UNIVERSITY OF VIRGINIA

MS THESIS

SocialText: A Framework for Understanding Mental Health from Digital Communication Patterns

Author:
Sanjana MENDU

Supervisor:
Dr. Laura BARNES

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Department of Engineering Systems and Environment
School of Engineering and Applied Science

May 2020

UNIVERSITY OF VIRGINIA

Abstract

Faculty Name
School of Engineering and Applied Science

Master of Science

SocialText: A Framework for Understanding Mental Health from Digital Communication Patterns

by Sanjana MENDU

Over 35% of the world's population uses social media. Platforms like Facebook, Twitter, and Instagram have radically impacted the way individuals interact and communicate. These platforms facilitate both public and private communication with strangers and friends alike, providing rich insight into an individual's personality, health, and wellbeing. In this work, we present a generalized framework that outlines a clear, comprehensive method for creating informative, organized feature spaces, used to analyze the semantics of social media discourse. We then demonstrate the efficacy of our framework by applying it to a sample of private Facebook messages in a college student population ($N = 103$). Our results reveal key individual differences in temporal and relational behaviors, as well as language usage in relation to validated measures of trait-level anxiety, loneliness, and personality. By leveraging the comprehensive structure outlined by our framework, we not only built more complete models of private social media discourse but also demonstrated the associated affordances with respect to classifying mental health. This work represents a critical step forward in linking features of private social media messages to validated measures of mental health and wellbeing.

Acknowledgements

I would like to express my deep gratitude to Dr. Laura Barnes, my research supervisor, for her patient guidance, enthusiastic encouragement, and useful critiques of this research work. This would have been impossible without her support and mentorship.

I would also like to thank Dr. Mehdi Boukhechba for his assistance in honing my research question and refining my analysis, and Dr. Gerald Clore for lending his domain expertise and facilitating this interdisciplinary project.

My grateful thanks are also extended to Anna Baglione and Sonia Baee for their help in writing this manuscript and cheering me on along the way, to Peter Wu, who helped me integrate his work on interpersonal networks into the methodology, and to Brandon Ng and Adi Shaked for their time and efforts in collecting the dataset used in this study.

Finally, I would like to thank my committee members for their continued support and encouragement: Dr. William Scherer, my committee chair; and Dr. Afsaneh Doryab. I offer my sincere appreciation for the learning opportunities they have provided me.

Contents

Abstract **i**

Acknowledgements **ii**

1 Introduction **1**

2 Related Work **4**

3 SocialText: A Unifying Framework **7**

4 Methods **11**

4.1 Data Collection 11

4.1.1 Participants 11

4.1.2 Psychological Measures 12

4.1.3 Facebook Messages 13

4.2 Feature Extraction 15

4.2.1 Lexical 15

4.2.2 Semantic 16

4.2.3 Temporal 17

4.2.4 Topological 18

4.3 Predictive Modeling 19

4.3.1 Feature Selection 19

4.3.2 Classification 20

5 Results **22**

5.1 Model Performance 22

5.2 Independent Predictive Models 24

5.2.1 Lexical 24

5.2.2 Semantic 26

5.2.3 Temporal 27

5.2.4 Topological 29

6 Discussion **31**

6.1 Understanding Mental Health from DTCs 31

6.2 Limitations 36

6.3 Implications and Future Work 37

7 Conclusion **39**

List of Figures

3.1 Visual representation of SocialText framework 7

4.1 Ditribution of Trait Psychological Measures 12

4.2 Visual representation of our modeling process 20

5.1 Lexical differences between high and low anxiety participants . . 24

5.2 Relationship between trait psychology and messaging latency . . 28

5.3 Relationship between anxiety and hourly messaging frequency . 28

List of Tables

4.1 List of features 15

5.1 Predictive model performance (F1 Score) 22

5.2 List of LDA topics 26

6.1 Categorization of existing approaches 37

Dedicated to my beloved parents, sister, and cat.

1 Introduction

In an age of political division and economic instability, college students are increasingly plagued with serious mental health issues (MHIs) such as loneliness, social anxiety, and depression [1]. As the era of MHIs has progressed, so too has social media. *Digital text communications (DTCs)* exchanged over social networks such as Facebook, Twitter, and Instagram and platforms such as Facebook Messenger, Twitter, and WhatsApp now form the the collective touchstone of modern communication for young adults. Recent work suggests that DTC platforms provide unique insight into the mental health and well-being of young adults [2, 3, 4], as they are often used for self-disclosure of mental illness. Moreover, existing literature suggests that DTCs strongly influence mental health outcomes, with Facebook standing out as a particularly powerful platform. Perceptions of social support (as opposed to actual social support) on Facebook have been found to be associated with depression [5], and Facebook use in general was associated with declines in well-being over time among college students [6]. The manner in which Facebook is used (e.g. actively versus passively) has a strong influence on affective well-being [7]. In general, the number of social media platforms used was found to be associated with anxiety [8].

Though the role of DTC platforms as largely positive or negative influencers of mental health outcomes remains disputed, these tools have been shown to facilitate easier access to informational and social support [9], especially for groups who may struggle to obtain support in the wild. Studies have shown that DTCs have been associated with improved mental health outcomes in vulnerable populations, including breast cancer patients [10], individuals with severe mental illness (e.g. schizophrenia and bipolar disorder) [11], homeless youth [12], and young adults with diabetes [13]. Among college students, greater perceived support on platforms like Facebook has been tied to less stress [14] and greater physical and emotional wellbeing [15, 16].

However, existing work has relied primarily on datasets of public and semi-public content from Facebook and Twitter. Limited analyses have been conducted on private messages (e.g. Facebook messenger datasets), which uniquely facilitate directed communication between one person and a limited number of social contacts. For example, Bazarova et al. found that Facebook users are more likely to share intense and private emotions in direct messages as opposed to network-wide status updates [17]. Furthermore, O'Leary et al. demonstrated that private direct communication in a formal and mental health focused setting online improved clinically significant levels of anxiety and depression [18].

Methodological approaches to classifying and predicting mental health issues (MHIs) from DTC datasets vary widely by dataset, platform, and MHI. Some rely on tools such as Linguistic Inquiry and Word Count (LIWC) to identify common language features of different MHIs. Others examine how variations in temporal communication patterns and social network topology influence MHI symptom prevalence [19, 20].

In theory, employing a combination of these diverse methodologies could reveal deeper insights about how, why, and when MHI symptoms present in digital communications and could help researchers transcend traditional disciplinary boundaries. In practice, however, approaches to analyzing DTC platform data remain divided largely along disciplinary lines. This divide poses a major challenge for both computing and psychology researchers alike, for differing reasons. Computing researchers often lack access to collaborators in the psychological sciences and tend to focus on the implementation of complex and innovative computational models, with little focus on human health and behavior theories that may drive DTC communication patterns. Conversely, researchers in the psychological sciences possess a strong grasp of human health and behavior theory, but may rely on more traditional modeling techniques. Research on the role of DTCs in mental health and wellbeing must be both technically innovative and grounded in psychological theory. There is thus a great need for unification of existing diverse approaches from both psychology and computing. To this end, we created a generalized feature extraction framework for DTC datasets and applied this framework to guide exploratory analyses of private DTCs from a college student population.

Our work provides three main contributions: 1) Establish a unifying hierarchy for DTC analysis methods, 2) Leverage qualitative and quantitative features of DTCs in both low- and high-level analyses, and 3) Identify individual differences in anxiety, loneliness, and personality within a college student population, as determined by these features. First, we provide a brief overview of the related literature. Then, we present our framework and explain how our feature extraction recommendations align with the related literature. Finally, we discuss an application of our framework to a private DTC dataset and highlight important findings afforded by our comprehensive approach.

2 Related Work

As social media platforms have grown to form the foundation of modern digital communication, DTC datasets have proliferated. These exchanges comprise a rich corpus of interpersonal exchanges, which can provide insight into how MHIs manifest in different social contexts. Of particular interest to the research community is how DTC lexica reflect individual communication styles and provide insight into personal traits, relationship quality, and mental state. Researchers have identified shared vocabularies and interpersonal differences in message semantics among individuals with MHIs [21, 22, 23, 24, 25, 26]. Text mining techniques can help us to represent DTC data and identify lexical patterns that relate to individuals' personality traits [27]. For example, Coppersmith et. al showed that a character language model can discriminate between MHIs, meaning that "spaces, punctuation, an emotico[n]" usage differs by condition [24].

Linguistic Inquiry and Word Count (LIWC) has proven popular among psychologists and HCI researchers alike for its ability to uncover links between personality, language, and MHIs. LIWC analysis has been used to predict personality traits [28, 29, 27], emotion [30], and MHIs like depression [31, 32], suicidality [33, 25], and disordered eating [34]. Sentiment analysis is another popular method for characterizing textual expression on DTC platforms [35, 36, 24, 30]. Alternatives to closed-vocabulary method include unsupervised, open-language approaches like topic modeling (i.e. latent dirichlet allocation (LDA))

[28, 29, 32], and word embeddings [35, 37]). These techniques are used to extract textual patterns that describe the relationship between different linguistic structures and their effect on the overall meaning of a given text. The understanding gained from this work enables researchers to uncover data-driven language structures rather than relying on pre-defined vocabularies. By examining both the syntax of messages and the context within which an individual is communicating, researchers uncover data-driven language structures rather than rely on pre-defined vocabularies.

