CLINICAL AND SOCIAL PATHWAYS TO CARE: A COMPUTATIONAL EXAMINATION OF SOCIAL MEDIA FOR MENTAL HEALTH CARE

A Dissertation
Presented to
The Academic Faculty

By

Sindhu Kiranmai Ernala

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing
CS

Georgia Institute of Technology

August 2021

CLINICAL AND SOCIAL PATHWAYS TO CARE: A COMPUTATIONAL EXAMINATION OF SOCIAL MEDIA FOR MENTAL HEALTH CARE

Thesis committee:

Dr. Munmun De Choudhury,
Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Eric P. S. Baumer
School of Computer Science and Engineering
Lehigh University

Dr. Moira Burke
Core Data Science
Facebook

Dr. Elizabeth Mynatt
Institute for People and Technology
Georgia Institute of Technology

Dr. Diyi Yang
School of Interactive Computing
Georgia Institute of Technology

Date approved: July 9, 2021

To my parents

and Prof. Navjyoti Singh

ACKNOWLEDGMENTS

A long list of people contributed to my intellectual, personal, and professional growth and made this thesis possible.

First, I would like to thank my advisor, Munmun De Choudhury, for her guidance, unwavering support, and dedication towards my growth as a scholar. Munmun has been an ideal advisor, appreciating my strengths and weaknesses, nurturing my curiosity, always being available, and giving THE best advice. Munmun, you are a constant source of inspiration. I have also benefited greatly from how you lead the SocWeb research group and create a supportive learning environment for all. In our lab, I found an academic family that is caring and constructive towards my research agenda. The culmination of our *guru-shishya-parampara* is apparent in my research approach and a love for good food, cooking, coffee, and decor (but not so apparent in my eye for formatting).

To my thesis committee members, Moira Burke, Beth Mynatt, Eric Baumer and Diyi Yang, thank you for your generosity, feedback and support. I am lucky that these wonderful people and brilliant researchers agreed to be on my committee.

Moira's mentorship during two summer internships at Facebook was transformative in my career. She has shown me how to do foundational, high impact, and meaningful research in the industry. Alongside everything that I have learned from you, including survey design, R programming, good science communication, etc., working with you reminds me why I love doing research, and for that I am grateful. Beth is a leader in HCI and Health Informatics, and I have immensely benefited from her guidance and wisdom. Beth, I am so thankful that you pushed me to think more deeply about patients' perspectives in my research; it has led to one of the most meaningful projects I have ever done. I wish I had sought your feedback much earlier in my PhD. Eric, your advice since my proposal has been instrumental in shaping how I think about research methods and the ethics of my work. I will recall your insightful conversations every time I think about the design of socio-technical systems. Diyi has been a great role model to emulate ever since she was a grad student. I am thankful for her kind support and encouragement and her expertise and thoughtfullness in combining the areas of NLP and social computing.

I am indebted to my research participants, without whom this work would not have been possible. Thanks to our team at Northwell Health, Michael Birnbaum, John Kane, Asra Ali, Elizabeth Arenare, Anna Van Meter, and Sabrina Yum-Chan, for all your efforts in initiating the THRIVE project. Everything I have learned about mental health care is from our fruitful collaboration.

To my SocWeb labmates: Stevie Chancellor, Koustuv Saha, Dong Whi Yoo, Vedant Das Swain, Sachin Pendse, and Matt Kim, thanks for your support, feedback, and friendship. You have made the ups and downs of the PhD journey manageable and fun. To my research assistants, Kathan Kashiparekh, Jordyn Seybolt, Domino Weir, Amir Bolous, Tristan Labetoulle, and Fred Bane, your hard work and enthusiasm greatly contributed to this thesis. I have learned so much from each one of you. To other friends and colleagues at GT, Umashanti Pavalanathan, Tanu Mitra, Youngwook Do, Eshwar Chandrashekaran, Sucheta Ghoshal, Ian Stewart, Sandeep Soni – thanks for the feedback and support.

Thanks to the Core Data Science team at Facebook and the Information Science Institute (ISI, USC) for hosting me during summer internships. Alex Leavitt, Alex Dow, Farshad Kooti, Justin Cheng, and Lada Adamic, at Facebook, Emilio Ferrara, Anna Sapienza, and Kristina Lerman at ISI: thank you for your support, for lending your expertise and shaping my research skills.

A group of co-authors, collaborators and mentors have shaped my thinking during this PhD. I have learned so much from Nicole Ellison; she is a deep thinker, a caring collaborator and writing papers with her is a tremendous joy. Sauvik Das provided early advice during my PhD qualifiers, and I have learned a lot from his innovative research approach. Our work on dynamic audience selection, with Sauvik, Rachel Chen, Kristen Wells, Yuxi Wu, and Stephanie Yang, shaped my thinking around privacy and social media. Social

Computing faculty at GT, Amy Bruckman, Eric Gilbert, and Jacob Eisenstein, thanks for your support and advice through the early stages of the doctoral process. Also, thanks to Rosa Arriaga, who has been a strong advocate for student well-being as the IC graduate co-ordinator. Thanks to Jan Morian, Renee Jaimeson, Danielle Shenise, Philicia Bellinger, and Cindy Parry, who patiently helped me navigate administrative processes, reimbursements, etc., at GT.

I had the amazing opportunity to get started on research right from undergraduate studies, thanks to the Center of Exact Humanities for developing my interest in the social sciences and other professors at IIIT Hyderabad for honing my skills in CS. The late Prof. Navjyoti Singh molded my wide-eyed enthusiasm and curiosity into a serious interest in interdisciplinary research. Prof. PRK Rao presented enough conundrums on the philosophy of science that will keep me a lifelong learner. I am also grateful for Paul Buitelaar and Shruthi Kunde, who took a chance on me and offered research internships while I was as an undergrad.

Moving away from India to the U.S. to pursue my PhD was possibly one of the hardest things I have ever done. Thanks to friends and family who made this transition smoother. My crew, Arpit Merchant, Monica Reddy, Shivani Poddar, Diksha Yadav, Aditi Gupta, Sahil Loomba, Atabak Ashfaq, Deepali Jain, and Anumeha Rai, have kept my sense of belonging and quirks intact.

This dissertation is indebted to the support of my family. To my parents: who I am today, personally and professionally, is a reflection of your love, strength, encouragement, and sacrifice. You nurtured my curiosity and ambitions early on, and your unshaken belief and confidence in me is what kept me going through the hard parts of this journey. I dedicate this thesis to you. To my sister, Pranathi, thank you for being my pillar. My in-laws have literally hustled with me as I tried to meet conference deadlines these past years. I am grateful for their love and support. Most of all, thanks to my partner in life, Sai, for your patience, support, and relentless pursuit in keeping me happy. You acted as a sounding

board for my research, particularly the practical implications of my work, and as a mentor and friend during the pandemic job search. This and what comes next, is our achievement!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In the last decade, powered by connectivity to large social networks and advances in collecting and analyzing digital traces of individuals from social media platforms, researchers have gleaned rich insights into individuals' and populations' mental health states and experiences, including their moods, emotions, social interactions, language, and communication patterns. Using these inferences, researchers have been able to study support-seeking behaviors, distinguishing patterns, risk markers, and diagnosis states for mental illnesses from social media data, promising a fundamental change in mental health care. What we need next in this line of work is for data and algorithms based on social media to be contextualized in people's pathways to mental health care. However, there are several challenges and unanswered questions that present hurdles.

First, gaps exist in the psychometric validity of social media based measurements of behaviors and the utility of these inferences in predicting clinical outcomes in patient populations. Second, if social media can act as an intervention platform, outside of discrete events, a holistic understanding on its role in people's lives along the course of a mental illness is crucial. Lastly, several questions remain around the ethical implications of research practices in engaging with a vulnerable population subject to this research.

This thesis charts out empirical and critical understandings and develops novel computational techniques to ethically and holistically examine how social media can be employed to support mental health care. Focusing on schizophrenia, one of the most debilitating and stigmatizing of mental illnesses, this thesis contributes a deeper understanding on pathways to care via social media along three themes: 1) prediction of clinical mental health states from social media data to support clinical interventions, 2) understanding online self-disclosure and social support as pathways to social care, and 3) intersection of social and clinical pathways to care along the course of mental illness. In doing so, this work combines theories from social psychology, computer-mediated communication, and clin-

ical literature with machine learning, statistical modeling, and natural language analysis methods applied on large-scale behavioral data from social media platforms. Together, this work contributes novel methodologies and human-centered algorithmic design frameworks to understand the efficacy of social media as a mental health intervention platform, informing clinicians, researchers, and designers who engage in developing and deploying interventions for mental health and well-being.

CHAPTER 1
INTRODUCTION

Social media has led to an unprecedented change in how individuals, clinicians, and online platforms consider mental health. Supported by anonymity and connections to large audiences, individuals are increasingly adopting online social platforms to share personal, sensitive stories about mental health. In the last decade, research investigations on mental health and social media use have ranged from understanding self-disclosure practices and goals [1], deciphering social support provisions to promote positive mental health outcomes [2, 3, 4], discovering community norms and behaviors [5], and exploring how these platforms can support intervention delivery [6]. At the same time, digital traces on these platforms have allowed clinical researchers to observe individuals' mental health attributes like mood, emotions, social interactions, language and communication patterns in a real-time, non-invasive, longitudinal fashion that was previously unimaginable for mental health care. Leveraging these digital traces, researchers have applied machine learning and natural language processing techniques to identify risk markers associated with several mental illnesses like depression [7], schizophrenia [8], and stress [9], and support clinical decision making. Beyond academic interest in this domain, products of this research are out in the world. For instance, Facebook developed tools to predict sensitive at-risk behaviors like suicidal ideation on their platforms[1] and the consequences of these predictions are impacting people's lives everyday. These directions collectively place mental health as one of the most notable topics associated with social media use.

This thesis examines the efficacy of social media in supporting mental health care. In order to resolve the role of social media in mental health care, what we need next in this line of work is for data and algorithms based on social media to be contextualized in

[1] https://www.facebook.com/safety/wellbeing/suicideprevention/

people's pathways to care. This could involve potentially using insights about mental health from online social platforms to connect people in need with timely and proactive help, in the form of interventions, or working with stakeholders like clinicians and public health organizations to harness these algorithmic insights to influence policy and decision-making. These directions could also mean designing and re-imagining online social platforms as safe, supportive spaces for disclosure of mental health experiences, raising awareness and seeking/providing social support.

In order to pursue these next steps and realize the potential of algorithmic insights on mental health, there are several challenges and unanswered questions that need to be addressed. A first set of questions revolve around the validity and domain utility of algorithmic insights into mental health based on digital traces from social media. Despite being promoted as a powerful means to shape interventions and impact mental health recovery and social care, there is little that we understand about how these approaches through social media fit into peoples' pathways to care. For clinical care, the psychometric validity of social media based measurements of behaviors and the utility of inferences in predicting actual clinical outcomes in patient populations is still unexplored. Gaps exist in theoretical understanding of how disclosure goals and support seeking behaviors change from an offline, private, one-on-one settings to online, public, networked contexts.

Second, current academic scholarship is heavily focused either on clinical care in the form of treatment, medication, etc., or on social support and self-management of the condition. For overall well-being of people with mental illness, both clinical treatment and social care are needed along the course of illness. Despite strong advocacy from paradigms such as the Recovery model [10] or Person-centered care [11], the intersection of social and clinical aspects of care for mental health has been unexplored. If social media can act as a mental health intervention platform, a holistic understanding of the role it plays in supporting clinical as well as social pathways to care along the course of a mental illness is crucial. How can we study the different journeys of those with mental health conditions combining their experiences of clinical recovery and social reintegration?

Third, there are several questions around the ethical implications of this research topic. How do we protect a vulnerable population who is subject to this research? What provisions should we have to prevent adverse outcomes? What is the social and moral responsibility of researchers in minimizing harms? These questions show that consideration to ethics and privacy need to be part of the research process itself, centering the expectations and needs of stakeholders.

Focusing on schizophrenia, one of the most debilitating and stigmatizing of mental illnesses, this thesis contributes a deeper understanding on pathways to mental health care via social media along three themes and ask these research questions:

1. **Clinical pathways to care: Prediction of mental health states from social media data to support interventions.** What is the efficacy and validity of social media based behaviors as diagnostic signals for the prediction of mental health states? Can social media data contributed by patient populations indicate risk to adverse clinical outcomes such as relapse events?

2. **Social pathways to care: Online self-disclosure and social support on social media.** Do the goals of offline self-disclosure such as therapeutic outcomes translate to the context of online broadcasting disclosures of mental illnesses? What are the mechanisms through which large invisible audience on social media provide support to individuals making mental illness disclosures? Do audience in a social network impact future disclosure behaviors of people?

3. **Intersection of social and clinical pathways along the course of illness.** Anchoring on psychiatric hospitalizations, this theme asks how can we study the different journeys of those with mental health conditions combining their experiences of clinical recovery and social reintegration? How are health status transitions characterized by social media use during the periods after hospitalization? What is the role of so-

cial technologies as individuals transition and reintegrate back to their social lives after psychiatric hospitalization?

Thesis Statement and Research Approach This thesis shows that social media, and algorithmic approaches informed by clinical and patient stakeholder perspectives, can support clinical and social pathways to care for mental health in the form of patient-provider interventions and social support provisions. Towards answering the above research questions, this thesis charts out empirical and critical understandings and develops novel computational techniques to ethically and holistically examine how social media can be employed to understand and support mental health care. In doing so, this work combines theories from social psychology, computer-mediated communication and clinical literature with machine learning, statistical modeling, and natural language analysis methods applied on large scale behavioral data from social media platforms.

Two senses of "social media" structure the questions central to this thesis. The first involves thinking of social media platforms as *instrumentation* [12], i.e., how can digital traces that people leave behind on social platforms be employed to characterize, measure and understand their mental health attributes and experiences? In a second sense, this thesis views social media as a new *object of study* [12], i.e, examining how self-presentation, self-disclosure and social support seeking behaviors related to mental health are transformed in the context of large scale, technology-mediated, broadcasting social networks. Overall, this thesis combines empirical insights from several social media platforms, including Facebook, Twitter, Reddit, Instagram, and communication technologies such as Whatsapp towards answering questions about mental health.

Understanding mental health and social media use in an ethical and holistic manner also necessitates certain human-centered research considerations that are employed throughout this thesis. First, involving people whose expertise lies outside of computing involved in the design of algorithmic systems is important while considering multi-stakeholder, high risk domains such as mental health. In addition to the people who might end up using the algorithmic decision support systems, this also includes individuals whose data is being analyzed to build such systems. Second, grounding computational approaches in theory and paying attention to the real world use of these algorithms is crucial in fostering interdisciplinary collaborations especially with clinicians. This means designing algorithms in a way that they are useful to the stakeholders in clinical settings and fit into existing practices. Finally, paying attention to privacy and ethics with respect to the people who contributed data for this research throughout the process is essential. The work described in this dissertation is done in close collaborations with clinical researchers, practitioners (psychiatrists, clinical psychologists) and people with lived experiences. The role of these stakeholders in the research process and how the above research considerations are employed in practice is highlighted at several sections in the subsequent chapters.

1.1 Contributions

Though the study of social media's efficacy as a mental health intervention platform, this thesis contributes to theory and practice in social computing with implications to mental health care and intervention design.

At a theoretical level, this dissertation clarifies "for whom" and "in what contexts" algorithmic systems based on digital traces from social media platforms can support mental health care. This work shows that if the broader research agenda is to inform clinical decision making, such as early diagnosis, treatment or patient-provider interventions, working with data contributed by clinically diagnosed patient populations is imperative. For people whose clinical diagnosis state is unknown, this work shows how self-disclosure goals and support seeking practices translate from a one-to-one, private, offline context to a networked, public online setting. For instance, work described in Chapter 4 demonstrates evidence for therapeutic outcomes of online self-disclosures and explains mechanisms through which a large invisible audience on Twitter provide reciprocity and support to disclosures. Different people might adopt social media platforms for differential needs during mental health transitions and experiences. Chapter 6 presents an empirically-derived taxonomy of heterogeneous behavioral patterns that characterize people's health status transitions around psychiatric hospitalizations. Based on the derived taxonomy, this work sheds light on the different recovery and reintegration journeys as exhibited on online social platforms. For instance, this work revealed that first hospitalization experiences lead most people to transition into the withdrawal focused status, whereas those re-hospitalized are able to maintain connectivity to their social networks. The support seeking goals and outcomes for the former group might be significantly distinct from the latter – insights that could contextualize how different individuals use social media to find help and advice around their mental distress.

To practice, this thesis contributes computational approaches for the study of mental health and social media use. Chapter 3 presents how patient-volunteered and patient-contributed social media data can be used to build machine learning models that can predict adverse outcomes such as relapse hospitalizations in schizophrenia. The feasibility of these approaches shows promise for future technology-mediated interventions for clinical outcomes such as relapse. This work also presents a model for social reintegration in mental health based on a combination of medical data and social media trace data as well as qualitative interviews with individuals with lived experience of psychiatric hospitalization. The computational methods described in Chapter 6 present a pipeline to characterize and measure heterogeneous behavioral patterns exhibited by people on social media platforms like Facebook during health status transitions. Empirical insights from this work inform the design of online social platforms to support people around major life events and health status transitions.

Outside of social computing, this thesis also provides domain specific contributions to the area of mental health. This work informs key stakeholders involved in developing and deploying technology based interventions for mental health and well-being. For clinicians who aim to provide early interventions for mental health, this thesis provides methodologies based on machine learning and predictive modeling that can be applied to improve evidence-based treatment and interventions. For instance, work described in Chapter 3 on prediction of relapse episodes in schizophrenia is designed to fit into a prospective real-world clinical setting. How one manages their illness outside of institutionalized clinical treatment strongly impacts both future clinical outcomes and overall well-being. Chapter 6 presents empirical insights on how people transition from the hospital back to their homes after psychiatric hospitalization and the role of social media and offline social networks in supporting recovery and reintegration. These findings inform clinicians and social workers about their own practices in supporting post-discharge care. Furthermore, the empirical insights on heterogeneous recovery and reintegration trajectories in mental health can act as new information in discursive therapy sessions and help in sensemaking of hospitalization experiences. For designers of mental health interventions, this research provides theoretically grounded, data-derived insights on individualized heterogeneity of mental health states that could factor into the intervention design for tailored support and care according to the person's context.

1.2 Overview of Thesis

The thesis is broadly organized into the three themes described above. Chapter 2 provides background on schizophrenia and prior work on social media use and mental health motivating the research topic. Chapter 3 presents work on employing machine learning techniques from social media data for prediction of clinical outcomes in patient populations. Chapter 4 focuses on pathways to social care and discusses two empirical studies on social media related to online self-disclosure and social support. In Chapter 5, a triangulation study is presented that overviews methodological approaches in this line of work surfacing challenges in employing social media for mental health care. Lastly, Chapter 6 addresses the intersection of social and clinical care for mental health combining empirical insights from computational modeling of health transitions and qualitative interviews with individuals with lived experiences. Chapter 7 presents concluding remarks, limitations and contributions of this work.

