (Woolston, 2017, 2019). By far the
most important stated reason for satisfaction was the intellectual challenge of graduate study. Other important reasons included the ability to work with interesting and bright people, enjoyment of the academic environment, and the ability to think creatively (Woolston, 2019). Most students also reported positive experiences overall in a survey of PhD students in the UK (Byrom et al., 2020). Most had good relationships with their supervisors and regarded them as mentors and academic and professional assets, felt well prepared for their academic work, had a good understanding of what they were supposed to be doing, were on track to complete on time, and had a good academic support network.

At the same time, most surveys of PhD students also indicate a high incidence of adverse mental health concerns at rates that exceed those prevalent in the general population. For instance, Levecque et al. (2017) found that more than half of their sample of PhD students in Belgium had experienced at least two mental health symptoms in recent weeks and were 2.43 times more likely to have or develop a common psychiatric disorder relative to the highly educated segment of the general population. Evans et al. (2018) found that the prevalence of moderate to severe depression, and of moderate to severe anxiety, was 41 percent and 39 percent among graduate students from 26 countries, compared to prevalence rates of 6 percent for both conditions in the general population. A US survey of PhD students found that 11 percent of students had suicidal thoughts, and that the incidence of depression and anxiety symptoms resembled that of incarcerated populations (Barreira et al., 2018). Woolston (2017) reported that 12 percent of science PhD students worldwide had sought help for anxiety and depression stemming from their PhD studies. There are variations by demography and program type: Allen et al. (2020) found that female students were more likely to report mental health problems than males, and students in professional doctoral degrees such as MD were significantly more likely to report experiencing stress and anxiety than students in academic doctoral degrees.

Interestingly, the prevalence of mental health issues often appears to coexist with positive sentiment towards graduate studies. Almost half of those who reported seeking help for anxiety and depression in the 2017 Nature survey, for instance, also reported being satisfied or very satisfied with their PhD program. Similarly, while Byrom et al. (2020) found that overall mental wellbeing of students in their sample was low, with a quarter of respondents reporting levels that indicated probable depression or anxiety, there was a preponderance of positive reports overall. This is a puzzling observation, but a potential explanation might be that adverse mental health symptoms are triggered by a variety of factors, only some of which operate at the faculty, department, or program level. Students who do not experience problems at program level may have positive sentiments towards the PhD process, but are still vulnerable to developing mental health conditions due to more systemic or psychological factors.

Common issues at department level include problems in the supervisory relationship, inadequacy of resource provision by universities, and ambiguity, lack of transparency, and closed decision-making in university rules and processes pertaining to PhD students (Byrom et al., 2020; Mackie & Bates, 2019). Accordingly, Hunter and Devine (2016) concluded that feeling supported and appreciated by the department, working with more experienced supervisors, and having more frequent meetings with supervisors contributed to higher wellbeing in students. Higher-level triggers include heavy workloads, the difficulty of maintaining work-life balance, low control over the work environment, the difficulty of accumulating research expertise and domain-specific expertise in a timely way to meet candidature requirements, lack of social support, and financial insecurity (Barry et al., 2018; Evans et al., 2018; Levecque et al., 2017; Mackie & Bates, 2019; Woolston, 2019). Uncertainty about future career prospects and directions is a particularly important stressor, particularly in view of sharp increases in PhD student numbers globally (Mackie & Bates, 2019; Marais et al., 2018; Woolston, 2019). Related to this is the strong focus on performance measurement: students are anxious about achieving high performance metrics on publications, funding, and impact in the early stages of their career (Nature, 2019). It is common for PhD students to struggle with rejection (for instance, of their manuscripts and funding applications) and experience feelings of isolation because people outside academia do not understand the challenges they are experiencing (Brown, 2013).
Finally, personal characteristics and situations also matter. Barry et al. (2018) found higher prevalence of stress, anxiety, and depression among students who are behind schedule, as well as students who are ahead of schedule, relative to those who are on schedule for PhD completion. Marais et al. (2018) concluded that students driven mainly by intrinsic or autonomous motivation are more likely to experience high levels of wellbeing, while those who are driven mainly by internal pressure to succeed are more likely to experience higher levels of stress, depression, and anxiety.

**METHODS**

**THE DATA**

We collected raw twitter data by searching Twitter’s API using the rtweet package (Kearney et al., 2020) in R (R is a free software environment for statistical computing and graphics) for tweets containing the hashtags #phdchat and #phdlife. We ran the searches once a week (on the same day and at the same time every week) for 4 weeks between September 1-22, 2020. This yielded a raw dataset of 26,115 tweets. Retweets were excluded in the data collection.

We then cleaned the raw data to remove duplicate entries, foreign language entries, tweets consisting solely of hashtags, images, or URLs, and tweets consisting of fewer than 3 words. There were many duplicate entries since many users use both hashtags in their posts. After this step, 16,928 tweets remained. These were used to generate a word cloud and to conduct a dictionary-based sentiment analysis using computational text analysis techniques (more details below). To prepare the data for computational text analysis, punctuation, numbers, URLs, and symbols were removed. All words were converted from upper case to lower case. Stopwords – common words without much inherent meaning such as “it”, “an” and “the” – were removed, as were hashtags commonly found in the dataset, such as “#gradschool”, “#phd” and “#academictwitter”. Finally, words were stemmed, which is the process of reducing words to their base or root form, such as “reveal” for “revealed” or “present” for “presenting”. All of these are common steps in implementing the bag-of-words approach, whereby each document (in this case, tweet) in the dataset is converted to a simple list of words that can be analysed.