Apart from content analyses, researchers have examined temporal patterns across DTCs, including communication around situational events [38], communication frequency overall [27, 39, 30], and communication frequency overall during different epochs [27]. Burke & Kraut, for example, used temporal and topological properties to understand social processes on Facebook following a job loss [20]. Researchers have also leveraged social network analysis methods to construct graphical structures of DTC data, abstracting individuals as nodes and their communications as edges [40, 41]. In the context of DTCs, relational patterns can be similarly inferred by constructing graphical networks from a dataset of directed messages. From these networks, researchers have found important links between structural patterns (e.g. network size [27, 34, 29, 39], betweenness [27], density [27], transitivity [27, 36], tie strength [36, 39], group associations [42, 39], persistence of social signature [43], turnover [43], rank dynamics [43, 30], interaction diversity [34]) and a diverse range of MHIs.

Methodological approaches to classifying and predicting MHIs from DTC datasets vary widely by dataset, platform, and MHI. While some researchers have used similar methodological approaches for relating DTC patterns to mental health, there exists a clear separation between the consideration of metadata features and content features in mobile sensing for mental health contexts. For example, researchers have explored DTC patterns in different temporal contexts, including daily [44, 25] weekly [19], and multi-month contexts [20, 26]. These variations in temporal resolution result in diverse feature extraction and modeling approaches. Furthermore, there exists a clear separation between qualitative and quantitative feature extraction approaches for modeling DTC patterns in the context of mental health. For example, while Gopalakrishna Pillai et al. [45] created a rich feature space that covered a broad range of communication behaviors, their findings primarily focused on textual patterns. Burke & Kraut [20], on the other hand, studied temporal patterns and interpersonal networks, but failed to integrate content-based insights.

Few studies have leveraged a combined feature space that affords insights from both qualitative and quantitative research practice. To effectively identify and analyze key underlying social contexts and related mental health factors from DTC data, researchers must extract a comprehensive corpus of features from raw textual data streams. In this paper, we present a generalized framework for extracting features from digital text communication datasets that leverages methodological approaches from diverse fields and provides an avenue for logically deconstructing DTC datasets.

3 SocialText: A Unifying Framework

To effectively identify and analyze the relationship between underlying social contexts and MHIs evidenced in digital text communication data, researchers must extract a comprehensive corpus of features from raw textual data streams. To this end, we propose *SocialText*, a feature extraction framework that unites diverse methodological approaches to analyzing the relationship between DTCs and MHIs. The goal of the *SocialText* framework is to provide a clear, comprehensive method for creating informative, organized feature spaces for analysis of DTC social semantics. Figure 3.1 provides a visual overview of *SocialText*. In the following section, we discuss the relevance of each of the framework's layers to social context and mental health states.

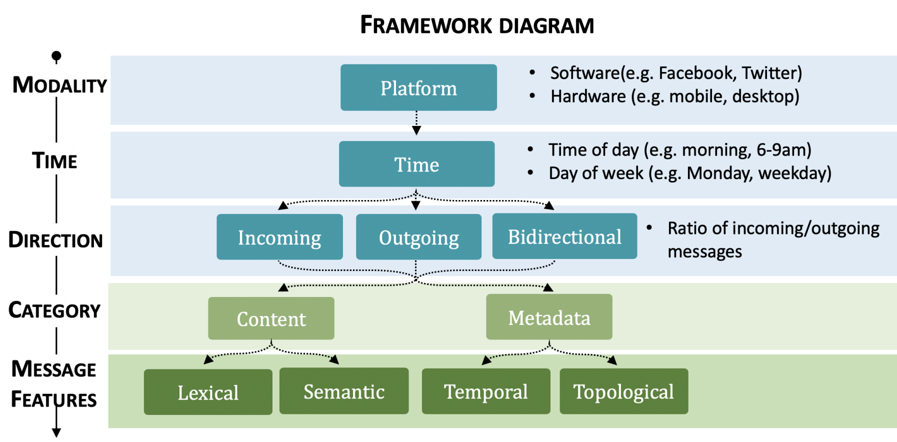


FIGURE 3.1: Visual representation of SocialText framework

Each layer of the *SocialText* framework intentionally highlights features that can be derived from DTC data and used to identify social context, thus improving prediction of MHIs from DTCs. While the upper layers define important variables for data partitioning, the lowest layer identifies categories of features that can be extracted from the messages themselves. Features pertaining to the semantics and lexicon of message content can characterize conversational context, while temporal and topological features can reveal social network ties and temporal messaging patterns. Considering all message features in combination provides a comprehensive characterization of the effect of the social dynamics of DTCs on participants' mental states, thus improving the performance of the resulting predictive models.

Modality pertains to both the software and hardware used send and receive DTCs. A unique modality can be defined in terms of the software platform (i.e. Facebook, SMS) and/or device used (i.e. laptop, phone). Grouping both platform and device together in the modality layer keeps the *SocialText* framework platform-agnostic and reduces the chance of bias. For example, messaging behavior (e.g. time, vocabulary, emojis) may vary across different platforms (e.g. cross-platform vs. mobile only), and the SocialText framework accounts for this.

Time refers to the time window of interest (i.e. hour, day, week) for analysis. The time at which individuals send and receive DTCs can reveal much about underlying social context, including interpersonal relationships and communication styles. Appropriate time windows vary depending on the desired outcome variable (e.g. momentary *state* vs. persistent *trait* anxiety). For example, the number of messages an individual sends in a week may remain relatively constant (following an individual's baseline trait anxiety) while daily messaging patterns vary (according to state anxiety).

Direction comprises three different message classifications: *incoming* messages, *outgoing* messages, and *bidirectional* messages. Incoming and outgoing message features reveal egocentric aspects of the underlying social context of a conversation. Outgoing message features, in particular, reveal relationships between an individual's communication practices and their mental state. For example, sending more messages in the morning vs. at night may be tied to MHIs such as loneliness and depression. Bidirectional message features, which describe all messages irrespective of whether they are incoming or outgoing, reveal factors like discussion quality and conversation dynamics (e.g. who is talking more).

Category distinguishes between two distinct categories of features: *content* features reveal shared vocabularies and interpersonal differences in message semantics between members of a social network; *metadata* features unveil relationships between the timing and frequency of message exchanges and the overarching network structure.

Message Features address the different content-based and metadata-based features of message subsets. This layer does not further partition the data but rather enumerates the aggregated features that can be calculated based on individual messages. We have defined four message feature domains, in total: *Lexical* features refer to vocabulary and term-related qualities of message content; *Semantic* features capture the relationships between words within a set of messages and the significance of these relationships to the overall tone and meaning. Semantic features of textual content describe the relationship between different linguistic structures and their effect on the overall meaning of a given text. Semantic features can be inferred by examining both the syntax of messages and the context within which an individual is communicating; *Temporal* features refer to time-sensitive message characteristics. The time at which individuals send and receive DTCs can reveal much about underlying social context, including interpersonal relationships and communication styles; *Topological* features refer to social network structures, commonly derived from social network analysis methods.

4 Methods

In this section, we present an application of the *SocialText* framework to a dataset of private Facebook messages collected from a sample of college undergraduates at a U.S. university. We examine the relationship between social media usage and mental health at the individual/trait level. By understanding the social strategies that people use in their everyday life, and whether different strategies may be most effective for people with different psychological traits and MHIs, we hope to achieve a better understanding of mental health for all.

4.1 Data Collection

4.1.1 Participants

Participants ($N = 103$) were recruited from undergraduate psychology classes at our university and received course credit as compensation. By recruiting young adults in a university setting, we obtained a relatively homogenous sample with respect to psychosocial stressors and life experiences, thereby eliminating many potential "nuisance factors". Our population was evenly sampled with respect to gender, with 51 female participants and 52 male participants. Participants' ages ranged from 18-22 years old, with the average age being 19 years old.

4.1.2 Psychological Measures

To assess participants' mental state, we administered clinically validated measures of anxiety, loneliness, and personality during an initial in-laboratory session. Each of the measures described below has been previously studied in a trait-level context [46, 47, 48].

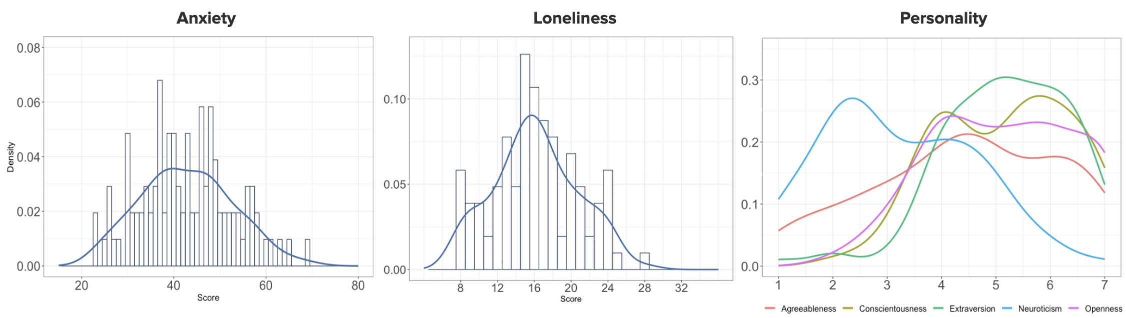


FIGURE 4.1: Distribution of Anxiety (*M* = 42.77; *SD* = 9.95), Loneliness (*M* = 16.16; *SD* = 4.57), and Personality Trait [*Openness*: (*M* = 5.13; *SD* = 1.31), *Extraversion*: (*M* = 5.18; *SD* = 1.16), *Agreeableness*: (*M* = 4.43; *SD* = 1.69), *Neuroticism*: (*M* = 3.21; *SD* = 1.39), and *Conscientiousness*: (*M* = 5.14; *SD* = 1.22)] levels among the participants

Anxiety: The State Trait Anxiety Inventory (STAI; [46]) assesses two distinct dimensions of anxiety: (1) *state anxiety* (a temporary condition resulting from an individual's current state) and (2) *trait anxiety* (a long-standing quality unique to the individual). In this analysis, we consider trait anxiety to be our proxy for anxiety on the individual level. Participants rated the degree to which they generally identified with each statement (e.g., "I feel satisfied with myself") from 1 ("almost never") to 4 ("almost always").