Table 1.1: Summary of Studies

Study	Thematic Area	Summary	Social Media Sites	Location
Collaborative approach to identify social media markers of schizophrenia	Pathways to clinical care via social media	Assessing the utility of social media as a viable diagnostic tool in identifying individuals with schizophrenia.	Twitter	Chp 3
Detecting relapse in schizophrenia using patient-contributed Facebook data	Pathways to clinical care via social media	Predicting imminent relapse in schizophrenia from social media activity of individuals receiving psychiatric care.	Facebook	Chp 3
Methodological gaps in predicting mental health states from social media	Pathways to clinical care via social media	Examining the quality of different social media-derived behavioral signals in predicting clinical diagnoses of mental illness.	Twitter, Facebook	Chp 5
Therapeutic outcomes of online self-disclosures of mental illnesses	Pathways to social care via social media	Examining how offline mental health disclosure goals like therapeutic benefits translate to the context of online broadcasting self-disclosures.	Twitter, Reddit	Chp 4
Audience engagement and its impact on online disclosures of mental illnesses	Pathways to social care via social media	Understanding the audience of online self-disclosures, and forecasting their impact on future disclosure behaviors.	Twitter	Chp 4
A social media study on mental health status transitions surrounding psychiatric hospitalizations	Intersection of social and clinical pathways to care via social media	Modeling social media use around health status transitions surrounding hospitalization	Facebook	Chp 6
The reintegration journey following a psychiatric hospitalization: Examining the role of social technologies	Intersection of social and clinical pathways to care via social media	Understanding the role of social technologies in reintegration after psychiatric hospitalization	Facebook. Twitter, Whatsapp, Instagram	Chp 6

CHAPTER 2
BACKGROUND

2.1 Schizophrenia

Schizophrenia is a devastating mental illness, affecting about 1% of the world's population [13]. It is characterized by distortions in thinking, perception, emotions, language, sense of self and behaviour including experiences such as hallucinations (hearing voices or seeing things that do not exist) and delusions (fixed, false beliefs) [14]. The condition is often described in terms of positive and negative (or deficit) symptoms [15]. Positive symptoms are those that most individuals do not normally experience, but are present in people with schizophrenia. They can include delusions, disordered thoughts and speech, and hallucinations. Negative symptoms are deficits of normal emotional responses or of other thought processes. They commonly include flat expressions or little emotion, poverty of speech, inability to experience pleasure, lack of desire to form relationships, and lack of motivation. Three-quarters of people with schizophrenia develop the disease during early age between 16 and 25 years of age[1] leading to considerable disability and interference with the establishment of healthy social, educational and occupational foundations.

Despite being a chronic illness and causing enduring disability, current psychiatric treatment paradigms are limited in leveraging benefits of early identification of risk markers, and interventions [16]. This is largely due to lack of timely contact and difficulty in gathering longitudinal data on patients' illness trajectory. The nature of data also presents challenges; current evaluation and treatment mechanisms heavily rely on self or family reported information that is subject to recall bias. Furthermore, schizophrenia is associated with high stigma in the society; historical accounts have considered lived experiences of

[1] https://sardaa.org/resources/about-schizophrenia/

schizophrenia as playing the role of a "sacrificial victim" [17]. These stigma perceptions lead to negative stereotypes and further inhibit people suffering with schizophrenia from receiving care and social support.

Into this amalgam of challenges, technology based interventions, especially based on social media or smartphone use, have started to show great potential by providing insights from behavioral data that is collected in a naturalistic, unobtrusive manner. Recent research has studied technology use by individuals suffering from schizophrenia and related psychotic disorders [18, 19]. Matthews et al. [19] found that technology use in this vulnerable population is often impacted by underlying mood, and that, these differential patterns in technology use may indicate incipient mood episodes. However, research on the use of social media platforms by this population is lacking. Focusing on social media use and schizophrenia, Mitchell et al. [8] present potential linguistic markers of schizophrenia using the tweets of self-identified schizophrenia sufferers. McManus et al. [20] mine Twitter data of individuals following a schizophrenia self-help account to improve detection of schizophrenia from social media use. But the extent to which social media based early detection techniques and interventions can support individuals with schizophrenia is largely unknown [21]. This thesis studies schizophrenia as a complex case of mental illnesses to examine how social media provides clinical and social pathways to care for individuals with schizophrenia.

Why schizophrenia for this thesis? Among complex mental illnesses, why is schizophrenia an excellent case study for examining the role of social media as a mental health intervention platform? I discuss three motivations framed around the clinical and social experiences of schizophrenia.

Early Detection and Interventions: A key challenge with schizophrenia is the high likelihood of relapse. Although the majority of patients initially achieve clinical remission of positive psychotic symptoms with pharmacological treatment, up to 80% schizophrenia patients relapse in five years [22]. Once a second episode has occurred, further episodes are likely and risks of continued functional decline increase. Although early interventions are known to help prevent escalation of symptoms, existing methods to recognize imminent relapse are significantly limited due to lack of timely contact and unavailability of data on patients' illness trajectories. Thus, the examination of social media for clinical pathways to care via early detection of risk markers and adverse episodes like relapse, has immense potential to transform interventions for schizophrenia.

Disclosure and Social Support: Schizophrenia is a highly stigmatized condition. Literature has recognized the value of candid disclosures resulting in improved well-being and therapeutic benefits among individuals challenged with this illness [23]. For instance, participation in offline self-help groups and advocacy organizations has been found to facilitate self-disclosure—such activities help challenge private shame about the illness, enhance self-esteem, enable people to be more resilient in response to stigma experiences, and thereby support symptomatic coping [24]. With the affordances of anonymity and connectedness to a large network, increasing number of people are appropriating social media and online communities for disclosures, raising awareness, breaking stigma and seeking social support for mental health. Therefore, the study of benefits and mechanisms for social support on social media provides opportunities to inform design of platforms for accommodating needs surrounding experiences of schizophrenia.

Computational Methods: The study of schizophrenia on social media also presents unique methodological challenges. To examine both social and clinical pathways to care on social media, it is imperative to identify the target population or behavioral signals indicating the pathways to care. This requires not only computational methods like machine learning employed at large scale, but also theoretically grounded measures from clinical and social science literature to identify *what needs to be measured*. For instance, impairment in verbal communication and language disturbances are characteristic diagnostic features of schizophrenia [25]. These features informed by theory, can be measured on social media using natural language analysis approaches. While employing social media as a data source, it is therefore crucial that the operationalization of online behaviors is grounded in theory (i.e. known evidence about the illness and people's experiences). This calls for a need for novel theoretically grounded methodologies with machine learning and computational linguistics approaches. Due to the interdisciplinary nature of the research topic, it is also important to consider implications of computational approaches to domain stakeholders. To the clinician community, whose primary source of diagnostic information comprises clinically validated questionnaires, scales, interviews, and symptoms reported by the patient [26], new forms of signals derived from social media, despite the right intentions, add complexities to the conventional psychiatric assessment method. Therefore, there is a need for accessible, interpretable, computational approaches and deep seated interdisciplinary collaborations in using social media data for clinical interventions.

2.2 Prediction of mental health states based on social media data

In recent years, a growing body of work has employed large scale social media data to model and infer mental well-being of individuals and populations [27]. These approaches have been used to identify and understand social media derived risk and psychological markers of other mental health conditions, ranging from postpartum depression [28], eating disorders [5, 29], post-traumatic stress [30], and other conditions [2, 31].

Burgeoning interest in this topic stems from the fact that social media data is readily available and archived, and can be unobtrusively gathered with low effort and cost [32]. These unique attributes help overcome many challenges in state-of-the-art clinical assessment of mental health that involves subjective recollection of historical facts—a method prone to retrospective recall bias [33]. However, appropriating social media data to inform clinical efforts around early diagnosis, tailoring treatment, or delivering interventions, suffers significant limitations. In a clinical setting, diagnostic information is available to the clinician via self-reported psycho-social signs and symptoms, theoretically and psychometrically validated clinical scales, interviews, questionnaires, and other diagnostic tools [26]. Social media data by itself, however, does not include such clinically validated signals to accurately identify and validate individuals' mental health states. Also, collecting clinically valid diagnostic signals from social media would require engagement with an at-risk patient population, a cohort that is stigmatized, sensitive, and vulnerable. This presents logistical challenges to identification of diagnostic signals, as well as privacy and data protection issues. Such a data collection approach can be difficult to scale, and is effort- and time-consuming, requiring carefully crafted clinical and risk management protocols, and involvement of clinical experts.

To circumvent these challenges, researchers have employed several online behaviors as gold standard information, or what we call *proxy diagnostic signals* to identify individuals' mental illness diagnoses. Through a systematic literature review [34] based on a keyword search of papers on predicting mental health states from social media, we identified three types of proxy diagnostic signals from the literature, which we elaborate below.

Proxy Diagnostic Signals in the Literature *Affiliation Behaviors:* A first category of research represents behaviors signaling engagement or association (via hashtags, account following, community membership) with content related to mental health resources on social media, as proxy diagnostic signals of an illness [35, 20, 36]. A prominent example is McManus et al. [20] who used following a Twitter account (@schizotribe) dedicated to conversations around lived experiences of schizophrenia as a signal for gold standard information that an individual might be suffering from schizophrenia. A complementary set of papers have operationalized membership in online mental health support communities such as Reddit and Livejournal as proxies for diagnostic information [37, 38, 39, 40].

Self-reports: Next, the most popular form of proxy diagnostic signals, this category operationalizes first-hand, public self-disclosures of diagnosis of a mental illness as indicators of a clinical mental illness [41, 42, 43, 30, 9, 44, 45, 46, 47, 48, 49, 50, 8, 51, 52, 53, 39, 54, 55, 56, 57]. A notable example, Mitchell et. al. [8] used regular expression search queries on Twitter ("I have been diagnosed with schizophrenia") to extract self-reports of schizophrenia diagnoses and then employed them for predicting their presence/absence.

External validation: Finally, this category represents human-in-the-loop, collaborative approaches that either seek self-reported information from the individual, or incorporate diagnostic scales and/or expert appraisal for identification of the proxy diagnostic signals [58, 59, 60, 45, 27, 61, 62, 63, 64, 65, 66].

Using Proxy Diagnostic Signals: Critical Challenges Appropriating these proxy diagnostic signals has overcome many challenges and barriers to gathering clinically valid diagnostic data on social media, particularly around scale and size [43], and these approaches continue to gain traction in the community. However they suffer from significant limitations, which we frame below, drawing upon the critical data literature [67, 68, 69].

Consider the case when affiliation to mental health resources is considered a proxy of a diagnosis. While including genuine patients, it likely also includes other stakeholders like mental health practitioners and experts, non-profits raising awareness campaigns, caregivers etc. As another example, although the act of self-disclosing a mental illness can be an indicator of a person's mental condition, there are gaps in understanding what an individual chooses to self-report, why, and when they decide to do so, or if they are being truthful.

In other words, there is lack of evidence that these proxy signals are accurately measuring what they intend to measure, also known as construct validity [70] (whether the signals accurately identify and represent individuals at-risk). A lack of contextualization in psychiatric practice [71] or theory [17] additionally reduces confidence in their construct validity—an issue recognized in prior critiques of big data approaches [67, 69]. Although proxy signals with expert validation attempt to tackle some of these theoretical and clinical gaps, because the approach is removed from direct interaction with the individual, their veracity can be questioned, and their "*claims to objectivity and accuracy can be misleading*" [67].

Further, individuals with unique attributes, attitudes, and characteristics, possibly distinct from patient populations, are likely to engage in the specific types of behaviors enumerated by the proxy signals. Apart from the inclusion of "noisy" data, the unique ways in which the proxy diagnostic signals are defined and construed can lead to a variety of biases in the predictions, despite the impressive sample sizes they promise. This resonates with what boyd and Crawford noted, that "*bigger data are not always better data* [67]" and what Olteanu et al. discuss at length surrounding methodological pitfalls of big data [70].

2.3 Self-disclosure, social support on social media

Self-disclosure & Stigma Management. Sociologist Erving Goffman emphasized the importance of "sympathetic others" in helping people cope with difficult experiences, as well in enabling self-disclosure [72]. Self-disclosure provides an opportunity to express one's thoughts and feelings, develop trust and build intimacy in personal relationships [73]. However, the act of self disclosure is a much more complex and critical process for people with a concealable, stigmatized identity such as mental illness [74]. On the one hand, the stigma around these conditions may risk unfavorable outcomes such as social rejection and discrimination and might be detrimental to well-being. Experimental manipulation studies found that participants do not experience the benefits of disclosure when confidant reactions are neutral or negative [75]. But on the other hand, positive outcomes of disclosure due to opening up, include a wide range of therapeutic benefits leading to both physical and mental well-being, such as lowered psychological distress [76]. For instance, studying the post-traumatic stress (PTSD) experiences of rape and sexual assault victims, Ullman and Filipas found that disclosures led to more positive and fewer negative social reactions [77]. This complex nature of both possibilities is nested within an ongoing process of *"stigma management"*—coping with the psychological and social consequences of one's identity [72].

A rich body of work in the Computer Mediated Communication (CMC) literature has studied self disclosures and the socio-cognitive processes centered around them. Through several experimental and anecdotal evidence, CMC and other internet-based behaviors have been characterized to exhibit high levels of self disclosure [78]. In fact, high self-disclosure has been recognized to lead to dis-inhibition on the Internet [79]. At the same time, self disclosure in CMC contexts is also argued to be beneficial, having been linked to trust and group identity [80], as well as playing an important role in social interactions by reducing uncertainty [81].

Turning to research on social media, an emergent line of research has investigated the nature of self disclosures on social media and online communities. Several quantitative studies have focused on identification, modeling and characterizing differences in multi-modal (textual, visual) forms of self disclosure on social media [82, 83, 84, 31, 85]. Similarly, from a qualitative perspective, prior work has studied how individuals undergoing gender transition appropriate Facebook for engaging in sensitive disclosures of their experiences [86]. Existing literature has also investigated unique design affordances of social media like "throwaway" accounts, in providing context-specific anonymity for first-time disclosures on abuse related posts on Reddit [1]. In another study, Andalibi et al. found that individuals struggling with negative emotions, such as those related to depression or self-harm, use Instagram to self-disclose and engage in social exchange and storytelling about their stigmatized experiences [2].

We note that work thus far has largely been around platforms where the others in the context of self-disclosures are sympathetic others, as Goffman (2009) posited it. However, the nature and impact of engaging with the audience of self-disclosures on public social media platforms is understudied. This work aims to fill this gap.

Social Support & Social Capital. There has also been a relevant line of research concerning online social capital and social support in the context of self-disclosures and well-being. Social capital allows an individual to draw on resources from other members in their social network through bonding and bridging [87]. Bonding social capital refers to strong ties or relationships amongst members of a network who are similar in some form [88]. Whereas, bridging social capital refers to weak ties or relationships amongst people who are dissimilar in a demonstrable fashion, such as age, socio-economic status, race/ethnicity and education [89]. While online social networks have been established to support building and maintaining both kinds of social capital (Ellison et al. 2007), scholars also refer to a related concept "social support", especially in the context of self-disclosure theories and studies of stigma. A large body of work reveals the support benefits people derive from their interpersonal relationships and social networks in relation to improved health and psychological well-being, self esteem, satisfaction with life, and reciprocity [91].

Specific to our focus on stigmatized experiences around mental health, both qualitative and quantitative studies have identified social capital and social support as necessary components in self-disclosure goals and outcomes [83, 92]. Nevertheless, gaps still exist in our understanding of how the expectations of social support and the benefits with respect to social capital translate when the audience of self-disclosures are invisible, public, or comprise largely of weak ties. Moreover, the role that the audience of stigmatized disclosures, through support provisioning and social feedback mechanisms, plays in encouraging (or constraining) future disclosure processes, is yet to be empirically investigated. This thesis extends prior work by providing robust data-driven studies of the audience of schizophrenia disclosures on Twitter.

2.4 Clinical studies of psychiatric hospitalization experiences

Since the deinstitutionalization movement [93], the role of the psychiatric hospital has shifted from a place for long-term stay and treatment to a community-based system of care emphasizing reducing feelings of dependency, and supporting integration [94]. Today, when patients are admitted to a psychiatric bed, the goals of clinical care revolve around crisis stabilization, diagnosis, and initiation of appropriate treatment [94]. Once patients' symptoms stabilize, this model encourages rapid discharge from the hospital so that individuals may continue receiving care in outpatient settings.

However, psychiatric hospitalizations are still life-altering, as the admission often implies that individuals are unequipped to manage their psychiatric needs and require removal from their existing environment to receive appropriate, urgent care [95]. When an individual is admitted to a hospital for mental illness, it is often due to an adverse event like self-harm or suicidal ideation, lack of insight or denial of illness, social crises such as relationship problems, or non-compliance with medication [96]. Based on the symptoms, treatment during hospitalization involves individual and group therapy, psychotherapy, pharmacotherapy, or other standing medication [97]. Such circumstances increase stress experienced by the patient and their caregivers, leading to perceptions of hospitalization-related anxiety, and fear of confinement [98]. Recovery journeys of people post-hospitalization are similarly challenging due to the high likelihood of re-hospitalization [99], lack of support, non-adherence and side effects to medication, and difficulty in managing the condition and reintegrating back to social life and roles [96]. Kent and Yellowlees [96] found that social factors contribute to 38.9% of re-hospitalizations, followed by factors related to psychiatric and physical illness. Paksarian et al. [98] found 69% of participants reporting at least one of their hospitalizations as traumatic, with the most common experiences related to rigidity and involuntary hospitalization, being put on restraints, and being forced to take medications. [100], on the other hand, found participants reporting re-hospitalizations to be less traumatizing than the first hospitalization, a necessary relief, as occurring by default and without progress, and as part of the recovery process.

Whether positive or negative, while re-hospitalizations reflect the clinical aspects of recovery journeys, social processes are found to be equally important. Based on a systematic review and narrative synthesis of the literature, Leamy et al. [10] identified five categories (CHIME): connectedness, hope and optimism about the future, identity, meaning in life, and empowerment to engender recovery processes. Hope, agency, opportunity for purposeful activity, and social inclusion are measured as outcomes of recovery in mental health [101]. In fact, successful social reintegration involving resuming "age, gender, and culture appropriated roles, statuses and activities" [102] and community participation [103] can help reduce stigma and improve overall well-being [104].

Underpinning this interplay of clinical and social processes is the recovery approach [105], which is today adopted as the guiding principle of mental health policy in many countries. Emerging from the deinstitutionalization period, the recovery approach emphasizes a person's potential for recovery – as "*a deeply personal, unique process of changing one's attitudes, values, feelings, goals, skills and/or roles. It is a way of living a satisfying, hopeful, and contributing life even with limitations caused by the illness. Recovery involves the development of new meaning and purpose in one's life as one grows beyond the catastrophic effects of mental illness.*" [105, 106] Together with recent approaches like person-centered model of care [107, 108, 109], these paradigms call for "*the promotion of health as a state of physical, mental, sociocultural, and spiritual well-being, as well as to the reduction of disease, and founded on mutual respect for the dignity and responsibility of each individual person*" [11]. Notable here is the work of Corey Keyes, who has advocated viewing recovery as flourishing in life despite having a mental illness and relying on two complementary reintegration experiences: the restoration from mental illness and the optimization of positive mental health [110]. These perspectives put the individual in the center and highlight the importance of understanding interrelationships over time and episodes like hospitalizations as part of "life course experiences with health" [109]. Combining medical records information (e.g., hospitalizations) and social media data of people suffering from mental distress, in this thesis, we adopt the recovery and person-centered care approach in our characterization of mental health status transitions to capture social and clinical aspects of peoples' hospitalization experiences.

2.5 Social media, health transitions and liminality

Transition is a concept widely used in the social science literature and is most commonly defined as "*a process of convoluted passage during which people redefine their sense of self and redevelop self-agency in response to disruptive life events* [111]." While some conceptualize transitions as linear processes with a clear beginning and end [112], others note that transitions can be "*complex, nonlinear, sometimes cyclical and potentially recurring*" [113, 114].