Next, we randomly selected 1994 tweets (about 12 percent of the data) for manual coding. Each tweet was coded as positive, negative, neutral, or indeterminate following Kharde and Sonawane (2016). We added the category ‘indeterminate’ after being unable to classify the tweets under other labels (see Table 1). The nature of coding was decided by the authors by identifying the sentiment expressed by a tweet. A tweet labelled positive clearly expressed a positive sentiment, while a tweet was labelled negative if it expressed a pessimistic sentiment. Tweets labelled as neutral were questions posed by users and did not lean towards either a positive or negative sentiment. The rest of the tweets with graphic interface format, or other symbols and signs for which the sentiment was not clear, were labelled as indeterminate. After agreeing on the codes to be assigned and their meaning, each tweet was independently coded by both authors, and the results were compared. Inter-rater reliability of almost 85 percent was achieved. The tweets that were coded differently by the two authors were discussed in greater detail until agreement was achieved. The final set consisted of 858 positive tweets (43 percent), 383 negative tweets (19 percent), 557 neutral tweets (28 percent), and 196 indeterminate tweets (10 percent).

The authors being PhD candidates and international students understand the journey of students undertaking a doctorate. Keeping in view their experiences, this paper has been written to explore the sentiments related to the PhD journey discussed on a social media platform like Twitter. This enabled them to understand the tweets and code them appropriately.
Table 1. Codes for manual coding

| Code      | Percentage of tweets (n=1994) | Meaning                                                      | Example of tweets                                                                 |
|-----------|------------------------------|--------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Positive  | 43%                          | Expresses a positive sentiment                               | I have two classes today and I’m excited to see people even if only on a screen! |
| Negative  | 19%                          | Expresses a negative sentiment                               | I am very demotivated early in the semester. Usually I would be energetic in the beginning but the pandemic and comps have left me out of energy #PhDLife #AcademicTwitter |
| Neutral   | 28%                          | Expresses no sentiment. Event announcements were coded as neutral, as were questions asked in a neutral tone | I need to cut 100 words in a manuscript. What are your tips to do this without losing content? |
| Indetermi-| 10%                          | Sentiment cannot be deciphered, sometimes because of the loss of accompanying images or emojis | When your puppy perfectly reflects how you feel on this Wednesday                  |

**USE OF TWITTER DATA AND ETHICS**

Since Twitter is a widely accessible and used public platform, tweets would be considered public conversation and hence publicly available information (Lachmar et al., 2017; Talbot et al., 2021). In this study, we do not specify usernames, user id, location, or any other identifiable information for our analysis. With the exception of the few examples that have been paraphrased in Table 1, we have avoided quoting tweets from the data set to avoid any personally identifiable information. Since we have protected privacy of Twitter users and confidentiality by retracting any identifiable information and rephrasing tweets in the paper, ethics approval to use the data was not required (Talbot et al., 2021; Trochim & Donnelly, 2001).

**DICTIONARY-BASED SENTIMENT ANALYSIS**

We start with a basic analysis of users and most frequent words in the dataset. Next, we conduct a dictionary-based sentiment analysis on the cleaned sample of 16,928 tweets. Dictionary-based sentiment analysis is based on the use of pre-existing dictionaries, i.e., pre-defined lists of words associated with particular sentiments. For instance, suppose we utilize a dictionary consisting of only two sentiments, positive and negative. The analysis process then essentially consists of comparing the actual words in every tweet against the lists of positive and negative words and assigning the entire tweet a sentiment score based on the proportions of positive and negative words. Since dictionary-based sentiment analysis has some known drawbacks, such as difficulties in handling context, nuance, or domain-specific vocabularies (Zimbra et al., 2018), we use it here simply to obtain a broad overview of sentiment in the dataset. As we explain below, for the more detailed thematic analysis of positive and negative tweets, we use the hand-coded sets of positive and negative tweets only, instead of relying on the categories generated by dictionary-based sentiment analysis.

We use the Bing and NRC dictionaries for our dictionary-based sentiment analysis. Both these dictionaries are commonly used for sentiment analysis, and are freely available with the tidytext package in R. The Bing dictionary categorizes words into those associated with positive and negative sentiment only. Words such as “benefit”, “fulfillment”, and “outstanding” are classified as positive, while
words such as “abnormal”, “dreadful”, and “misbehave” are classified as negative. The NRC dictionary categorizes words into eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and two sentiments (positive and negative). We use only the latter two categories in our analysis.