Loneliness: The UCLA Loneliness Scale (ULS-20; [49]) is a widely used loneliness measure. We used an alternative short-form measure in this study (ULS-8; [50]). Participants rated the degree to which they generally identified with each statement (e.g. "I feel isolation from others") from 1 ("I never feel this way") to

4 ("I often feel this way").

Personality: The Ten-Item Personality Inventory (TIPI; [47]) provides measures of the "Big Five" (i.e. Five-Factor Model) dimensions of personality: *Openness*, *Extraversion*, *Agreeableness*, *Neuroticism*, and *Conscientiousness*. Participants rated the degree to which they agreed with each statement, specifically the extent to which each pair of traits applied to them (e.g., "extraverted, enthusiastic"), from 1 ("disagree strongly") to 7 ("agree strongly").

4.1.3 Facebook Messages

We requested that participants provide us with their Facebook messages since the time of account creation and, optionally, their public Facebook logs. Participants who opted not to provide us with their logs still received full credit for participating in the study. Those who opted to provide their logs downloaded them from the Facebook website during an in-laboratory session. Due to the lack of download configuration options available at the time the logs were downloaded, logs dated back to the creation of the account.

To account for individual differences in account creation date, we calculated the number of days of available data for the participant using the most recent account creation date ($T \approx 5$ months) and used that as a uniform time interval to compare all participants fairly. Overall, the dates used in this analysis span from June 8 to November 7, 2016. All data falling outside this specified time range was omitted from the current analysis. Our final dataset is comprised of 1,051,858 messages across all participants, with an average of 10,212 messages (σ = 27,869) and 48 unique chats (σ = 37) per participant.

Ethical Considerations

Aggregation of private data into large, readily-available datasets has come under intense scrunity in the wake of events such as the Cambridge Analytica scandal. Though the debate over the extent to which private information may be ethically collected continues, ethical researchers agree that participant privacy must take utmost precendence in all studies involving sensitive data. We took careful steps to protect participants' privacy at each step of the research process. Participants signed a consent form at the beginning of the study and a material release form at the end of the study. A member of the research team was present for all lab sessions to explain the consent process and to answer the participants' questions.

This study required the use of private data for several reasons. Firstly, the public and private selves are often quite different, especially with regard to DTCs. Free disclosure of mental health concerns in public online spaces (e.g. example in public Tweets and Facebook posts) may be met with lack of response from one's network due to the hypothesized "positivity bias" against negative status updates [51]. Researchers for this study hypothesized that private DTCs are more likely to contain dynamic depictions of how MHIs manifest in daily life (e.g. through sustained communication with other individuals in one's network). Moreover, when DTC platform users feel more able to discuss health concerns freely (i.e. in private communications), the quantity of messages and thus the size of the dataset is hypothesized to increase. Having more data allows researchers to make more accurate observations about DTC communication pattern phenomena, such as density of messages by time of day and how this differs according to personality trait and MHI intensity.

4.2 Feature Extraction

In accordance with the SocialText framework structure, the features we extracted cover a broad range of DTC properties which we divide into four distinct categories: *Lexical*, *Semantic*, *Temporal*, and *Topological*. Table 4.1 provides a comprehensive list of extracted features.

Message Feature	Name	#	Direction
Lexical	LIWC	184	↑↓
Semantic	TF-IDF	6,348	↓
	LDA Topic Usage	100	↓↑
Temporal	Latency	2	↑↓
	Hourly Proportion	72	⇕↑↓
Topological	Number of Individual Alters	3	⇕↑↓
	Number of Group Alters	3	⇕↑↓
	Maximum Edge Weight	3	⇕↑↓
	Entropy of Edge Weights	3	⇕↑↓
	Mean/SD persistence	6	⇕↑↓
	Mean/SD turnover	6	⇕↑↓

TABLE 4.1: List of features. Direction (⇕: bidirectional, ↑: outgoing, ↓: incoming)

4.2.1 Lexical

DTC lexica reflect individual communication styles and provide insight into personal traits, relationship quality, and mental state, among other factors. We extracted lexical features using the popular *Linguistic Inquiry and Word Count (LIWC)* method, which has been rigorously validated in the context of psychometric analysis of textual data [52].

4.2.2 Semantic

Semantic features of DTCs describe the relationship between linguistic structure and the meaning of a given text. We created a set of linguistic structures in the corpus using the Natural Language Toolkit (nltk) TweetTokenizer [53] to split each message into unigrams. We also extracted bigrams and trigrams (e.g. phrases) - two and three-word sequences that occur at rates much higher than chance (e.g. "happy birthday", "I love you") - by calculating the pointwise mutual information (PMI) [54, 55] of each phrase (i.e. a ratio of the joint-probability to the independent probability of observing the phrase within the aggregated corpus of messages):

$$\text{PMI}(\text{phrase}) = log\frac{p(\text{phrase})}{\prod_{w \in \text{phrase}} p(w)} \tag{4.1}$$

We retained all bigrams and trigrams with PMI values greater than 3 times the number of words in the phrase. The resulting vocabulary consisted of 6,348 words and phrases. To reduce the number of features, we kept words and phrases that were used at least once by at least 10% (n=10) of the population. We calculated the *Term Frequency - Inverse Document Frequency (TF-IDF)* of each term in the vocabulary described above in order to measure each term's usage within each participants' set of messages. TF-IDF serves as a useful measure for between-subjects analyses such as ours because it accounts for the relevance of terms across multiple documents.

We also identified *topics* - clusters of frequently co-occurring words in our corpus - using Latent Dirichlet Allocation (LDA) [56]. The generative LDA model assumes that documents (i.e. a participant's complete set of private Facebook messages) contain a combination of topics, and that topics are a distribution of words (i.e. observations) for which the latent variables can be estimated through Gibbs sampling [57]. For this analysis, we leveraged the implementation of this algorithm provided in the Mallet package [58] to produce 100 naturally-occurring topics, each consisting of many words with relative weights. We then calculated each individual's use of each topic, defined as the probability of using a topic:

$$p(\text{topic}|\text{user}) = \sum_{\text{word} \in \text{topic}} p(\text{topic}|\text{word}) * p(\text{word}|\text{user}) \tag{4.2}$$

where $p(word|user)$ is the individual's normalized word use.

4.2.3 Temporal

The time at which individuals send and receive DTCs can reveal much about underlying social context, including interpersonal relationships and communication styles. We calculated the *hourly distribution* of messaging activity (i.e. the proportion of messages sent during each hour of the day) from the aggregated collection of each participants' Facebook message logs. We also calculated *latency* for both outgoing and incoming messages, where outgoing latency is the average amount of time (in minutes) that a participant takes to respond to a message they receive, and incoming latency is the average amount of time (in minutes) it takes for a participant to receive a response to a message they sent.

4.2.4 Topological

The topology of the ego-centric network formed by an individual's social circle can provide significant insight into the individual's personality traits [59]. Suppose, over a time period (5 months in this study), a subject (ego) exchanged (i.e., sent and/or received) at least one message with K unique alters. The K alters can be partitioned into K_1 individual alters representing individual recipients and K_2 group alters representing two or more recipients giving $K = K_1 + K_2$. We extracted the following features to capture the size of individuals' social networks: *number of individual alters* K_1 (i.e., the number of contacts representing an individual with whom a subject exchanged at least one message); and *number of group alters* K_2, (i.e., the number of contacts representing at least two people with whom a subject exchanged at least one message).

Straightforwardly, the messages exchanged between the subject and an alter constitute the edges in the network, and we define *edge weight* as the proportion of messages exchanged with an alter (individual or group) among all alters. We denote as p_r $r \in \{1, \ldots, K\}$ the r-th highest proportion of messages exchanged with an alter among all alters, and the distribution of proportions/edge weight over all alters as $P = \{p_1, \ldots p_K\}$. We extracted the following features to capture differences in exchanges in the context of an individual's social network: *entropy of edge weight* $H(P) = -\sum_{p \in P} p \log(p)$ (i.e., the Shannon entropy of the proportions of messages exchanged with all alters a subject had). This measure quantifies how a subject distributes their time across multiple threads of conversations; and *maximum edge weight* p_1 (i.e., the proportion of messages exchanged with the alter with whom the subject exchanged the most messages).

We also sought to characterize the variation of social dynamics over more granular time intervals. We calculate two measures: the persistence of social signatures and the turnover in ego-centric networks. These measures come from existing works on ego-centric network dynamics [60, 43], which proposed and applied these measures on phone call and Bluetooth encounter networks. To calculate we first divide the 5-month observation period into 21 week-long periods $\{w_1, \ldots, w_{21}\}$. For each pair of consecutive periods (w_i, w_{i+1}) $\forall i \in \{1, \ldots, 20\}$ we calculate the following features: (1) *persistence of social signature*, defined as the Jensen-Shannon divergence between the P's calculated from w_i and w_{i+1},

$$\text{persistence}(w_i, w_{i+1}) = H\left(\frac{P_{w_i} + P_{w_{i+1}}}{2}\right) - \frac{H(P_{w_i}) + H(P_{w_{i+1}})}{2}, i \in \{1, ..., 20\} \tag{4.3}$$

; and (2) *turnover of ego-centric network*, defined as the Jaccard difference between the two sets of alters, $A(w_i)$ and $A(w_{i+1})$, corresponding to w_i and w_{i+1} for a subject, concretely:

$$\text{turnover}(w_i, w_{i+1}) = \frac{|A(w_i) \cap A(w_{i+1})|}{|A(w_i) \cup A(w_{i+1})|}, i \in \{1, ..., 20\} \tag{4.4}$$

We obtain 20 values for each measure and calculate the mean and standard deviation, producing 4 features in total: mean persistence, standard deviation of persistence, mean turnover, and standard deviation of turnover.

