"The Smartest Decision for My Future": Social Media Reveals Challenges and Stress During Post-College Life Transitions

CRYSTAL GONG, Northwestern University, USA
KOUSTUV SAHA, Georgia Tech, USA
STEVIE CHANCELLOR, University of Minnesota - Twin Cities, USA

The post-college transition is a critical period where individuals experience unique challenges and stress before, during, and after graduation. Individuals often use social media to discuss and share information, advice, and support related to post-college challenges in online communities. These communities are important as they fill gaps in institutional support between college and post-college plans. We empirically study the challenges and stress expressed on social media around this transition as students graduate college and move into emerging adulthood. We assembled a dataset of about 299,000 Reddit posts between 2008 and 2020 about the post-college transition from 10 subreddits. We extracted top concerns, challenges, and conversation points using unsupervised Latent Dirichlet Allocation (LDA). Then, we combined the results of LDA with binary transfer learning to identify stress expressions in the dataset (classifier performance at F1=0.94). Finally, we explore temporal patterns in stress expressions, and the variance of per-topic stress levels throughout the year. Our work highlights more deliberate and focused understanding of the post-college transition, as well as useful research and design impacts to study transient cohorts in need of support.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: social support, social media, life transitions, stress, Reddit, machine learning, language, disclosure

ACM Reference Format:
Crystal Gong, Koustuv Saha, and Stevie Chancellor. 2021. "The Smartest Decision for My Future": Social Media Reveals Challenges and Stress During Post-College Life Transitions. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 298 (October 2021), 29 pages. https://doi.org/10.1145/3476039

1 INTRODUCTION

In 2019, 3.9 million people graduated from college in the United States, about half of whom received bachelor's degrees [68]. Although graduation is often celebrated, this time in peoples' lives can be very stressful. Students are finishing coursework, often finding and securing post-graduate plans like employment or graduate school, and moving to new locations and living situations.

The *post-college transition* is an important life transition or an event that notably alters the course of one's life. Life transitions include leaving home, starting and ending relationships, getting a new job, getting married, or having children. These transitions alter individuals' sense of being, roles and responsibilities, and physical surroundings [37, 62]. In particular, the post-college transition

Authors' addresses: Crystal Gong, crystalgong2020@u.northwestern.edu, Northwestern University, 633 Clark St, Evanston, IL, USA; Koustuv Saha, koustuv.saha@gatech.edu, Georgia Tech, Atlanta, GA, USA; Stevie Chancellor, steviec@umn.edu, University of Minnesota - Twin Cities, Minneapolis, MN, USA.

2573-0142/2021/10-ART298 $15.00
https://doi.org/10.1145/3476039

is important as young adults face a crucial psychological maturation stage, struggling between the desire to meet expectations and come to terms with the realities of emerging adulthood [34]. Research has shown that this transition is punctuated with short-term stress and decreased well-being [34, 50, 62, 83].

Despite it being a predictable transition, assessing psycho-social challenges during this time is difficult. Most university surveys of students focus on procedural outcomes, like graduation rates and employment placement, and not on well-being markers like stress and overall life satisfaction. Furthermore, the abrupt discontinuation of collegiate support caused by graduation means that there is an absence of formal support for the well-being of new graduates. This leaves many individuals in a proverbial "no-man's land" with ambiguous (and possibly absent) support and assistance [89]. This gap of care hinders longitudinal studies that may inform proactive interventions and support for individuals undergoing the post-college transition [62, 89]. What is needed, then, is an observational and comprehensive study of this transition to overcome these gaps.

Online support communities present a unique viewpoint for the post-college transition because they overcome traditional barriers to studying this cohort. People express emotion and stress in these communities [6, 30] – and also seek support for these transitions [21, 23]. Literature in Computer Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) has a keen interest in spaces like these, including transitions between jobs and unemployment [12], the transition from high school to college [29], and collegiate civil to military transitions [30].

Our Contributions. We use social media data to study the well-being of individuals in post-college transition. Recent research has demonstrated the potential of online communities as a source of information about well-being and mental health, specifically stress expressions [39, 78, 81]. We use Reddit, a popular social media and area of interest for prior work on college students [8, 78, 81] and for helping those who seek social support [26, 29]. By focusing on stress and well-being expressed through language in social media, we can pinpoint topics and times of stress as students seek support, providing a more comprehensive view of the process of post-college transitions.

In this paper, we quantitatively examine online communities that support individuals through post-college transitions. Thus, our research questions are:

RQ1: What challenges and themes do people discuss in online communities for post-college transitions?

RQ2: Can we evaluate stress and how it changes over time in online communities about post-college transitions?

- **RQ2a:** What topics and challenges map to high-stress expressions?
- **RQ2b:** How does stress temporally shift for specific topics and challenges in post-college transitions?

We collected a dataset of about 299,000 posts on Reddit between 2008 and 2020 about post-college transitions. We design and deploy a two-step data gathering process of identifying college subreddits then filtering posts using word embeddings to find conversations on post-college transitions. In RQ1, we use unsupervised topic modeling with Latent Dirichlet Allocation (LDA) combined with human annotations to identify key challenges and thematic takeaways. In RQ2, we use transfer learning to build a classifier that identifies low and high stress posts, with performance at 0.94 F1. We find that temporal stress patterns correspond with both the academic school year, and that reported stress is increasing over time in our dataset (+29.64% year over year). Combining the topic models and stress classification together, we relate the challenges to stress over time. We find that specific post-college challenges exhibit different patterns of stress expression throughout the year.

Our work empirically assesses the challenges of the post-college transition and the types of support people seek out. This aligns with and expands on work in the social sciences on social support in transitions [21, 22], specifically the post-college transition [62, 65, 89]. Our work contributes a taxonomy of the major concerns and challenges experienced by individuals undergoing post-college transition. Additionally, we offer methods contributions for filtering datasets on Reddit. This work can better inform on-campus stakeholders in their recent efforts to prioritize student health and well-being and to conduct tailored interventions and policy changes. Finally, we contribute to CSCW and HCI research in considering what ways online communities can fill the gaps of traditional support mechanisms for those in transition.

Privacy, Disclosure, and Researcher Perspectives. Because our work uses publicly accessible Reddit data and we do not interact interaction with users, our study did not qualify for ethics board approval at our institutions. However, we are committed to ethical research practices in large-scale social media research. Please see our Discussion for a more extensive treatment of our practices.

We also understand the importance of disclosing our position as researchers, as our perspectives influence our methods, presentation, and interpretation of data [13]. Two authors are experts in social media, well-being, and mental health. One author has expertise in young adults and college student populations. Collectively, the researchers have undergone post-college transitions at both U.S. and international colleges. We are limited in perspective as we have primarily have attended large research institutions for our undergraduate and graduate training. We believe these experiences help us situate the findings on this population, problem, and circumstances. When referring to "the research team", we reference all co-authors working collaboratively on tasks.

2 BACKGROUND AND RELATED WORK

2.1 Life Transitions and the Post-College Transition

All people go through life transitions, or life-changing events that alter an individual's sense of being. These include leaving home, starting and ending relationships, getting a new job, getting married, and having children. Research on transitions is extensive in sociology and organizational studies, as transitions are an important life process. They alter the roles and responsibilities, surroundings, and sense of identity of individuals [34, 37, 62]. Mikal et al. identifies four main transitional domains: health-illness (individual), development (familial), organization (community), and situational (societal) [62], whereas other frameworks rest on whether the life transition was anticipated or not [61]. In either framework, those in transition encounter both decreased well-being in the short-term and long-term personal growth [62].

The post-college transition has been of particular interest for social scientists, due to its connection to career outcomes and life satisfaction [43]. The literature mostly takes two perspectives – analyzing final year students (or college seniors) and their concerns as they search for post-graduate opportunities [49, 51] or evaluating students as they transition to gainful employment in private industry [50, 65]. Traditional approaches to understand these transitions include surveys and interviews [10, 89]. While these methods are accurate and provide deep insights, these approaches do not scale and are susceptible to biases caused by the difficulty of surveying a transient cohort [85]. Closest to our work, Wendlandt and Rochlen conducted a literature review highlighting the challenges associated with the college-to-work transition. They identified three stages of transition in this time period – anticipation, adjustment, and achievement and also proposed implications for career counselors [89]. A common theme in this research was the role of social support for a smooth transition and establishing a sense of well-being. Our work contributes to this area by providing insights into the major challenges during post-college transitions by using large-scale social media data.

2.2 Transitions, Social Support, and Social Media

The use of social technologies like social media and online communities during life transitions and disruptions has been a major theme of social computing and CSCW research [12, 29, 40, 42, 58, 59, 79, 86]. Studies on these events are as diverse as the aforementioned transitions research, and include the birth of a child [25], losing a job [12], relationship changes [28], and death of loved ones [47]. Recent studies have provided comprehensive overviews of how people discuss major life events via social media [41, 79].

Social media can provide the needed social and material resources to help individuals adapt, reconfigure their life, and rebuild connections through times of transition [48, 59]. This is closely related to the notion of "social support" – the use of social resources like friends, family, and peers to provide a sense of belonging, reassurance, information, and connection with others [23, 48]. Social support helps individuals develop coping strategies from stressful events and assists with psychological adjustment [19, 60]. For students, social support can positively affect academic adjustment [29]. Online communities can be analogs of offline support groups, where individuals seek support through sensitive self-disclosures [26, 46] and receive support in return. De Choudhury and De showed that social media facilitates candid self-disclosures and drawing support related to mental health and stigma, and can help individuals connect with others who have had related experiences [26].

In the context of life transitions, research has found that online communities are a source of peer support [7, 58, 66]. Users turn to social media to seek advice, diminish feelings of isolation, provide support, and to share coping strategies and experiences [66]. Social support can help manage transitional outcomes and stress [62]. For example, van Ingen et al. measured coping after negative life events and found a small positive association between social media use and problem-focused coping [86]. In times of life transitions, the Internet enables an individual to stay connected with established networks while bringing together new, relevant support networks, without relying on geographic proximity [62]. In the case of gender transition, social media can be both a source of stress because of sensitive information disclosure as well as a platform to help alleviate such stress by providing support from one's social network [40]. Burke and Kraut studied social media use following job loss and found that communication with strong ties was generally associated with positive effects, such as improvements in stress and social support, and increased likelihood of reemployment [12]. Closest to our research, DeAndrea et al. studied students transitioning from high school to college and found that social media gives students access to an online community and provides a means for students to give and receive help [29].

Post-college transition is a critical phase when people seek support and advice from others. Thus far, there is little research that explores the use of social media for support seeking at this crucial time. Our work builds on the above research to understand what challenges and concerns people express during this period of time.

2.3 College Students, Mental Well-being and Social Media

College students undergo mental well-being challenges [52]. Stress is a major and prevalent concern among college students, and as many as three out of four college students consider themselves to be stressed [53]. Many personal and academic life factors and environmental stressors precipitate college student stress [75]. Unfortunately, there are major impediments to seeking mental health services for college students, such as stigma and lack of timely and proactive care [31, 32].