**Thematic Analysis**

To determine the main themes or topics discussed in tweets with positive and negative sentiment, we carried out separate thematic analyses of the positive and negative tweets in the hand-coded dataset. Our original intention was to use supervised machine learning methods to code the entire set of tweets into positive, negative, neutral, and indeterminate categories, and to follow that with an unsupervised topic modelling approach run separately on the positive and negative categories to identify the key themes in each category. However, the Naïve Bayes supervised machine learning algorithm we used to categorize tweets by sentiment did not yield very accurate results when compared against our hand-coded set, reflecting the difficulties associated with using these methods on short texts such as tweets. Consequently, we decided to undertake manual coding only for the thematic analysis. Although this set has lower volume of tweets, it has the advantage of being accurately coded.

A visual depiction of the data collection and analysis process is shown in Figure 1.
RESULTS

OVERVIEW OF DATA
The final sample of 16,928 tweets had 7,772 distinct users. To determine the profile of users, we examined the twitter profiles in the hand-coded set of 1,994 tweets. 58.3 percent of them were current PhD students, 17 percent were individuals who had already completed their PhD, and 0.6 percent were pre-PhD. An additional 6.5 percent of tweets were from non-commercial group accounts: online Twitter communities with the objective of providing resources and support to graduate student communities. 5.5 percent were from institutional accounts, such as universities or research organizations. Accounts offering commercial services such as copy-editing, proof-reading, or coaching made up 5.2 percent of the total. In the remaining 7 percent of cases, we were unable to conclusively determine the status of the user from their Twitter profile. Most users (>94 percent) tweeted five or fewer times during the time period. At the other end of the spectrum, ten users tweeted more than fifty times. The distribution of tweets is shown in Table 2. The ten most prolific users included three commercial accounts and one community account, with the other users being individuals.

Table 2. Number of tweets by user

| Number of tweets | Number of users | Total tweets |
|------------------|----------------|--------------|
| 1-5              | 7,329          | 11,038       |
| 6-10             | 284            | 2,088        |
| 11-20            | 99             | 1,396        |
| 21-30            | 35             | 863          |
| 31-40            | 12             | 418          |
| 41-50            | 3              | 132          |
| 51-100           | 7              | 527          |
| 101-200          | 2              | 243          |
| >200             | 1              | 223          |
| **Total users**  | **7,772**      | **16,928**   |

RESULTS FROM SENTIMENT ANALYSIS
We preface the sentiment analysis with a word cloud of the top hundred words in the dataset (Figure 2). The word cloud shows that negative words do not predominate; the most common words are either neutral (“work”, “research”, “write”) or positive (“like”, “good”, “great”).

520
Results from the dictionary-based sentiment analysis using the Bing and NRC dictionaries are shown in Figure 3. The figure shows the distribution of the sentiment scores, with the dashed line indicating the average value of the net sentiment. The results of both are similar and indicate that the net sentiment of the data is positive on average.

![Figure 3. Net sentiment of all tweets](image)

The results from the dictionary-based sentiment analysis are qualitatively consistent with the breakdown of the hand-coded set of 1,994 tweets in Table 1, as both indicate a preponderance of positive sentiment. Method triangulation through dictionary-based sentiment analysis and hand-coding each of the tweets has helped us in validating the findings.

**Thematic Analysis of Positive Tweets**

We identified the following main themes in the positive tweets: updates about work and life, giving and asking for advice, appreciating others and making connections on Twitter, sharing feelings and achievements, and expressing enthusiasm for research. These are summarized in Table 3 and elaborated below.
Table 3. Key themes in positive tweets

| Theme                          | Explanation                                                                                   |
|-------------------------------|-----------------------------------------------------------------------------------------------|
| Updates about work and life   | Creating work goals and to-do lists; reporting progress on data collection, experiments, and writing; working, and studying from home |
|                               | Sharing leisure activities; taking a break from work; self-care                               |
|                               | Humorous tweets about student life                                                             |
| Advice                        | Soliciting advice on PhD applications, academic writing, preparing for presentations and the PhD defence etc. |
|                               | Offering advice based on own experiences                                                       |
|                               | Sharing useful resources, e.g., data management tools, useful software                         |
|                               | Mental health advice on how to cope with work pressure and isolation, maintain motivation, and avoid burnout |
| Appreciation posts            | Expressing gratitude and appreciation for friends, family, supervisors, professors, research collaborators, and colleagues |
| Connecting with others        | Introducing themselves and their research interests to the academic community on Twitter       |
|                               | Promoting social events for graduate students                                                  |
|                               | Tweeting requests for followers and retweets                                                   |
| Sharing feelings and achievements | Sharing rewarding moments of PhD life                                                           |
|                               | Celebrating their own and others’ achievements                                                 |
|                               | Announcing dissertation submission and PhD completion                                           |
| Inclination towards research  | Sharing their enthusiasm for their own research                                                |

**Updates about work and life**

Updates on work and personal life comprised the largest category of positive tweets. Work-related tweets focused on updates about experiments, data collection, writing manuscripts, and setting work agendas and to-do lists. For instance, one of the users posted about their to-do list in a day comprising of data analysis, writing the results for a case analysis, and expressing their love for the research work. Users uploaded pictures of successful laboratory experiments and even shared music playlists that helped in being productive.