Psychiatric hospitalization as a liminal period. Many major life transitions, including health transitions, are marked by rituals. Van Gennep's [112] liminality framework refers to transitory processes as comprising preliminal, liminal, and postliminal stages, that relate to separating from a previous identity, making the transition, and incorporating back into the social world after transition, respectively. In the context of health transitions, Kaziunas and colleagues examined the interconnections between information and emotion work performed by bone marrow transplant caregivers by adopting a liminality lens [115]. The findings from this work highlight the usefulness of the liminality framework to make visible the work involved in navigating multiple social lives (as a part of everyday life and as a medical caregiver.) Following Van Gennep, hospitalizations for mental illnesses, the duration of which can range between a day to 4 months [116], can also be considered as institutionalized rituals and periods of liminality. While the individual transitions to their role as a patient in the hospital, there are many rules and rituals set by the hospital that they have to follow, including initially being in a locked ward that they cannot leave at will, and following a schedule for their meals, treatments, and activities. Importantly, in many cases, there is a lack of access to technology, social support, and offline connections [95, 98]. Finally, post-discharge from the hospital, individuals need to manage their new treatment plans and reintegrate back to professional, personal, or social lives [103].

Reintegration in mental health The recovery model, emerging from the deinstitutionalization movement in the late 20th century, is the guiding principle of mental health policy in many countries [105]. According to this model of care, the hospital is no longer considered an institution for long-term stay but is seen as a community-based system of care focused on overall well-being of individual and community integration [94]. Corey Keyes, who has advocated adopting the model of mental health as a "complete state" [110] suggests viewing recovery as flourishing in life despite having a mental illness and relying on two complementary reintegration experiences: the restoration from mental illness and the optimization of positive mental health [110]. These perspectives put the individual in the center and highlight the importance of the relationship between clinical recovery and social reintegration in mental health.

Clinical researchers and scholars in social work and nursing have focused attention on social reintegration and rehabilitation in mental health, alongside the well-established area of clinical recovery. Based on a systematic review and narrative synthesis of 366 papers on personal recovery in mental health, Leamy et al. [10] provide five categories (CHIME): connectedness, hope and optimism about the future, identity, meaning in life, and empowerment to engender recovery processes. Newman et al. point out that dimensions of reintegration like hope, agency (a sense of control over their lives), opportunity for purposeful activity, and social inclusion are in fact, outcomes of recovery in mental health [101]. Silva et al. found a 20% lower risk of re-hospitalisation for patients referred to community-based psychosocial support units following inpatient care compared to patients referred to the usual formats of outpatient care [117]. Ådnanes et al. emphasized the importance of meaningful social activities and community participation as well as support from peers and family members in reducing psychiatric re-hospitalizations [103]. Successful social reintegration, often defined as resuming "age, gender, and culture appropriated roles, statuses and activities" [102], is also known to help reduce stigma and improve overall well-being [104].

The current emphasis on reintegration outcomes in existing work over-weighs the understanding of processes and transformations that people undergo after psychiatric hospitalization. Furthermore, today, a significant portion of social activities and community participation happen over technology-mediated channels. However, our current understanding of reintegration and the measurement of social functioning as a clinical outcome only focuses on face-to-face interactions in the offline world. To realize opportunities and challenges in reintegration in mental health, we argue that it is crucial to also consider people's online social lives. By examining people's first hand experiences after psychiatric hospitalization, in this thesis we contribute insights into people's reintegration journeys in the offline and online context, furthering the understanding of aspects that support or hinder reintegration in mental health.

Social technology use following major life transitions Social technologies play an important role around major life transitions by helping individuals establish a "new normal" [118, 119], conduct identity work [120, 121] and reach out to similar others [122]. For instance, Semaan et al. [122] found that in the context of veterans returning to civilian life, technologies like social media enable people to re-integrate into society by developing identity awareness and connecting to similar others to understand post-military life and receive support. Prior research has examined several life transitions such as engagement [121], marriage [123], parenthood [124], loss of a job [125], divorce [126], the loss of a loved one [127, 128, 129, 130], and transition to college [131] in the context of different social technology use [125, 121, 119]. This body of work shows that people actively shape their digital footprints on these platforms by curating the self-presentation signals in the context of shifting identities [119] and show changes in language use [130] and behaviors surrounding transitions [28, 7].

Psychiatric hospitalization, that is characterized by institutionalized rituals and rules set by the hospital, and the shift from one's role as a patient in the hospital to home can also be considered as a major life transition. In contrast to other life transitions described above, psychiatric hospitalization is also cyclical and potentially recurring due the the high likelihood of relapse. How does social technology use change during psychiatric hospitalization? What are the ways in which these technologies are appropriated during reintegration periods during which people re-establish social connections? This dissertation contributes to this literature by examining psychiatric hospitalizations, an unexplored event under the lens of life transitions. We extend prior literature by identifying shifts in individuals' social technology use during reintegration after psychiatric hospitalization, relative to a previous 'normal' in their lives.

Modeling and Understanding Health Transitions Prior work in HCI and CSCW has explored different approaches to characterize and model health status transitions [132, 133, 134]. Relevantly, MacLeane et al. [135] developed a taxonomy of phases of addiction on Forum 77, an online health forum, using the transtheoretical model of behavior change [136]. Hayes et al. [137] defined the concept of a 'personal cancer journey' drawing from an in-depth study of cancer communities, while Jacobs et al. [138] presented a holistic framework describing the cancer journey from patient-centered perspectives (also see [139]). Eschler and Pratt identified the tasks related to challenges and responses in different phases of young adult cancer during diagnosis, treatment and survivorship [140]. Wen and Rose [141] developed machine learning methods to extract cancer event trajectories from messages in online breast cancer support groups. Liu et al. [133] similarly combined domain knowledge and machine learning methods to form a hierarchical classification of Twitter data that resolves different stages of drinking behavior. Feuston et al. studied how people get back to their social lives following traumatic brain injury. and they introduced the concept of social re-emergence as "a non-linear process of developing a new social identity that involves withdrawing from social life, developing goals for social participation, disclosing health information for social support and acceptance, and attaining social independence." [142]. Drawing on the notion of illness trajectory introduced by Strauss and colleagues [143, 144], Chen et al. discuss chronic care cycles, the repeated cycles between routine medical visit and subsequent homecare period [145]. Burgess et al. [146] studied how patient information work shifts over time and highlight two distinct but often overlapping phases, 'learning' and 'living with' a chronic condition.

In the case of mental health status transitions around hospitalizations, complex, cyclical and potentially recurring transitions are likely due to the high likelihood of relapse. As a first step into the study of recovery and reintegration transitions in mental health, we focus on individual hospitalization events in a person's journey with mental illness and propose a transition model to understand heterogeneous recovery and reintegration trajectories after hospitalization.

2.6 Social technologies & health management

CSCW and Health Informatics researchers have paid extensive attention to social technologies, particularly online health communities, to investigate their role in caregiving, informational exchange and peer support for health conditions. Researchers have investigated the role of online health communities and social technologies like Tumblr [147], Facebook [148], Wechat [149], and Instagram [150] in various health conditions including eating disorders [147, 151], fertility [152], and vulvodynia [153].

Significant work in this area focuses on studying people's behaviors on social technologies, i.e. seeking and providing social support, learning about coping mechanisms, and building peer networks, etc., to better understand health trajectories and outcomes. Researchers have conducted qualitative and participatory research to understand how people with health conditions use online health communities [154]. Focusing specifically on eating disorders, Pater et al. investigated how people suffering from eating disorders reveal their conditions on different social technologies (Twitter, Tumblr, and Instagram) [151] and provided evidence that eating disorder-related content can negatively affect people suffering from eating disorders even if they do not actively consume such content [147]. Huh and Ackerman studied how diabetes patient support groups help one another find individualized strategies and coping mechanisms for managing diabetes [155]. From a quantitative perspective, Wen and Rose [141] developed machine learning methods to extract cancer event trajectories from messages in online breast cancer support groups. Yang et al. investigated how communication on online health communities affected commitment and tenure of participants [156], and modeled different social roles like seekers, providers, storytellers, etc., they take up in online health communities [157]. Closely relevant to our focus on social reintegration, Feuston et al. studied how people get back to their social lives following traumatic brain injury. They introduced the concept of social re-emergence as "a non-linear process of developing a new social identity that involves withdrawing from social life, developing goals for social participation, disclosing health information for social support and acceptance, and attaining social independence" [142]. Burgess et al. examined how people with depression connect with others for support, largely via interactions mediated through locations and communication channels, and highlight the importance of sociality for self-management of depression [158].

In another line of work, researchers focused on improving the design of online health communities to better facilitate people's interactions and goals. Hartzler et al. developed prototypes that provide a health interest summary extracted from users' profiles to facilitate peer matching processes on these platforms [159]. O'Leary et al. designed a peer support chat system to enable peers to chat online using effective principles of talk therapy [160].

We situate our contributions in this body of work and the varied ways social technologies have supported or hindered people's health goals and management. While previous research has provided rich insights related to social technologies and how people with health conditions utilize such technologies, the topic has been under-explored in cases where people suffering from the health conditions are socially isolated (both online and offline), which is a common experience in people who are suffering from mental illnesses requiring intensive care. Work described in Chapter 6 expands these efforts by providing empirical evidence concerning mental health patients' use of social technologies focusing on psychiatric hospitalization and the period after discharge from in-patient care settings.

CHAPTER 3

PATHWAYS TO CLINICAL CARE THROUGH SOCIAL MEDIA

In this chapter, I introduce the first theme of this thesis: pathways to clinical care via prediction of mental health states from social media data. This theme addresses challenges to clinical interventions for mental health that arise due to lack of longitudinal data on patients' illness trajectories and difficulties in maintaining timely contact with patients. In this area, social media plays a role as an unprecedented, low cost and unobtrusively accessible data source that captures naturalistic behaviors of individuals from large, diverse populations.

I discuss two studies in this chapter to demonstrate pathways to clinical care using social media. First, as a pilot analysis, I discuss a "collaborative approach to identify social media markers of schizophrenia" employing clinical appraisals and machine learning. This study aims to move from noisy self-reports of schizophrenia on social media to a more accurate identification of diagnoses by exploring a human-machine partnered approach, wherein computational linguistic analysis of content is combined with clinical appraisals. Examining Twitter timeline posts of 671 individuals with self-disclosed diagnoses of schizophrenia, we found significant linguistic differences including greater use of interpersonal pronouns, decreased emphasis on friendship, and greater emphasis on biological processes. The machine learning classifier distinguished individuals with disclosures of schizophrenia from a control group with a mean accuracy of 88% using linguistic data alone. Compared to clinicians on new, unseen data, the classifier's precision, recall, and accuracy measures were 0.27, 0.77, and 0.59, respectively. This work provides evidence of distinguishing patterns in the language of social media (Twitter) posts between individuals who disclose about schizophrenia diagnosis and those who do not make any disclosures (control group).

In the second study, "Detecting relapse in schizophrenia using patient-contributed Facebook data," we collected 52,815 Facebook posts across 51 participants with recent onset of schizophrenia and applied anomaly detection methods to explore linguistic and behavioral changes associated with psychotic relapse. We built a one-class classification model that makes patient-specific personalized predictions on risk to relapse. The classifier achieved a specificity of 0.71 in predicting relapse hospitalizations. Results from this study indicate that social media activity captures objective linguistic and behavioral markers of psychotic relapse in young individuals with recent onset of schizophrenia and that machine learning models are capable of making personalized predictions of imminent relapse hospitalizations at the patient-specific level.

3.1 Identifying social media markers of schizophrenia employing a collaborative approach involving machine learning and clinical appraisals

Social media provides an unprecedented opportunity to transform early psychosis intervention strategies, especially for youth who are significant users of social media and at the greatest risk for the emergence of a psychotic disorder. Globally more than 2 billion users engage with social media regularly[1]. Youth with newly diagnosed schizophrenia in particular report frequently utilizing social networking sites throughout the course of illness development and treatment, engaging in social media activity several times daily, and spending several hours per day online [161].

Social media data has thus become a unique source for capturing personalized and population data in the forms of language, behaviors and frequency of use. Prior work in speech and text analysis has identified reliable linguistic markers associated with schizophrenia, including significant differences in word frequency, word categories, and use of self-referential pronouns [162, 163, 164, 165, 166]. These approaches have also been applied to demonstrate significant linguistic differences in posts written by individuals with

[1] https://www.webcitation.org/6rKzygEBi

schizophrenia compared to individuals with depression, physical illness, and healthy controls [167]. Furthermore, machine learning approaches employing social media data have achieved success in distinguishing participants with psychotic disorders from healthy controls based on linguistic differences in writing samples [166] and speech [164, 168]. For instance, computational models have achieved more than 80% and 90% accuracy [8, 20] in correctly identifying users with self-reported schizophrenia from healthy controls.

However, it is challenging to confirm the authenticity of online self-disclosures and as demonstrated by prior work, words that might have been automatically identified as self-disclosure such as ’psychosis,’ ’schizophrenia,’ and ’delusion’ are often used inappropriately online [169] and may represent a major limitation to computational models. To date, limited efforts have involved expert input to evaluate the authenticity of diagnostic self-disclosures. To move from noisy diagnostic inferences to accurate identification, we propose a human-machine partnered approach, wherein linguistic analysis of content shared on social media is combined with clinical appraisals. **This project aims to explore the utility of social media as a viable diagnostic tool in identifying individuals with schizophrenia.**

3.1.1 Data and Methods

Data. Data acquisition involved extracting publicly available Twitter posts from users with self-disclosed diagnoses of schizophrenia. We chose Twitter as a data source based on previous work identifying self-disclosure practices around mental illnesses on the social media platform. Adopting filtering techniques from prior work [170, 43], we used case-insensitive examples like “I am diagnosed with schizophrenia,” “told me I have schizophrenia,” and “I was diagnosed with schizoaffective disorder” (Refer Table 4.1) as search queries for data gathering. These search queries resulted in 21,254 posts by 15,504 users between 2012 and 2016. For each user, Twitter timeline data from 2012 to 2016 were collected using a Web-

Table 3.1: Search queries for Twitter data collection.

Diagnosed me with (schizophrenia/ psychosis)
Diagnosed schizophrenic
I am diagnosed with (psychosis/ schizophrenia)
I am schizophrenic
I have been diagnosed with (psychosis/ schizophrenia)
I have (psychosis/ schizoaffective disorder/ schizophrenia)
I think I have schizophrenia
My schizophrenia
They told me I have schizophrenia
I was diagnosed with (psychosis/ schizoaffective disorder/ schizophrenia)
Told me I have (psychosis/ schizophrenia)

based Twitter crawler called GetOldTweetsAPI[2], which scrapes public Twitter profiles to obtain historical Twitter data in a structured format. The data included tweet text, username, posting time, hashtags, mentions, favorites, geolocation, and tweet ID. A subsample of 671 users from the primary dataset was randomly selected (each user had equal probability of being selected) and provided to two clinicians for appraisal. As a control group, a random sample of Twitter users was collected from individuals without any mentions of ’schizophrenia’ or ’psychosis’ in their timeline.

Clinician Appraisal. To eliminate noisy data (disingenuous, inappropriate statements, jokes, and quotes) and obtain a cleaner sample of schizophrenia disclosures likely to be genuine, a psychiatrist and a graduate-level mental health clinician from Northwell Health’s Early Treatment Program, with extensive expertise in early stage schizophrenia, annotated the data. Each schizophrenia disclosure was annotated by categorizing them into one of three classes. Class “yes” contained users who appeared to have genuine disclosures. Class “no” contained users who had inauthentic posts, including jokes, quotes, or were from accounts held by health-related blogs. Class “maybe” contained users for whom the experts could not confidently appraise the authenticity of the disclosure. The annotation task for 671 users resulted in 146 yes, 101 maybe, and 424 no users (Cohen Kappa = 0.81 between yes and no classes). These three classes of users shared 1,940,921, 1,501,838,

[2]https://www.webcitation.org/6q8zxN1qp

and 8,829,775 tweets, respectively, with a mean (SD) of 13,293.98 (18,134.83), 14,869.68 (19,245.88), and 20,824.94 (45,098.07) tweets per user.

Classification Method. ***Data Preparation:*** To distinguish users with disclosures deemed genuine from the regular Twitter stream, the problem was modeled as a machine learning classification task. Users who had been annotated with class yes, formed the positive examples (class 1) for the classifier. A sample of same size collected from the control group formed the negative examples (class 0). Given the ambiguity of the "maybe" class, it was left out of this initial model. The training dataset, constructed by combining both positive and negative examples resulted in 292 users. The classifier was built and evaluated by applying 10-fold cross-validation, an established technique in supervised machine learning [171].

Classification Framework: Using the training datasets described previously, a supervised learning framework was used to build the classifier. The classification framework involved three steps: featurizing training data, feature selection to improve predictive power, and classification algorithm.

Featurizing Training Data The textual data from Twitter timelines was used to generate features for the classifier. Each tweet in the user's timeline was represented using the following features: n-Gram language model: a language model of 500 top unigrams, bigrams, and trigrams (ie, sequences of one, two, and three words) was generated from the entire timeline data of all users. Each tweet was represented as a feature vector of normalized term frequency-inverse document frequency (tf-idf) frequency counts of the top 500 n-grams. Linguistic inquiry and word count (LIWC): The widely validated LIWC lexicon [172] was employed, which identifies linguistic measures for the following psycholinguistic categories: (1) affective attributes, including positive and negative affect, anger, anxiety, sadness, swearing; (2) cognitive attributes, including both cognition categories comprising of cognitive mechanisms, discrepancies, inhibition, negation, causation, certainty, and tentativeness, and perception categories comprising of see, hear, feel, percept, insight, and relative; and (3) linguistic style attributes, including lexical density (verbs, auxiliary verbs, adverbs, prepositions, conjunctions, articles, inclusive, and exclusive), temporal references (past, present, and future tenses), social/personal concerns (family, friends, social, work, health, humans, religion, bio, body, money, achievement, home, sexual, and death), and interpersonal awareness and focus (first-person singular, first-person plural, and second-person and third-person pronouns). Each tweet was represented as a vector of normalized LIWC scores for each of the preceding 50 categories. Thus, the feature space for the classifier was 550; 500 n-grams and 50 LIWC categories.

Feature Selection to Improve Predictive Power **:** As the linguistic attributes of text contain several correlated features, the classification model tends to be unstable. To improve the predictive power of the model, feature scaling and feature selection methods were employed. Adopting the ANOVA F test reduced the feature space from 550 features to k –best features (where k=350) by removing noisy and redundant features.

Classification Algorithm: Finally, training data represented by the top k features was fed into a model to learn the classification task. The model was trained over several algorithms including the Gaussian naïve Bayes, random forest, logistic regression, and support vector machines [171]. Among these, the best performing algorithm on cross-validation was used for analysis.

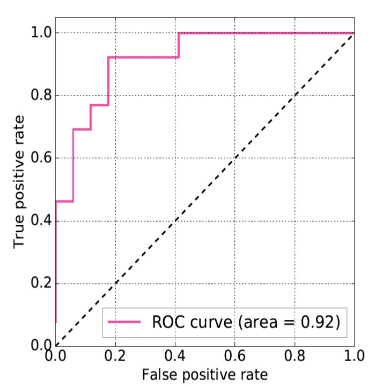


Figure 3.1: Receiver operating characteristic (ROC) curves for the classification task

Table 3.2: Confusion matrix showing agreement and disagreement between the machine learning classifier and the experts.