Social media has the potential to demonstrate naturalistic patterns of mood, behavior, cognition, psychological states, and social milieu [38, 54, 82]. Prior work has leveraged social media data at scale to quantitatively identify mental health attributes such as stress and depressive symptoms [15, 20, 25, 35, 39, 76]. We look to this work to inform our approach in empirically studying Reddit and the post-college transition.

Because social media use is so popular among young adults, researchers have begun to use these platforms to understand the well-being of college students [55, 57, 64, 81]. For college students, social media is a rich source of information, given their digital literacy and access [71]. The seminal work of Ellison et al. revealed the positive association between college students' social media use and maintenance of social capital [33]. Social media helps this demographic to satisfy their psychosocial needs [56]. In related work, Bagroy et al. not only found that Reddit college communities are also representative of the offline college communities but also can reveal population-scale mental well-being of college campuses [8]. Close to our work, Saha and De Choudhury quantified the severity of gun violence crises on college campuses using linguistic indicators and machine learning on social media data [78]. Another study revealed that stress expressions in online college communities increased after exposure to hateful speech [77].

We build off this prior work to study challenges and stress during the post-college transition in social media. We use NLP and machine learning techniques to study discussions on Reddit about the challenges and stress during post-college transitions.

3 DATA

We conducted our study on post-college transitions using data from Reddit. Reddit is a pseudo-anonymous platform that hosts over 2 million online communities, where users can post content on subreddits – user-created boards centered around a certain topic (designated with r/communityname). Reddit is one of the most popular social media platforms and it caters to people between 18-29 years: Pew Research found that 65% of Reddit users are young adults [71]. This age demographic aligns well with the typical college student and transition time period highlighted in the success of prior work that uses Reddit to study college students [8, 77, 78, 81]. Additionally, prir work has shown that Reddit facilitates candid disclosures through affordances and norms on its site, such as topical focus, anonymity with throwaway accounts, and community-driven moderation to maintain discussion quality [5, 6, 26, 87], making it a great place to study the post-college transition.

Given Reddit's community structure, the most intuitive source of data about post-college transitions are specific subreddit that facilitate discussions on a central topic. However, in gathering data on any focused topic on Reddit, there is an inherent tension in identifying posts and comments that are both *specific* and *comprehensive* to a topic while managing *data noise*. This poses some challenges for targeted data collection that impacted our study.

For our work, there was no one comprehensive subreddit dedicated to the post-college transition, and analyzing all of Reddit for specific posts about this subject is non-trivial and a research project unto itself. There are very large subreddits about college, such as r/college, which has been studied in prior work [8, 78, 81] and covers a wide breadth of subjects related to the college experience. However, we found that large subreddits like this can be "noisy" given our focused research interest in studying post-college transition and not all topics about college. Additionally, there are smaller subreddits like r/LifeAfterSchool that are less noisy and clearly focused on the post-college transition. However, r/LifeAfterSchool is very small, at 1700 posts, and may not capture broad discussions about post-college transitions.

Consequently, our data collection strategy balanced these two goals of both specificity and comprehensiveness, while trying to minimize noise. To do so, we adopted a two-step approach in data gathering and filtering. To focus on comprehensiveness, we first identified a strong candidate set of subreddits related to the general college experience. Then, we devised a filtering mechanism we call *lexico-semantic similarity filtering* (LSSF), which used word embeddings [24] built on highly

Table 1. List of subreddits considered in our study, which cater to discussions on post-college transitions.

Subreddit	Description	#Mbrs.	#Posts	First Post
r/gradadmissions	Advice for graduate school applications.	49.6K	33,524	2/2012
r/LifeAfterSchool	Discussion for life transition and challenges after school, primarily focused on college transitions	37.9K	1,727	4/2019
r/findapath	Determining a desirable career and life path	95.8K	29,548	2/2014
r/careerguidance	Career choices and advice	128K	104,526	7/2013
r/Adulting	Learning skills of adulthood	37.6K	4,843	7/2015
r/youngadults	Discussion for people in early and young adulthood	16K	2,861	1/2013
r/movingout	Moving out and into one's own place	2.7K	1,057	9/2011
r/GetEmployed	Advice for obtaining and keeping employment	31.7K	10,666	6/2012
r/college	Discussion of college related content	267K	166,566	4/2008
r/GradSchool	Discussion of graduate school related content	86.6K	55,030	8/2009

specific post-college transition subreddits to filter our candidate set of general college subreddits. We describe our approach below.

3.1 Finding Candidate Subreddits

To identify candidate subreddits related to college experiences, we adopted a community identification and curation inspired by Chancellor et al.'s curation of subreddits. We began by considering r/LifeAfterSchool, a subreddit that directly addressed life transitions after college. We leveraged r/LifeAfterSchool's curated list of over 400 related subreddits found in the subreddit's Wiki page [1], and manually visited them all. To determine if subreddits primarily catered to the undergraduate college experience, two members of the research team designed a rating task. To be included, subreddits had to meet three criteria, inspired by prior work [8, 16, 78]: 1) the discussions were about undergraduate college experience; 2) there was at least one post in the last month when we gathered our data (April 2020); 3) the size of the subreddit was larger than 2,500 subscribers. Two researchers independently annotated whether a subreddit met these criteria and met to resolve disagreements.

Most subreddits were cut for topical relevance (e.g. r/personalfinance is important for the post-college transition but not focused on college specifically). We deliberated over whether to include specific industry (i.e. r/cscareerquestions) or admissions (r/lawschooladmissions or r/premed) subreddits. After more closely reading posts from these subreddits, we found that these subreddits featured posts across the career spectrum, from new hire to senior and management, and did not focus on post-college employment specifically. Table 1 shows our shortlist of 10 subreddits after this deliberation.

3.2 Data Collection

After identifying the subreddits in Table 1, we used the Pushshift.io Reddit API[1] to gather all posts from the subreddits. We collected all historical data in late April 2020. The oldest subreddit, r/college, was started on January 25, 2008, allowing us over 12 years of data. Descriptive statistics of these 349k posts are summarized in Table 2. Posts that had been deleted or removed, or had deleted or removed authors, were excluded from the dataset to respect the privacy of authors who remove their content.

[1] http://github.com/pushshift/api

Table 2. Descriptive statistics of our subreddit dataset.

Post Statistics	
Metric	**Value**
Number of posts	349,188
Number of unique posters	203,421
Mean words per posts	144.94
Median words per post	99

3.3 Finding Specific Posts using a Lexico-Semantic Similarity Filter (LSSF)

As mentioned, the 10 subreddits contained advice for most of college and young adulthood that spanned beyond our goals of studying post-college transitions. To filter our dataset to more appropriately study post-college transitions, we used what we call a *lexico-semantic similarity filter* (LSSF) to specifically identify posts in this target group.

LSSF leverages word embeddings, a popular and powerful technique from the natural language processing literature [63]. Word embeddings represent words as vectors in higher dimensional latent space, where words that are lexically or semantically similar tend to be closer in vector space [63, 70]. For instance, synonyms tend to appear near each other in embeddings. Within the word embedding vector space, linguistically similar words or documents can be estimated using distance measures [63]. We leveraged these benefits of word embeddings to find relevant discussions about post-college transitions from our 10 subreddits.

First, we manually identified subreddits directly related to post-college transitions in our 10 subreddits. These subreddits included r/gradadmissions, r/LifeAfterSchool, and r/findapath, which the research team verified were overwhelmingly focused on the post-college transition. Using `word2vec`, we created a word embedding vector of the 65k posts from these three subreddits. Data for the embedding tokenized the dataset, removed stop words, and applied stemming. The word embedding used the bag-of-words model in word2vec with all words included (no minimum presence of words required). This word embedding constitutes our LSSF.

We then applied the LSSF to each post from the remaining seven subreddits, looking for comparable posts to the word embedding about post-college transitions. For this threshold, we used cosine similarity (at a threshold of 0.85) for each discrete post to the LSSF embedding vector. The research team manually validated our threshold for plausible cosine similarity scores between 0.7 and 0.9 in 0.05 increments, motivated by prior work [16, 73, 76]. We found that a cosine similarity of 0.85 appropriately balanced filtering noise and identifying posts related to the post-college transition.

Finally, to verify that the LSSF is tailored to post-college transitions, we compared the word embeddings to a random dataset of posts from r/popular pre-gathered from 2011-2014[2]. r/popular is an algorithmically-generated subreddit that curates trending content across Reddit, can serve as a "control" for general discussion across the site. This dataset has 100,219 posts and often includes current events/news, gaming, pop culture references, humorous posts, and topical content. Our LSSF displayed much weaker average cosine similarity to r/popular data (mean=0.5047, std 0.0381) than to the overall dataset of 10 subreddits found earlier (mean=0.8449, std=0.15). This indicates that our LSSF is well-tuned to the issue of post-college transitions compared to general conversations on Reddit.

Using this LSSF, we filtered all 284k posts from these 7 subreddits to 234k posts, meaning 82% of posts made the cutoff. Taking our two datasets together (65k from the three subreddits, and our

[2]The r/popular dataset includes submissions made from July 1, 2011 to November 31, 2014. This data is later used for training the stress classifier in 4.2.1.

Table 3. Perplexity scores for each of the LDA models with topic numbers of 5 to 50, in increments of 5. For perplexity, lower (i.e. more negative) scores are better. For coherence, higher scores are better.

#Topics	Perplexity
5	-9.20
10	-9.53
15	-9.89
20	-10.21
25	-10.51
30	-10.83
35	-11.16
40	-11.42
45	-11.74
50	-12.00

filtered 234k), we identified 299,026 high-precision posts related to post-college transitions. For all analyses below, we use this dataset of 299k posts to study post-college transitions.

4 METHODS

4.1 RQ1: Topic Model Inference and Thematic Annotations

Our first research goal was identifying discussion topics about the post-college transition. To automatically extract high-quality topics, we used Latent Dirichlet Allocation (LDA) [11]. LDA is an ideal choice because it often produces stable, human interpretable topics, and has been used in prior work for health prediction and social media [15, 35, 74]. We concatenated the title and text body of the 299k posts to create a bag-of-words representation of the dataset, removing stop words and stemming the words. Then, we applied online LDA with 10 passes and alpha and eta values to defaults for topic numbers of 5 to 50, in increments of 5[3].

To evaluate the quality of the topic models, we look to the recommendations of Wallach et al. and Chang et al., and used perplexity scores as well as human evaluation. Perplexity score measures how well a probability distribution predicts on a held-out sample, and is a common evaluation metric for topic models. In Table 3, we report the perplexity scores of each topic model. In alignment with Wallach et al.'s findings, we found perplexity scores to be somewhat useful in guiding our selection process — the topics with the lowest perplexity (45 and 50 topics) were less semantically coherent to the research team, with a substantial increase in noisy and irrelevant words. During mutual discussion with the research team, we found the topic keywords for the 35 and 40 topic LDA models to be most semantically coherent, while still providing topic diversity and comparable performance when evaluated strictly with perplexity.