Since data collection occurred during the COVID-19 pandemic, working from home was commonplace. Tweets about setting up a place to work at home, purchasing ergonomic devices for work, and attending classes and work engagements on Zoom were frequent. Users also tweeted about their leisure time. They shared their elation after meaningful recreational activities (like drawing and painting) and described their plans for taking a day off or enjoying a mid-day break. They tweeted about self-care, describing how they deleted work email from the phone or were devoting time to physical fitness, listening to music, and indulging in hobbies they enjoyed. One of the tweets described a day in the life of a user of spending time cooking, listening to music, and staying away from computer screens which felt better than watching television series.

Humorous tweets about daily life while pursuing a PhD were also common. Related to PhD life, these were mostly to have a jab at the life of graduate students when they are completely devoted to their research which disrupts their work-life balance. Tweets featuring snippets of conversations between the user and others in daily life was included in this category of positive tweets. Users shared videos or links (to pictures/memes) describing one's state of mind, pictures of view from the
workplace, motivational posts, and observations during the pandemic that could be interesting research questions.

Advice
A major proportion of positive tweets were related to advice or counsel on work, the PhD journey, mental health of graduate students, and the relationship with the PhD supervisor. Advice was mainly based on the own experiences of graduate students. Users shared tips on presentation skills, viva-voce for the PhD dissertation, preparation for writing academic papers, writing manuscripts or the dissertation, and searching for post-doctoral positions. They also provided guidance for a successful partnership with PhD supervisors, such as being pro-active in meeting supervisors and preparing written questions before meetings.

Users also shared links to blogposts on useful resources for graduate students. These resources included tools for literature review and making illustrations, resources for data management and citations, techniques for time management and writing, technical advice for accessing online resources while working from home, shortcuts on programs and templates for building software, and reference papers on specific research topics.

Tweets with mental health advice included tips for relaxation, self-care, navigating the PhD journey, work life balance, and reinforcing the motivation to carry on with work. Key themes included suggestions to participate in extra-curricular activities, the importance of taking adequate breaks, how to deal with burnout, and relying on friends and family to counteract isolation in PhD life. For example, a user tweeted an encouraging message for those in the long and difficult PhD journey to reach out and not feel alone with people around who would be willing to help in different ways.

Many users also used Twitter to solicit advice on different aspects of PhD work, such as the process of writing manuscripts, preparing for presentations for viva and PhD defense, application to PhD programs, and communicating research progress to the supervisor. Related to the pandemic, users also sought advice on doing a PhD defense online and about coping strategies for students.

Appreciation posts
Tweets expressing appreciation for friends, family, colleagues, and professors comprised a major proportion of positive tweets. Users expressed gratitude to their advisors for their encouragement in the PhD journey and after completion of the dissertation. Appreciation messages were also tweeted for research collaborators. The help provided by individual colleagues in writing and publishing manuscripts was also mentioned in tweets under this theme. Gratitude messages for family members for being a support system during the PhD journey was a common theme in the tweets. Several users also acknowledged help and feedback from participants and panellists in conferences for their research. Appreciation tweets also included congratulatory messages for candidates who successfully defended their dissertation or were selected for fellowships.

Connecting with others
Making connections with other Twitter users in the “Academic twitter” community was another common theme. Most such tweets introduced individual users, their research field, and their particular research interests. Users also tweeted requests for more followers and retweets. To build more connections, users also promoted social and recreational events for graduate students.

Sharing feelings and achievements
Users utilized Twitter extensively to share their feelings while going through the PhD journey, and to share their own or others’ achievements. They described their motivations for PhD studies, the merits of pursuing a PhD and/or post-doctorate, and the rewarding moments of PhD life. Users also posted on Twitter about the excitement of having a productive day, giving a successful talk, or
coming back refreshed after a vacation. Achievements like joining a PhD program, successfully completing candidacy exams or the proposal defence, publishing a paper, and hitting a new record number of reads for a paper also made up a major proportion of these tweets. Apart from individual users, networks and institutes also shared their accomplishments such as developing an agreement for supervision for improving working conditions and quality of supervision for PhD students.

Users announced dissertation submission and PhD completion on Twitter. Such tweets expressed the happiness of completion and gratitude for the support provided by advisors, colleagues, friends, and family. Similar sentiments were expressed for fixing a date for dissertation defence and completion of a manuscript or a paper draft. The relief of completing data collection, data analysis, or sections of the dissertation was also discussed. Within this theme, some users also expressed their relief for successfully completing a PhD during a difficult time like the COVID-19 pandemic.

Inclination towards research

Users used Twitter as a platform for expressing their interest in research and analysis. Most tweets expressed excitement for research by communicating different perspectives garnered by reading literature, at conferences, and classes. PhD candidates who are also parents expressed their interest in work despite handling family-related responsibilities that took up time during the day. Users tweeted about the motivation that reading papers provided and the significance of being curious about a particular topic. They expressed fondness for their research work, indicated by the tweets that captured the excitement for coding on a particular day or research work that encompasses all interests of a researcher like nature and travel.

Thematic Analysis of Negative Tweets

The following main themes were identified in the negative tweets: mental health concerns; overwork and fatigue; difficulty of the work; COVID-related issues; academic culture; and other issues. The key themes are summarized in Table 4 and elaborated below.