4.3 Predictive Modeling

4.3.1 Feature Selection

We began our evaluation by testing predictive models for each feature feature category independently. To reduce the effect of irrelevant features and mitigate the curse of dimensionality, we used a random forest classifier to select a subset of the 10 most relevant features to the given outcome for each feature domain independently, based on the mean decrease in Gini impurity when a feature is used to partition the data.

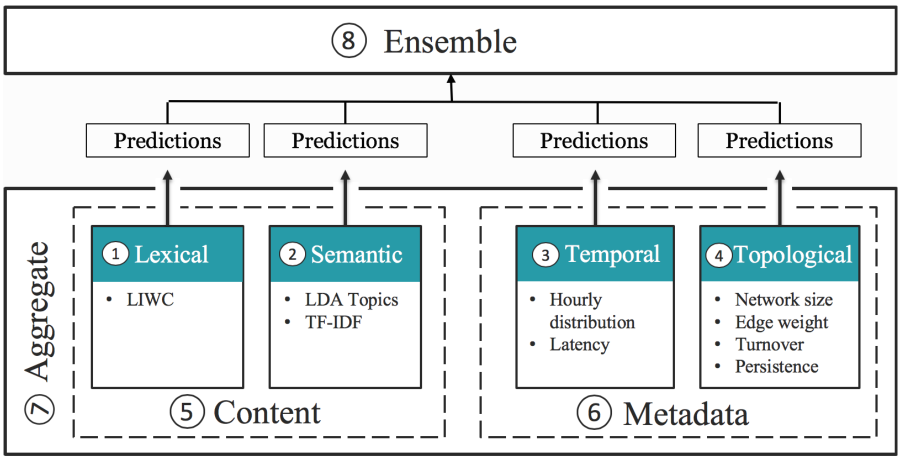


FIGURE 4.2: Visual representation of our modeling process

4.3.2 Classification

We used a Support Vector Machine (SVM) and leave-one-subject-out cross validation (LOSOCV) to perform a binary classification rask for each psychological measure using unique groupings of the messages features as input. We investigated the effect of combining the content feature spaces (i.e. Semantic & Lexical) and metadata feature spaces (i.e. Temporal & Topological) on model performance. Finally, we used two approaches to combining features across all four message feature domains: ensemble and aggregated. For the ensemble model, we used stacked generalization [61] to predict psychological characteristics. This approach is advantageous because it overcomes the potential for features from larger domain spaces (i.e. Semantic & Lexical) to overpower smaller domain feature spaces (i.e. Temporal & Topological), since the representation of knowledge from each domain is condensed in the form of the each independent model's prediction. For the aggregated model, we combined features across message feature domains into a single feature space. We then applied the same Random Forest approach used for the independent domain models to reduce the dimensionality of the cummulative feature space.

5 Results

5.1 Model Performance

		Anx.	Lon.	Extra.	Agree.	Open.	Neuro.	Consc.
Message Features	Lexical	**0.765**	0.678	0.627	0.547	**0.762**	**0.667**	0.615
	Semantic	0.701	**0.732**	**0.717**	**0.634**	0.734	0.594	**0.716**
	Temporal	0.698	0.705	NA*	0.580	0.548	0.641	0.500
	Topological	0.239	0.684	0.413	0.522	0.486	0.636	0.376
Category	Content	**0.694**	**0.743**	**0.660**	0.574	**0.796**	**0.646**	**0.687**
	Metadata	**0.694**	0.698	0.455	**0.611**	0.619	0.614	0.574
Combined	Aggregate	0.707	0.692	0.667	0.587	**0.789**	0.660	0.720
	Ensemble	**0.752**	**0.793**	**0.698**	**0.743**	0.774	**0.708**	**0.768**

TABLE 5.1: The above table shows each model's performance as measured by F1 score. * denotes an undefined F1 score resulting from a zero-valued recall and precision measure for the given model.

The semantic model outperformed all other independent models for predicting 4 out of the 7 psychological measures, with performance ranging from 0.634 to 0.732 overall. The lexical model performed better than the semantic model for predicting anxiety (F1 = 0.765), openness (F1 = 0.818), and neuroticism (F1 = 0.667) with a relative improvement of 0.064, 0.028, and 0.073 for each measure respectively. While the temporal model achieved moderate performance overall (Mean F1 = 0.524), it performs particularly poorly for predicting extraversion. Notably, extraversion is also the only measure for which the topological model (F1 = 0.413) outperforms the temporal model (F1 = NA). This is particularly interesting given the topological model poor performance overall (Mean F1 =

0.479). This discrepancy suggests that interpersonal dynamics outweighs temporal factors with respect to characterizing extraversion as manifested in private social media discourse.

The content-based model, which used only lexical and semantic features as predictors, outperformed the metadata-based model for predicting the majority of the psychological measures. This result is supported by findings from prior studies that used only content features to predict personality traits [29, 28]. Notably, the metadata model performed better than the content model for predicting agreeableness. Additionally, the content model performed relatively poorly for predicting agreeableness (0.574) as compared to the other psychological measures (0.646 to 0.796), suggesting that agreeableness relates less to what people say and more to when and with whom people engage in private social media discourse.

Additionally, our results highlight that using all feature domains to predict trait measurements outperforms the independent models. While the relative performance improvements vary from measure to measure, the average performance of the aggregate and ensemble models (0.689 and 0.749 respectively) exceeds the average performance of any given independent model (0.479 to 0.690). Furthermore, the ensemble model outperformed the aggregate model for predicting the majority of the psychological measures (6 out of 7). This suggests that considering each of the different dimensions of DTC data (i.e. message features) in equal measure, rather than heavily weighting any given one, not only supports a more comprehensive consideration of underlying factors but also improves the performance of predictive modeling approaches.

5.2 Independent Predictive Models

5.2.1 Lexical

Participants in the "high" and "low" anxiety groups exhibited differences in their use of 1st person plural pronouns, 3rd person plural pronouns, relativity (spatial, temporal), and male references (see Figure 5.1). Participants' levels of anxiety were also discriminated by the authenticity of language used by those in their network, as well as incoming content containing relativity, netspeak, certainty, and informal language.

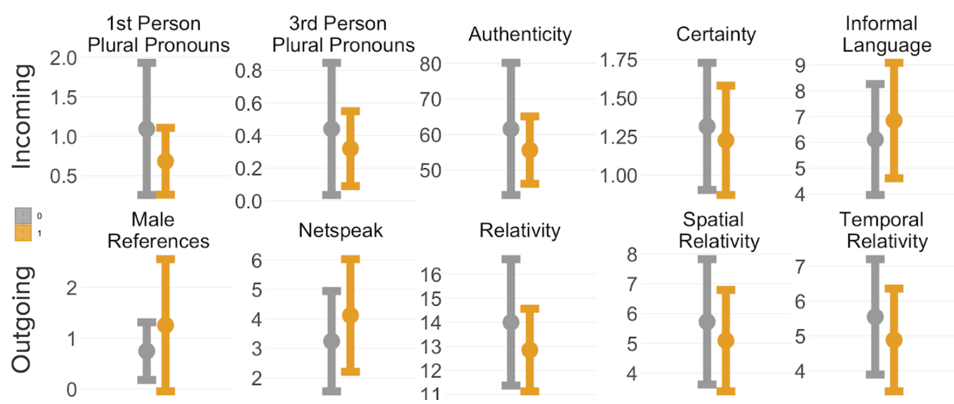


FIGURE 5.1: Difference in incoming and outgoing message content between "high" (i.e. 1) and "low" (i.e. 0) anxiety classes in terms of LIWC measures

Participants in the "high" and "low" depression groups used words related to sadness, friends, and negative emotions to different extents. Individual differences in the depression measure were also discriminated by participants' use of netspeak and 3rd person singular pronouns. Participants' levels of depression were also discriminated use negations and language pertaining to perceptual processes in messages they received.

Participants' levels of loneliness were most significantly discriminated by their outgoing language (i.e. the words they used in messages they sent to others). Specifically, participants in the "high" and "low" loneliness groups exhibited differences in terms of their use of 3rd person plural pronouns, periods, and adverbs. They also differed in their discussion of friends and other affiliations. Discussion related to certainty and interrogative topics, as well as level of authenticity in outgoing language further distinguished the two groups.

Extroverts and introverts (i.e. participants in the "high" and "low" extraversion groups) varied in their use of personal pronouns and discussion of social processes, as well as their social contacts' use of function words, exclamation marks, netspeak, personal pronouns, pronouns, achievement, 1st person singular pronouns. Participants in the "high" and "low" conscientiousness groups were more readily discriminated by the content of messages individuals received vs. sent. Specifically, participants in the "high" and "low" conscientiousness groups were discriminated by use of informal language, punctuation, netspeak, and words longer than 6 letters contained in messages they received within the independent lexical model. Incoming message language pertaining to assent, affiliations, past focus, risk, and drives was also used to discriminate between participants in the "high" and "low" conscientiousness groups. This suggests that participants' levels of conscientiousness mediate the formality of language used by their social contacts on Facebook Messenger. Surprisingly, the extent to which participants discussed biological processes was an important discriminating factor between "high" and "low" classes for both extraversion and agreeableness.