Machine label	Expert label	
	Yes	No
Yes	14	37
No	4	45

3.1.2 Results

To evaluate the performance of the classification model, a 10-fold cross-validation method was used. During each fold (iteration), the data was split into a 70% training set and 30% validation set. A model was then constructed on the 70% data and tested on the remaining 30%. Among the several classification algorithms that were applied, a random forest performed best with an average receiver operating characteristic (ROC) area under the curve (AUC) score of 0.88. The best performance for the classifier was 0.95 by the same AUC metric. The ROC curve is presented in Figure 3.1.

Verification in Unseen Data. To test the models for predicting new, unseen data, a sample of 100 users was passed through the classifier. The same sample was also provided to clinicians for appraisals. The confusion matrix displaying agreement between the two labels (machine and expert) is presented in Table 3.2. By taking the expert annotations as true outcome and the machine labels as predicted outcome, true positive, true negative, false positive, and false negative scores were computed. The resulting precision, recall, and accuracy measures were 0.27, 0.77, and 0.59, respectively.

3.1.3 Discussion

Consistent with prior trials [162, 164, 166], first-person pronouns were found to be significantly increased in the psychosis group, suggesting greater interpersonal focus. Additionally, these data replicate findings that biological processes, including words such as "body" and "health," are more frequently used in psychosis [167], suggesting a greater awareness or focus on health status. Furthermore, the psychosis group was significantly less likely to use words from the "friends" category, possibly associated with social withdrawal. Although language dysfunction, and specifically thought disorder, is an established core symptom of schizophrenia, these data suggest that subtle, more granular changes may additionally be associated with schizophrenia.

To date, the majority of studies have used a computational approach to flag publicly available social media profiles of users who self-disclose with limited input from mental health clinicians to assess the authenticity of online disclosure. In this study, expert appraisal eliminated more than 70% of Twitter profiles that might have otherwise been recognized by computational models as belonging to users with schizophrenia related disclosures. These data reinforce the need for ongoing collaborations integrating expertise from multiple fields to strengthen our ability to accurately identify and effectively engage with online traces of mental illness. A major challenge in treating schizophrenia remains the lengthy delay between symptom onset and receiving appropriate care [173]. At the same time, there is compelling evidence to suggest that linguistic and behavioral changes manifest on the pages of social media before they are clinically detected, providing the prospect for earlier intervention [45, 7, 174]. Although the potential beneficial impact of social media integration could be transformative, new critical questions regarding clinical expectations and responsibilities will require resolution. The degree of agreement between the classifier and the experts in this study suggests that the classifier performs well at eliminating inauthentic noisy samples, but was over-inclusive in labeling true cases of schizophrenia. For example, although the post "My parents are convinced I have schizophrenia," was labeled by the classifier as a genuine disclosure, clinicians deemed it to be a noisy sample, reflecting a more careful and conservative approach. Therefore, the classifier can theoretically assist in triaging digital data to capture signals of authentic disclosures of schizophrenia, for the purpose of early identification and clinical interventions.

3.2 Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook

Schizophrenia and other psychotic disorders can be associated with significant impairment [175]. Although the majority of patients with first-episode psychosis initially achieve clinical remission of hallucinations and delusions, up to 80% experience at least one relapse within the first 5 years [176]. Each new episode can be associated with costly emergency room visits, psychiatric hospitalizations, family burden, medical complications, legal issues, and suicide [177, 178]. There is substantial evidence, suggesting that psychotic symptom exacerbation is preceded by periods of anxiety, low mood, sleep pattern irregularity, trouble concentrating, social withdrawal, strained interactions with others, altered psychomotor activity, and attenuated psychotic symptoms [179, 180]. Clinical interview, patient self-report, and family observation remain the primary sources for gathering early warning signs [181]. Unfortunately, the utility of these strategies is severely limited by the need for direct, frequent, and timely contact with trained professionals, as well as accurate and insightful patient and family recall. Continuous, objective monitoring of burgeoning psychotic symptoms could facilitate the initiation of early and proactive relapse prevention strategies [182, 183].

The dramatic rise in social media use could provide an opportunity to inform early relapse identification. There is compelling evidence suggesting that subtle changes manifest in social media activity before they become clinically apparent, providing the potential for earlier identification and intervention. Changes in social media-based linguistic and behavioral activity, for example, have been shown to reliably predict future episodes of depression [7], postpartum mood disorders [184], binge drinking behavior [174], and self-disclosures of schizophrenia [58, 185] with high degrees of accuracy.

Although promising, this line of research is limited by the fact that it has been conducted primarily using publicly available social media data, has relied largely on anonymous self-disclosed or self reported diagnoses of mental illness, and has rarely been validated for its theoretical and clinical grounding and validity [186]. Importantly, in order to make clinical use of social media data, it is crucial that these initiatives include collaborations with mental health clinicians, using data from known patients with confirmed diagnoses. There are currently few studies that combine the expertise of both computer scientists and mental health professionals to assess the generalizability and robustness of these data and machine-learning models built on them, in clinical contexts.

We conduct an ecologically valid investigation into the relationship between social media activity and behavioral health. Specifically, **we aimed to identify and predict early relapse warning signs in social media activity collected from a cohort of individuals receiving psychiatric care for schizophrenia and other primary psychotic disorders**. To achieve this goal, we tested a machine-learning model to predict relapse events by differentiating temporal periods preceding hospitalizations for symptomatic exacerbations from periods of relative health. The model leverages patient Facebook data and dates of hospitalizations from their medical record, and was designed to make predictions at an individual level, consistent with a personalized approach to medicine [187].

Human-centered algorithmic design: Our methodology is grounded in clinical theory about schizophrenia relapse experiences and the practical use of an algorithm predicting relapse for patient-provider interventions. The modeling approach is further guided by iterations with clinical researchers, psychiatrists and clinical psychologists. Below we discuss the considerations that went into our modeling approach.

First, a key challenge in predicting relapse hospitalizations is the relative rarity of these events compared to periods of health, causing a class imbalance when binary classification approaches are adopted. Further, while most periods of relative health are similar, each relapse hospitalization can be unique, even within the same individual [188, 189]. Therefore, recognizing each patient is different is an important consideration. Lastly, when clinicians want to use this information from a prediction algorithm, their focus is less on what is indicate of relapse in general (information provided by discriminative machine learning models), but on the likelihood of imminent relapse for a specific patient they are treating.

With these considerations, we adopted supervised anomaly detection techniques – specifically one-class classification algorithms for prediction [190, 191], which distinguishes between "normal" and "anomalous" observations [192]. This methodological framework can enable efficient intervention by predicting anomalies or exacerbations indicative of relapse in a personalized manner based on learned patterns of behaviors during healthy periods.

We compiled hospitalization dates and Facebook archives from 110 consenting participants with a psychotic disorder. Using the hospitalization dates as markers, each participant's Facebook data was segmented into periods of relapse and periods of relative health. The one-class classification algorithm was then trained on periods of relative health to identify distinguishing patterns of inliers. The best performing model was then tested on an unseen sample of both periods of relapse and relative health with the goal of predicting healthy periods as inliers and relapse periods as outliers (Refer flowchart in Figure 3.2). We assessed the validity of the model on patient-specific predictions based on the inferential ability (specificity, sensitivity). Finally, we conducted an error analysis by accessing data from medical records to understand the specific instances of mislabeled data or incorrect predictions by the model.

3.2.1 Data and Methods

Participant Recruitment: Participants between the ages of 15 and 35 years old who had been diagnosed with a primary psychotic disorder screened for eligibility from Northwell Health's inpatient and outpatient psychiatric departments. Most were recruited from the Early Treatment Program (ETP), Northwell Health's specialized early psychosis intervention clinic. Individuals with secondary psychiatric comorbidities were included. Eligible participants were approached by a local research staff member and offered the opportunity to participate. Recruitment occurred between March 2016 and December 2018. The study

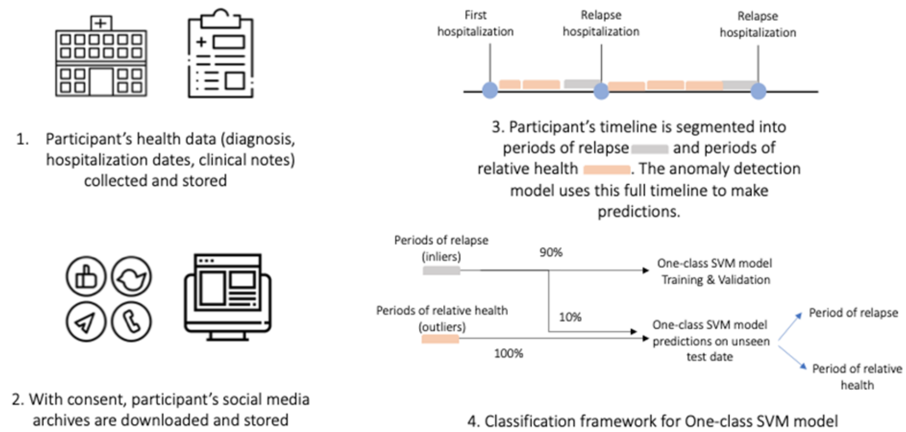


Figure 3.2: Flowchart of the relapse prediction machine learning methodology.

was approved by the Institutional Review Board (IRB) of Northwell Health (the coordinating institution), as well as local IRBs at participating sites. Written informed consent was obtained locally for adult participants and legal guardians of participants under 18 years of age. Assent was obtained for participating minors. None of the participants were involved in intervention research and all were receiving treatment as usual.

All participants were asked to extract their Facebook archive by logging on to their Facebook account and requesting their history accessible in their settings. Participation involved a single visit at the time of consent during which all historical social media data was downloaded and collected. Archives include all uploaded content (comments, messages, shares, likes, photos, etc.) since account creation. All user-generated social media content and activity was available for analyses. Clinical data including dates of hospitalizations and diagnoses were obtained through medical records.

Data description. A total of 52,815 Facebook posts (mean = 71.08, SD = 366.78) were collected across 51 participants (mean = 71.08, SD = 366.78) who had been diagnosed with a primary psychotic disorder (mean age = 23.96 years; 70.58% male) and had at least one relapse hospitalization. There was an average of 2.4 relapse hospitalizations per participant with a median hospitalization stay of 13 days.

Each participant's Facebook timeline data comprising self-generated posts from the day of the first hospitalization to the day of most recent hospitalization for a relapse was segmented into temporal periods (Refer Figure 3.2). Using the hospitalization dates per participant as markers, temporal periods 1 month prior to a relapse hospitalization were labeled as periods of relapse, as we expected to see symptom exacerbation most distinctive closer to the hospitalization. Excluding the 1-month preceding a relapse hospitalization, all other time periods were considered periods of relative health and representative of a person's baseline behavior. Healthy periods were segmented at varying granularity ranging from 1 to 3 months to understand the tradeoffs between availability of data and performance of the model (Refer to Figure 2 in [193]).

Classification framework. We built three models based on each of the data configurations described above: 1-month model, 2-month model, and 3-month model.

Preparing training data: For the 1-month model, inliers correspond to 1-month temporal periods of relative health (n = 719) and outliers correspond to 1-month periods of relapses (n = 49). For the 2-month model, inliers comprises 2-month temporal periods of relative health (n = 421) and outliers comprises 1- month periods of relapse (n = 49). Finally, for the 3-month model, inliers comprises 3-month temporal periods of relative health (n = 312) and outliers comprises the same 1-month periods of relapse (n = 49). The training data used for the three models (1-month, 2-month, and 3- month) overlapped.

Preparing unseen test data. Each of the three models was trained on 90% of the inliers and the remaining 10% of inliers alongside 100% of the outliers were held out as unseen data to test the classifier. Therefore, the held out data for the 1- month, 2-month, and 3-month model comprises 72, 42, and 31 periods of relative health, and 49 periods of relapse.

Features. We used linguistic features such as word usage (through an n-gram language model) and psycholinguistic attributes (via LIWC) [194] as a rich body of literature has identified associations of these attributes to emotion and behavior, including mental health states [195]. To capture structural aspects of language in social media, we used linguistic

Table 3.3: Class distributions and model performance on unseen test data for the one-class SVM models. PPV stands for positive predictive value and NPV indicates negative predictive value.

	#Periods of relative health	#Periods of relapse	Sensitivity	Specificity	PPV	NPV
1-Month model	719	49	0.47	0.65	0.66	0.46
2-Month model	419	49	0.57	0.28	0.41	0.44
3-Month model	312	49	0.90	0.04	0.37	0.4
Ensemble model	719	49	0.38	0.71	0.66	0.44

readability, word repeatability, and word length as features to the model (details in Supplement [193]). To capture behavioral measures on social media, providing insight into social functioning, diurnal patterns, sleep, and interests, we extracted volume and timing of posts, and Facebook activities such as check-ins, co-tagging, liking, sharing content, and using third-party apps. We applied a feature selection method based on the coefficient of variance [196] and filtered a final set of 79 features (details in Supplement [193]).

3.2.2 Results

Exploratory analysis. Comparing linguistic and behavioral features during periods of relative health to periods of relapse, randomly sampled per participant, identified significant differences across several categories (Refer Table 2 in [193]). We observed increased usage of words belonging to the anger ($p < 0.001$, Wilcoxon signed rank test), death ($p < 0.0001$), swear ($p < 0.0001$), negative affect ($p < 0.001$), hear ($p < 0.0001$), and feel ($p < 0.01$) categories during periods preceding a relapse hospitalization. We also observed an increased usage of pronouns during the period preceding a relapse hospitalization, including first-person plural ($p < 0.0001$) and second-person ($p < 0.01$) compared to periods of relative health. Among the social media activity-based features, we observed an increase in cotagging ($p < 0.001$) and friending ($p < 0.0001$) behaviors, as well as heightened posting activity between 05:00 a.m. and 12:00 p.m. ($p < 0.01$) and between 22:00 p.m. and 05:00 a.m. ($p < 0.01$) prior to a relapse hospitalization. Additionally, we observed significantly decreased use of words belonging to the work ($p < 0.01$), achievement ($p < 0.05$), friends ($p < 0.0001$), body ($p < 0.01$), and health ($p < 0.0001$) categories during periods of relapse.

Machine-learning model to predict relapse events. We built three one-class support vector machine (SVM) models [191] for three different data configurations: (1) periods of relapse and periods of relative health as 1-month temporal periods (1-month model), (2) periods of relapse as 1-month temporal periods and periods of relative health as 2-month periods (2-month model), (3) period of relapse as 1-month temporal periods and periods of relative health as 3-month periods (3-month model). A 1-month relapse period was selected as it represents a period of time prior to hospitalization during which early relapse warning signs typically become clinically apparent [197, 198]. Each one-class SVM model is trained on temporal periods of relative health as inliers (positive class) and then tested on an unseen sample of both periods of relapse (outliers/negative class) and relative health (Table 3.3). We then compared the performance of all three models based on their sensitivity and specificity. We found that the 1-month model had the highest specificity of 0.65 when compared to the 2-month or 3-month model (specificity of 0.28 and 0.04, respectively). This affirmed our expectation that behaviors characteristic to relapse would be most dominant during the 1-month period preceding a relapse (closer to the hospitalization). On the other hand, we found that the 1- month model performed worst in correctly predicting the healthy periods (sensitivity of 0.47) when compared to the 2-month or 3-month model (sensitivity of 0.57 and 0.90, respectively). This trend shows the trade-off between availability or volume of data and predictive performance revealing that incorporating longer periods of relative health (higher volume of data) helps in correctly predicting healthy periods but the performance on relapse prediction worsens. Given that the goal of this initiative was to predict relapse, and the clinical value in identifying symptomatic exacerbations, we emphasized the significance of specificity over sensitivity.

In order to improve the performance of the 1-month model, we built an ensemble one-class support vector machine algorithm (details in Supplement [193])). Ensemble methods are algorithms that combine multiple machine-learning models into one to reduce errors and decrease variance in predictions. The ensemble model was trained on 90% of 1-month periods of relative health as inliers and was tested on an unseen sample of 10% 1-month periods of relative health and all of the periods of relapse. The model predicts whether a given time period will have an adverse outcome such as relapse hospitalization. This ensemble model correctly predicted unseen relapse periods as outliers with a specificity of 0.71 and sensitivity of 0.38 (Refer Table 3.3). We find that the ensemble model performs better than the individual models in predicting periods of relapse with the highest specificity. However, the performance lowered in terms identifying periods of relative health.

Error analysis: evaluation via clinical chart review Given that the goal of the classifier is to predict periods of relapse, we conducted a deeper analysis of the misclassifications made by the model, specifically false negatives (periods of relative health wrongly predicted as a relapse). Note that the models consider periods of health as positive (inliers) and periods of relapse as negative examples (outliers). For each misclassified time period, two co-authors reviewed the accompanying clinical records. For 20 out of the 45 false-negative time periods (44%), data was available from the patient's medical record. In 18 of these 20 instances, the presence of psychotic symptoms during periods defined as relative health was documented, and six of these participants had known non-adherence to medication during this time which can contribute to symptomatic exacerbations [176]. Thus, of 20 periods for which symptom status could be verified from the medical record, 18 represented periods during which there was significant psychotic symptom exacerbation, even though the severity threshold necessitating hospitalization was not reached. There were also five instances that were incorrectly predicted by the model to be periods of relapse (false positives); however, a relapse hospitalization did indeed occur within the subsequent 2- month window or the participant was admitted into an intensive day treatment program. These periods may therefore represent true periods of relapse.

3.2.3 Discussion

This research aimed to identify early psychosis relapse warning signs from linguistic and behavioral features extracted from Facebook. With our machine learning approach, we have demonstrated that personalized methods to longitudinally forecast the likelihood of imminent adverse mental health outcomes, like a relapse event, is feasible. We believe this is a significant step toward the goal of leveraging social media activity to improve mental health services [199, 7, 43, 20].

We identified significantly increased use of words belonging to the swear, anger, and negative emotion categories in the period of time preceding a relapse hospitalization consistent with escalating irritability and depression known to be associated with emerging relapse [197, 200]. We also found increased use of words belonging to the hear and feel categories in the month preceding a relapse hospitalization, consistent with emerging perceptual disturbances, commonly experienced by individuals with psychosis [198, 200]. This is also consistent with prior work in those at risk for developing psychosis, suggesting that words related to auditory perception, such as voices and sounds, predicted conversion to psychosis [201]. Increased use of first-person pronouns may also be indicative of emerging self-referential thinking, a common psychotic experience contributing to delusions, whereby neutral environmental stimuli are perceived to be personally meaningful [202].

In addition to linguistic changes, we additionally identified several features that proved critical to our relapse classifier, including the total amount of friending, tagging, photo uploads, reposts, and likes, as well as nighttime posting, and information sharing in the late evening and very early morning. These features most likely represent digital representations of behavioral changes associated with escalating psychotic symptoms, including disruptions in sleep and circadian rhythm, disturbances in social functioning, and shifting interests and activities [203, 198].

Most research to date has focused on the association between objectively recorded smartphone sensor data, including geolocation, physical activity, phone usage, and speech and clinical state or symptom fluctuations [162, 204, 205]. Our results demonstrate that user-generated social media activity represents an equally critical source of digital data contributing to relapse identification.

Combining linguistic and behavioral features resulted in a classifier that predicted relapse with an accuracy of 71%, however, low sensitivity (0.38) limits the clinical utility of our model. Performance was likely impacted by our definition of relapse, which was defined as a hospitalization due to psychotic symptoms. Relapse, however is a complicated phenomenon, and has other definitions, including symptomatic exacerbations that do not result in hospitalization [206]. Furthermore, the decision to hospitalize is often multifactorial and may not always be a reliable indicator of psychotic symptoms. Our error analysis suggested that several periods believed to be incorrectly identified as periods of relapse did in fact have documented evidence for the presence of psychotic symptoms, although they did not necessarily result in a hospitalization. As we continue to explore digital manifestations of psychotic symptom exacerbation, researchers will need to identify models that have both high specificity and high sensitivity in predicting relapse. To be clinically useful, models will need to be capable of accurately predicting emerging relapse while avoiding false positives that would unnecessarily increase clinician burden and could negatively impact patient outcomes. False negatives could also be detrimental, particularly if clinicians relied on model prediction and failed to intervene in spite of concerning clinical changes.