**Table 4. Key themes in negative tweets**

| Theme                    | Explanation                                                                                                                                 |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Mental health concerns   | • Expressing feelings of imposter syndrome, self-doubt, lack of motivation, procrastination, stress, anxiety, isolation, and depression    |
| Overwork and fatigue     | • Resenting the culture of overwork in academia                                                                                             |
| Difficulty of the work   | • Venting frustration when research and writing were not going well • Finding it difficult to handle publication pressure as novice researchers |
| COVID-related issues     | • Frustrations with remote working, isolation, lack of face-to-face academic contact, and interruptions to laboratory work • Coping with increase in household and childcare responsibilities • Facing a slowdown in the academic job market |
| Academic culture         | • Expressing anger about overt or covert forms of discrimination in academia, exploitation of graduate students, and gatekeeping practices in research publications • Concerns about “ivory tower” behaviour of academics |
| Other                    | • Inadequate supervision; financial worries; poor job market prospects; experiences of rejection                                              |
Mental health concerns

By far the single largest category of negative tweets were about phenomena with adverse mental health connotations, including stress, anxiety, procrastination, imposter syndrome, lack of motivation, depression, and isolation. The most mentioned phenomenon was imposter syndrome, with several users commenting on it as a constant backdrop to their PhD journey. Comparing their own work to that of others and submitting their work for evaluation (for instance, to a journal) appeared to trigger particularly strong feelings of imposter syndrome. These feelings did not necessarily disappear after PhD completion: one user reported feeling like an imposter at the prospect of supervising their first PhD student, while another commented that imposter syndrome was triggered every time they wrote a conference paper or journal article.

Feelings of self-doubt were also common. Unsurprisingly, they surfaced particularly strongly when users’ work was going badly (e.g., their lab experiments were going wrong) or when they were submitting their work for evaluation (e.g., applying for a major scholarship). However, they also emerged during “business as usual” PhD research and study: students experienced self-doubt as they struggled to gain a deeper understanding of their chosen area, or when they critiqued the extant literature despite feeling like a newcomer to the field. The incremental nature of one’s unique research contribution, when set against the field as a whole, also triggered self-doubt.

Lack of motivation was another common problem and appeared to hit particularly hard when it came to tackling academic writing. Many users responded to loss of motivation by procrastinating, putting off research-related tasks in favour of doing household chores, napping, or using social media. Ironically, Twitter was a common choice for procrastinators, with some users even resorting to deleting the app from their phones to focus on their work (not always successfully).

Many users reported experiencing stress and anxiety. Writer’s block often contributed to this, as did tight submission deadlines or awaiting results of evaluations (such as examiner reports of theses). Users also struggled with feelings of isolation. Sadly, some users reported experiencing full-blown depressive episodes, for which mental health support at the institutional level did not always appear to be available.

Overwork and fatigue

Many users tweeted about the long hours they put in as graduate students, often sacrificing good eating, sleeping, and self-care habits to achieve this. Several noted that the academic work culture normalized overwork. For some students who were unable to sustain patterns of overwork, this in turn led to doubts about whether they were really cut out for PhD studies. At the same time, several users noted that the culture of overwork was ultimately unproductive, leading to stress, exhaustion, and burnout.

Difficulty of the work

Several users took to Twitter simply to vent their frustrations about difficulties they were experiencing in their PhD studies. For Science, Technology, Engineering and Math (STEM) students, problems in the lab were a common source of frustration: equipment broke, cultures did not grow, or lab procedures were tedious and repetitive. Several users mentioned that they found academic writing difficult, particularly writing discussion sections of papers or theses that required them to zoom out from the details of their research to the bigger picture. Sorting through a large amount of literature to find relevant references was also challenging.

Handling the pressure to publish was something that many users found particularly difficult. In the first place, some PhD students were skeptical about setting publication targets for students, arguing either that this generated the wrong incentives for students, was difficult to reconcile with the learning process during the PhD, or created expectations that were impossible to meet. For those students who did submit manuscripts to journals, responding to reviewer comments proved very difficult,
especially when reviewers were critical without providing constructive suggestions for improving the manuscript, had contradictory opinions, or made suggestions that seemed to stem from personal biases.

COVID-related issues
As the data for this study were collected during the pandemic, it is unsurprising that COVID-related concerns featured heavily in the tweets with negative sentiment. Most often, users took to Twitter to express their frustrations with Zoom. Students were also frustrated with remote working and the lack of face-to-face academic contact. Problems were particularly acute in STEM, where crucial experiments were on hold indefinitely because students were not allowed back on campus. Other concerns mentioned in relation to the pandemic included the increase in household and childcare responsibilities because of lockdowns, the slowdown in the academic job market, and the mental health toll of trying to complete a dissertation in complete isolation.

Academic culture
One problematic aspect of academic culture, the tendency to normalize overwork, has already been mentioned above. Apart from this, Twitter users also expressed concerns about discrimination, student exploitation, and “ivory tower” behaviour. They expressed their anger about overt or covert forms of discrimination against women and people of colour in academia. Users reflected on the lack of an inclusive environment in universities and about gatekeeping practices in publication by western scholars. Users also expressed anger about the exploitation of graduate students, especially those who are inadequately renumerated for teaching at their institutions. Finally, some users expressed their concerns about how impenetrable scientific research papers had become for non-experts, highlighting a need for better science communication.