4.1.1 Interpretation of Topics with Human Labels. To systematically decide which topic model to use (either the 35 or 40 topic models) as well as extract themes, we designed an interpretive annotation task for the LDA models. Two members of the research team inductively and independently coded each topic for implied themes about the post-college transition. Examples include "financial aid and tuition questions" and "how to move out."

The two researchers met and compared which model was the most semantically coherent, considering within-topic coherence, between-topic separation, and thematic presence. The researchers agreed that the topic model with 35 topics was most semantically coherent. Finally, the researchers

[3]We used *gensim.models.LdaMulticore* and its default parameters

mutually assigned a theme and title for each topic. We use the 35 topic model and these themes for the remainder of the paper.

4.2 RQ2: Stress Classification and Temporal Analysis

4.2.1 Building a Classification Approach. Inspired by past work in using social media to understand mental well-being [15, 16, 78], we used transfer learning to assess stress expressions in our dataset. Transfer learning is a popular approach in machine learning that trains and tunes a supervised model on a closely related but different dataset — the resulting model is then "transferred" to the new, target dataset [84]. Transfer learning works well in unsupervised scenarios where there is a lack of ground truth labels in the target data, but a closely related, labeled dataset already exists. Similar transfer learning approaches have been successful in identifying the language of symptomatic mental health outcomes in other work [8, 80].

We drew on the approach introduced by Saha and De Choudhury to develop a binary transfer learning classifier for **high** and **low** stress on Reddit [78]. Their class labels were based on the Perceived Stress Scale (PSS), a clinically established psychometric stress assessment [18]. In its 10-item version, PSS generates a score between 0-40, where scores around 13 are an average level of stress. In a factor analysis done on PSS scores, the scoring indicated that there were two major classes of stress, high and low stress [44]. Our work adopted Saha and De Choudhury's binary classification using PSS for high and low stress and expands on their approach to our specific questions around post-college transitions.

Source Data. To train our classifier, we obtained Reddit data from specific subreddits related to high and low stress.

For *high stress* (class label = 1), data came from three subreddits: 1) r/Stress (self discussion of stressors and stress management); 2) r/badroommates (self discussion of bad roommate experiences); and 3) r/CollegeRant (self discussion about negative college experiences). These subreddits were selected for two reasons: r/Stress is a source of precise information on people who self-identify as being stressed and has been used in prior work for similar purposes [78, 80]. To tune our classifier to the experiences of college students, we identified and incorporated the data of two college-related subreddits, r/badroommates and r/CollegeRant, into the positive class labels. In transfer learning, complementary datasets have shown to improve transfer learning performance on the final task [16, 78], and our experiments with this approach also showed the same. In total, these three subreddits contributed 16,256 posts.

For *low stress* (class label = 0), we obtained data from 1) r/popular (Reddit's algorithmically-generated subreddit for popular posts), 2) r/collegeadvice (advice related to college), 3) r/GetStudying (discussion of non-stressful studying), and 4) r/DecidingToBeBetter (discussion about self improvement). We used r/popular data because it serves as an appropriate control data for non-stressed discussions [78] from 2011-2014[4]. The other three subreddits were similar in scope to our interests in college students, and were selected because the research team agreed that they contained substantially less stressful conversations. Through experiments, we found that scaling this dataset to be about 4 times the size of the positive dataset produced a robust and stable classifier, and we randomly sampled from the 4 subreddits above in proportional amounts. The low stress dataset contained 65,024 posts.

On the high and low stress datasets, we cleaned and pre-processed the data by removing stop words, stemming the words, and normalizing the text. For features, we experimented with the inclusion of uni-, bi-, and tri-grams through term frequency-inverse document frequency (TF-IDF),

[4]The COVID-19 pandemic and the 2020 U.S. presidential election season was featured prominently in more recent r/popular data. Informative features of low stress data were dramatically skewed towards these two topics.

Table 4. Summary of F1 scores across k-fold cross-validation.

Classifier	F1 Across Folds	Mean.	Std.
Multinomial Naive Bayes	0.89, 0.89, 0.89, 0.89, 0.89	0.89	0.003
Bernoulli Naive Bayes	0.78, 0.79, 0.78, 0.78, 0.79	0.78	0.004
Logistic Regression	0.93, 0.94, 0.94, 0.94, 0.94	0.94	0.001
SGD Classifier	0.93, 0.94, 0.93, 0.94, 0.93	0.93	0.002
SVM (RBF Kernel)	0.94, 0.94, 0.94, 0.95, 0.94	0.94	0.002
SVM (Linear Kernel)	0.93, 0.94, 0.94, 0.94, 0.94	0.94	0.002

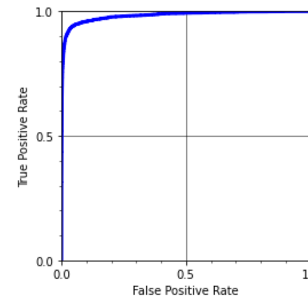


Fig. 1. ROC Curve of the SVM (linear) Classifier.

as well as the number of features used in the model. Our initial experiments indicated that uni- and bi- grams at n=10,000 features produced the most stable model.

We tested the following classifiers: Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, SGD Classifier, and SVMs (using a Linear and Radial Basis Function [RBF] kernel), using python's `sk-learn` and default parameters for initial evaluation. We used an 80-20 training-testing split using 5-fold cross-validation.

The RBF SVM classifier, Linear SVM classifier, and Logistic Regression classifier had the highest average F1 scores for the initial evaluation and balanced performance between precision and recall. Although the RBF SVM had the absolute highest F1 score, we opted to use the Linear SVM classifier because performance was very similar, and linear models produce naturally intuitive and interpretable features for analysis. On our final run of the Linear SVM, we achieved an accuracy of 0.96, precision 0.96, recall 0.93, and F1 of 0.94. The performance is strong, with balanced performance between precision and recall. In Figure 1, we provide the receiver operating characteristic (ROC) curve of our classifier.

Finally, two researchers independently labeled a sample of 75 posts to further verify the performance of the classifier, using PSS to guide their ratings of high and low stress. We found a Cohen's Kappa of 0.76 between the two raters on stress expression. Comparing the raters and the classifier's labels, we found an F1 of 0.78 and accuracy of 0.78, with balanced precision and recall at 0.79/0.78. This shows the convergence and divergence validity of our stress classifier in our problem.

4.2.2 Understanding the Predictive Features. Interpreting and contextualizing models are an important component of human-centered perspectives in quantitative work [13]. We present the most predictive features that indicate high and low stress in our classifier. We evaluated feature importance using the recursive feature elimination algorithm provided through the `sklearn` implementation of Linear SVM. This measure provides the relative importance of each feature in influencing high and low stress classifications — larger magnitude implies a stronger influence, and sign dictates the direction of the influence on positive (class = 1) or negative (class = 0) prediction. We then hand-selected meaningful and domain-relevant features and present them in Table 5.

Many words associated with low stress showed signs of sharing jokes and information (URLs to Twitter and Amazon) and support-seeking efforts that may assist in stress relief ("psychologist," "tip"). In particular, we note the difference in low stress words that seem to imply resolution of topics ("view," "instill," "psychologist," "tip") rather than active conflicts that may be causing stress ("stepfath," "bankrupt," "argument"). Examples of these kinds of support-seeking labeled as low stress by our classifier include the following excerpts:

Table 5. 48 of the top 100 features connected to low and high stress classification. Low stress features appear with negative classifier coefficients and **green** bars represent magnitude of coefficients. High stress features appear with positive coefficients and **pink** bars represent magnitude of coefficients. Words appear as root words because of language normalization.

Low Stress Feature				High Stress Feature			
Feature	**Coefficient**	**Feature**	**Coefficient**	**Feature**	**Coefficient**	**Feature**	**Coefficient**
https://twitter	-6.64	huge	-1.55	uneven	7.52	tylenol	2.12
amazon.com	-3.07	instill	-1.53	stepfath	5.18	heartless	1.95
weaken	-2.98	euphoria	-1.51	relief	4.28	inquiri	1.92
weav	-2.82	clutch	-1.49	bankrupt	4.00	whomev	1.91
intol	-2.70	psychologist	-1.44	trajectori	3.54	smartphon	1.87
stake	-2.10	specul	-1.38	recount	3.02	heartbroken	1.87
honor	-1.73	autom	-1.37	friendlier	2.80	morbid	1.82
view	-1.66	slump	-1.34	arguement	2.74	let_fade_away	1.80
bitter	-1.64	filthi	-1.34	egocentr	2.60	rape_victim	1.75
taunt	-1.61	ground	-1.32	seper	2.33	grit	1.74
tip	-1.59	construct	-1.31	undo	2.18	smack	1.73
lucrat	-1.57	prior	-1.30	need_advice	2.17	deepen	1.70

> "Never thought I'd be asking this, but how do I make friends again? Any tips?"

> "Like most people, I'm not working in my dream field. I'm hoping for some huge changes this year"

On the other hand, high stress feature words as understood by our classifier related to stressful situations ("argument," "bankrupt"), life events and common stressors ("stepfath," "heartbroken"), as well as strong language on seeking support from the community ("seper-," "need_advice," "inquiri"). We saw similar trends in the 75 post evaluation task earlier, and interpreted the high-stress posts to indicate active signs of acute distress that aligned with PSS. Similar excepts from example posts highlight the differences in acute stress:

> "My roommate is an absolute nightmare. We have countless conversations and arguments about his behavior, but none have gone anywhere"

> "Even my stress-relief hobbies are stressing me out. I'm falling into depression, which I'm already on medication for"

These quotes and the features indicate that the classifier can distinguish between low-stress behavior about seeking general advice, and high-stress posts where the user describes acutely stressful experiences. In sum, (Table 5) and corresponding examples help us evaluate the face and construct validity of our classifier [69].

4.3 RQ2a and 2b: Explaining Temporal Stress Variation with Topics

In RQ2, our goal was to explore the volume of high-stress content changed over time, and explore its connection to temporal factors in the post-college transition. To do this, we used our Linear SVM classifier to classify our dataset of 299,026 posts for low and high stress, and we use these labels to organize RQ2's analysis.

4.3.1 RQ2a: Temporal Analysis of High-Stress Posts. After applying our stress classifier to the whole dataset, we filtered our dataset to extract posts labeled as high-stress (class label = 1) by our classifier. This allows us to focus on the high-stress content in these communities.

For our temporal analysis, we aggregated posts in weekly buckets over a multi-year time frame to account for calendar changes. We then calculated the normalized percentage of stress posts across time by week, represented as the number of high-stress posts in a week divided by the square root of the total number of posts in a week. This is a variant of the TF-IDF (term frequency-inverse document frequency) technique which accounts for both raw counts as well as proportion of posts [78]. This process accounts for time series irregularities in posting volume, both in a week and over time. We matched up the data from 2011 - 2019 by week and graphed the trend of stress over the school year.