CHAPTER 4

PATHWAYS TO SOCIAL CARE THROUGH SOCIAL MEDIA

In this chapter, I present the second theme of the thesis: pathways to social care for mental health through self-disclosure and social support on social media. Social media, in this context acts as an online space supporting stigmatizing experiences related to mental illnesses through self-disclosures, connectedness to similar others and social support. To understand how social media supports these pathways to care, it is imperative to address theoretical gaps in the understanding of benefits and outcomes of online self-disclosure and mechanisms through which large, invisible audience on social media provide social support.

I discuss two studies in this chapter to demonstrate pathways to social care for mental health on social media. In the first study, "Therapeutic outcomes of online self-disclosures of mental illnesses", I discuss a theoretically grounded approach that provides evidence that online self-disclosures of mental illnesses have indicators of therapeutic benefits similar to offline disclosures made to therapists or family members. Specifically, this work contributes linguistic markers from social media data that are indicative of therapeutic outcomes after disclosures. We found that when people make such sensitive disclosures online, they show indications of therapeutic outcomes such as improved readability and coherence in language, future orientation, lower self preoccupation, and reduced discussion of symptoms and stigma perceptions that are traditionally seen in face-to-face disclosures. However, self-disclosures on social media do not exist in isolation and occur within a networked context. As a follow up study, in "Audience Engagement and its impact on online disclosures of mental illnesses", we examined how a large, public, unknown audience on Twitter engages with self-disclosures of schizophrenia. This work demonstrates evidence of topical and temporal reciprocity in the engagement between disclosers and their audience. Further, using a time series forecasting technique, in this study, we show evidence that audience on social media impact future disclosure behaviors on the platform. Together this chapter uncovers the mechanisms through which individual behavioral changes around disclosures, and social support from an audience constitute social care for individuals with schizophrenia.

4.1 Therapeutic outcomes of online self-disclosure of mental illnesses

Self disclosure, a process of "making the self known to others" [81], is identified as an important therapeutic element in the achievement of physical and mental well-being [80]. In individuals experiencing conditions associated with high stigma, like mental health challenges, self disclosure is a widely adopted mechanism for coping. Historically, "opening up" and disclosing about mental health experiences has been an established phenomena in psychotherapy, an activity that is situated between a therapist and client [207]. In stark contrast to such dyadic disclosures to a carefully selected receiver (the therapist), today, social media platforms have emerged as new arenas for *"broadcasting self-disclosures"* [73, 31]. The concept of broadcasting self-disclosures refers to sharing personal, sensitive information in public contexts, often to invisible audiences [208], and supported by the affordances of anonymity or semi-anonymity in these platforms [83]. Being quite distinctive from dyadic disclosures, whose prominent goal is relational development and deriving therapeutic benefits, broadcasting self-disclosures have impression management as a salient goal [209], including alleviating inhibitions [29], identifying confidants [210], building trust and intimacy [73], and finding a mechanism for emotional release [147].

Despite the pervasive adoption of these new broadcasting self-disclosure practices, particularly around stigmatized mental health concerns [30], how these disclosures lead to behavioral changes on social media platforms, and if they help an individual meet their therapeutic goals, are less explored. Recent research has studied mental health disclosures shared on social media platforms such as Reddit and Instagram, exploring the ways in which linguistic attributes such as affect, cognition and linguistic style may reveal cues about one's psychological state [5, 83]. The attributes of support seeking nature of anonymous disclosures on Reddit among sexual abuse victims and depression sufferers has also been examined [1, 2]. Together, these works reveal how the unique needs around stigmatized mental health experiences can be met when one self discloses on a public platform like social media. We contribute to this line of research by examining how one of the most prominent goals of offline mental health disclosures, therapeutic benefits, that typically happen in offline therapist-client settings, translate to the context of online broadcasting self-disclosures. Specifically, we ask the question: ***Can we identify specific linguistic markers that indicate behavioral changes following disclosures of schizophrenia on social media? Are these changes indicative of any therapeutic outcomes?*** We address these questions in this study by drawing from the literature on psycholinguistics and the expressive writing paradigm [76].

Table 4.1: Clinician-contributed key-phrases for Twitter data collection

Diagnosed me with (schizophrenia / psychosis)
Diagnosed schizophrenic
I am diagnosed with (psychosis / schizophrenia)
I am schizophrenic
I have been diagnosed with (psychosis / schizophrenia)
I have (psychosis / schizoaffective disorder / schizophrenia)
I think I have schizophrenia
My schizophrenia
They told me I have schizophrenia
I was diagnosed with (psychosis / schizoaffective disorder — schizophrenia)
Told me I have (psychosis / schizophrenia)

4.1.1 Data

To obtain data on self-disclosures of schizophrenia as expressed on Twitter, we compiled a list of key-phrases indicative of self-reported diagnoses of schizophrenia, which serve as search queries. In consultation with two clinical psychiatrists, we used phrases listed in Table 4.1 as search queries on Twitter.

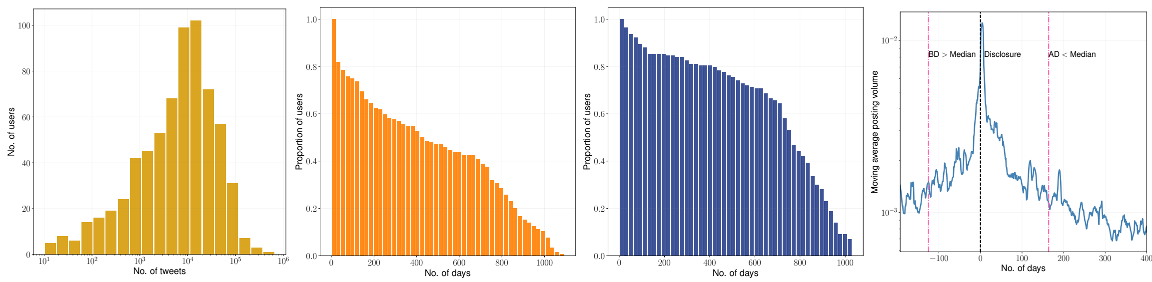
This resulted in a total of 21,254 posts authored by 15,504 unique users between 2012 and 2016. Since our research goal involves temporal analysis of Twitter content shared before and after the self-reported schizophrenia diagnoses, we selected disclosure posts and their authored users from the year 2014 (middle of time period of our collected data). For each filtered user, we extracted their Twitter timeline data from 2012 to 2016 using a web based Twitter crawler[1]. This timeline data for each filtered user included tweet text, username, posting time, hashtags, mentions, favorites, geo-location and tweet ID. We report basic descriptive statistics of this acquired data in Table 1 from [185].

Although the key-phrases involved first-person reports of schizophrenia experiences and diagnoses, several filtered tweets included noisy data in the form of disingenuous, inappropriate statements, jokes, and quotes. For example, note the tweet: "*I wish I had schizophrenia. So I can escape reality*". To obtain an accurate sample of genuine disclosures we designed an annotation task for expert (psychiatrist) validation on the authenticity of the disclosures. The final data that we used for our analysis is a sample of 146 users annotated to have made genuine disclosures (Class "Yes") by the annotators. For a detailed description of the annotation task, refer to Chapter 3 (section 3.1).

Compiling Timeline Data on Genuine Disclosures For each of the 146 users, the posting time of the disclosure tweet is taken as the *disclosure date*. To handle different disclosure dates by different users, we adopted an empirical approach by generating cumulative density functions (CDFs) of the number of Twitter posts shared by the users before and after their respective disclosure dates. These CDFs are shown in Figure 4.1(b) and (c). Based on these figures, we observe that most users (around 80%) have posts for at least 200 days before and 400 days after the disclosure dates, spread over 2014. Therefore, we choose each of the 146 genuine disclosure users timeline data spanning 200 days before and 400 days after their disclosure dates as a fixed length time period for analyses.

Matched Control Data To allow robust statistical comparisons between Twitter users

[1]https://github.com/Jefferson-Henrique/GetOldTweets-python



(a) (b) (c) (d)

Figure 4.1: (a) Distribution of number of users over number of tweets. (b) CDF of post distribution over the 146 genuine disclosure users preceding the disclosure dates. (c) CDF of post distribution over the 146 genuine disclosure users following the disclosure dates. (d) Temporal phases identified around disclosure using a moving average model of posting volume. The central vertical line indicates the disclosure event, while the vertical lines on its two sides indicate the boundaries of the BD and AD phases.

who choose to self-disclose regarding schizophrenia, and those who do not we collect data of matched control users. This also allows us to establish causation between the schizophrenia disclosures and the linguistic changes we seek to see preceding and succeeding them—statistical matching is a established technique to demonstrate causation in observational data, like in this study [211].

For each user who made a genuine disclosure on day d, as identified by the above expert annotation task, we identify a "matched control user" who had posted on Twitter, in the same year, on either of the days $d-1$, d, or $d+1$: this allows us to simulate a "control disclosure". Additionally, we ensure that the matched control user does not have any mentions of schizophrenia disclosures in their posts shared on their timeline. In this way, we compile the timeline data of 146 matched control users for the disclosure year 2014, and thereafter 200 days of pre- and 400 days post- control disclosure data for each of them. This resulted in 832,052 posts from the 146 matched controls, with a mean of 5699 posts (σ = 6984.25) per control user.

4.1.2 Methods

Identifying Temporal Phases around Disclosures Analyzing the behavioral changes that surround self disclosures of schizophrenia necessitates identifying data spanning pre- and post-disclosure phases where the symptoms of schizophrenia are most likely to be manifested. We draw from findings in clinical literature, specifically around the prodromal and active phases of schizophrenia(which denotes the beginning or complete manifestation of symptoms). We devise an approach to identify two phases in each (genuine) disclosing user's pre- and post-disclosure timeline data (compiled above) during which the symptoms of schizophrenia are most likely to be manifested: one preceding the disclosure (referred to as *"Before Disclosure"* or BD), and the other following it (*"After Disclosure"* or AD). Per clinical literature referred above [212], during these BD and AD phases, we expect the users to show markers of social withdrawal on social media. Abrupt declines in posting activity on social media are noted to be a sign of social withdrawal in prior work [7]. We utilize measures of changes in posting volume of an individual (normalized number of posts per day) as a way to identify these BD and AD phases around the genuine disclosures.

Using a median split method on the rate of posting behavior, we adopt the following day demarcations to define the BD and AD phases: BD = d_{-137} to d_{-1}; and AD = d_1 to d_{156}, assuming Disclosure = d_0 (Refer Figure 4.1 d)[2]. To allow meaningful comparison, we mapped these BD and AD phases to the extracted data of the matched control cohort as well, to obtain control BD and AD phases.

Linguistic Markers Around Schizophrenia Disclosures Using a theoretically grounded approach, drawing from prior work in clinical psychology and psycholinguistics, we quantify the linguistic markers of Twitter users around their disclosures.

Psycholinguistic Measures To quantify such psycholinguistic changes in the phases around disclosure, we use three categories of measures: (1) Affective attributes *positive and negative affect, anger, anxiety, sadness and swear*, (2) Cognitive attributes *cognitive mechanisms, discrepancies, inhibition, negation, causation, certainty, and tentativeness, see, hear, feel, percept, insight, and relative* and (3) Linguistic style attributes. We use the following four measures: (a) Function words: *verbs, auxiliary verbs, adverbs, prepositions,*

[2]We assume day 0 as the day of schizophrenia disclosure.

conjunctions, articles, inclusive, and exclusive (b) Temporal references: *past, present and future tense* (c) Social and Personal concerns: *family, friends, social, work, health, humans, religion, bio, body, money, achievement, home, sexual, and death* and (d) Interpersonal awareness and focus: *1st person singular, 1st person plural, 2nd person, and 3rd person pronouns*. All of the above measures are calculated based on the well-validated psycholinguistic lexicon Linguistic Inquiry and Word Count (LIWC) [172]. Using the textual content of posts of each user during the BD and AD phases respectively, we calculated the average LIWC score (normalized) per category during BD and AD phases.

Linguistic Structures Sentence structures and boundaries form an important aspect of written language [213]. We define measures of change characterizing linguistic structural attributes of Twitter posts spanning the BD and AD phases.

Readability. The relation between thought or meaning and forms of grammatical organization have been extensively studied as symptoms of schizophrenia [23]; individuals with schizophrenia use simpler grammatical forms in spoken and written communication, as well as exhibit a lack of spontaneity and fluency [214].

To capture this, we use the Coleman-Liau Index (CLI), a readability assessment test based on character and word structure within a sentence [215]. It approximates a U.S. grade level required to understand the text and is calculated using the formula: $CLI = 0.0588L - 0.296S - 15.8$, where, L is the average number of letters per 100 words of content and S is the average number of sentences per 100 words. In our case, the CLI is calculated from the day-wise aggregated content of posts by each user during the BD and AD phases respectively.

Stereotypy. Next, we consider two measures of stereotypic thinking in the posts shared by users during the BD and AD phases: (1) *Word repeatability*, and (2) *Word complexity*. Per the socio-cognitive model [213], sufferers exhibit signs of impoverished speech and content, word repetitions, decrease in usage of complex words or sentence verbosity, in favor of a greater number of simple ones [216].

In our data, we measure *word repeatability* by calculating the normalized count of non-unique words (or unigrams) in a Twitter post of a user during the BD or the AD phase, while *word complexity* is computed by estimating the normalized length of a word (or a unigram) in a disclosing user's posts during the BD or the AD phases.

Domain-Specific Content Measures Twitter is largely used as a microblogging platform where people share a wide range of everyday experiences and happenings. However, beyond the everyday experiences, individuals challenged with schizophrenia are likely to share content specific to their experiences of symptoms of the condition. E.g., over-representation of abstract and metaphysical termini or verbal abuse of death, power and hostility themes are known to have a strong bearing with the schizophrenic vision of the world [216]. To understand linguistic usage specific to the diagnosis or experiences of schizophrenia, we build a domain-specific lexicon from Reddit's online mental health communities. The lexicon consisting 1981 unigrams and bigrams relevant to schizophrenia was used to quantify differences in domain-specific content around disclosure. For each token in this lexicon, normalized occurrence is finally calculated per user during the BD and AD phases.

Topical Measures Discourse coherence disturbances like tangential responses, derailments and non sequitur responses are known to be related to language disturbances in schizophrenia [216]. We employ topic modeling [217], which is a useful, established approach to identify themes in data that are not captured by textual analysis at the level of tokens and sentences. To build a topic model, we run Latent Dirichlet Allocation using MALLET: MAchine Learning for LanguagE Toolkit[3], which has been an established method in prior work on mental health and social media [5].

Theme Variation. To identify thematic variation manifested in the Twitter posts of users around the disclosure event, we employ a qualtiative annotation task to combine LDA topics into broader interpretable themes. Then, we calculate the z-scores of the average

[3]http://mallet.cs.umass.edu/

probability of each theme per day across all users; this allows us to identify theme-specific variation manifested in the BD and AD phases.

Topical Coherence. For calculating the topical coherence measure during the BD and AD phases, we consider the topic distribution of a user's posts on day t, and compare it with the mean topic distribution over all posts shared by the same user in the previous week, i.e., days $t-1$ through $t-7$. Since we are comparing distributions, we employ the cosine similarity metric. Thus, a higher cosine similarity would indicate that the content shared on day t is topically coherent with respect to the same in the week before.

4.1.3 Results

Changes in Psycholinguistic Measures Table 4.2 gives a summary of psycholinguistic changes; we indicate the mean value of each measure in the BD and AD phases, their mean difference across the Twitter posts of all self-disclosing users, as well as the results of Wilcoxon signed-rank tests comparing the measures across the BD and AD phases.

Affective attributes. For the affective attributes, there is a significant increase in overall *negative affect* (mean difference 5.6%) and decrease in overall *positive affect* (mean difference 2.4%) right after the disclosure. This relates to the expressive writing literature, which associates an immediate increase in negative affect and decrease in positive affect after opening up about emotionally distressing topics rather than immediate relief of emotional tension [218]. The exposure to distress and confrontation of stigmatized conditions like schizophrenia might also implicate the increase in anger, sadness (mean differences 4.7%, 2.9% and 7.1% respectively). E.g., consider the paraphrased tweet: *"I'm sad sad sad sad"*.

Cognitive attributes. Among the cognitive attributes, we observe an increase in *certainty* words after disclosure, demonstrating heightened emotional stability indicative of the therapeutic nature of self disclosure. On the other hand, there is also an increase in *inhibition* (7.1% increase) which relates to the restraint and self-consciousness around disclosing

Table 4.2: Differences in psycholinguistic measures between the BD and AD phases, based on Wilcoxon signed rank tests. Only significant measures, following Bonferroni correction, are included.

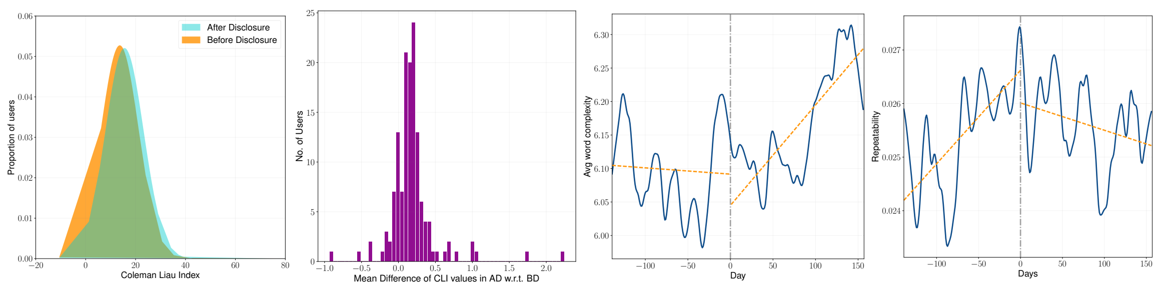
LIWC	BD	AD	t	p	Mean diff
Affective attributes					
Positive Affect	0.0410	0.0400	3048.0	***	-0.0243
Negative Affect	0.0095	0.0100	1832.0	***	0.0569
Anger	0.0124	0.0130	3086.0	***	0.0475
Anxiety	0.0029	0.0029	2086.0	***	0.0043
Sadness	0.0046	0.0048	1591.0	***	0.0298
Swear	0.0083	0.0077	904.0	***	-0.0715
Cognitive attributes					
Cognition					
Cognitive mech	0.0986	0.0963	3007.0	***	-0.0232
Inhibition	0.0035	0.0037	335.0	***	0.0710
Causation	0.0098	0.0093	508.0	***	-0.0497
Certainty	0.0094	0.0098	1003.0	***	0.0412
Negation	0.0162	0.0150	2888.0	***	-0.0696
Tentativeness	0.0150	0.0139	1836.0	***	-0.0769
Perception					
See	0.0076	0.0076	894.0	***	0.0023
Hear	0.0054	0.0051	1711.0	***	-0.0604
Feel	0.0058	0.0055	1864.0	***	-0.0538
Percept	0.0203	0.0194	3249.0	**	-0.0434
Insight	0.0150	0.0144	282.0	***	-0.0413
Relative	0.0869	0.0924	136.0	***	0.0637
Temporal References					
Past Tense	0.0194	0.0187	3600.0	*	-0.0338
Present Tense	0.0774	0.0739	436.0	***	-0.0453
Future Tense	0.0067	0.0068	1546.0	***	0.0034

LIWC	BD	AD	t	p	Mean diff
Lexical Density and Awareness					
Auxiliary Verbs	0.0681	0.0670	2205.0	***	-0.0158
Preposition	0.0725	0.0780	138.0	***	0.0750
Adverbs	0.0346	0.0357	905.0	***	0.0322
Verbs	0.1095	0.1076	1930.0	***	-0.0171
Article	0.0335	0.0335	566.0	***	0.0014
conjunction	0.0317	0.0313	2167.0	***	-0.0127
Inclusive	0.0212	0.0217	2801.0	***	0.0241
Exclusive	0.0173	0.0160	1800.0	***	-0.0739
Social/Personal Concerns					
Money	0.0037	0.0037	378.0	***	0.0125
Humans	0.0078	0.0076	618.0	***	-0.0299
Home	0.0030	0.0028	1535.0	***	-0.0582
Religion	0.0021	0.0022	1459.0	***	0.0344
Health	0.0081	0.0069	1792.0	***	-0.1478
Bio	0.0302	0.0286	592.0	***	-0.0551
Social	0.0670	0.0639	2826.0	***	-0.0469
Body	0.0109	0.0100	419.0	***	-0.0856
Death	0.0027	0.0029	3120.0	*	0.0887
Friends	0.0017	0.0015	1570.0	***	-0.1073
Achievement	0.0096	0.0098	458.0	***	0.0159
Work	0.0267	0.0273	2863.0	***	0.0246
Interpersonal focus					
1st p. singular	0.0347	0.0318	2096.0	***	-0.0853
2nd p.	0.0173	0.0165	3267.0	**	-0.0492
3rd p.	0.0064	0.0062	1414.0	***	-0.0288
1st p. plural	0.0029	0.0030	1369.0	***	0.0324
indefinite pronoun	0.0311	0.0306	3371.0	**	-0.0142

about a stigmatized condition (schizophrenia diagnosis) on a public social platform. For instance: *"Awkwardly waiting"*. Moving to the set of perception attributes, we observe that a majority of the measures show a decrease (e.g., *hear, feel, percept, insight*), characteristic of the emergence of a personal narrative writing style following a disclosure [76].