Topic Label	ID	Terms
Election	40	trump, vote, election, president, voted, people, america, voting, rip, pretty, debate, votes, wins, country, politics, everyone, years, republican, winning, scared
Recreation	41	omg, haha, night, weekend, going, week, last, next, back, fun, come, coming, hahaha, dude, man, yay, party, visit, meet, nice
Emotional Processes	44	feel, talk, things, like, really, sorry, know, okay, think, someone, person, time, talking, life, something, sad, tell, anything, felt, miss
Pokemon GO	46	pokemon, caught, level, team, anyone, catch, gym, tonight, game, walk, find, house, found, mystic, around, valor, blue, playing, egg, people
Sports	61	game, play, team, playing, last, yeah, played, ball, soccer, games, beat, though, hit, football, hard, basketball, lost, time, damn, pretty
Spiritual Music	78	song, listen, music, man, time, pretty, could, never, high, every, family, without, made, makes, ever, times, though, great, different, god
Social Support	88	thanks, thank, great, miss, school, day, work, best, luck, fun, well, send, class, excited, aww, awesome, start, uva, summer, already
Alcohol	90	drink, party, drunk, drinking, night, alcohol, drinks, parties, people, beer, tonight, fun, sober, getting, drank, frat, shots, boy, wine, gone

TABLE 5.2: Topics present in private Facebook messages

5.2.2 Semantic

As mentioned in the *Model Performance* section, the independent semantic models overwhelmingly outperformed the other independent models, producing the highest performance for predicting five out of the seven psychological measures. This suggests that the semantic features we extracted from the private Facebook message corpus were most representative of individual differences in personality traits and MHIs within this college student population. Differences in topic usage across the "high" and "low" groups for the psychological measures yielded a number of interesting findings. Table 5.2 provides examples of meaningful topics we extracted from our corpus.

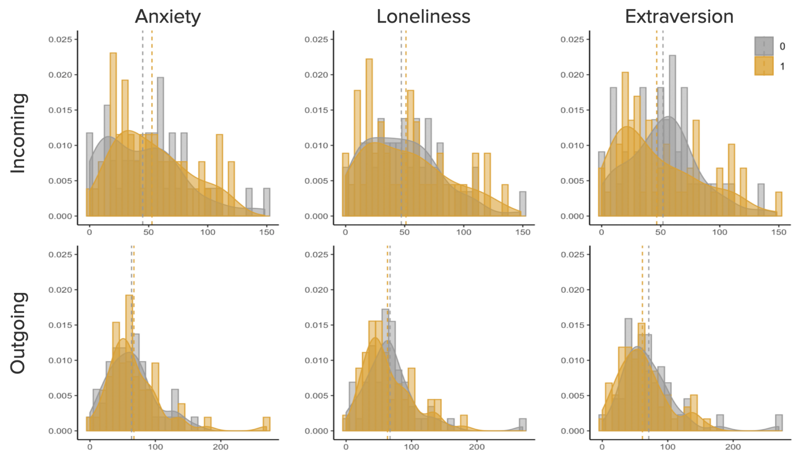
Discussion of Pokemon GO (i.e. topic ID = 46), a mobile game released in the United States on July 6, 2016 [62], emerged as a meaningful discriminator for anxiety, loneliness, and neuroticism. Discussion of sports (i.e. topic ID = 61) emerged as a meaningful discriminator for anxiety and neuroticism, supporting the idea that engaging in physical activity plays a key role in college students' emotional stability. Discussion of spiritual music (i.e. topic ID = 78) also emerged as a meaningful discriminator for anxiety. Use of words related to social support (i.e. topic ID = 88) provided meaningful context for differentiating participants with "high" and "low" levels of agreeableness. Discussion of alcohol and partying (i.e. topic ID = 90) was a meaningful for differentiating individual with "high" and "low" levels of openness. Extroverts and introverts (i.e. individuals with "high" and "low" levels of extraversion) exhibited notable differences in their discussion of emotional processes (i.e. topic ID = 44) via private messages on Facebook. Furthermore, extroverts and introverts also varied in their use of words related to recreation (i.e. topic ID = 41), which may reflect existing psychological associations between extraversion and positive affect.

5.2.3 Temporal

Surprisingly, outgoing latency was *not* a discriminating feature with respect to anxiety, loneliness, or personality traits. On the other hand, incoming latency was one of the more important temporal features for predicting four out of the seven psychological measures. As shown in Figure 5.2, anxious and lonely individuals' friends took longer to respond to them (i.e. anxious and lonely populations' communications exhibited a greater incoming latency). Furthermore, extroverts and introverts took about the same amount of time to respond to messages they received, on average. However, introverts' friends took longer to respond to them than did extroverts' friends.

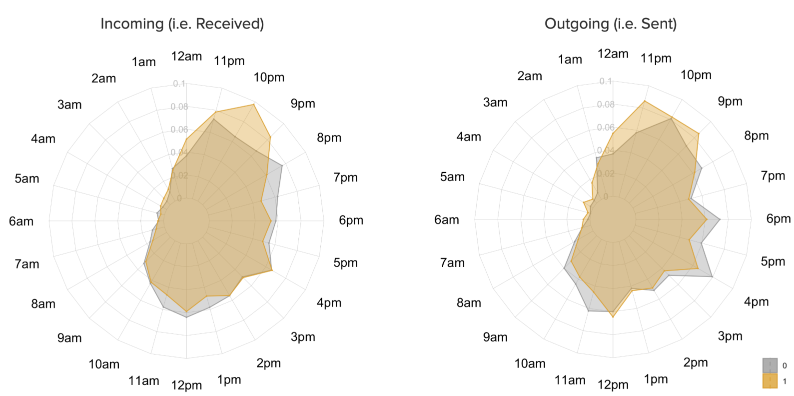
Individual differences in psychological attributes also mediated when participants engaged in conversations on Facebook Messenger. Anxious participants showed notable variation in evening DTC activity compared to non-anxious individuals, especially between the hours of 9pm and 12am. More specifically,

FIGURE 5.2: Difference in incoming and outgoing latency between "high" (i.e. 1) and "low" (i.e. 0) anxiety, loneliness, and extraversion groups



anxious participants sent more messages at 9pm and 11pm and received more messages at 10pm than non-anxious participants, as highlighted in Figure 5.3. Lonely participants exhibited a similar divergence, although less consistently

FIGURE 5.3: Comparison between "high" (i.e. 1) and "low" (i.e. 0) anxiety classes with respect to the average proportion of messages received/sent during each hour of the day



and during a slightly shorter time window (9pm to 11pm). These results suggest a marked diurnal shift in communication patterns in our anxious participant population. Moreover, participants exhibited notably different patterns of Messenger usage at the beginning and end of the standard work day (i.e. 8am, 5pm) when compared to emotionally stable participants.

5.2.4 Topological

Turnover in ego-centric network and persistence of social signature were found to be important factors within the aggregate models for neuroticism. Specifically, individuals with high vs. low levels of loneliness exhibited noticeably different levels of persistence in general. This suggests that level of loneliness mediates the extent to which participants engage in consistent messaging behavior over the course of the five-month interval we studied. Neurotic and emotionally stable individuals exhibited a similar divergence in interpersonal dynamics as measured by average bidirectional persistence social signature. This divergence suggests that level of neuroticism mediated the extent to which individuals exchanged messages with the same groups of people each week during the five-month interval.

Maximum edge density measure proved to be an effective proxy for biased communications (i.e. concentrating messages primarily within a single chat) within our population. Participants in the "high" and "low" agreeableness groups were characterized by differences in maximum outgoing edge density, while participants in the "high" and "low" conscientiousness groups were characterized by differences in maximum incoming edge density. Notably, participants in the "high" and "low" loneliness group were characterized by differences in both incoming and outgoing edge weight entropy. This indicates that individuals' level of loneliness mediates the consistency of their messaging behavior.

The number of alters also proved to be a valuable discriminator within the metadata models for extraversion and conscientiousness, contributing to an improvement in model performance 0.042 and 0.198 respectively when compared to the independent models. The overall number of alters over the five-month interval was found to be one of the more discriminating features between extroverts and introverts. This trend extends to individuals with “high” vs. “low” levels of loneliness. Participants in the “high” and “low” conscientiousness groups were characterized by differences in number of individual alters to which messages were sent.

6 Discussion

6.1 Understanding Mental Health from DTCs

Political Factors. Semantic results reveal new insights about anxiety and neuroticism in relation to factors such as political unrest, social activities, social support, and even musical preference. The emergence of Topic 46 (“Election”) is unsurprising, given that we collected baseline measures in early November 2016, but nevertheless affirms the relevance of political tensions to college students’ mental wellbeing. Hoyt et al. found evidence of increased negative affect and cortisol levels in a US young adult population around the time of the 2016 US presidential election [63], pointing to the significant detrimental effects the election had on young Americans’ mental health. Whether this effect is unique to the 2016 election remains undetermined. Moreover, in the context of our work, we foresee an opportunity to investigate the relationship between political discussion, communication patterns, and short-term mental health outcomes in our population.

Games & Recreation. Discussion of games, both virtual (e.g. Topic 46: “Pokemon GO”) and physical (e.g. Topic 61: “Sports”) as discriminators conditions such as anxiety and neuroticism reveals much about the role of social games in mental health and wellbeing. Existing research supports our result that Pokemon GO served an important role in the mental wellbeing and emotion regulation, particularly among college students. For example, Kari et al. found that many participants reported using the game as self-treatment for helping with anxiety [64]. Our findings are more novel, however, for neurotic individuals, who have been shown to use variants of the words "depressed" and "lonely" more often [65], use anger words frequently in their posts, and use “social interaction words” more sparingly [27]. Neurotic individuals’ lack of discussion around sports, which are naturally social activities, further solidifies this evidence that neurotic individuals may be socially isolated or withdrawn.