To understand what topics were most common in high-stress posts, we applied the LDA model onto each high-stress post to extract the distribution of topics present in the post. This allowed us to determine relevant topics and generate a topic relevancy "score" (the number of high stress posts that were associated with the topic divided by the total number of high stress posts in total) for each of the 35 topics.

4.3.2 RQ2b: Temporal Analysis of Relevant Topics to High Stress Posts. Finally, we wanted to examine temporal changes in topic relevance and discussion over time, bringing our three areas of focus together. Using the high stress posts, we binned the posts by calendar year (12 buckets for each calendar month, January - December). Then, we took the most discussed topics in the high stress posts and calculated the normalized volume of topical presence for each month. This was calculated as the number of high stress posts in a month in which a given topic appeared divided by the square root of the total number of high stress posts in that month (the same normalization technique applied before).

5 RESULTS

5.1 RQ1: Topic Model and Thematic Analysis of Challenges in Post-College Transitions

In this section, we discuss the results of our LDA topic model and relevant topics. In Table 6, we feature 8 topics out of 35 that are exemplary of the themes we found and the diversity of posts in these subreddits. We provide the complete list of topics and themes in Appendix Table A1.

To begin, we found many requests for informational advice about post-college transitions for **recruiting into new employment and career situations**. This included questions about best practices for interviewing, recruiting, and breaking into private industry (Topics 19 and 31), highlighted by words like "interview," "recruit," and "phone." We also saw general questions and advice about jobs and evaluating employment opportunities in Topic 29. We saw support-seeking for **post-undergraduate education and finding a research advisor** (Topic 18), with words like "advisor," "professor," "project," "academia," and "mentor." This included getting recommendations, funding, conducting research, and evaluating programs. One student returned to the subreddit they participated in – *"after total rejection last cycle, I just committed to a program that I'm really excited about!!...I wanted to share my experience."* These posts mirror information-seeking support practices through stressful experiences found in the past literature [22].

In addition to questions about gainful employment after college, people asked questions to evaluate their options. In Topic 11, students asked for support and advice on **locations**, including living situations, cities to move to, and moving for post-secondary school/work opportunities, indicated by words like "state," "city," "location," and specific place words like "california" and "new york." These evaluations were often related to different choices in school or work, evaluating locations based on their desirability (e.g. "rank," "better," "cost").

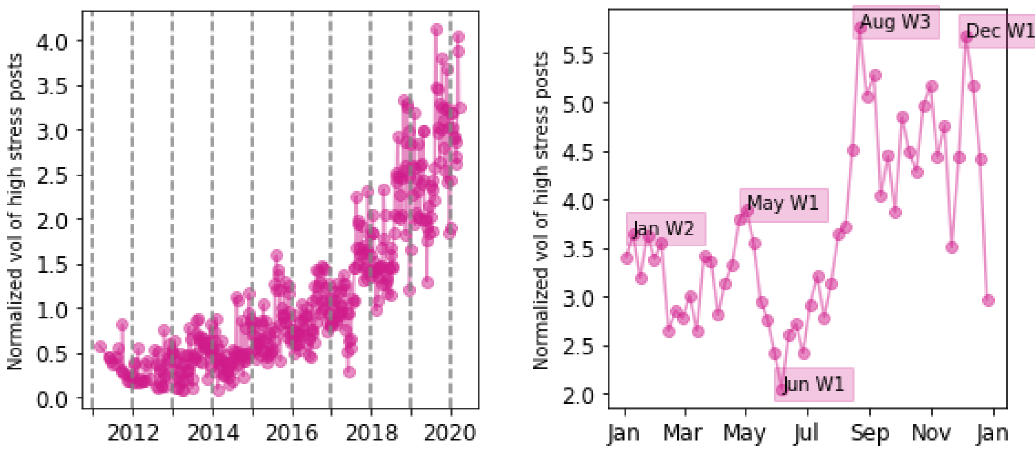
As collegiate support trickles off, new questions about post-college lifestyle emerge, what Wendlandt and Rochlen describe as both "pre-entry knowledge and [setting] expectations" [89] [p. 159]. One important concern was changing **social dynamics and relationships** from the loss of social opportunities, as seen in Topic 12. These posts explored socializing, meeting new people,

Table 6. Examples of LDA derived topics from the 35-topic model LDA applied to post-college transition data. Each topic is annotated with a topical theme, representative words, and example post snippet. Appendix Table A1 provides an extensive list of thematic topics observed in our data.

Topic 9: **Health and Mental Well-being**
Rep. Words: health, medical, mental, nurs, doctor, help, hospital, issue, care, healthcare, psychology, counseling, medicine, physician, clinic, social, therapy, family, disable, anxiety, depression, patient, med, therapist, ill
Example: "A month ago, campus health started me on SSRIs to help with my long-standing depression and anxiety. While they have helped somewhat, they have been causing fatigue — the higher they set my dose, the worse the fatigue gets."
Topic 11: **Locations**
Rep. Words: univers, state, city, live, california, rank, school, location, texas, attend, new, program, york, florida, install, area, chicago, look, better, boston, transfer, cyber_security, cost, washington, current
Example: "What I mainly do is sift through the schools based on location (mainly California and New York since I have friends and family there) and then within those schools I'd search for faculty members who may fit my interests."
Topic 12: **Social and Relationships**
Rep. Words: friend, people, like, talk, social, know, go, school, feel, meet, want, home, new, lot, family, group, person, think, life, live, party, city, college, year, join
Example: "After university, it will get harder to meet and talk to new people."
Topic 17: **Cost of School, Loans, and Debt**
Rep. Words: pay, year, money, live, save, school, loan, need, cost, time, parent, debt, famili, work, month, financial, expense, go, student, income, home, help, insurance, afford, able
Example: "I am aware that social workers do not make a ton generally. I'm not interested in having a career if I don't pursue this MSW, but the smart thing to do seems to be not to go into so much debt."
Topic 18: **Finding a Research Advisor**
Rep. Words: student, project, advisor, depart, work, supervisor, meet, group, professor, thesis, advise, member, lab, graduate, talk, academia, faculty, committee, discuss, dissertation, research, mentor, propose, defense, university
Example: "After total rejection last cycle, I just committed to a program that I'm really excited about!... I have received a lot of good advice from this community, so I wanted to share my experience."
Topic 19: **Interview Advice**
Rep. Words: interview, ask, question, know, thank, answer, like, want, advice, think, help, recruit, phone, person, expect, go, tip, sure, good, say, guy, people, hear, talk, look
Example: "I don't have anything relevant on my resume. I'd really appreciate some advice on how to get started as a paid writer or something similar."
Topic 29: **Job Considerations**
Rep. Words job, work, company, position, month, year, current, time, experience, get, look, pay, employment, application, new, leave, start, want, offer, resume, hire, interview, salary, internship, graduate
"My academic career has been ideal. I graduated Magna Cum Laude, however, I cannot seem to land a job! It is now a month after graduation and I am still unemployed. I am in serious need of advice."
Topic 31: **Stress, Scheduling, and Time**
Rep. Words: work, hour, day, time, week, home, schedule, start, spend, night, weekend, stress, need, shift, break, go, long, sleep, minute, like, month, stay, office, sick, commute
Example: "Since switching to nights, I cry before every shift. I know I should be putting my mental health first but I just have such a hard time giving things up, because I have worked so hard to get where I am now."

making friends, and navigating relationships, indicated by words like "friend," "talk," "social," "meet," "family," "party," and "group." A common frustration was loneliness and challenges in making friends, especially in new cities, as described by one person, *"After university, it will get harder to meet and talk to new people."*

In addition to managing expectations of the transition, we also saw recent graduates discussing immediate and pressing concerns about adjusting to post-college life. This came through in discussion of the **cost of school, loans, and resulting debt**, indicated by words like "pay," "money,"



(a) Across Years

(b) Monthly

Fig. 2. Temporal patterns of normalized volume of high stress posts in post-college transition subreddits. The left plot shows the volume of high stress posts from 2011 to 2020. The right plot shows the monthly average volume of high stress posts.

"save," "loan," "cost," and "debt" (Topic 17). Posters wrote about financing post-secondary education, paying for student debt when out of school, covering living expenses, budgeting, loans, and other debts.

Finally, we saw conversations about **stress, mental health, and well-being** of new graduates in Topic 9. One common point of discussion was concern about mental health and getting proper treatment. This includes words that reference health care, including "health," "doctor," "healthcare," "counseling," "anxiety," "depression," and "therapist." The concerns in this topic align with past work on the post-college transition, which ties this life phase to short-term decreased stress and well-being markers [50, 62].

These topics and the topic model overall reveal the major challenges faced by individuals undergoing post-college transitions, as well as seeking support during periods of isolation and loneliness. These align with prior findings of the concerns of college students and young adults [50] and those entering new workplaces [89].

5.2 RQ2: Stress and Temporal Analysis of Challenges in Post-College Transition

5.2.1 RQ2a: Temporal Analysis of High Stress Posts. Next, we report on the results of applying the high stress classifier to the post-college transition data for temporal patterns. Figure 2a shows the normalized volume of stress over time. Figure 2a shows a gradually rising trend from year to year of normalized stress volume throughout the years from 2011 to 2020. Recall that these amounts are normalized to control for variation in posting frequencies.

Beginning with the monthly stress evaluation in Fig 2a, we noted that the normalized volume of high stress posts increased over time. To capture this rate of change, we calculated the overall *momentum* of change in stress in the data. This was calculated as the difference between the volume of stress in the last full week of data collection (March 2020) and the first week in the dataset (February 2011), divided by the number of years in the dataset. The momentum indicates that normalized stress is going up by 29.64% per year. One possible reason for this could be an increase of people turning to Reddit for advice as Reddit became more popular — posters may become more comfortable sharing their personal, stressful situations after seeing others post. This may also be caused by “dropout” in online communities around health and wellness – individuals who are not stressed leave communities that they perceive to be misaligned with their own incentives [58].

We also examined stress in weekly buckets to understand the annual seasonality of stress in the dataset. Figure 2b shows the weekly normalized volume of high stress posts in our dataset. In January, the normalized volume of high stress started relatively low, before peaking in the middle and at the end of the month. This could indicate returning to school or work after the winter holidays. We found a peak in stress around late April, a few weeks before the end of the school year and potentially graduation. During this period, students may be finalizing their summer or post-graduation plans. This aligns with our hypothesis that stress is appearing in the dataset focused on the life transition — students may exhibit higher stress towards the end of the quarter or semester as they are finishing their exams and moving towards graduation.

Moving into the summer (May to July), stress was the lowest around early to mid-summer. This period usually coincides with the students’ summer break, so their schedules may be quite different than what they typically experience during the school year. Stress peaked in August, around the start of the school year. Typically, this may be stress caused by students entering their final year, as they begin preparation for transition [89]. However, this is also an interesting time period for post-graduates, because the average new graduate can take three to six months after graduating to find a job [2]. It may be that post-graduates face a challenging stage of finding employment during August as well.