Linguistic style attributes. Finally, among the linguistic style attributes, we see variations in pronoun usage before and after disclosure (*1st person singular, 1st person plural, 2nd person* and *3rd person*) reflecting a transformation in the way people think about themselves in relation to others and the world. Following disclosure, individuals tend to show reduced self-attentional focus (mean difference for 1st pp. singular is -8.5%) as well as lowered social interactivity and orientation, as indicated by reduced usage of 2nd and 3rd pp (4.9% and 2.8% decreases respectively). Reduction in self preoccupation is a known attribute of improved psychological functioning [219]. Through greater use of 1st p. plural (mean difference 3.2%), we observe the emergence of a collective identity succeeding disclosures, which prior work has observed to be linked to therapeutic outcomes following psychological crises [220].

Social and Personal concerns. Among the attributes of Social and Personal concerns, an increase in usage of *achievement* words (mean difference 1.5%) indicates improved self esteem following engaging in disclosure of a stigmatized illness like schizophrenia. E.g., consider the tweet: *"I tried", "Mission success!"*. Additionally, according to the social cognitive behavioral model [213], "mastery experience", which involves providing an individual with ample opportunities to succeed, is an underlying positive health behavior change mechanism. This change may also show a tendency of the users to work towards goal-oriented activities following disclosure, which is often attributed to be a reduction in the symptoms of schizophrenia [23]. Further, there is a decrease in the *health*, *body* and *bio* categories (mean differences -14.7%, -8.5%, -5.5% respectively), signaling reduced self-consciousness of their wellness status or perceptions of their physical health. Reduced cognition of these topics is linked to improved therapeutics in individuals with mental illnesses [221]. Next, reduced use of *death* words indicates improved self-efficacy and the evolution of more positive attitudes towards life [222]. Finally, a negative mean difference in the usage of *social*, *home* and *friends* words (mean difference = -0.0469, -5.8%, -0.1073 respectively) reflects a detachment and isolation from the social realm after the disclosure, as also revealed earlier in the lowered use of 2nd and 3rd person pronouns. This may indicate a desire for the users to engage in solitude perhaps due to disclosing a stigmatized condition.

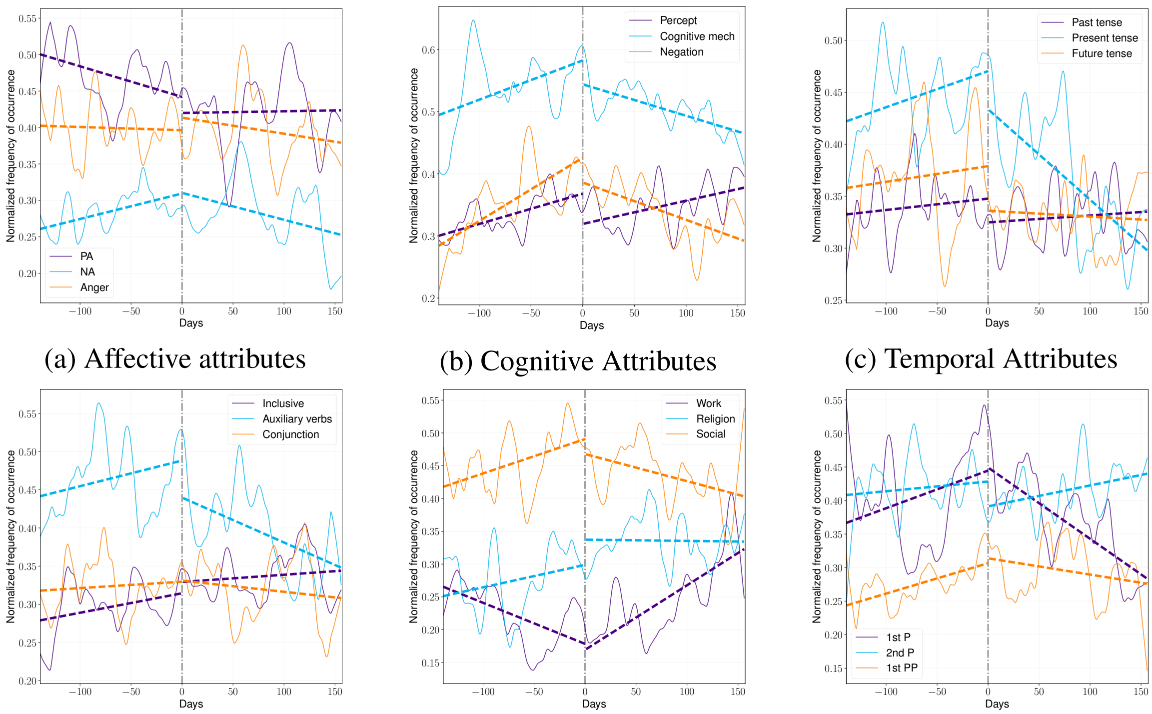


(a) (b) (c) (d)

Figure 4.2: (a) Distribution of CLI scores (readability) over number of users in the BD and AD phases. (b) Distribution of mean differences in CLI index (readability) in the AD phase, compared to the BD period. (c) Temporal changes and linear trend in the word complexity measure in the AD phase, compared to the BD phase. (d) Temporal changes and linear trend in word repeatability in the BD and AD phases. 0 indicates disclosure date.

Temporal Changes. Next, we provide finer grained temporal analyses of these psycholinguistic measures around the disclosure events. In Figure 4.3, we show the mean time series distribution of three psycholinguistic measures from each category; we pick a sample of the statistically significant measures. We overlay these time series data with their respective linear trends (based on a fitting polynomial models of degree 1).

Despite an overall decrease and overall increase in *positive affect* and *negative affect* respectively after disclosure, the temporal analysis shows an increasing trend in *positive affect* and decreasing trend in *negative affect* over time. This improvement in affect over time is identified as one of the long-term health benefits of self-disclosure. Additionally, schizophrenia sufferers are characterized by the inability to experience pleasure [216], and



(a) Affective attributes (b) Cognitive Attributes (c) Temporal Attributes

(d) Lexical density (e) Social/Personal Concerns (f) Interpersonal focus

Figure 4.3: Time series distribution and trend (based on fitting a linear model) of psycholinguistic attributes spanning the BD and AD phases. Selected statistically significant measures per Table 4 are shown. 0 indicates disclosure date.

therefore an increasing trend in *positive affect* may indicate a reduction in anhedonia, and improvement in overall functioning. Similarly, the increasing trend in usage of *work* related words also finds place as a long term behavioral outcome of expressive writing—following a sensitive disclosure, reference to multifaceted topics spanning one's everyday life is a known feature [216]. Further, we observe an increasing trend in first person plural pronouns, which relates to prior findings that among members of stigmatized groups, writing about being a group member changes the sense of self worth one derives from group membership. Finally, the decreasing trend in *auxiliary verbs* and self referential pronouns is indicative of lower self preoccupation.

Changes in Linguistic Structures To characterize structural differences in the language of Twitter posts before and after the disclosure (i.e., the BD and AD phases), we present an analysis of the three measures, *readability, word complexity*, and *word repeatability*.

Figure 4.2(a) shows a distribution of the mean difference in readability scores (AD compared to BD) over number of users, as measured by the Coleman-Liau index; individual distributions of users over CLI scores in the BD and AD phases are also shown in Figure 4.2(b). An overall positive mean difference with an average of 0.18 (σ = 0.32) is observed indicating an increase in readability after disclosing regarding one's diagnosis of schizophrenia. More elaborately, we find that overall 80% of the users show a mean CLI difference value greater than 0, indicating that largely, disclosing users show an improvement in the language framing limitations characteristic of people with schizophrenia, which is considered to be a therapeutic change [216]. Literature suggests that reorganizing and structuring traumatic memories or experiences helps in developing a complex and coherent narrative [223], suggesting that with these observed changes in our readability measure, disclosing users show evidence of reduced sentence framing limitations in the AD period.

Next, there is an increasing trend observed in word complexity as measured by the normalized length of words in the users Twitter posts (mean difference 0.01). This signifies an increase in usage of complex words and sentence verbosity moving away from the stereotypy symptoms of schizophrenia. Prior work says that people with the schizophrenia illness are limited in their ability to think with any degree of complexity. They are able to think in very simple terms, but generally are unable to solve complex problems, plan ahead or organize their thoughts [216]. The evidence of emergence of complexity in the language of Twitter users, therefore, reveals a reduction of an important negative symptom of schizophrenia. Finally, repeatability in terms of the proportion of non-unique words in Twitter posts has a decreasing trend after disclosure. Individuals suffering from schizophrenia often have repetitive thoughts that interfere with their ability to think, per the socio-cognitive model of schizophrenia [213]. Reduction of word repeatability is, therefore, likely indicative of lesser word repetitions and better articulation via language, as well as more concrete thinking and functioning among the disclosing users, a finding also observed in the case of the psycholinguistic measures [224].

Table 4.3: Theme keywords derived from topic modeling and human annotation analysis.

Theme	Topics	keywords
Mental illnesses	Topic 5, 12	mental, fighters, stigma, people, health, hospital, crazy, problems, doctor, illness, schizophrenia, donate, pndchat, pndhour, support, amazing, work, pnd
Symptoms	Topics 15,2,22,3,20	r/paranormal, r/ufos, r/creepy, ufo, house, ghost, ass, shit, fuck, lol, dick, fuck, shit, people, hate, stop, life, stupid, talking, god, hell, damn, holy, friends, anymore, love, jesus, hell, world, real, angel, christ, heaven, lord, soul, trust, fight, bless, sleep, night, bed, day, tomorrow, morning, work, time, tonight, hours, asleep, tired, today, sleeping, wake
Functioning	Topics 4, 7, 9, 10,	time, day, years, today, happy, week, times, past, months, minutes, eyes, face, back, hand, head, dear, lips, touch, felt, smile, deep, pain, care, anymore, people, time, whats, wrong, make, feel, hurt, wanna, love, talk, life, live, world, die, heart, mind, time, real, thing, true, words, end, rest, dream
Stigma	Topic 5	today, mental, fighters, stigma, people, drugs, health, days, hospital, left, bad, crazy, lot, real, problems, doctor, profit, illness, schizophrenia, donate

Changes in Domain-Specific Content Measures We observe that the usage of tokens related to symptoms and medication for schizophrenia such as *'experience hallucinations'*, *'voices really'*, *'meds work'* primarily appear only during the BD phase. Hallucinations, delusions, and paranoia are among the most distinctive negative symptoms of schizophrenia [23]. Usage of these tokens in the posts of the disclosing users, such as, *"i never sleep alone, hallucinations are troubling me"* and *"I miss hearing voices that tell me to stand in the rain at five in the morning"* reveal that prior to the disclosure, the users in our dataset were appropriating social media to engage in discourse on these topics and personal experiences.

Whereas, tokens related to treatment or help (*'going doctor'*, *'inpatient'*, *'seeking help'*), self-care (*'rehabilitation'*, *'self care'*) appear only after the disclosure. This indicates that, following disclosure, the users feel comfortable and less restrained in talking about their treatment experiences around schizophrenia; e.g., *"Everyone who was an inpatient with me at the hospital has moved on.."*.

Changes in Topical Measures Table 4.3 shows four major themes that appeared from the semi-open coding task involving two clinical psychiatrist annotators, the set of topics that define the theme and the most contributing words in each theme. The themes primarily appear to revolve around the clinical attributes of schizophrenia and are prevalent in both the BD and AD phases.

First, the theme "Symptoms" includes words such as */r/paranormal*, */r/ufos* and */r/creepy* which are Reddit communities for discussions about paranormal thoughts and activities. Together with terms like *ghost*, *ufo* and *house*, these words capture disorganized thinking and delusional attitudes which are notable markers of schizophrenia; e.g., as demonstrated in this tweet: *"An orange 'UFO' story from /r/Paranormal"*. Additionally, terms related to sleeplessness like *tired*, *sleep*, *waking* appearing in tweets like *"I'm tired, in pain, and cranky. Someone please make it stop, I swear to God."* are also established negative symptoms of schizophrenia. In fact, sleep disturbances and exhaustion, fatigue have significant impact on quality of life in individuals with schizophrenia [15]. Next, words like *jesus*, *god*, *holy*, *angel*, *heaven* reveal spirituality and religiousness, which are notable in schizophrenia sufferers [224]. We observe that shortly prior to the disclosure event, there is a reduction in the discussion of the theme, and it persists to be low for sometime in the AD phase as well. It may indicate improved functioning of the disclosing users; absence of the negative symptoms of schizophrenia are known to be linked to therapeutic outcomes [23].

Next, "Functioning" includes words like *happy, touch, smile, pain, hurt, die, care, wrong*; an example tweet says: *"im in a bad mood like im ready to either hurt myself or someone else"*. Typically, the socio-cognitive model of schizophrenia [213] indicates that reduced functioning, as is indicated by the words in this theme, is an important attribute of the schizophrenia experience, including behaviors such as neglect of social, emotional, physical, and cognitive aspects of life, as well as a lack in overall sense of purpose in an individual.

More generally, beyond schizophrenia related themes, we also observe the disclosing users to share content about other mental illnesses on Twitter such as around hospital visits, both during the BD and the AD phases. This might demonstrate comorbidity in the user group or expression of group identity related to stigmatized mental health conditions. For example, the words *pnd*, *pndhour*, *pndchat*, *stigma* point to a support community of people affected by postnatal depression. This theme shows a noticeable dip prior to the disclosure event, and then shows a stable, but overall reduced activity in the AD period. The higher values of this theme in the BD period might be due to the presence of the premorbid/active phases of schizophrenia prior to the disclosure event.

Finally, "Stigma" appears as a theme in itself comprising terms related to fighting the stigma of experiences of schizophrenia—*fighters, hospital, bad, doctor, illness*, also demonstrated in tweets like: *"my schizophrenia has gotten worse ever since I started living alone"*. We observe lowered stigma leading up to the disclosure event, although there is a peak right after the disclosure, which likely conveys the difficulty in disclosing about one's stigmatized conditions like schizophrenia.

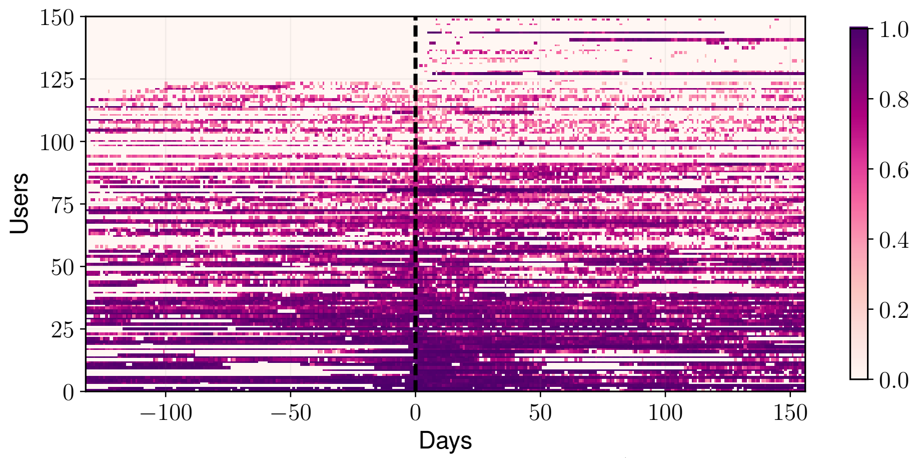
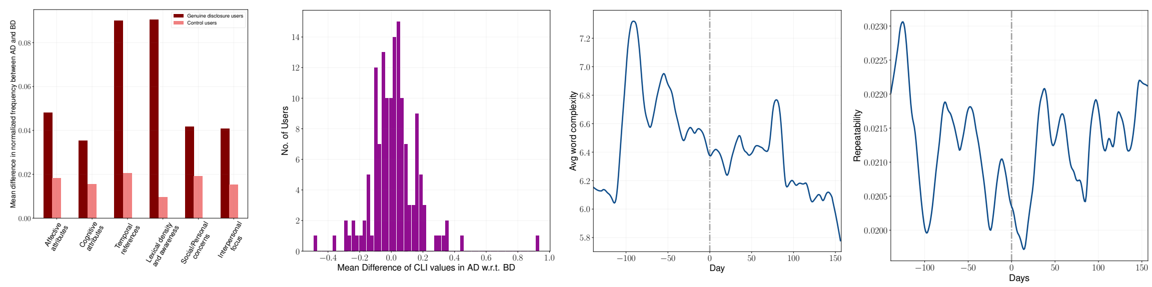


Figure 4.4: Changes in topical coherence in the BD and AD phases. Higher intensity cells in the heatmap indicate higher topical coherence. 0 indicates disclosure date.

To examine changes in topical coherence before and after disclosure, the average coherence value per user, per day is plotted as a heatmap in Figure 4.4. The figure reveals a gradual increase in topical coherence for most users following their disclosure of the schizophrenia diagnosis or experience. This result is also found in psychotherapy literature [222, 225] and the expressive writing paradigm [76], where self disclosure is known to help people organize and remember events in a coherent fashion while integrating their thoughts and feelings. The therapeutic nature of self disclosure are noted by Joinson [79] also identifies an increase in coherency and articulation after disclosing and confronting difficult experiences.