Other issues
Several other issues were mentioned in negative tweets. Users often took to Twitter to express their frustration with hardware or software glitches. Interactions with supervisors left some anxious, while others were unhappy with lack of input from their supervisor. Some users reported instances of unethical behaviour on the part of supervisors, such as accepting co-authorship without contributing to the project or even misappropriating grant proposals. Parents tweeted about the difficulties of caring for children while simultaneously trying to work on their PhD. Financial worries were mentioned in tweets, as were poor job market prospects, with users noting that post-doctoral fellowship positions were shrinking while the entry requirements for junior tenure track positions were continually becoming more onerous. Finally, several users tweeted about how rejection was common in academia, but how difficult they found it to personally experience rejection.

DISCUSSION
Our analysis of tweets from the hashtags #phdchat and #phdlife finds that most users express positive sentiments on Twitter. This is consistent with findings from most survey-based studies, that have also tended to find a preponderance of positive sentiments (Byrom et al., 2020; Woolston, 2015, 2017, 2019). Analysing the key themes of positive tweets, we find that PhD students use Twitter to seek and share advice, announce important information, and share their achievements in the PhD journey. Some users use Twitter as an accountability mechanism, recording their agenda and to-do lists as well as the tasks they accomplish. Twitter also serves as a medium for graduate students to find a community where useful resources on graduate studies, academic life, and mental health concerns can be accessed. In addition, it is a platform where graduate students can reach out to professors, experts, and potential research collaborators. To sum up, they appear to use Twitter to build both professional and human connections and create a sense of community as found by Bowman (2015) and Bennett and Folley (2014).
Tweets that express negative sentiments are fewer in number than those that express positive sentiments. Some negative tweets simply express frustration with the sheer difficulty of graduate work and can be seen as a healthy coping mechanism, although these tweets do suggest that providing graduate students training in skills such as academic writing would be valuable, as would tempering publication expectations. Others, however, are more troubling. Mental health concerns predominate, with stress, anxiety, self-doubt, procrastination, lack of motivation, depression, and isolation mentioned frequently. The imposter syndrome appears to hit PhD students particularly hard. The frequency with which users complain about the culture of overwork in academic life is also striking, which results in fatigue and adversely affects relationships with family and friends. Tweets also indicate the adverse impact of COVID on the mental health and daily life of users.

Our results on the prevalence of mental health issues in PhD students mirror the results obtained from PhD student surveys (Barreira et al., 2018; Evans et al., 2018; Levecque et al., 2017; Woolston, 2017). While some of the contributory factors commonly reported in surveys, such as troubled relationships with supervisors, feeling unsupported by departments, or lack of funding from universities, are not found frequently in our data, others such as heavy workloads and the difficulty of maintaining work-life balance are common to both and appear to contribute significantly to stress and exhaustion levels (Barry et al., 2018; Evans et al., 2018; Levecque et al., 2017; Mackie & Bates, 2019; Woolston, 2019).

There are certain limitations to our findings, most of them stemming from our use of Twitter as the data source. Methodologically, while our intention was to segregate tweets with positive, negative, and neutral sentiments, it was difficult to gauge sentiments from tweets due to the word limit on the platform, and we were unable to determine the overall sentiment in about ten percent of the tweets we hand-coded. There were instances when a single tweet had both positive and negative sentiments. Sarcasm was also difficult to interpret.

More broadly, we caution that the views uncovered here do not necessarily generalize to all PhD students: tweeting is a voluntary activity and there may well be systematic differences between those who choose to go on Twitter and those who do not. Some PhD students have been found to deliberately limit their engagement on Twitter to manage time pressures and avoid distractions (Dowling & Wilson, 2015; Veletsianos et al., 2018). There is also a potential self-censorship issue: users might hesitate to post tweets publicly expressing dissatisfaction with their own working environment, department, colleagues, or supervisors, which might be easier to express in anonymized surveys. On the other hand, because Twitter does not require users to use their own names when creating Twitter handles, users who desire anonymity can employ non-identifiable pseudonyms if they choose to; Peddinti et al. (2014) noted that over a quarter of Twitter users are anonymous or semi-anonymous. It is important to note that most graduate student surveys are usually based on non-random convenience samples, prone to self-selection bias, and likely to over-state the prevalence of mental health conditions (Byrom et al., 2020; Duffy et al., 2019; Evans et al., 2018). To the extent that the two data sources may be biased in different directions, Twitter data can serve as a useful complement to survey data.

**CONCLUSION**

The aim of this article was to examine the content of PhD student tweets to determine whether they expressed mainly positive or negative sentiments and the factors that contributed to both positive and negative sentiment. We found that PhD students expressed positive sentiments on Twitter much more often than they expressed negative ones. On the positive side, they went on Twitter to share updates and achievements in their work and personal life, ask for and request advice, express appreciation for others, make connections and build community, and share their passion for research. On the negative side, most tweets discussed mental health concerns such as stress, anxiety, self-doubt, procrastination, lack of motivation, depression, isolation, and the imposter phenomenon. Negative
tweets also pointed to problems with academic culture, especially the normalization of overwork, and the effects of the pandemic on students.