Last, there were various peaks of stress from August to December. This corresponds with times when students start searching for future opportunities, whether applying to professional positions or graduate school. We see this reflected in the dataset, where students ask for assistance with applications (see LDA, Topics 18 and 19). We also observed a local peak in stress in the first week of December. The volume of high stress posts drops in December to January, which typically corresponds with the winter holiday. This analysis revealed that the latter half of the year was a prime time of stress for students in transition. This contrasts with findings of general college students, that stress ebbs and flows throughout the year because of academic deadlines and major exams at least twice a year.

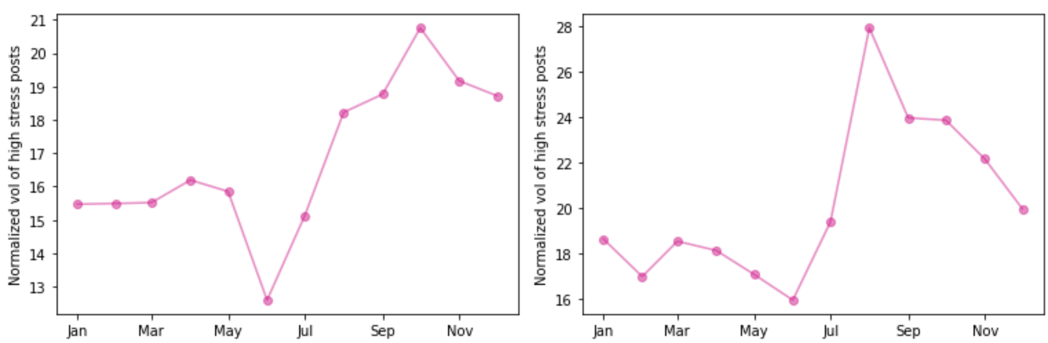
5.2.2 RQ2a: Identifying High-Stress Challenges in Social Media Data. Next, we report on the prevalence of challenges and themes that we identified in the LDA topic model in high stress posts. Recall that we calculated the top themes from the LDA by prevalence in the high stress posts.

Table 7 shows the top themes/challenges in this high stress dataset, taken from the whole LDA in decreasing order. We interpret these as the most common and prevalent stressors in individuals seeking support in these subreddits. These included **Finding a New Career** (Topic 3), **Finding a Place to Live** (Topic 8), **Admissions Chances** (Topic 4) and **Funding Applications** (Topic 32). Many topics were related to anticipation and uncertainty of the future and new things that are yet to be decided [89]. For example, a user posted about several of these topics that were worrying them: *“I’m worried about how I’m going to pay for living expenses in grad school if I move.”*

In addition to these topics anticipating future stress, we also saw evidence that students were dealing with specific stressful events during the post-college transition. These included collegiate concerns like **Stressful Classes** (Topic 0) and questions about employment and assistance with **Organizational Hierarchy** (Topic 25) and **Workplace Communication** (5). We believe that this analysis points to two interesting outcomes. One, it demonstrates that our approach can see across the post-college transition process, which includes wrapping up collegiate coursework and requirements and transitioning to new employment and opportunities afterward. Second, it highlights what topics are stressing out people seeking support during the post-college transition.

Table 7. Top 10 topical themes in the sorted order of the percentage prevalence of high stress posts.

Id	Topical Theme	% High Stress Posts
26	Expression of Emotion	84.77
3	Finding a Good Career	77.71
32	Applications for Funding	59.62
8	Finding a Place to Live	59.47
12	Social and Relationships	55.27
25	Organizational Hierarchy	54.35
4	Admissions Chances	49.62
0	STEM Coursework	48.63
5	Workplace Communications	46.91
28	Seeking Advice & Support	43.66



(a) Admission Chances

(b) Funding Applications

Fig. 3. Temporal patterns of normalized volume of high stress posts per topical theme. Note that the axes change slightly.

5.2.3 RQ2b: Temporal Changes in High Stress Challenges. Finally, we combined high-stress posts with the topic modeling and the temporal analysis to see temporally-specific high stress topics and their rates of discussion on Reddit.

We observed a common pattern that stress tends to increase as the year progresses. That is, stress was typically lower in the months of January to July and was higher during the fall and winter months, similar to the findings in RQ2a. We hypothesize that these temporal patterns likely reflect when and how people face challenges during the post-college transition.

However, this analysis can also highlight whether certain topics have atypical temporal patterns of stress outside of the normal ebbs and flows of a year. We examined two exemplary topics that demonstrate this in conversational volume over time, as shown in Figure 3. For **Admissions Decisions** shown in Figure 3a, we found that stress around admissions peaks in November and December, typically when graduate admissions applications are due at many universities. Rather than a corresponding downturn in stress after that, we saw that stress plateaus in January through May, with another small peak in April as we imagine that admissions decisions are being returned to students. Similarly, we found a major peak in stress related to **Funding** and money (Figure 3b) in August, which dramatically decreases afterward.

6 DISCUSSION

This work offers empirical insight to understand the stress and challenges of the post-college transition on Reddit. We investigated how people in transition interact with others in online communities. We found students seeking support for the post-college transition: starting discussions, sharing personal experiences or struggles, offering tips, and complaining together during a stressful time period of leaving college and entering emerging adulthood. Our work aligns with research on life transitions and disruptions [12, 29, 40, 42, 45, 59, 86], and we find similar results to other research about social support, life transition, and online communities [4, 27]. In short, we find that that social media sites like Reddit can be a source of assistance in challenging times. In the Discussion, we better contextualize our results, connecting our results to the prior work, and discuss methods and design implications for online communities, and for stakeholders in the post-college transition.

6.1 Theoretical Implications to Social Support and Stress

Provisioning Social Support for Post-College Transitions. We first explored the topics and challenges that individuals discussed in their Reddit posts during the post-college transition. Many topics from our LDA model in RQ1 indicate the provisioning or seeking social support through Reddit. We observed that people sought support and discussed procedural and informational details, like requirements for graduation, post-baccalaureate academics, applications and recommendation letters for graduate school, and how to interview for jobs.

In addition to information-seeking behaviors, participants also discussed deeper questions, seeking assistance with decision-making that moved beyond procedural aspects of the transition. Students often discussed how to balance their passions with job prospects, and how to combine multiple interests to develop a career path. One post captured the general intention of these post well, saying:

> "I want to make the smartest decision for my future."

These topics align with prior work about challenges during the post-college transition [50, 65] and also focus on *emotional support* for listening, encouragement, and understanding [22].

The acquisition of both informational and emotional support through transition has been identified as a crucial part of the support process laid out by Cutrona and Suhr [22] as well as frameworks around the different needs during life transitions [62]. In our dataset, individuals appreciated help from others with whom they shared a sense of belonging, and could relate to having similar experiences, i.e. receiving *network support* in these subreddits [27].

> "I received two responses to my concern, one response encouraged me to do [my work] nicely, the other told me to get my ass up... I'm thankful to both because I finally goddamn finished it 4 days earlier!!"

Prior work has found esteem and network support to be most efficacious in online (mental) health communities on Reddit [27], and we see these themes and theoretical connections to the support literature in our quantitative analyses.

Stress and Expanding Transitions Research. In RQ2, we studied high-stress topics and temporal facets of them. Our stress classifier can estimate when students are stressed, and see what times of the year could potentially be more stressful. Tracking levels of stress over time allowed us to pinpoint times of the year when there was a higher influx of stress posts. Finally, we connect the classifier to changes in stress over topics and over time, highlighting what may be the cause of such stress.

In addition to social support, temporal factors of stress for the post college transition tie into the literature on transitions, stress, and temporal change. Wendlandt and Rochlen propose a three-stage framework for conceptualizing the transition process of college-to-workplace - "anticipation" of transition facets before they occur, "adjustment" as individuals move through the event of concern, and "achievement"[89]. Although Wendlandt and Rochlen applied this to transition in the workplace, we see similar patterns in the post-college transition example using the temporal analysis in RQ2.

Combining LDA with the stress classifier in RQ2a, we can examine what topics cause the most stress. We see that these map to many anticipatory concerns, which "occur before individuals actually enter into an organization"[89], including information about moving, getting into graduate and professional programs, and finding a good career. This method also shows adjustment concerns, as it identified pressing questions about adjusting to work environments, organizational hierarchy, and communication in the workplace.

Finally, our empirical approach provides a temporal perspective on stress during transitions, something often missing from quantitative studies of online communities and important to Wendlandt and Rochlen's approach for transition. In RQ2b, we studied the challenges students face over time; our method documented when specific stressors were more or less pressing at a given time. For example, we saw the stress of being admitted to graduate programs lingered from Jan to April — the temporal analysis in RQ2 would not have shown this given its noticeable dip in stress during that time. Quantitative approaches and research on online communities often collapse representation across time into a singular setup, thereby losing temporal fidelity [13]. Our approach allows for a more fine-grained temporal study of the challenges of post-college transitions. In sum, our research supports the use of social media data for modeling stress [39, 77, 78] and to temporally forecast and track behavior [25, 72].

6.2 Research and Methods Implications

6.2.1 Augmenting Surveys of Post-College Transitions. Surveys are a primary method to gather data about the post-college transition, and they often focus on recent graduates. These surveys report on academic outcomes (i.e. percentage or number of students who graduate, degrees conferred, final GPA) as well as placement data (i.e. salary information, and employment location). Additionally, external organizations also collect statistics on college students and their post-graduate outcomes, such as the National Center for Education Statistics [67] or the Computing Research Association's "Data Buddies" project [3].

Although surveys are an important source of information about graduates, surveys of this cohort have a few notable limitations. Many surveys happen at the end of or after college students have left the university, and many surveys do not explicitly ask about or report mental well-being markers — the closest they may capture is satisfaction with the overall college experience. Survey response rates are also affected by many factors such as survey length, accessibility, timing, engagement, and presence of an incentive [85]. Post-graduates are in a life transition that often entails a change of geophysical location and occupation, making students difficult to physically locate and ensure that reliable survey responses can be gathered.

Our work offers a unique perspective to traditional sources of survey data as we leverage social media as an alternative data source of people's post-college experiences. Social media data is forthrightly shared, meaning that the responses do not need to be sought from individuals. Furthermore, the pseudoanonymous nature of Reddit enables individuals to candidly share personal content. We show the utility of this method through RQs 1 and 2, as we can extract topics related to the post-college transition experience, and also study temporal changes in stress over time.

We imagine our methods augmenting and improving student surveys in a few ways. First, the topics from RQ1 and our LDA may facilitate new questions about student satisfaction and post-collegiate placement – we wonder if questions about "Satisfaction in Transitioning to Post-College" may be appropriate for student outcomes in addition to typical markers of success like job placement. Additionally, we envision deploying surveys with an eye to the temporal methods and results of RQ2b – surprising places where stress peaks may lead survey designers or campus employees to target survey timing. We look forward to future work that may integrate our insights into research and practice in survey design and development and improving student outcomes.