Social Support. The relationship between "Social Support" (Topic 88) and to agreeableness is also informative, given that agreeableness may be influenced by MHI symptoms. Social support, both perceived and tangible, has been shown to strongly influence mental health outcomes, both positively and negatively. For example, Grieve et al. found that connectedness on Facebook correlated with reduced anxiety and depression [66]. Further, Indian and Grieve found that greater perceived social support on Facebook was associated with higher subjective wellbeing among high-socially anxious users [67]. Additionally, Burke and Kraut showed that inbound directed communication (e.g. Facebook Messenger texts, comments, wall posts) with strong ties is associated with increases in wellbeing, for Facebook users at large [68]. Future research with our dataset could investigate manifestations of social support beyond the Semantic domain, as well as the interplay between social support and personality trait expression. We also note that Topic 78 (“Spiritual Music”) provides an opportuntiy for further exploration, given the great relevance of both spirituality and music (separately) to mental health outcomes [69, 70].

Pronouns. Knowledge of language patterns linked to MHIs is often critical for early intervention and prevention of worsening symptoms. For example, De Choudhury et al. found that individuals who first posted to mental health subreddits, then later to a suicide-specific subreddit (r/SuicideWatch), had posts with poorer linguistic structure in general [33]. They also received fewer comments, a phenomenon De Choudhury et al. observed among suicidal individuals on Reddit [33]. Additionally, Eichstaedt et al. found that words associated with loneliness, hostility, and rumination were the strongest predictors of depression in medical records [32]. Further, previous LIWC analyses have shown that greater use of first person singular pronouns has been associated with depression (including postpartum depression) and suicidality [31, 71, 33, 32]. Interestingly, our lexical results showed that first person plural pronouns (e.g. "we", "us", "our") were discriminative for anxious individuals, but first person singular pronouns were not. This finding suggests the need for deeper exploration of the relationship between pronoun usage and MHIs beyond depression and suicidality.

Diurnal Shifts. Our temporal results reveal that communication at late night hours is highly discriminative for both loneliness and anxiety. These MHIs have been shown to be comorbid with sleep disorders such as insomnia or hypersomnia [72], especially in young adults [73]. This dovetails off of existing literature documenting a general relationship between loneliness and adverse sleep outcomes [74, 75, 76]. The relationship between loneliness and adverse sleep outcomes has also been documented in adolescents as well [77]. Similar relationships have been found between those with anxiety disorders and diminished sleep quality as well [78]. In short, the present study lends support to the potential link between loneliness and anxiety and sleep, which may manifest itself temporally through diurnal shifts in communication (e.g., lonely individualsbeing more likely to communicate at night).

Responsiveness. We also note that the anxious and introverted groups had longer inbound message latency, on average, which could indicate diminished engagement or involvement from their friend networks. Anxious individuals were especially prone to this phenonomenon, as their friends used more filler words and netspeak. Diminished engagement has historically been found in depressed groups. For example, De Choudhury et al. observed lowered overall engagement among depressed individuals including mothers with PPD; both groups tend to share less and interact less with their peers on Facebook [31, 79]. Moreover, the PPD group also exhibited sharp, sudden changes in their level of activity over time [79]. Collectively, our data and previous literature point to the need for deeper investigation into temporal communication pattern disruption and the relationship between latency and network engagement in depressed, anxious, and introverted populations.

Social Inclusion. Broadly, the exchange of DTCs within a social network represents an ever-shifting exchange of *social support*. Previous works have emphasized the importance of *directed communications*, a subset of DTCs, for accumulating social support (*social capital*) [80, 81] and increasing tie strength [39]. Social anxiety and loneliness have both been associated with having fewer friends on Facebook [82, 83], indicating a possibly hampered ability of socially anxious or lonely individuals to grow their online social support network. Moreover, both groups exhibit distinct styles of sharing personal information that may affect their ability to gather and retain social support.

The differences between high and low-loneliness individuals in topological characteristics, is also grounded in prior literature within social psychology, particularly within social exclusion. Baumeister & Leary [84] found that the need to belong is so imperative to human existence that humans have developed a complex regulatory system to re-establish belonging when it is threatened (deemed the Social Monitoring System [85]). Further research has shown that social exclusion motivates heightened attention and vigilance to others' social cues [86] in the service of repairing belonging, a relationship which has also interestingly been shown for individual differences in levels of loneliness such that higher amounts of loneliness are related to higher amounts of attention to social cues [87].

However, recent work has shown that rejection based on a stigmatized identity (i.e., a potential chronic source of rejection) caused decreased attention to social cues [88]. In this way, participants high in loneliness may have impaired motivation to re-affiliate, and thus demonstrate decreased persistence relative to less lonely participants. For example, Fernandez et al [82] found that individuals with greater social anxiety tend to include more information in their personal profiles, which may signal a need for validation through oversharing. Meanwhile, Jin et al [83] showed that lonely individuals tend toward negative self-disclosure and less "communicating activity" (such as making comments on others' posts), both of which could deter potential network friends from offering social support outright.

6.2 Limitations

Our findings carry several limitations that should be addressed in future research. First, our study population was relatively small and homogenous ($N = 103$). Communication behaviors may be impacted by age and/or cohort effects, thereby limiting generalizability of our findings to an older population. Future work should look at a more diverse sample and examine whether age is a covariate. Second, our analysis focuses on private messaging patterns and does not account for public social media behavior and other social interactions (i.e. in-person, phone, SMS). By not accounting for social interactions our participants may have engaged in on these other platforms, our experimental results and conclusions are biased toward private messaging behaviors on a specific type of platform (Facebook messenger). Furthermore, prior work suggests that texting behavior (e.g. sharing intense and private emotions) varies across different platforms [17]. Thus, our findings from private messaging patterns may not necessarily translate to dynamics on public-facing DTC platforms. Third, as evidenced by the poor performance of the topological predictive models, our measures of interpersonal dynamics were not as reflective of individual differences in psychological measures as we hypothesized. This is likely due to the egocentric, generalized manner in which we extracted these features from our dataset. A more detailed picture of social network dynamics beyond the egocentric properties (e.g. interactions between participants' social contacts) would allow for more effective characterization of the quality of interpersonal exchanges on this platform.

6.3 Implications and Future Work

Each layer of the *SocialText* framework intentionally highlights features that can be derived from DTC data and used to identify social context, thus improving prediction of MHIs from DTCs. While the upper layers define important variables for data partitioning, the lowest layer identifies categories of features that can be extracted from the messages themselves. Features pertaining to the semantics and lexicon of message content can characterize conversational context, while temporal and topological features can reveal social network ties and temporal messaging patterns. Considering all message features in combination provides a comprehensive characterization of the effect of the social dynamics of DTCs on participants' mental states, thus improving the performance of the resulting predictive models. Researchers can use *SocialText* to identify and leverage multiple methodologies for characterizing or predicting mental health states.

Ref.	**Modality**	**Time**	**Category**	**Direction**	**Message Features**	**Health Outcome**
[20]	Facebook	Months	Metadata	↓	Temporal, Topological	Stress, Social Support
[45]	Twitter	Month	Content	↑	Semantic, Lexical	Stress
[26]	Twitter	Months	Content	↑	Semantic, Lexical	Mood
[44]	SMS	Day	Metadata	↑	Temporal	Communication Satisfaction
[22]	SMS	All times	Content	↑	Semantic	Neuroticism
[89]	SMS	Month	Metadata	↑	Temporal	Social Anxiety, Loneliness
[19]	SMS	Week	Metadata	↓↑	Temporal	Depression
[25]	SMS	Day	Content	↓↑	Semantic, Lexical	Depression, Suicide

TABLE 6.1: Existing Literature Table. Direction (↑: outgoing, ↓: incoming)

Table 6.1 provides a list of selected, relevant studies that utilize DTC data to study mental health outcomes. In this table, we map each study onto the *SocialText* hierarchy, demonstrating its flexibility. Moreover, this mapping highlights important methodological overlaps in the existing literature. For example, Elhai et. al. [19] studied depression with respect to temporal patterns in SMS data, while Nobles et. al. [25] studied the semantic and lexical features of a similar dataset. While these studies choose different time windows (or, rather, *Time* layer selections), they are similar along all other dimensions of *SocialText*'s structure. By using *SocialText* to identify similar studies, such as [19] and [25], researchers can streamline the process of creating new methodological approaches from the best aspects of existing approaches. Thus, *SocialText* facilitates the development of novel methodologies for mobile mental health sensing.

7 Conclusion

Analysis of digital text communications (DTCs) remains an open research area at the intersection of mental health and computing. DTCs are feature-rich characterizations of social context, yet remain largely underexplored in existing mobile sensing frameworks. Previous approaches to analyzing DTC features often address quantitative and qualitative features separately. In this paper, we have proposed the *SocialText* framework, which defines a hierarchical structure for extracting features from DTC datasets. Features pertaining to the semantics and lexicon of message content can characterize conversational context, while temporal and topological features can reveal social network ties and temporal messaging patterns. Considering all message features in combination provides a comprehensive characterization of the effect of social dynamics of DTCs on participants' mental states, and allows researchers to leverage DTC feature extraction methodologies across academic disciplines. Our results corroborate previously established results and reveal novel individual differences in temporal and relational behaviors, as well as in vocabulary usage and topics of discussion, on Facebook Messenger. This work provides a novel path forward for future analysis and discussion of the role of DTCs in personality, mental health, and wellbeing online.