**Recommendations for Practice**

Our research found that heavy workloads and the difficulty of maintaining a healthy work-life balance were common sources of negative sentiments expressed by PhD students in their tweets. These are systemic but not necessarily insuperable factors and should prompt calls for change in academic culture. At a departmental or institutional level, there should be adequate provision of mental health support for PhD students. PhD program administrators and faculty should be sensitized in how to identify signs of emotional distress and direct students to available support. Given the ubiquity of the imposter phenomenon in PhD students, institutions should take steps to address it. In this regard, Chakraverty (2020) has provided useful recommendations including honest conversations that start early and help new PhD students identify such feelings, closer mentoring relationships, and paying attention to the professional development needs of students. Taking cognizance of instances of exploitation of graduate students and discrimination against them, universities should ensure that they have formal regulations and safe channels to address these problems. Finally, universities should pay attention and respond sensitively to the challenges that the pandemic has created for PhD students, including slowdowns in research productivity, the stresses posed by isolation, and worry and pessimism about the state of the academic job market.

**Recommendations for Future Research**

Future research related to analysing sentiments of PhD tweets can explore various issues. The specific challenges and opportunities faced by PhD students in specific disciplines and locations could be compared using tweets. Apart from PhD students, networks/organizations and individuals who have completed their PhD tweet about the PhD experience as well, and examining tweets from these particular groups can yield insights about the differing nature of Twitter use. The time period used here includes the time when the world was engulfed by the pandemic. Future research could track post-pandemic sentiments to observe if they have changed. Such insights can aid the provision of targeted support to PhD students.

**References**

Allen, H. K., Lilly, F., Green, K. M., Zanjani, F., Vincent, K. B., & Arria, A. M. (2020). Substance use and mental health problems among graduate students: Individual and program-level correlates. *Journal of American College Health, 1-9*. [https://doi.org/10.1080/07448481.2020.1725020](https://doi.org/10.1080/07448481.2020.1725020)

Barreira, P., Basilico, M., & Bolotnyy, V. (2018). *Graduate student mental health: Lessons from American economics departments* [Working paper]. Harvard University. [https://scholar.harvard.edu/bolotnyy/publications/graduate-student-mental-health-lessons-american-economics-departments](https://scholar.harvard.edu/bolotnyy/publications/graduate-student-mental-health-lessons-american-economics-departments)

Barry, K., Woods, M., Warnecke, E., Stirling, C., & Martin, A. (2018). Psychological health of doctoral candidates, study-related challenges and perceived performance. *Higher Education Research & Development, 37*(3), 468-483. [https://doi.org/10.1080/07294360.2018.1425979](https://doi.org/10.1080/07294360.2018.1425979)

Bennett, E., & Folley, S. (2014). A tale of two doctoral students: Social media tools and hybridised identities. *Research in Learning Technology, 22*, 23791. [https://doi.org/10.3402/rlt.v22.23791](https://doi.org/10.3402/rlt.v22.23791)

Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., & Bucci, S. (2017). # WhyWeTweetMH: Understanding why people use Twitter to discuss mental health problems. *Journal of Medical Internet Research, 19*(4), e107. [https://doi.org/10.2196/jmir.6173](https://doi.org/10.2196/jmir.6173)

Bowman, T. D. (2015). *Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter* [PhD thesis, Indiana University]. [http://www.tdbowman.com/pdf/Dissertation.pdf](http://www.tdbowman.com/pdf/Dissertation.pdf)
Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brown, P. (2013). Loneliness at the bench: Is the PhD experience as emotionally taxing as it is mentally challenging? *EMBO Reports, 14*(5), 405-409. https://doi.org/10.1038/embor.2013.35

Byrom, N. C., Dinu, L., Kirkman, A., & Hughes, G. (2020). Predicting stress and mental wellbeing among doctoral researchers. *Journal of Mental Health, 1*-9. https://doi.org/10.1080/09638237.2020.1818196

Chakraverty, D. (2020). PhD student experiences with the impostor phenomenon in STEM. *International Journal of Doctoral Studies, 15*, 159-180. https://doi.org/10.28945/4513

Dahal, B., Kumar, S. A., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining, 9*, 1-20. https://doi.org/10.1007/s13278-019-0568-8

Dowling, R., & Wilson, M. (2017). Digital doctorates? An exploratory study of PhD candidates’ use of online tools. *Innovations in Education and Teaching International, 54*(1), 76-86. https://doi.org/10.1080/14703297.2015.1058720

Duffy, M., Thanhouser, C., & Derry, H. (2019). A lack of evidence for six times more anxiety and depression in US graduate students than in the general population. *Nature Biotechnology, 37*, 711-712. https://doi.org/10.1038/s41587-019-0179-y

Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., & Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. *Nature Biotechnology, 36*(3), 282-284. https://doi.org/10.1038/nbt.4089

Ferguson, H., & Wheat, K. L. (2015). Early career academic mentoring using Twitter: the case of #ECRchat. *Journal of Higher Education Policy and Management, 37*(1), 3-13. https://doi.org/10.1080/1360080x.2014.991533

Gomez-Vasquez, L., & Romero-Hall, E. (2020). An exploration of a social media community: The case of #AcademicTwitter. Paper presented at the International Conference on Human-Computer Interaction.