6.2.2 Methods Contributions. In addition to augmenting surveys, we believe some of our methods approaches are valuable to CSCW and HCI.

First, the lexico-semantic similarity filtering (LSSF) is a powerful tool for filtering and finding more specific and relevant information from social media communities. This technique builds on recent work in CSCW and HCI that has adapted word embeddings for filtering and data discovery [16, 76, 77], and we offer LSSF as a useful technique in the toolkit of empirical researchers in HCI. Finding specific experiences or details in large and broad subreddits is challenging, and can lead to substantial noise. Our high-dimensional word embedding of data known to be relevant to the post-college transition let us filter larger, less specific communities for our signal of interest. This "semi-supervised" approach of a word embedding with hand-curated data makes it easier to gather larger datasets on more nuanced questions from online communities.

In future applications of LSSF, we recommend close attention by the research team to the overall process of dataset design and filtering. The word embedding and its training data are just as important as evaluating cosine similarity thresholds required for sufficient filtering for a given research context. We are particularly excited about lexico-semantic similarity filtering and its ability to assist qualitative research in directed sampling for topics of interest in online communities. This approach may also help to mitigate some of the negative privacy implications of targeting users from their metadata, characteristics, or self-described traits in online communities. In short, we encourage future research teams to adopt this approach as a larger push towards human-in-the-loop decision-making for quantitative research and data gathering.

Second, the combination of LDA and transfer learning fuses computational techniques to study temporal changes of stress over time. We hope that the lexicon we provide through the LDA and resulting stress classification can be used by researchers interested in the post-college transition in online communities. Stress manifests across life transitions and events broadly, and we envision alternative transfer learning protocols being fused with topic models, regression techniques, or more robust time series modeling for quantitative research in online communities.

6.3 Implications for Stakeholders in Post-College Transitions

6.3.1 Implications for Online Communities. Last, this work points to design implications for online communities. For example, we envision structural conversation features on Reddit that could support our findings. These may be adopted and implemented by community moderators and administrators, more than just on Reddit. The LDA methods we introduce point to tools that could be leveraged by subreddits interested in supporting the post-college transition. For instance, subreddits could adopt weekly or monthly threads based on the time of year and known stressors at that time (November for graduate school applications, February for employment interviews). We also imagine better recommendations for topical "ask me anything" threads from professionals and experts that align with popular topics and areas of concern, like preparing for interviews (Topic 19) or financial costs and logistics (Topic 11 and 17).

As these online communities significantly rely on peer support, it is important to delineate what constitutes effective support responses in these subreddits. Drawing on research like ours, moderators can prepare guidelines and recommendations for better responding to people's expressed challenges and stress. Given the importance of information sought and shared, it is also essential that these communities create norms that guide sharing credible sources of information. Centering these guidelines around the topical themes found in our work could be one way to start improving these online communities of post-college transition.

Given that we have many subreddits that contribute to discussions about this topic, we also imagine larger initiatives on social media sites that cross-cut communities. Subreddits with similar goals and user bases could share a Wiki or resources that could be collaboratively managed. Subreddits could also coordinate buddy matching efforts of students with similar transition pathways or similar concerns, using the temporal matching we provide in RQ2 as an additional source of matching information. Finally, we also imagine that inter-community dynamics could drive the creation of new content in Wikis or new subreddits along topical interests. These may emerge from the topics we find in RQ1 or interesting applications of the LSSF we propose in the paper.

6.3.2 Implications for College and Support Staff. Our work also bears on stakeholders that support students through these transitions. First, we hope our approaches could support college career and development staff to anchor and develop their programming. Counselors and staff could use the topic suggestions from RQ1 to identify opportunities for programming to support students while they transition their careers, or provide information through blogs, webinars, or workshops on these questions. These workshops can be strategically planned using our temporal analysis from RQ2, where unexpected time and sources of stress may pop up. For example, one common concern is blending together passions and job prospects, and college counseling centers can introduce workshops that tackle this with a panel of professionals with unconventional career paths.

Although we expect the content of the LDA and stress analysis to complement what is known about the post-college transitions, we believe our analysis helps college career and development staff more effectively and efficiently deploy their resources. Online social media analysis gives stakeholders on campuses a real-time pulse on changing situations and student stress. This can assist in making better decisions about timing for resources and their deployment. This matches the intuitions we have about deploying programming already, like job interview workshops in Jan/Feb, but this gives us concrete evidence about what is stressing students and providing immediate responsiveness to stress demand.

With some work to tweak the outputs, campus administrators could become more familiar with their campus subreddits to better understand students' unique needs. This could allow tailored insights for stress management and responsiveness by administrators given the predictable yet unique circumstances of each academic year and campus environment. Complementary to this, staff may consider how social media and interaction there may support the goals and aims of the center, through a better understanding of their own student populations or through more strategic social media interactions. To be clear, such systems would need to be tempered with the necessary privacy and ethical awareness that these systems could be used to violate privacy expectations and norms of self-disclosure in these communities.

That being said, colleges cannot be the unilateral solution for managing the "no man's land" of support in the post-college transition. Colleges are certainly the most prepared to educate students before they graduate, but they should not be wholly responsible for managing what can be a multi-year and geographically diverse transition [50]. The question remains what institutions could provide more comprehensive support for these individuals, outside of universities and online communities. This is a complex question, and the answer is out of bounds of this paper, but essential to addressing the challenges we find in our work.

6.4 Ethics and Privacy Considerations

Our research examined public disclosures on Reddit and the research team had no interactions with users – for this reason, we did not seek ethics board approval for our research. Despite this, we heed recent calls for more human-centered practices in quantitative research on sensitive topics like mental health despite a lack of ethics board approval or oversight [9, 13, 14]. We reflect on important study design decisions that impact ethical and privacy concerns.

First, we explored aggregated experiences across subreddits, instead of precise user modeling or targeting. Given that our study did not need individual user information, we used community-wide analysis to avoid biasing our models to individual factors. Second, we de-identified usernames, links, and subreddits in our data before computational analyses to minimize the risk of individual traits driving model development. Quotes have been edited to help mitigate the re-identification of individuals, and we removed identifying language from LDA and model reporting to avoid spotlighting certain posts. Last, we provide details for model replication of our machine learning and unsupervised language analysis rather than release models after publication. We acknowledge this as a trade-off in our approach between privacy preservation and replicability and benchmarking.

Despite our precautions, no research approach can outright remove all harms and risks from research, only minimize and attempt to mitigate them. If a Redditor were to face a privacy invasion due to our models or their participation [59], this could lead to increased stress, unwillingness to continue engaging with the community, and harmful ramifications for their personal or professional lives if sensitive details are disclosed [40]. We also recognize the possibility of expectation mismatches between people's self-conceptualization of social media use and their data, and inferences on this data without their consent or awareness [36]. In applying our design implications, we highly encourage critical reflection on best practices about using these techniques on social media data. This needs to be addressed from a multi-stakeholder perspective spanning technology designers, researchers, ethicists, policy-makers, well-being interventionists, and centering the individuals who are the intended beneficiaries of systems like this one.

6.5 Limitations and Future Work

Our research is not without limitations. First, our work cannot clinically validate the stress expressions on social media or guarantee that they are indeed going through the transition phase we identified. Although we draw on prior work on understanding stress to assemble this classifier [78], we caution and discourage extending our classifier to clinical contexts. Our study does not account for prior or baseline mental health status. This baseline behavior can be accounted for in future work by conducting longitudinal analyses on individual's long-term data on the social media platform with consent and integration into a larger, human-centered study. Similarly, we assume the training datasets we assembled are well-filtered and noise is managed by both the LSSF and the aggregated text representations. There is some risk that remaining noise in training data influences the classifier, and we hypothesize that this risks our model underperforming the true task.

Second, there are limitations in using Reddit as a source of data about college student behavior. Even though students and their age demographic are some of the most common users of social media [8], we cannot guarantee our sample on Reddit is representative of diverse college experiences. It is likely that this dataset skews towards U.S perspectives and the experiences of those in "traditional" four-year university experiences because of the demographic make-up of Reddit or bias towards the experiences of certain gender and racial groups. Further, social media is prone to self-selection bias, i.e., we only study those who express their concerns on social media.

Despite these limitations, we are excited about opportunities for future work. We identify an important trend that stress is increasing over time, and we hypothesize that this may be tied to the rising stress of college students overall. Future work could causally determine if this increase is caused by "dropout" of non-stressed people leaving the platform or other factors like an overall rise in stress. We also see the potential for studies of other communities in transition using our temporally-minded empirical framework, as well as combining these insights with labels from users about their stress expressions on Reddit.

7 CONCLUSION

We used Reddit data from ten subreddits related to the post-college transition and empirically studied concerns and stress in the dataset. Our dataset of 299,026 Reddit posts gave us insight into the important role that social media plays in providing support during a difficult life transition. Using our LDA, we identified subjects such as job and career considerations, living situations, financing education, and new social concerns. Next, we trained a stress classifier to classify posts for high and low stress with performance at 0.94 F1 score. We used these insights to explore temporal patterns in stress expressions and the variance of per-topic stress and challenges as they change throughout the year. In conclusion, our research offers insight into the concerns of students during the important post-college transition, and we hope this inspires future work.

8 ACKNOWLEDGMENTS

Chancellor completed the majority of this work while at Northwestern University.

REFERENCES

[1] [n.d.]. Wiki: Related Subreddits. https://www.reddit.com/r/LifeAfterSchool/wiki/related_subreddits

[2] 2019. What can students do to improve their chances of finding employment after college? https://www.washington.edu/doit/what-can-students-do-improve-their-chances-finding-employment-after-college

[3] 2021. The Data Buddies Project.

[4] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2018. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 1–35.

[5] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proc. CHI*.

[6] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: the case of# depression. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1485–1500.

[7] Jaime Arguello, Brian S. Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Rosé, and Xiaoqing Wang. 2006. Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 959–968. https://doi.org/10.1145/1124772.1124916

[8] Shrey Bagroy, Ponnurangam Kumaraguru, and Munmun De Choudhury. 2017. A Social Media Based Index of Mental Well-Being in College Campuses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1634–1646. https://doi.org/10.1145/3025453.3025909

[9] Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. 94–102.

[10] Alison Blackwell, Lindsey Bowes, LEE Harvey, Anthony J Hesketh, and Peter T Knight. 2001. Transforming work experience in higher education. *British Educational research journal* 27, 3 (2001), 269–285.

[11] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *JMLR* 3, Jan (2003), 993–1022.

[12] Moira Burke and Robert Kraut. 2013. Using Facebook after losing a job: Differential benefits of strong and weak ties. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1419–1430.

[13] Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. 2019. Who is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.

[14] Stevie Chancellor, Michael L Birnbaum, Eric D Caine, Vincent MB Silenzio, and Munmun De Choudhury. 2019. A taxonomy of ethical tensions in inferring mental health states from social media. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 79–88.