Comparison with Matched Controls Figure 4.5(a) shows the mean relative differences of psycholinguistic attributes for the genuine disclosure and control users, spanning the BD and AD phases. Over all the categories, we observe a greater change in psycholinguistic measures for the disclosed user group as compared to the control group. For example, lexical density and awareness attributes (e.g., adverbs, auxiliary verbs, prepositions), affective attributes, and attributes of interpersonal focus (e.g., first person singular) showed the largest change in the AD phase compared to the BD phase for the genuine disclosure group; however for the control group, the change across the time phases was minimal. In fact, based on independent sample t-tests (that adopted Bonferroni correction), we observe that the changes in the case of the genuine disclosure group were statistically significant across the different attributes, compared to the control cohort ($p < 10^{-15}$). Further, we notice minimal changes in the linguistic structure measures for the control group. The mean difference in readability (CLI measure) between AD and BD phases was 0.18 for the genuine disclosure group; whereas, the control group had a difference of 0.018. This is shown in Figure 4.5(b). Similarly, changes in word complexity, and word repeatability measures for genuine disclosure group were 5%, 22% greater in magnitude when compared to the control group.

Together, these observations show that the patterns of linguistic differences we observe in the case of the individuals disclosing schizophrenia on Twitter, can be attributed to the disclosure event itself, since such changes are absent in individuals who do not engage in a similar disclosure.



(a) (b) (c) (d)

Figure 4.5: (a) Comparison of LIWC differences BD, AD of genuine disclosure users with matched control users. (b) Distribution of mean differences in CLI index (readability) in the AD phase, compared to the BD period (c) Temporal changes in the word complexity measure in the AD phase, compared to the BD phase (d) Temporal changes in word repeatability in BD and AD phases. 0 indicates disclosure date for the matched users.

4.1.4 Discussion

Based on this analytical methodology, our results indicate significant behavioral differences before and after the disclosures, many of which align with known markers of reduction in the negative syndromes of schizophrenia. As a way to establish causation, we observe these differences to be minimal in a matched control group. Specifically, we find that following disclosures on Twitter, individuals express lowered stereotypy such as word repetitiveness, and demonstrate improved readability, linguistic complexity, and topical coherence in the content shared on Twitter. They also show greater future orientation and an increasing positive affect trend, as well as lowered attention to the self. Interestingly, following disclosures, the individuals tend to engage in reduced discussion of symptoms and stigma perceptions on social media. Situating our analyses within the socio-cognitive model of schizophrenia [213], the expressive writing paradigm [76], and feedback from clinician experts, these observations characterize the prominent effects of “opening up” about a stigmatized condition like schizophrenia. Summarily, this work signals therapeutic outcomes following disclosures made on a public platform like Twitter.

4.2 Audience engagement and its impact on online disclosures of mental illnesses

A variety of motivations and intents underlie people's decisions to self-disclose stigmatizing experiences like in the case of mental health challenges. One established reason is that people need 'sympathetic others', as Goffman (2009) posited: those who share the same social stigma, have had similar experiences, and those who "*share with him the feeling that he is human and 'essentially' normal in spite of appearances and in spite of his own self doubt*". The sympathetic others in an online social platform can, however, be varied. On platforms like Reddit, where there are dedicated support communities for mental health challenges, the others are often experts and peers with similar experiences. On social networking sites like Facebook, the others are likely social ties embedded in the offline context. Yet, "broadcasting self-disclosures" refer to sharing personal, sensitive information in a public social media context such as Twitter, to somewhat nebulous, less defined others [209].

Unlike online support communities, even if the disclosing individual has a mental conceptualization of their audience [226], they are likely to be 'invisible' and large, consisting not necessarily of experts or of peers undergoing similar experiences, but perhaps a wide variety of people with different backgrounds, interests, identity profiles, and purposes of social media use. Unlike social networking sites, the audience might also largely comprise weak ties [227]—those that the individual might not know or ever encounter offline.

Initially, disclosure of sensitive, stigmatized mental illnesses to such an invisible or even imagined audience can seem puzzling. However, the prevalence of the phenomenon, as shown in prior work [228, 185], suggests that the discloser might gain certain social benefits from such an audience. ***How can we better understand these audience, the ways they engage with stigmatized content, and the manner they impact the disclosure process on an otherwise general purpose, social media platform?*** Addressing these questions will help us understand the social benefits a discloser derives over time by continuing to disclose

Table 4.4: Descriptive statistics of disclosers & audience data.

Number of disclosers	395
Total tweets of disclosers	1,491,623
Mean tweets per discloser	3776.26
Mean tweets per day per discloser	17.48
Median tweets per discloser	1338
Distinct number of retweets audience	124,630
Distinct number of favorites audience	169,041
Distinct number of mentions audience	80,090
Total number of audience	373,761
Mean distinct audience per discloser	1218.4

to this audience.

Building on this motivation, we present a quantitative methodology to understand audience and their engagement to stigmatized self-disclosures on Twitter. We focus on the following two research questions:

RQ1: *What are the patterns in which social media audience are engaging with the self-disclosing individuals?*

RQ2: *How does the audience engagement impact the future disclosure process? In other words, is audience engagement predictive to future intimacy of disclosures?*

4.2.1 Data

Twitter Data on Schizophrenia Disclosures

Employing the same data collection strategies from Section 4.1.1, we identified 579 Twitter users who engaged in self-disclosures of schizophrenia. To investigate the patterns of audience engagement around the Twitter content of these users we focus on an year long period of Twitter activity succeeding the users' self-disclosures. Over the year-long period, we found 395 users to have shared 1,491,623 tweets with an average of 3776.26 tweets per user and 17.48 tweets per day per user. We report a summary of these descriptive statistics in Table 4.4.

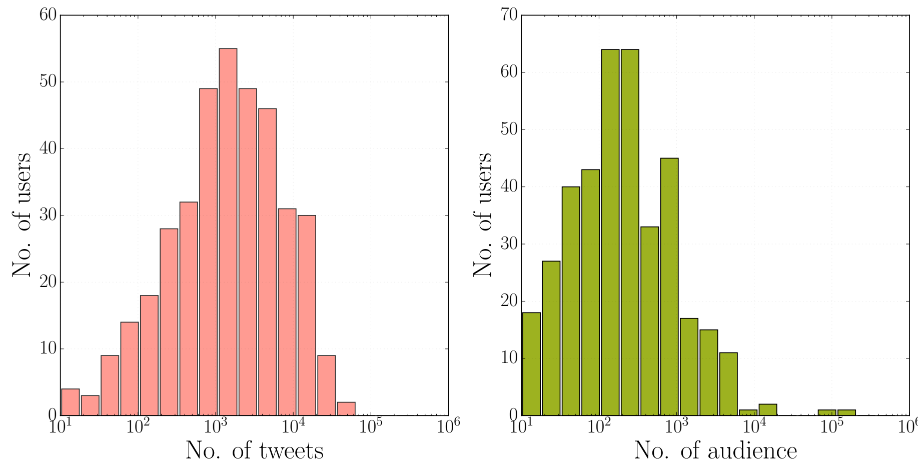
Definitions: Throughout this section, a '*discloser*' is an individual who has self disclosed (revealed) their diagnosis of schizophrenia by publicly posting on Twitter, on day d, the day of disclosure. The '*audience*' of these disclosures is the set of Twitter users who have interacted with the discloser's Twitter posts viz-a-viz the platform's functionalities—retweets, favorites or 'likes', mentions over the period of one year after day d. We operationalize '*audience engagement*' as any instance of such an interaction between a member of the audience and the discloser. Retweets, mentions, favorites or 'likes' constitute the various markers of audience engagement.

Audience and Audience Engagement Data Our audience engagement data collection proceeded by first collecting data on the various engagement markers surrounding the disclosers' Twitter content—retweets, favorites or 'likes', and mentions, and then compiling audience information from this data.

Retweets Data. We collected this data by identifying the Twitter users who have interacted with the disclosers by retweeting their content during the one year after disclosure. Across all 395 disclosers we obtain 124,630 distinct Twitter users (retweets audience) who retweeted the disclosers' content 2,895,118 times.

Favorites Data. We identified Twitter users who interacted with the disclosers' data through favorites (liking) during the one year after disclosure. Overall, we obtained a set of 169,041 Twitter users (favorites audience) who favorited the disclosers' content 4,592,890 times.

Mentions Data. We also collected data on those Twitter users who have interacted with the disclosers using the mentions (or @-replies) functionality. On Twitter, when an individual replies to another (say with username B), the tweet is automatically appended with the '@B' string. We used this stylistic convention of tweets to compile a list of search queries by appending an '@' symbol before the username of each of our disclosers. This provided us with all tweets that were incoming mentions to the disclosers including Twitter users who mentioned them and the textual content of the mention tweets. The final data consisted of 80,090 distinct users (mentions audience) who mentioned the disclosers in their 348,456



(a) (b)

Figure 4.6: (a) Distribution of #disclosers over #tweets. (b) Distribution of #disclosers over #distinct audience.

mention tweets.

Audience Data. To compile the final set of audience, we collated the list of users in the three datasets above— 124,630 retweets audience, 169,041 favorites audience and 80,090 mentions audience, and extracted overall 373,761 users. At a discloser level, on average, the audience size was 1218.4. The distribution of audience (size) and its descriptive statistics are provided in Figure 4.6 and Table 4.4.

4.2.2 Methods

RQ1: Characterizing Audience Engagement Per RQ1, we propose methods to characterize audience engagement around disclosers' Twitter data based on two attributes: the content of engagement and its markers.

Thematic Representation of Disclosers' Data. First, we develop a thematic representation of the data shared by the disclosers over the year-long period following their day of disclosure d. This representation is used to examine the dynamic interaction between the disclosers and their audience in terms of content sharing. We begin by employing topic modeling on Twitter timelines of our 395 disclosers. After preprocessing the tweets to remove URLs and stopwords, we run Latent Dirichlet Allocation using MALLET: MAchine Learning for LanguagE Toolkit. We perform hyper-parameter optimization over the sampling iterations to extract 30 topics. Using the topic model, we compute the topic distribution via posterior probabilities for each tweet.

Next, to identify semantically interpretable, broader themes from the 30 topics, we employed qualitative labeling. Two human raters who were social media and mental health experts performed semi-open coding on the extracted topics and combined them into semantically interpretable, broader themes. They also labeled whether or not each theme was related to the diagnosis and experiences of schizophrenia. We used the theme annotations and computed z-scores of the average probability of each theme per day across all disclosers.

Characterizing Engagement Content. Using the same topic modeling and qualitative theme annotation approach as above, we characterized the engagement content (i.e. dataset of the mention tweets), corresponding to each discloser.

Characterizing Engagement Markers. To characterize the engagement markers, we use the dataset of retweets, favorites and mentions received by each discloser per day during the one year period following day of disclosure. For each day d, ranging from $d = 0$ to $d = 365$, we find the average number of retweets, favorites and mentions received by all the disclosers and transform the average values into z-scores. This transformation gives us the variation in engagement markers received by the disclosers as a function of time and allows relative comparison. We obtain three time series, one for retweets, favorites and mentions from this step.

Discovering Patterns of Audience Engagement. To study the variations in engagement indicators (markers and content) with respect to that in disclosers' data, we make the following categorization. Based on the thematic annotations over disclosers' data and their corresponding engagement content, we categorize the theme labels into: themes related to the diagnosis and experiences of schizophrenia, and those unrelated. For both theme categories, we adopt time series comparison techniques (e.g., the cross correlation measure) to understand how the z-score distributions of the engagement markers and the themes of the engagement content vary with the disclosers' theme distributions over time.

RQ2: How Audience Engagement Predicts Future Intimacy of Disclosures

To begin, we describe how we operationalize intimacy of disclosures, and then propose and evaluate a time series forecasting model to predict these values accurately from the engagement markers and content.

Operationalizing Intimacy of Disclosures. To operationalize the disclosure process, we refer to the Social Penetration theory that models self-disclosure as a process of building intimate interpersonal relationships [229]. We adopt one of the measures proposed by the theory i.e. depth of disclosure or *intimacy* to operationalize disclosure in our work. The depth of disclosure relates to the degree of intimacy i.e. "how open or close someone can become with another person despite their anxiety over self-disclosure". In the context of mental health related self-disclosures on Twitter, depth of disclosure would denote the extent to which the discloser continues to share information about their experiences specific to their stigmatizing condition.

Given the lack of availability of ground truth data on disclosure intimacy and because discrete human judgments from a specific post may not be applicable across all users, to measure intimacy of disclosures from the textual content of disclosers' tweets, we use the following hybrid approach leveraging topic modeling and human annotations [5].

I. Manual annotation of disclosers' topics: Adopting the results from topic models built over disclosers' data as a thematic representation of their content (RQ1), we employed three human raters to analyze the top contributing keywords per topic and then label the level of intimacy disclosed via the topic. We defined the levels of intimacy to span a three-point Likert scale—low (1), medium (2) and high (3) motivated by prior work [230]. First, the raters manually browsed a sample of tweets by the disclosers to familiarize themselves with the content. Then, corresponding to this rating scale, they created a set of rules to annotate each topic with one of the three levels.

High intimacy of disclosure (score of 3). This included topics specific to the experiences of schizophrenia, information that is rarely expressed on a public social media platform like Twitter. For example, topics around symptomatic expressions, related to mental illnesses were included in this category.

Medium intimacy of disclosure (score of 2). This category included behavioral expressions related to functioning, social interactions, temporal planning that were not unusual to be shared on Twitter.

Low intimacy of disclosure (score of 1). This included topics that were totally unrelated to the disclosure of schizophrenia and consisted casual social media conversations such as political issues, entertainment, etc.

Following the manual annotation task, the raters had a high inter-rater reliability of 0.78 given by the Fleiss κ measure. Out of the 30 topics belonging to disclosers' data, this annotation task yielded 8 topics with high (3) intimacy, 7 with medium (2), and 15 with low (1) intimacy score.

II. Calculating tweet-level and time series measures of intimacy of disclosure. Given a tweet posted by the discloser, its posterior topic distribution given by the topic model (in RQ1), and the intimacy label (in RQ2) we calculate the intimacy of the tweet as a weighted sum of all topic probabilities by their intimacy labels to obtain a single score of intimacy of disclosure. We aggregate these tweet-level intimacy values per day and per discloser throughout our analysis period; we use z-scores of these aggregated values to capture their relative variation over time.

Predicting Future Intimacy of Disclosures from Audience Engagement. Given the intimacy of disclosure expressed by the disclosers and the associated engagement markers and content of the audience over time, we describe the prediction task as a time series forecasting problem. Since historical values of intimacy can also assist in predicting future intimacy values, we adopt an auto-regressive time series forecasting model. The dependent (or response) variable that is being forecasted is the time series representing daily measurements of intimacy of disclosure (obtained above). The exogenous variables (or predictors) are the engagement markers received from the audience as characterized by the following time series—number of retweets, favorites, mentions, and theme distribution of engagement content. Note that all timeseries are expressed as z-scores of average daily measurements of the variable.

Data Preparation. First, we process the data to verify stationarity assumptions of time series forecasting methods. We execute the following steps: 1) We apply a moving average transformation with a window size of 14 days to check for changes in the mean and variance over time. 2) We apply the Augmented Dickey Fuller (ADF) test, a standard test for stationarity in a series [231]. For the series that do not pass the ADF test, we apply a first order shift in the data and re-evaluate conditions for stationarity.

Model Fitting. We propose an Auto Regressive Integrated Moving Average with Exogenous Input (ARIMAX) model to predict the dependent variable (future intimacy) from the exogenous variables (audience engagement data). Our model is meant to forecast on day t, the intimacy of disclosure based on the exogenous variables spanning n days before t. We perform grid search over a maximum lag of 20 days for the autoregressive (p) and the moving average (q) parameters to find candidate models. Applying maximum likelihood estimation, we use log-likelihood, Akaike & Bayesian information criterion (AIC, BIC) measures to assess goodness of fit. We validate the final model by performing in-sample rolling predictions and assessing model performance using metrics like the root mean squared error.

Table 4.5: Theme descriptions obtained via topic modeling and qualitative annotations on disclosers' and audience's engagement data. n stands for number of topics per theme.

	Disclosers' Data		**Engagement Content (Audience)**	
Theme	n	*Top Words*	n	*Top Words*
MHSS	1	mental health depression illness pndhour anxiety mentalhealth issues submitted stigma today schizophrenia meds disorder cancer hospital support pain	2	hcsmca social support public info issue important system kids personal care health experience pndhour mental health support depression meds pain issues awareness illness anxiety story loss
Appearance	2	hair wear shirt white red clothes dress blue pants shoes fashion color back eyes head hand face softly arms neck lips smile kiss hair	2	hair wear red black dress nice clothes shirt blue body pants shoes back head eyes hand face neck smiles mouth softly cheek hugs arms lips butt
Functioning	4	love lot make time care talk anymore friends people dont life women social men thing good human work change money kids company job tax day sleep night week	3	good life hard work watch times thing love lot live make money pay food free people lot job low rich high business woman married relationship single engaged miracle divorced
Emotions	2	happy good hope today great beautiful amazing lovely year sweet make good feel bad people time life lot lol thought pretty today weird	4	care anymore worry hurt ill trust mad reason treat fuck person good bad feel life makes wrong find nice love wtf happy beautiful hope love talk fake
Sexuality	2	girl man guy hes shes sex cute love youre boy years baby dad friend mom gay woman child	1	lol girl shit girls youre fuck man ass hes funny fucking cool pretty shes guy weird guys cute
Symptoms	4	r/paranormal ufo r/creepy shit ass fuck bitch house back door night angels gods soul hell saved world	0	–
Temporal References, Planning	1	time day sleep work night today tomorrow back school home bed week hours days ill morning tonight ago gonna	3	night sleep time tomorrow week work today late hours home days morning year ago time long past day sunshine fab weekend friday
Communication	0	–	2	back text reply message lol tweet word tweets didnt haha talking forgot answer thought question funny doctor isnt meant english wrong swear correct
Others	14	cats dogg standwithrand tedcruz video football war government israel campaign police	13	law power gamergate superbowl stories club bro party school parents

4.2.3 Results

Comparing Disclosers' Themes and Audience's Themes. We present results from the thematic annotations on audience's engagement content and discuss them in the context of the themes derived from disclosers' data (Ref. Table 4.5). This juxtaposition of themes helps us understand the audience response with respect to what the disclosers' are sharing on Twitter.

First, among the engagement content themes that relate to experiences of schizophrenia, we begin by considering the theme "Mental Health Support/Stigma" (MHSS) that also surfaces in the disclosers' data. For instance, we notice the usage of words such as 'hcsmca', 'pndhour', 'awareness', 'issue' referring to online communities dedicated to exchanges around health care, mental illness and spreading awareness. also includes overlapping words like 'depression', 'anxiety', 'meds', 'mental health', 'pain', relating to the stigma and challenges around experiences of schizophrenia. This shows that the audience, in response to the schizophrenia content of the disclosers share their experiences and resources related to mental health care, providing solidarity.

Next, we consider another common schizophrenia related theme, 'Functioning'. We observe overlapping keywords, such as 'people', 'life', 'good', 'work', 'money', 'job', 'love', 'sleep'. Relatedly, we also find the theme 'Appearance' (words: 'hair', 'wear', 'red', 'clothes', 'arms', 'softly') that surfaces in the tweets of both the disclosers and their audience. Taken together, these themes relate to the everyday experiences capturing behaviors around the social, emotional, physical, and cognitive aspects of life. Their co-occurrence as themes reflects the utility of engagement content as a mechanism to converse about everyday aspects of life, communicate, plan, and exchange thoughts and ideas.