Bibliography

[1] P. LeViness, C. Bershad, K. Gorman, L. Braun, and T. Murray, “The Association for University and College Counseling Center Directors Annual Survey - Public Version 2018”, en, p. 73, 2018.

[2] S. Bagroy, P. Kumaraguru, and M. De Choudhury, “A Social Media Based Index of Mental Well-Being in College Campuses”, en, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*, Denver, Colorado, USA: ACM Press, 2017, pp. 1634–1646.

[3] A. Shensa, J. E. Sidani, C. G. Escobar-Viera, K.-H. Chu, N. D. Bowman, J. M. Knight, and B. A. Primack, “Real-life closeness of social media contacts and depressive symptoms among university students”, *Journal of American College Health*, vol. 66, no. 8, pp. 747–753, 2018.

[4] C.-c. Yang and Y. Lee, “Interactants and activities on facebook, instagram, and twitter: Associations between social media use and social adjustment to college”, *Applied Developmental Science*, pp. 1–17, 2018.

[5] J. Park, D. S. Lee, H. Shablack, P. Verduyn, P. Deldin, O. Ybarra, J. Jonides, and E. Kross, “When perceptions defy reality: The relationships between depression and actual and perceived facebook social support”, *Journal of Affective Disorders*, vol. 200, pp. 37–44, 2016.

[6] E. Kross, P. Verduyn, E. Demiralp, J. Park, D. S. Lee, N. Lin, H. Shablack, J. Jonides, and O. Ybarra, “Facebook use predicts declines in subjective well-being in young adults”, *PLOS ONE*, vol. 8, no. 8, e69841, 2013.

[7] P. Verduyn, D. S. Lee, J. Park, H. Shablack, A. Orvell, J. Bayer, O. Ybarra, J. Jonides, and E. Kross, “Passive facebook usage undermines affective well-being: Experimental and longitudinal evidence.”, *Journal of Experimental Psychology: General*, vol. 144, no. 2, p. 480, 2015.

[8] B. A. Primack, A. Shensa, C. G. Escobar-Viera, E. L. Barrett, J. E. Sidani, J. B. Colditz, and A. E. James, “Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among us young adults”, *Computers in Human Behavior*, vol. 69, pp. 1–9, 2017.

[9] D. C. DeAndrea, N. B. Ellison, R. LaRose, C. Steinfield, and A. Fiore, “Serious social media: On the use of social media for improving students’ adjustment to college”, *The Internet and higher education*, vol. 15, no. 1, pp. 15–23, 2012.

[10] D. J. Attai, M. S. Cowher, M. Al-Hamadani, J. M. Schoger, A. C. Staley, and J. Landercasper, “Twitter social media is an effective tool for breast cancer patient education and support: Patient-reported outcomes by survey”, *Journal of Medical Internet Research*, vol. 17, no. 7, e188, 2015.

[11] J. A. Naslund, S. W. Grande, K. A. Aschbrenner, and G. Elwyn, “Naturally occurring peer support through social media: The experiences of individuals with severe mental illness using YouTube”, *PLOS ONE*, vol. 9, no. 10, e110171, 2014.

[12] E. Rice, S. Kurzban, and D. Ray, “Homeless but connected: The role of heterogeneous social network ties and social networking technology in the mental health outcomes of street-living adolescents”, *Community Mental Health Journal*, vol. 48, no. 6, pp. 692–698, 2012.

[13] G. Fergie, S. Hilton, and K. Hunt, “Young adults’ experiences of seeking online information about diabetes and mental health in the age of social media”, *Health Expectations*, vol. 19, no. 6, pp. 1324–1335, 2016.

[14] K. B. Wright, “Emotional Support and Perceived Stress Among College Students Using Facebook.com: An Exploration of the Relationship Between Source Perceptions and Emotional Support”, en, *Communication Research Reports*, vol. 29, no. 3, pp. 175–184, Jul. 2012.

[15] N. B. Ellison, C. Steinfield, and C. Lampe, “The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites”, en, *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, Jul. 2007.

[16] R. L. Nabi, A. Prestin, and J. So, “Facebook Friends with (Health) Benefits? Exploring Social Network Site Use and Perceptions of Social Support, Stress, and Well-Being”, en, *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 10, pp. 721–727, Oct. 2013.

[17] N. N. Bazarova, Y. H. Choi, V. Schwanda Sosik, D. Cosley, and J. Whitlock, “Social sharing of emotions on facebook: Channel differences, satisfaction, and replies”, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 2015, pp. 154–164.

[18] K. O’Leary, S. M. Schueller, J. O. Wobbrock, and W. Pratt, “ suddenly, we got to become therapists for each other”: Designing peer support chats for mental health”, in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 2018, p. 331.

[19] J. D. Elhai, M. F. Tiamiyu, J. W. Weeks, J. C. Levine, K. J. Picard, and B. J. Hall, “Depression and emotion regulation predict objective smartphone use measured over one week”, *Personality and Individual Differences*, Examining Personality and Individual Differences in Cyberspace, vol. 133, pp. 21–28, Oct. 2018.

[20] M. Burke and R. Kraut, “Using facebook after losing a job: Differential benefits of strong and weak ties”, in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ACM, 2013, pp. 1419–1430.

[21] M. Skowron, M. Tkalčič, B. Ferwerda, and M. Schedl, “Fusing social media cues: Personality prediction from twitter and instagram”, in *Proceedings of the 25th International Conference Companion on World Wide Web*, International World Wide Web Conferences Steering Committee, 2016, pp. 107–108.

[22] T. Holtgraves, “Text messaging, personality, and the social context”, en, *Journal of Research in Personality*, vol. 45, no. 1, pp. 92–99, Feb. 2011.

[23] G. Coppersmith, M. Dredze, and C. Harman, “Quantifying mental health signals in twitter”, in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 51–60.

[24] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, “From ADHD to SAD: Analyzing the language of mental health on twitter through self-reported diagnoses”, in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado: Association for Computational Linguistics, 2015, pp. 1–10.

[25] A. L. Nobles, J. J. Glenn, K. Kowsari, B. A. Teachman, and L. E. Barnes, “Identification of imminent suicide risk among young adults using text messages”, in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 2018, p. 413.

[26] J. Bollen, H. Mao, and A. Pepe, “Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena”, in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[27] G. Farnadi, S. Zoghbi, M.-F. Moens, and M. De Cock, “Recognising personality traits using facebook status updates”, in *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.

[28] H. A. Schwartz, M. Sap, M. L. Kern, J. C. Eichstaedt, A. Kapelner, M. Agrawal, E. Blanco, L. Dziurzynski, G. Park, D. Stillwell, M. Kosinski, M. E. Seligman, and L. H. Ungar, “Predicting Individual Well-Being Through The Language of Social Media”, in *Biocomputing 2016*, Kohala Coast, Hawaii, USA: World Scientific, Jan. 2016, pp. 516–527.

[29] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, “Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach”, *PLOS ONE*, vol. 8, no. 9, e73791, Sep. 2013.

[30] S. R. Pendse, “Sochiatrist: Inferring the relationship between emotion and private social messages”, 2018.

[31] M. D. Choudhury, M. Gamon, S. Counts, and E. Horvitz, “Predicting depression via social media”, *ICWSM*, vol. 13, pp. 1–10, 2013.

[32] J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley, D. Preoţiuc-Pietro, D. A. Asch, and H. A. Schwartz, “Facebook language predicts depression in medical records”, en, *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11 203–11 208, Oct. 2018.

[33] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, “Discovering shifts to suicidal ideation from mental health content in social media”, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, Santa Clara, California, USA: ACM Press, 2016, pp. 2098–2110.

[34] T. Wang, M. Brede, A. Ianni, and E. Mentzakis, “Detecting and characterizing eating-disorder communities on social media”, in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, ACM, 2017, pp. 91–100.

[35] S. Volkova, K. Han, and C. Corley, “Using social media to measure student wellbeing: A large-scale study of emotional response in academic discourse”, in *International Conference on Social Informatics*, Springer, 2016, pp. 510–526.

[36] X. Wang, C. Zhang, and L. Sun, “An improved model for depression detection in micro-blog social network”, in *2013 IEEE 13th International Conference on Data Mining Workshops*, IEEE, 2013, pp. 80–87.

[37] A. Benton, M. Mitchell, and D. Hovy, “Multi-Task Learning for Mental Health using Social Media Text”, en, *arXiv:1712.03538 [cs]*, Dec. 2017.

[38] A. Reissmann, J. Hauser, E. Stollberg, I. Kaunzinger, and K. W. Lange, “The role of loneliness in emerging adults' everyday use of facebook-an experience sampling approach”, *Computers in Human Behavior*, vol. 88, pp. 47–60, 2018.

[39] M. Burke and R. E. Kraut, “Growing closer on facebook: Changes in tie strength through social network site use”, en, in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, Toronto, Ontario, Canada: ACM Press, 2014, pp. 4187–4196.

[40] A. Sano, A. J. Phillips, Z. Y. Amy, A. W. McHill, S. Taylor, N. Jaques, C. A. Czeisler, E. B. Klerman, and R. W. Picard, “Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones”, in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, IEEE, 2015, pp. 1–6.

[41] C. Wu, M. Boukhechba, L. Cai, L. E. Barnes, and M. S. Gerber, "Improving momentary stress measurement and prediction with bluetooth encounter networks", *Smart Health*, 2018.

[42] Y. Bachrach, T. Graepel, and P. Kohli, "Your Digital Image: Factors Behind Demographic And Psychometric Predictions From Social Network Profiles (Demonstration)", en, p. 2, 2014.