[15] Stevie Chancellor, Zhiyuan (Jerry) Lin, Erica Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and Predicting Mental Illness Severity in Online Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM conference on Computer supported cooperative work & social computing*. ACM.

[16] Stevie Chancellor, George Nitzburg, Andrea Hu, Francesco Zampieri, and Munmun De Choudhury. 2019. Discovering Alternative Treatments for Opioid Use Recovery Using Social Media. In *Proc. CHI*.

[17] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei. 2009. Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems* 22 (2009), 288–296.

[18] Sheldon Cohen, T Kamarck, R Mermelstein, et al. 1994. Perceived stress scale. *Measuring stress: A guide for health and social scientists* (1994).

[19] Sheldon Cohen and Thomas A Wills. 1985. Stress, social support, and the buffering hypothesis. *Psychological bulletin* 98, 2 (1985), 310.

[20] Glen Coppersmith, Craig Harman, and Mark Dredze. 2014. Measuring post traumatic stress disorder in Twitter. In *International AAAI Conference on Web and Social Media*.

[21] Carolyn E Cutrona. 1990. Stress and social support-in search of optimal matching. *Journal of social and clinical Psychology* 9, 1 (1990), 3–14.

[22] Carolyn E Cutrona and Julie A Suhr. 1992. Controllability of stressful events and satisfaction with spouse support behaviors. *Communication Research* 19, 2 (1992), 154–174.

[23] Carolyn E Cutrona and Beth R Troutman. 1986. Social support, infant temperament, and parenting self-efficacy: A mediational model of postpartum depression. *Child development* (1986), 1507–1518.

[24] Vedant Das Swain, Koustuv Saha, Manikanta D Reddy, Hemang Rajvanshy, Gregory D Abowd, and Munmun De Choudhury. 2020. Modeling Organizational Culture with Workplace Experiences Shared on Glassdoor. In *CHI*.

[25] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13)*. Association for Computing Machinery, New York, NY, USA, 3267–3276. https://doi.org/10.1145/2470654.2466447

[26] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.

[27] Munmun De Choudhury and Emre Kıcıman. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *ICWSM*.

[28] Munmun De Choudhury and Michael Massimi. 2015. " She said yes"–Liminality and Engagement Announcements on Twitter. *iConference 2015 Proceedings* (2015).

[29] David C. DeAndrea, Nicole B. Ellison, Robert LaRose, Charles Steinfield, and Andrew Fiore. 2012. Serious social media: On the use of social media for improving students’ adjustment to college. *The Internet and Higher Education* 15, 1 (2012), 15 – 23. https://doi.org/10.1016/j.iheduc.2011.05.009 Social Media in Higher Education.

[30] Bryan Dosono, Yasmeen Rashidi, Taslima Akter, Bryan Semaan, and Apu Kapadia. 2017. Challenges in transitioning from civil to military culture: Hyper-selective disclosure through ICTs. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–23.

[31] Daniel Eisenberg, Marilyn Downs, Ezra Golberstein, and Kara Zivin. 2009. Stigma and Help Seeking for Mental Health Among College Students. *Medical care research and review : MCRR* 66 (06 2009), 522–41. https://doi.org/10.1177/1077558709335173

[32] Daniel Eisenberg, Ezra Golberstein, and Sarah E Gollust. 2007. Help-seeking and access to mental health care in a university student population. *Medical care* 45, 7 (2007), 594–601.

[33] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook “friends”: Social capital and college studentsúse of online social network sites. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1143–1168.

[34] Erik H Erikson. 1968. *Identity: Youth and crisis*. Number 7. WW Norton & company.

[35] Sindhu Kiranmai Ernala, Asra F Rizvi, Michael L Birnbaum, John M Kane, and Munmun De Choudhury. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–27.

[36] Casey Fiesler, Cliff Lampe, and Amy S Bruckman. 2016. Reality and perception of copyright terms of service for online content creation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1450–1461.

[37] Linda K George, Dan G Blazer, Dana C Hughes, and Nancy Fowler. 1989. Social support and the outcome of major depression. *The British Journal of Psychiatry* 154, 4 (1989), 478–485.

[38] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.

[39] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes Eichstaedt, and Lyle Ungar. 2018. Understanding and Measuring Psychological Stress using Social Media.

[40] Oliver L. Haimson, Jed R. Brubaker, Lynn Dombrowski, and Gillian R. Hayes. 2015. Disclosure, Stress, and Support During Gender Transition on Facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1176–1190. https://doi.org/10.1145/2675133.2675152

[41] Oliver L Haimson, Albert J Carter, Shanley Corvite, Brookelyn Wheeler, Lingbo Wang, Tianxiao Liu, and Alexxus Lige. 2021. The major life events taxonomy: Social readjustment, social media information sharing, and online network separation during times of life transition. *Journal of the Association for Information Science and Technology* (2021).

[42] Oliver L. Haimson, Bryan Semaan, Brianna Dym, Joey Chiao-Yin Hsiao, Daniel Herron, and Wendy Moncur. 2019. Life Transitions and Social Technologies: Research and Design for Times of Life Change. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing (CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 480–486. https://doi.org/10.1145/3311957.3359431

[43] Paul Hettich. 2000. Transition Processes from College to Career Symposium: Employment with a Psychology BA. *Annual Conference of the American Psychological Association* (2000), 24.

[44] Paul L Hewitt, Gordon L Flett, and Shawn W Mosher. 1992. The Perceived Stress Scale: Factor structure and relation to depression symptoms in a psychiatric sample. *Journal of Psychopathology and Behavioral Assessment* 14, 3 (1992), 247–257.

[45] Terence Hicks and Samuel Heastie. 2008. High school to college transition: A profile of the stressors, physical and psychological health issues that affect the first-year on-campus college student. *Journal of cultural diversity* 15 (02 2008), 143–7.

[46] Jina Huh. 2015. Clinical questions in online health communities: the case of see your doctor threads. In *Proc. CSCW*.

[47] Jialun "Aaron" Jiang and Jed R Brubaker. 2018. Tending Unmarked Graves: Classification of Post-mortem Content on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–19.

[48] Berton H Kaplan, John C Cassel, and Susan Gore. 1977. Social support and health. *Medical care* 15, 5 (1977), 47–58.

[49] Jessie Koen, Ute-Christine Klehe, Annelies van Vianen, Jelena Zikic, and Aukje Nauta. 2010. Job-search strategies and reemployment quality: The impact of career adaptability. *Journal of Vocational Behavior* 77 (08 2010), 126–139. https://doi.org/10.1016/j.jvb.2010.02.004

[50] Lisa Kuron, Sean Lyons, Linda Schweitzer, and Eddy Ng. 2015. Millennials' work values: differences across the school to work transition. *Personnel Review* 44 (09 2015), 991–1009. https://doi.org/10.1108/PR-01-2014-0024

[51] Joel A Lane. 2016. Attachment, well-being, and college senior concerns about the transition out of college. *Journal of College Counseling* 19, 3 (2016), 231–245.

[52] Peter LeViness, Carolyn Bershad, and Kim Gorman. 2017. The association for university and college counseling center directors annual survey. *Retrieved fro m https://www. aucccd. org/assets/documents/Governance/2017% 20aucccd% 20surveypublicapr26. pdf* (2017).

[53] Cindy H Liu, Courtney Stevens, Sylvia HM Wong, Miwa Yasui, and Justin A Chen. 2019. The prevalence and predictors of mental health diagnoses and suicide among US college students: Implications for addressing disparities in service use. *Depression and anxiety* 36, 1 (2019), 8–17.

[54] Jason Liu, Elissa R Weitzman, and Rumi Chunara. 2017. Assessing behavioral stages from social media data. In *CSCW*.

[55] Sam Liu, Miaoqi Zhu, and Sean D Young. 2018. Monitoring freshman college experience through content analysis of tweets: observational study. *JMIR public health and surveillance* 4, 1 (2018), e5.

[56] Adriana M Manago, Tamara Taylor, and Patricia M Greenfield. 2012. Me and my 400 friends: the anatomy of college students' Facebook networks, their communication patterns, and well-being. *Developmental psychology* (2012).

[57] Gloria Mark, Yiran Wang, Melissa Niiya, and Stephanie Reich. 2016. Sleep Debt in Student Life: Online Attention Focus, Facebook, and Mood. In *Proc. 2016 CHI Conference on Human Factors in Computing Systems*. 5517–5528.

[58] Michael Massimi, Jackie L. Bender, Holly O. Witteman, and Osman H. Ahmed. 2014. Life Transitions and Online Health Communities: Reflecting on Adoption, Use, and Disengagement. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1491–1501. https://doi.org/10.1145/2531602.2531622

[59] Michael Massimi, Jill P. Dimond, and Christopher A. Le Dantec. 2012. Finding a New Normal: The Role of Technology in Life Disruptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 719–728. https://doi.org/10.1145/2145204.2145314

[60] Marifran Mattson and Jennifer Gibb Hall. 2011. *Health as communication nexus: A service-learning approach*.

[61] Sharan Merriam. 2005. How Adult Life Transitions Foster Learning and Development. *New Directions for Adult and Continuing Education* 2005 (11 2005), 3 – 13. https://doi.org/10.1002/ace.193

[62] Jude P. Mikal, Ronald E. Rice, Audrey Abeyta, and Jenica Devilbiss. 2013. Transition, stress and computer-mediated social support. *Computers in Human Behavior* 29, 5 (28 Jan. 2013), A40–A53. https://doi.org/10.1016/j.chb.2012.12.012

[63] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[64] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on Facebook: depression disclosures by college students on a social networking site. *Depress. Anxiety* (2011).

[65] Kerri Murphy, David Blustein, Amanda Bohlig, and Melissa Platt. 2010. The College-to-Career Transition: An Exploration of Emerging Adulthood. *Journal of Counseling & Development* 88 (04 2010). https://doi.org/10.1002/j.1556-6678.2010.tb00006.x

[66] John A. Naslund, Stuart W Grande, Kelly A. Aschbrenner, and Glyn Elwyn. 2014. Naturally Occurring Peer Support through Social Media: The Experiences of Individuals with Severe Mental Illness Using YouTube. *PLoS ONE* 9 (2014).

[67] US Department of Education. [n.d.]. National Center for Education Statistics. https://nces.ed.gov/

[68] "US Department of Education". 2018. Degrees conferred by postsecondary institutions, by level of degree and sex of student: Selected years, 1869-70 through 2028-29.

[69] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2016. Social data: Biases, methodological pitfalls, and ethical boundaries. (2016).

[70] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[71] Pew. 2018. pewinternet.org/fact-sheet/social-media.

[72] Andrew G. Reece, Andrew J. Reagan, Katharina L. M. Lix, Peter Sheridan Dodds, Christopher M. Danforth, and Ellen J. Langer. 2017. Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports* 7 (2017). Issue 1. https://doi.org/10.1038/s41598-017-12961-9

[73] Navid Rekabsaz, Mihai Lupu, and Allan Hanbury. 2017. Exploration of a threshold for similarity based on uncertainty in word embedding. In *European Conference on Information Retrieval*. Springer, 396–409.