Hunter, K. H., & Devine, K. (2016). Doctoral students’ emotional exhaustion and intentions to leave academia. *International Journal of Doctoral Studies, 11*, 35-61. https://doi.org/10.28945/3396

Kearney, M. W., Heiss, A., & Briatte, F. (2020). Package ‘rtweet’: https://cran.r-project.org/web/packages/rtweet/rtweet.pdf

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.

Lachmar, E. M., Wittenborn, A. K., Bogen, K. W., & McCauley, H. L. (2017). #MyDepressionLooksLike: Examining public discourse about depression on Twitter. *JMIR Mental Health, 4*(4), e43. https://doi.org/10.2196/mental.8141

Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., & Gisle, L. (2017). Work organization and mental health problems in PhD students. *Research Policy, 46*(1), 868-879. https://doi.org/10.1016/j.respol.2017.02.008

Lopez, J. C. A. D., Collignon-Delmar, S., Benoi, K., & Matsuo, A. (2017). Predicting the Brexit vote by tracking and classifying public opinion using Twitter data. *Statistics, Politics and Policy, 8*(1), 85-104. https://doi.org/10.1515/spp-2017-0006

Mackie, S. A., & Bates, G. W. (2019). Contribution of the doctoral education environment to PhD candidates’ mental health problems: A scoping review. *Higher Education Research & Development, 38*(3), 565-578. https://doi.org/10.1080/07294360.2018.1556620

Malik, A., Heyman-Schrum, C., & Johri, A. (2019). Use of Twitter across educational settings: A review of the literature. *International Journal of Educational Technology in Higher Education, 16*(1), 1-22. https://doi.org/10.1186/s41239-019-0166-x

Marais, G. A., Shankland, R., Haag, P., Faault, R., & Juniper, B. (2018). A survey and a positive psychology intervention on French PhD student well-being. *International Journal of Doctoral Studies, 13*, 109-138. https://doi.org/10.28945/3948
A Sentiment Analysis of the PhD Experience

Nature. (2019). The mental health of PhD researchers demands urgent attention. *Nature, 575*, 257-258. https://www.nature.com/articles/d41586-019-03489-1

Peddinti, S. T., Ross, K. W., & Cappos, J. (2017). User anonymity on Twitter. *IEEE Security & Privacy, 15*(3), 84-87. https://doi.org/10.1109/msp.2017.74

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology, 68*(9), 2037-2062. https://doi.org/10.1002/asi.23833

Talbot, J., Charron, V., & Konkle, A. (2021). Feeling the void: Lack of support for isolation and sleep difficulties in pregnant women during the COVID-19 pandemic revealed by Twitter Data Analysis. *International Journal of Environmental Research and Public Health, 18*(2), 393. https://doi.org/10.3390/ijerph18020393

Tomeny, T. S., Vargo, C. J., & El-Toukhy, S. (2017). Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. *Social Science & Medicine, 191*, 168-175. https://doi.org/10.1016/j.socscimed.2017.08.041

Trochim, W. M., & Donnelly, J. P. (2001). *Research methods knowledge base* (Vol. 2). Atomic Dog Pub.

Veletsianos, G., Kimmons, R., Belikov, O., & Johnson, N. (2018). Scholars’ temporal participation on, temporary disengagement from, and return to Twitter. *First Monday*. https://doi.org/10.5210/fm.v23i11.8346

Woolston, C. (2015). Graduate survey: Uncertain futures. *Nature, 526*(7574), 597-600. https://doi.org/10.1038/nj7574-597a

Woolston, C. (2017). Graduate survey: A love–hurt relationship. *Nature, 550*(7677), 549-552. https://doi.org/10.1038/nj7677-549a

Woolston, C. (2019). PhDs: The tortuous truth. *Nature, 575*(7782), 403-407. https://doi.org/10.1038/d41586-019-03459-7

Yu, A. (2020). For some graduate students, the cost of doing science is their mental health. *WHYY*. https://whyy.org/segments/for-some-graduate-students-the-cost-of-doing-science-is-their-mental-health/

Zhu, Y., & Procter, R. (2012). Use of blogs, Twitter and Facebook by PhD students for scholarly communication: A UK study. *Proceedings of the 2012 China New Media Communication Association Annual Conference*, Macao https://www.escholar.manchester.ac.uk/api/datastream?publicationPid=uk-ac-man-scv:187789&datastreamId=FULL-TEXT.PDF

Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems (TMIS), 9*(2), 1-29. https://doi.org/10.1145/3185045

**AUTHORS**

*Panchali Guha* is a PhD candidate at the Lee Kuan Yew School of Public Policy, National University of Singapore. Her research interests are in education policy, international development, the non-governmental sector, and policy evaluation. She would like to thank Dani Madrid-Morales for his course on quantitative text analysis at NUS.
Devyani Pande is a PhD candidate at the Lee Kuan Yew School of Public Policy, National University of Singapore (NUS). Her core research area is governance of new and emerging technologies. She is passionate about advocating for mental well-being among graduate students and has been trained as a peer student supporter at NUS.