[43] S. Centellegher, E. López, J. Saramäki, and B. Lepri, "Personality traits and ego-network dynamics", *PLOS ONE*, vol. 12, no. 3, e0173110, 2017.

[44] A. Angster, M. Frank, and D. Lester, "An exploratory study of students' use of cell phones, texting, and social networking sites", *Psychological Reports*, vol. 107, no. 2, pp. 402–404, 2010-10.

[45] R. Gopalakrishna Pillai, M. Thelwall, and C. Orasan, "Detection of stress and relaxation magnitudes for tweets", in *Companion of the The Web Conference 2018 on The Web Conference 2018*, International World Wide Web Conferences Steering Committee, 2018, pp. 1677–1684.

[46] C. D. Spielberger, R. L. Gorsuch, R. Lushene, P. R. Vagg, and G. A. Jacobs, *Manual for the state-trait anxiety inventory*, English, Consulting Psychologists Press, 1983, 1983.

[47] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains", *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, Dec. 2003.

[48] P. Shaver, W. Furman, and D. Buhrmester, "Transition to college: Network changes, social skills, and loneliness.", 1985.

[49] D. Russell, L. A. Peplau, and C. E. Cutrona, "The revised ucla loneliness scale: Concurrent and discriminant validity evidence.", *Journal of personality and social psychology*, vol. 39, no. 3, p. 472, 1980.

[50] R. D. Hays and M. R. DiMatteo, "A short-form measure of loneliness", *Journal of Personality Assessment*, vol. 51, no. 1, pp. 69–81, 1987.

[51] M. Ziegele and L. Reinecke, "No place for negative emotions? the effects of message valence, communication channel, and social distance on users' willingness to respond to sns status updates", *Computers in Human Behavior*, vol. 75, pp. 704 –713, 2017.

[52] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods", *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[53] E. Loper and S. Bird, "Nltk: The natural language toolkit", *arXiv preprint cs/0205028*, 2002.

[54] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography", *Computational Linguistics*, vol. 16, no. 1, pp. 22–29, 1990.

[55] D. Lin, "Extracting collocations from text corpora", in *First workshop on computational terminology*, Citeseer, 1998, pp. 57–63.

[56] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[57] T. Griffiths, "Gibbs sampling in the generative model of latent dirichlet allocation", 2002.

[58] A. K. McCallum, "Mallet: A machine learning for language toolkit", http://mallet.cs.umass.edu 2002.

[59] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, "Friends don't lie: Inferring personality traits from social network structure", in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM, 2012, pp. 321–330.

[60] A. Alshamsi, F. Pianesi, B. Lepri, A. Pentland, and I. Rahwan, "Network diversity and affect dynamics: The role of personality traits", *PLOS ONE*, vol. 11, no. 4, e0152358, 2016.

[61] D. H. Wolpert, "Stacked generalization", *Neural Networks*, vol. 5, pp. 241–259, 1992.

[62] J. Wilson. (2016). Pokémon go launches in us on ios and android, [Online]. Available: `https://venturebeat.com/2016/07/06/pokemon-go-launches-worldwide-on-ios-and-android/` (visited on 07/06/2016).

[63] L. T. Hoyt, K. H. Zeiders, N. Chaku, R. B. Toomey, and R. L. Nair, "Young adults' psychological and physiological reactions to the 2016 us presidential election", *Psychoneuroendocrinology*, vol. 92, pp. 162–169, 2018.

[64] T. Kari, J. Arjoranta, and M. Salo, "Behavior change types with Pokémon GO", in *Proceedings of the International Conference on the Foundations of Digital Games - FDG '17*, Hyannis, Massachusetts: ACM Press, 2017, pp. 1–10.

[65] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, gender, and age in the language of social media: The open-vocabulary approach", *PLOS ONE*, vol. 8, no. 9, pp. 1–16, Sep. 2013.

[66] R. Grieve, M. Indian, K. Witteveen, G. A. Tolan, and J. Marrington, "Face-to-face or facebook: Can social connectedness be derived online?", *Computers in Human Behavior*, vol. 29, no. 3, pp. 604–609, 2013.

[67] M. Indian and R. Grieve, "When facebook is easier than face-to-face: Social support derived from facebook in socially anxious individuals", *Personality and Individual Differences*, vol. 59, pp. 102–106, 2014-03.

[68] M. Burke and R. E. Kraut, "The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength: Facebook AND Well-Being", *Journal of Computer-Mediated Communication*, vol. 21, no. 4, pp. 265–281, Jul. 2016.

[69] S.-T. Lin, P. Yang, C.-Y. Lai, Y.-Y. Su, Y.-C. Yeh, M.-F. Huang, and C.-C. Chen, "Mental health implications of music: Insight from neuroscientific and clinical studies", *Harvard Review of Psychiatry*, vol. 19, no. 1, pp. 34–46, 2011. eprint: `https://www.tandfonline.com/doi/pdf/10.3109/10673229.2011.549769`.

[70] K. S. Seybold and P. C. Hill, "The role of religion and spirituality in mental and physical health", *Current Directions in Psychological Science*, vol. 10, no. 1, pp. 21–24, 2001. eprint: `https://doi.org/10.1111/1467-8721.00106`.

[71] S. Park, I. Kim, S. W. Lee, J. Yoo, B. Jeong, and M. Cha, "Manifestation of depression and loneliness on social networks: A case study of young adults on facebook", in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, Vancouver, BC, Canada: ACM Press, 2015, pp. 557–570.

[72] D. Neckelmann, A. Mykletun, and A. A. Dahl, "Chronic Insomnia as a Risk Factor for Developing Anxiety and Depression", *Sleep*, vol. 30, no. 7, pp. 873–880, 2007.

[73] D. J. Buysse, J. Angst, A. Gamma, V. Ajdacic, D. Eich, and W. Róssler, "Prevalence, Course, and Comorbidity of Insomnia and Depression in Young Adults", *Sleep*, vol. 31, no. 4, pp. 473–480, 2008.

[74] J. T. Cacioppo, L. C. Hawkley, G. G. Berntson, J. M. Ernst, A. C. Gibbs, R. Stickgold, and J. A. Hobson, "Do lonely days invade the nights? potential social modulation of sleep efficiency", *Psychological Science*, vol. 13, no. 4, pp. 384–387, 2002.

[75] J. T. Cacioppo, L. C. Hawkley, L. E. Crawford, J. M. Ernst, M. H. Burleson, R. B. Kowalewski, W. B. Malarkey, E. Van Cauter, and G. G. Berntson, "Loneliness and health: Potential mechanisms", *Psychosomatic medicine*, vol. 64, no. 3, pp. 407–417, 2002.

[76] L. C. Hawkley and J. T. Cacioppo, "Loneliness matters: A theoretical and empirical review of consequences and mechanisms", *Annals of Behavioral Medicine*, vol. 40, no. 2, pp. 218–227, 2010.

[77] R. A. Harris, P. Qualter, and S. J. Robinson, "Loneliness trajectories from middle childhood to pre-adolescence: Impact on perceived health and sleep disturbance", *Journal of Adolescence*, vol. 36, no. 6, pp. 1295–1304, 2013.

[78] R. M. Chase and D. B. Pincus, "Sleep-related problems in children and adolescents with anxiety disorders", *Behavioral Sleep Medicine*, vol. 9, no. 4, pp. 224–236, 2011.

[79] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, Paris, France: ACM Press, 2013, p. 3267.

[80] M. Burke, R. Kraut, and C. Marlow, "Social capital on facebook: Differentiating uses and users", en, in *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems - CHI '11*, Vancouver, BC, Canada: ACM Press, 2011, p. 571.

[81] N. B. Ellison, C. Steinfield, and C. Lampe, "Connection strategies: Social capital implications of facebook-enabled communication practices", *New Media & Society*, vol. 13, no. 6, pp. 873–892, 2011.

[82] K. C. Fernandez, C. A. Levinson, and T. L. Rodebaugh, "Profiling: Predicting Social Anxiety From Facebook Profiles", en, *Social Psychological and Personality Science*, vol. 3, no. 6, pp. 706–713, Nov. 2012.

[83] B. Jin, "How lonely people use and perceive Facebook", *Computers in Human Behavior*, vol. 29, no. 6, pp. 2463–2470, Nov. 2013.

[84] R. F. Baumeister and M. R. Leary, "The need to belong: Desire for interpersonal attachments as a fundamental human motivation.", *Psychological Bulletin*, vol. 117, no. 3, p. 497, 1995.

[85] C. L. Pickett and W. L. Gardner, "The social monitoring system: Enhanced sensitivity to social cues as an adaptive response to social exclusion.", 2005.

[86] C. L. Pickett, W. L. Gardner, and M. Knowles, "Getting a cue: The need to belong and enhanced sensitivity to social cues", *Personality and Social Psychology Bulletin*, vol. 30, no. 9, pp. 1095–1107, 2004.

[87] W. L. Gardner, C. L. Pickett, V. Jefferis, and M. Knowles, "On the outside looking in: Loneliness and social monitoring", *Personality and Social Psychology Bulletin*, vol. 31, no. 11, pp. 1549–1560, 2005.

[88] L. S. Richman, J. Martin, and J. Guadagno, "Stigma-based rejection and the detection of signs of acceptance", *Social Psychological and Personality Science*, vol. 7, no. 1, pp. 53–60, 2016.

[89] D. J. Reid and F. J. Reid, "Text or Talk? Social Anxiety, Loneliness, and Divergent Preferences for Cell Phone Use", en, *CyberPsychology & Behavior*, vol. 10, no. 3, pp. 424–435, Jun. 2007.