[74] Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 99–107.

[75] Shannon E Ross, Bradley C Niebling, and Teresa M Heckert. 1999. Sources of stress among college students. *Social psychology* 61, 5 (1999), 841–846.

[76] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 95.

[77] Koustuv Saha, Eshwar Chandrasekharan, and Munmun De Choudhury. 2019. Prevalence and Psychological Effects of Hateful Speech in Online College Communities. In *Proceedings of the 10th ACM Conference on Web Science (WebSci ’19)*. Association for Computing Machinery, New York, NY, USA, 255–264. https://doi.org/10.1145/3292522.3326032

[78] Koustuv Saha and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 92 (Dec. 2017), 27 pages. https://doi.org/10.1145/3134727

[79] Koustuv Saha, Jordyn Seybolt, Stephen M Mattingly, Talayeh Aledavood, Chaitanya Konjeti, Gonzalo J Martinez, Ted Grover, Gloria Mark, and Munmun De Choudhury. 2021. What Life Events are Disclosed on Social Media, How, When, and By Whom. In *Proc. CHI*.

[80] Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kıcıman, and Munmun De Choudhury. 2019. A Social Media Study on the Effects of Psychiatric Medication Use. In *ICWSM*.

[81] Koustuv Saha, Ingmar Weber, and Munmun De Choudhury. 2018. A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. In *Twelfth International AAAI Conference on Web and Social Media*.

[82] H Andrew Schwartz, Maarten Sap, Margaret L Kern, Johannes C Eichstaedt, Adam Kapelner, Megha Agrawal, Eduardo Blanco, Lukasz Dziurzynski, Gregory Park, David Stillwell, et al. 2016. Predicting individual well-being through the language of social media. In *Pac Symp Biocomput*, Vol. 21. 516–527.

[83] Inge Seiffge-Krenke. 2013. *Stress, coping, and relationships in adolescence*. Psychology Press.

[84] Lisa Torrey and Jude Shavlik. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 242–264.

[85] Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response*.

[86] Erik van Ingen, Sonja Utz, and Vera Toepoel. 2016. Online Coping After Negative Life Events. *Soc. Sci. Comput. Rev.* 34, 5 (Oct. 2016), 511–529. https://doi.org/10.1177/0894439315600322

[87] Jessica Vitak. 2012. The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media* 56 (2012).

[88] Hanna M Wallach, Iain Murray, Ruslan Salakhutdinov, and David Mimno. 2009. Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning*. 1105–1112.

[89] Nancy M Wendlandt and Aaron B Rochlen. 2008. Addressing the college-to-work transition: Implications for university career counselors. *Journal of Career Development* 35, 2 (2008), 151–165.

A APPENDIX

Below we include the full list of LDA topics, as referenced in the (5) section.

Table A1. Full list of LDA derived topics from the 35-topic model LDA applied to post-college transition data. Each topic is annotated with a topical theme and representative words.

Topic 0: **STEM Coursework**
Rep. Words: math, scienc, major, physic, chemistri, biolog, mathemat, cours, statist, take, class, subject, like, econom, minor, studi, level, econ, stat, good, calculus, bio, interest, stem, think
Topic 1: **Evaluating graduate school offers**
Rep. Words: offer, accept, school, decis, get, program, choic, receiv, wait, hear, reject, choos, decid, appli, fund, want, know, excit, declin, dream, better, chanc, second, option, recent
Topic 2: **Class Logistics**
Rep. Words: class, grade, semest, exam, fail, pass, final, get, drop, studi, assign, test, professor, take, go, know, end, time, bad, cours, withdraw, second, point, think, need
Topic 3: **Finding a Good Career**
Rep. Words: want, work, job, like, life, love, year, know, think, thing, career, time, good, school, enjoy, idea, peopl, start, money, go, live, feel, passion, help, old
Topic 4: **Admissions Chances**
Rep. Words: school, year, gpa, graduat, high, get, appli, grad, good, univers, grade, chanc, go, senior, colleg, look, current, take, major, low, semest, averag, junior, program, want
Topic 5: **Workplace Communications**
Rep. Words: tell, say, week, email, ask, day, send, get, go, month, boss, time, today, check, want, ago, meet, call, know, give, happen, come, leav, receiv, start
Topic 6: **Writing for Applications**
Rep. Words: write, test, score, gre, essay, page, statement, take, section, sit, person, writer, word, sampl, help, draft, verbal, read, requir, servic, good, need, thank, edit, act
Topic 7: **Research and Publication**
Rep. Words: paper, write, research, present, review, publish, confer, journal, author, articl, data, public, topic, read, thesi, poster, propos, submit, sourc, project, collect, dentist, work, analysi, cite
Topic 8: **Finding a Place to Live**
Rep. Words: live, room, hous, car, apart, move, clean, buy, drive, home, roommat, place, get, wear, look, walk, away, mom, cloth, boyfriend, campus, need, dog, bring, stuff
Topic 9: **Health and Mental Well-being**
Rep. Words: health, medical, mental, nurs, doctor, help, hospital, issue, care, healthcare, psychology, counseling, medicine, physician, clinic, social, therapy, family, disable, anxiety, depression, patient, med, therapist, ill
Topic 10: **Degree**
Rep. Words: degre, major, busi, want, bachelor, year, graduat, job, master, school, think, scienc, like, financ, field, go, know, career, current, work, get, pursu, program, colleg, finish
Topic 11: **Locations**
Rep. Words: univers, state, city, live, california, rank, school, location, texas, attend, new, program, york, florida, install, area, chicago, look, better, boston, transfer, cyber_security, cost, washington, current
Topic 12: **Social and Relationships**
Rep. Words: friend, people, like, talk, social, know, go, school, feel, meet, want, home, new, lot, family, group, person, think, life, live, party, city, college, year, join
Topic 13: **Skills and Career Goals**
Rep. Words: career, field, work, look, path, interest, like, experi, skill, advic, option, help, thank, current, educ, pursu, job, consid, appreci, good, think, degre, know, idea, relat
Topic 14: **College Advice**

Rep. Words: colleg, communiti, transfer, year, school, go, want, dont, know, student, start, high, help, advic, ive, need, attend, local, freshman, plan, best, senior, electrician, like, thank

Topic 15: **Problem Solving**

Rep. Words: problem, thing, understand, way, peopl, use, differ, person, question, exampl, issu, point, mean, import, think, give, inform, reason, number, case, result, explain, word, certain, includ

Topic 16: **Recommendation Letters and Applications**

Rep. Words: letter, recommend, refer, professor, cover, ask, write, resum, appli, know, address, email, contact, good, send, prof, thank, profession, person, rec, need, use, lor, want, gift

Topic 17: **Cost of School, Loans, and Debt**

Rep. Words: pay, year, money, live, save, school, loan, need, cost, time, parent, debt, famili, work, month, financial, expense, go, student, income, home, help, insurance, afford, able

Topic 18: **Finding a Research Advisor**

Rep. Words: student, project, advisor, depart, work, supervisor, meet, group, professor, thesis, advise, member, lab, graduate, talk, academia, faculty, committee, discuss, dissertation, research, mentor, propose, defense, university

Topic 19: **Interview Advice**

Rep. Words: interview, ask, question, know, thank, answer, like, want, advice, think, help, recruit, phone, person, expect, go, tip, sure, good, say, guy, people, hear, talk, look

Topic 20: **Questions about Graduate School**

Rep. Words: program, research, grad, phd, master, school, work, year, lab, experi, graduat, undergrad, want, appli, field, scienc, like, start, know, interest, look, time, think, current, get

Topic 21: **Finances**

Rep. Words: account, bank, food, buy, card, invest, sell, cook, financ, access, meal, eat, code, expir, pro, price, cat, win, restaur, hack, hot, banker, exec, athlet, rip

Topic 22: **Teaching and Education**

Rep. Words: teach, english, teacher, learn, book, read, languag, educ, studi, write, spanish, tutor, speak, skill, literatur, use, self, histori, note, know, like, librari, need, level, help

Topic 23: **Software Engineering and Technology**

Rep. Words: engin, softwar, develop, program, design, data, learn, tech, web, technolog, mechan, cod, electr, scienc, project, industri, technic, comput, system, like, machin, build, code, cours, electron

Topic 24: **Marketing and Media**

Rep. Words: market, art, design, media, creativ, communic, digit, social, video, film, product, freelanc, write, portfolio, advertis, edit, industri, content, creat, photographi, game, journal, produc, entertain, writer

Topic 25: **Organizational Hierarchy**

Rep. Words: work, manag, compani, role, year, busi, team, sale, posit, current, new, experi, project, job, boss, custom, promot, servic, industri, skill, offic, like, market, employe, train

Topic 26: **Expression of Emotion**

Rep. Words: feel, like, know, think, go, get, thing, time, want, tri, year, start, peopl, leav, tell, way, bad, lot, right, come, make, good, hard, life, lose

Topic 27: **Academic Workload**

Rep. Words: class, cours, take, onlin, semest, credit, summer, requir, spring, time, fall, need, complet, enrol, hour, plan, schedul, finish, start, drop, current, unit, extra, quarter, elect

Topic 28: **Seeking Advice & Support**

Rep. Words: help, post, onlin, websit, com, use, look, need, thank, googl, reddit, free, search, link, app, know, list, comment, https_www, guy, share, tri, best, linkedin, https

Topic 29: **Job Considerations**

Rep. Words job, work, company, position, month, year, current, time, experience, get, look, pay, employment, application, new, leave, start, want, offer, resume, hire, interview, salary, internship, graduate

Topic 30: **Travel and Immigration**

Rep. Words: studi, countri, univers, abroad, canada, histori, travel, intern, student, europ, live, american, usa, languag, visa, year, cours, london, australia, cultur, foreign, germani, canadian, german, english
Topic 31: **Stress, Scheduling, and Time**
Rep. Words: work, hour, day, time, week, home, schedule, start, spend, night, weekend, stress, need, shift, break, go, long, sleep, minute, like, month, stay, office, sick, commute
Topic 32: **Applications for Funding**
Rep. Words: applic, appli, process, deadlin, submit, admiss, award, fellowship, reject, list, earli, grant, send, januari, receiv, decemb, app, wait, select, requir, march, late, candid, status, novemb
Topic 33: **Policy and Government Positions**
Rep. Words: law, polit, intern, public, mba, govern, internship, polici, econom, firm, lawyer, organ, consult, relat, student, non_profit, scienc, nonprofit, sector, master, administr, nation, india, financ, global
Topic 34: **Gap Year**
Rep. Words: uni, scholarship, con, pros, gap, pros_con, organis, placement, whilst, marin, merit, realis, atm, scheme, honour, modul, centr, bag, mat, ride, uber, vest, slide, shall, redditor

Received October 2020; revised April 2021; accepted May 2021