Lastly, we consider the theme 'Emotions' that appears in the engagement content with words like 'love', 'happy', 'good', 'hope', 'lovely', 'miss', 'sweet', 'beautiful'. While this theme is also present in disclosers' data, we note a higher prevalence of emotional content in the engagement content than that of the disclosers based on the number of topics contributing towards the theme. This particular imbalanced overlap characterizes the emotional support provisioning nature of the engagement that the disclosers gather from their audiences; a form of support found in the literature to be key to improved mental health state and in supporting therapeutic outcomes from disclosures of stigmatized conditions [232].

Nevertheless, despite the thematic reciprocity noted above, we note a sharp distinction between the tweets of the disclosers and audience—shown by the theme 'Symptoms'. In the case of the disclosers, this theme ('r/paranormal', 'r/creepy', 'ufo') reveals a predominant occurrence of words that have symptomatic relevance to schizophrenia. We do not observe such patterns in the themes extracted from the audience's engagement content. This indicates that, although the disclosers are sharing their first person experiences of the illness, the audiences do not respond with similar personal accounts. This brings to light the distinction in broadcasting disclosures on platforms like Twitter, where, unlike support communities, the audience need not necessarily consist of peers undergoing similar experiences.

By juxtaposing the thematic annotations from the disclosers and their audiences, we find evidence of reciprocal conversations around shared themes related to experiences of schizophrenia. We situate this discussion in the social penetration theory that gives a distinctive emphasis to self-disclosing behaviors being maintained by the "gradual overlapping and exploration of their mutual selves by parties to a relationship" [233].

Patterns of Changes in the Engagement Content. Here, we are interested in the question—how do the above (schizophrenia related and other) themes from the disclosers and the audience co-vary over time?

Inspecting Figure 4.7(a-d), we observe that there is a close temporal alignment between the disclosers' and the audiences' themes relating to schizophrenia experiences. Specifically, by analyzing the cross correlation between the two, we find that the highest correlation of 0.125 between the two time series occurs at a negative lag of 4. This positive corre-

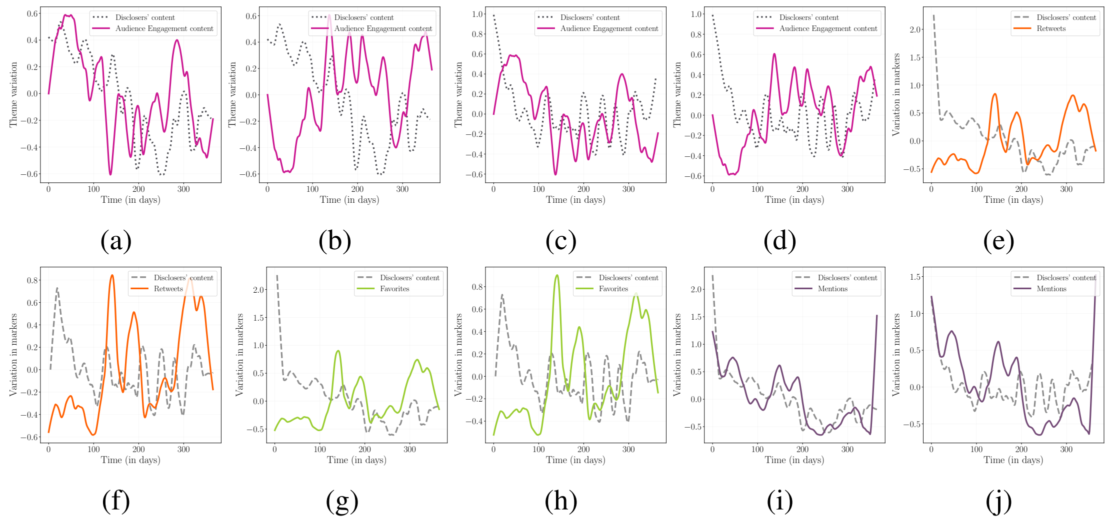


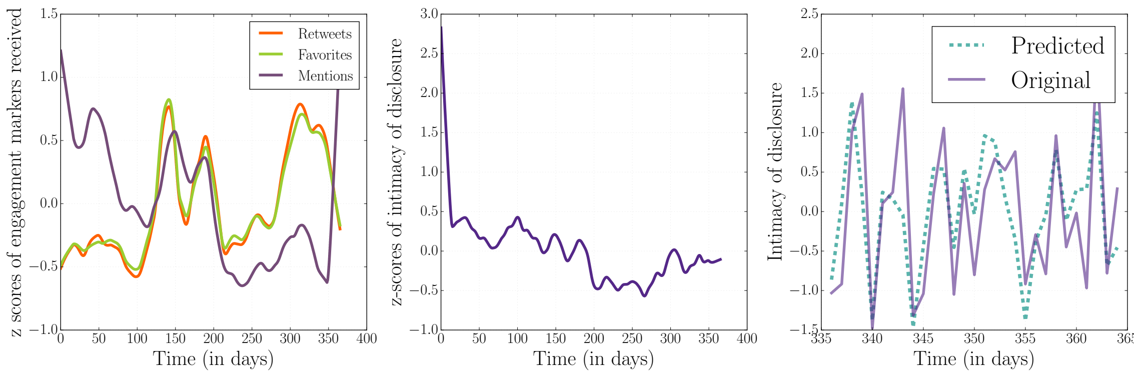
Figure 4.7: Patterns in audience's engagement content and engagement markers with respect to Disclosers' data. We show these patterns for 10 cases: (a) disclosers' data & audience engagement content both related to schizophrenia experiences; (b) disclosers' data related to schizophrenia experiences & audience engagement content unrelated to schizophrenia experiences; (c) disclosers' data unrelated to schizophrenia experiences & audience engagement content related to schizophrenia experiences; (d) disclosers' data & audience engagement content unrelated to schizophrenia experiences; (e) disclosers' data related to schizophrenia experiences & retweets; (f) disclosers' data unrelated to schizophrenia experiences & retweets; (g) disclosers' data related to schizophrenia experiences & favorites; (h) disclosers' data unrelated to schizophrenia experiences & favorites; (i) disclosers' data related to schizophrenia experiences & mentions; (j) disclosers' data unrelated to schizophrenia experiences & mentions. The discloser' data is plotted with the lag at maximum correlation.

lation at a negative lag provides indications of reciprocity in the disclosure process—as the disclosers increasingly talk about their schizophrenia experiences at time $t - 4$ (in days), it correlates with the audience talking about similar themes related to these experiences at t. Reciprocity has been identified as a major norm in self-disclosure research [234]. In contrast, we find that as the disclosers increasingly talk about their experiences, the audience begin limiting posts on other unrelated topics in the future (maximum correlation of -0.125 at a negative lag of 4).

Patterns of Changes in the Engagement Markers. We ask the question—how does the audience, with the help of various platform functionalities, respond to disclosers, and how do different engagement markers co-vary with disclosers' themes.

Figure 4.8a shows the z-score distribution of these markers over time. We observe two findings. First, beginning at the day of disclosure, there is a peak in mentions indicating an increase in incoming engagement from the audience. However, there is lowered audience engagement during this early period through retweets and favorites. This could indicate that the audience find the disclosers' content out of place and take time to modulate their engagement around it. Second, there is a very close alignment between the temporal variation in retweets and favorites received from the audience. This may be attributable to the similar functionality between both actions i.e. they both indicate some form of acknowledgement or endorsement, and have a lower barrier for content production (at the click of a button), compared to mentions which have a higher barrier to content production, requiring consciously drafted replies.

Next, in Figure 4.7(e-j), we present an analysis of the temporal variation in the three engagement markers in relation to the disclosers' themes—both the schizophrenia related ones as well as the rest. Upon visual inspection, we notice that the alignment between the daily measurements of engagement markers is higher with disclosers' data related to schizophrenia experiences as compared to other unrelated content. For the time series representing thematic variation in schizophrenia related experiences, the maximum correlation with retweets and favorites is -0.09, -0.08 observed at cross correlation lags of 5, 5 respectively. The negative correlation at a positive lag denotes that as the disclosers increasingly talk about their condition and experiences, it correlates to receiving fewer retweets and favorites in the days following. This is likely explained by the perception that the actions of retweet or favorite signal information sharing intentions and do not convey an appropriate response to stigmatized disclosures. On the other hand, we observe a stronger alignment between the disclosure content related to experiences of schizophrenia and the mentions received. The correlation of disclosure related content with mentions is the strongest with a lag 0 with a positive value of 0.17. This shows that as the disclosers increasingly talk



(a) (b) (c)

Figure 4.8: (a) Engagement markers over time. (b) Intimacy of disclosure, across all 395 disclosers' data over time. (c) Predicted and original measures of intimacy over time.

about their experiences, it correlates to receiving more mentions (on the same day). However, in the case of unrelated themes, we observe a delayed response via mentions from the audience (maximum correlation of 0.14 at lag -7). Summarily, our findings from RQ1 suggest reciprocity, temporally in the number of engagement markers received and topically, in the themes received viz-a-viz the audience engagement content.

RQ2. Impact of Audience Engagement Figure 4.8b shows the temporal variation in intimacy of disclosure, combined across all disclosers' data. We observe a peak representing heightened levels of intimacy of disclosure on the day of disclosure ($d = 0$) and the immediately succeeding days. Since topics related to the experiences of schizophrenia were rated with an intimacy score of 3, it appears that the short period immediately following the day of disclosure continues to include high intimacy content.

With this time series of intimacy of disclosure as our response variable, we proceed to results on the forecasting model.

We found the best lag order for the ARIMAX process i.e. the auto-regressive and moving average parameters to be p=8 and q=3. Including the differencing parameter $d = 1$ we fit an ARIMAX(8,1,3) model on the time series data (intimacy of disclosure, engagement markers and content) for forecasting. The goodness of fit of this model in terms of log-likelihood, AIC and BIC were found to be -351.9, 751.9 and 845.0 respectively.

Table 4.6 summarizes the ARIMAX model in terms of point mass estimates of the external variables, their 95% confidence intervals, and the corresponding p-values. We refer to this information, to examine the variables that provide the most explanatory power in the forecasting problem i.e. we ask what engagement markers and engagement content shared by the audience have high predictive power in forecasting future intimacy levels of the disclosers. We assess statistical significance here at the p=0.05 level.

First, we observe that the number of mentions received is a significant predictor of future intimacy. This affirms our previous findings that mentions indicate a strong incoming engagement in ways of conversing, sharing experiences and resources with the disclosers.

Next, we find two themes within the audience engagement content that are statistically significant to future intimacy levels. The first such theme is 'Emotions' with keywords such as 'care', 'worry', 'trust', 'life'. Emotional support received in cases of stigmatized conditions has been shown to help with coping and provide satisfaction in online support communities by previous studies [235]. Prior work has also linked intimacy to satisfaction with social support received during crisis [236]. This relates with our finding that emotional content received through audience engagement can be linked to intimacy and predict future disclosure behaviors. The second significant theme is 'Sexuality'. Discussions on one's sexuality are often considered to be sensitive in nature. When they happen on a public social media platform like Twitter, they indicate the audience's intent to reciprocally converse with the disclosers about topics that are otherwise personal. This reciprocity might also motivate the disclosers to reveal more intimate aspects of their illness experiences to their audience.

Finally, to validate the model, we compute in-sample rolling predictions for the model on an out-of-sample data over the last 30 days in our year-long period of analysis. Note that the ARIMAX model forecasts the differenced intimacy of disclosure and therefore, the predicted values are compared to the original differenced values of intimacy (Ref. Fig-

Table 4.6: Summary of point estimates of the exogenous variables in the intimacy forecasting ARIMAX model. Note that the estimates of exogenous variables in the model need to be interpreted conditional to the lags in response variable.

Exogenous variable	estimate	P>z	95% C.I.	
mentions	**-0.0266**	**0.014**	-0.048	-0.005
retweets	-0.0197	0.748	-0.140	0.100
favorites	0.0278	0.666	-0.098	0.154
themes:appearance	0.0031	0.868	-0.033	0.039
themes:communication	0.0022	0.893	-0.030	0.035
themes:functioning	-0.0182	0.411	-0.062	0.025
themes:emotions	**0.0581**	**0.0006**	0.025	0.091
themes:mhss	0.0156	0.408	-0.021	0.052
themes:sexuality	**0.0356**	**0.0306**	0.003	0.068
themes:temporal	0.0354	0.103	-0.007	0.078
themes:other	0.002	0.918	-0.036	0.039

ure 4.8c) We observe that our model is able to closely forecast the actual intimacy levels of disclosure. Assessing model performance, we find the Root Mean Square Error, Mean Absolute Error and Symmetric Mean Absolute Percentage Error measures as 0.66, 0.52 and 6.8 respectively. These values statistically establish the satisfactory performance of the model. As a final validation step, we check the residuals of the model for absence of serial correlation. We compute the Durbin-Watson statistic which tests for the null hypothesis that there is no serial correlation [237]. We find the test statistic (Durbin-Watson's d) as 1.8, which is close to the ideal value of 2 in case of no serial correlation.

4.2.4 Discussion

We began this study questioning the puzzling nature of stigmatized self-disclosures made to an invisible audience on a public microblogging platform. By characterizing the audience engagement towards disclosures of schizophrenia on Twitter, we found evidence of reciprocity, both topically and temporally, in the interactions between the audience and disclosers. We also observed that using the functionalities of favorites, retweets, and mentions, the audience is able to engage with the disclosers in a variety of ways: providing support, advice, and solidarity, sharing personal experiences and online help resources, and conversing about everyday aspects of life. While these attributes are key characteristics of online support communities, their occurrence on Twitter is revealing as it lacks many critical components of an online community such as norms, moderation, roles etc. Similarly, strong social ties are considered to be the hallmark of quality support and psychological wellbeing [238]. However, despite lacking many aspects of a social network [227] Twitter seems to be providing positive outcomes to individuals with a highly stigmatized condition, schizophrenia.

Further, we examined how audience engagement impacts future disclosure behavior, to understand if the disclosers gather interpersonal and social benefits through this public disclosure process. The results from our forecasting model demonstrate that key predictors, such as number of mentions, emotional support, and discussions on personal, sensitive topics can successfully forecast future intimacy of disclosures. This finding indicates that the disclosure process supports not only bridging social capital, that is, finding new acquaintances who provide access to new information and help resources, but also over time, in *bonding* social capital, in the form of reciprocity, support, and companionship [90]. Although the nature of audience providing these social capital resources is nebulous, i.e. the disclosers may not necessarily know *who* this audience is, even if they have an imagined mental conception of who it might be [226], the reciprocal engagement that the audience provides over time confirms observations about online social platforms facilitating formation and maintenance of social capital. Nevertheless, as argued in the literature [239], one might expect that disclosing about stigmatized, sensitive issues like mental illnesses to such an invisible and imagined audience might increase the likelihood of a context collapse that can hinder future disclosures. However, we find that, despite the risk of context collapse, the disclosers do not employ counteractive strategies, but rather continue to engage in schizophrenia related intimate exchanges with their audience over time.

CHAPTER 5

DIFFERENTIATING USES OF SOCIAL MEDIA FOR MENTAL HEALTH: A TRIANGULATION STUDY

In previous chapters, I have discussed the clinical and social pathways to care for individuals with schizophrenia through social media. For clinical pathways to care, advances in machine learning approaches combined with social media data from patient populations is shown to have the potential to predict adverse clinical outcomes like imminent relapse. For social pathways to care, affordances of social media platforms showed evidence of therapeutic benefits from online self-disclosures and social support and reciprocity from a large audience. So far, research examining social media and mental health has taken a monolithic view towards the participants in their study; they either belong to *patient* populations or are *users* of social media sites whose clinical status is unknown. *Who are all the people leveraging these pathways to care? Are they all clinically diagnosed with schizophrenia? Are the individuals whose social media traces are leveraged to predict clinical outcomes i.e. those who are clinically diagnosed patients, also using social media for disclosure and support seeking behaviors?* In this chapter, I present a triangulation study based on how social media data is used for the prediction of mental health states. Findings from this work call for a differentiating analysis of the uses of social media for mental health and highlight methodological gaps in the area of research.

For prediction of mental health states from social media data, on the methodological front, supervised machine learning techniques have gained prominence, providing promising predictive outcomes of mental health states [240]. The success of these techniques, however, hinges on access to ample and high-quality gold standard labels for model training. In mental health, gold standard labels often comprise *diagnostic signals of people's clinical mental health states*, for instance, whether *an individual might be suffering from a* *specific mental illness*, or *at the cusp of experiencing an adverse episode* like a relapse or suicidal thoughts.

Unlike conventional machine learning tasks in fields like computer vision and natural language processing, extensive, high quality gold standard data for predicting clinical diagnoses of mental illnesses from social media is not readily available. Literature has advocated the use of clinically validated diagnostic information collected from *patient populations* for building such predictive models [58, 240]. However, undertaking such efforts presents many practical and logistical challenges. These range from the difficulties in recruiting a sensitive and high risk population, to the myriad privacy and ethical concerns that accompany engaging directly with vulnerable individuals. Because of the effort- and time-consuming nature of such data acquisition approaches and the need for deep-seated cross-disciplinary partnerships, particularly with clinicians, researchers have noted such data acquisition efforts to not scale easily and quickly to large and diverse populations [43].

Consequently, researchers have operationalized a variety of *online behaviors* as diagnostic signals to build machine learning approaches that predict mental illness diagnoses. These "proxies" are easily accessible and inexpensively gathered from social media, without the need to directly engage with the individuals themselves. We define binary indicators of the presence or absence of these social media behaviors that might correspond to their clinical mental health state as ***"proxy diagnostic signals"***. One notable example from existing literature consists of public self-reports of mental illnesses made by individuals in their social media feeds [43, 8].

This study posits a significant challenge in using these *proxy diagnostic signals* revolving around their lack of clinical grounding, theoretical contextualization, and psychometric validity—concerns noted by psychiatrists and computational researchers alike [6, 241]. In other words, drawing on boyd and Crawford's critique [67], despite gains in scale, gaps exist in our understanding of how these signals are defined, where their theoretical underpinnings are, whether they objectively and accurately measure what they claim to measure (that is, the clinical mental illness diagnosis), and whether the patterns of behaviors they exemplify are truly representative of the behaviors of patients. [69, 242]. for treatment and patient-provider interventions.

5.1 Data

We use public and non-public data (gathered using appropriate protocols) from two prominent social media sites, Twitter and Facebook, for the purposes of this study. We begin by introducing four datasets used in the study, followed by a description of how they were collected.

Gathering Proxy Diagnostic Signal Data The first three datasets correspond to the three proxy diagnostic signals we adopt based on the topical focus and the existing literature, and which were introduced above. We consider them as proxies (or "proxy datasets") of schizophrenia diagnoses in individuals.

Affiliation Data. Our first dataset is motivated from prior literature that used behaviors signaling affiliation (e.g. following, hashtag usage) to mental health resources, related to schizophrenia, as diagnostic information. Adopting the approach of McManus et al. [20] ($N = 96$), we used a Twitter account named @sardaa (Schizophrenia and Related Disorders Alliance of America), a support organization for people with schizophrenia and their caregivers, as our starting point to build this affiliation dataset. As operationalized by McManus et al. [20] and following verification of the account's trustworthiness with our clinical coauthors, we considered all followers of the account @sardaa as individuals with a schizophrenia diagnosis. We obtained the list of all followers of @sardaa ($N = 1847$) and consistent with McManus et al. [20] collected their timeline data for the year 2014. We also collected profile information of these individuals including number of posts, chosen language on Twitter (filtering for English), number of followers and number of followees, leading to a final sample of 861 Twitter users. Descriptive statistics of this data are reported in [186].

Self-report Data. For the second dataset, we adopt the proxy diagnostic signal of mental illness self-reports utilized in many prior works (e.g., (most prominently [8]), introduced in the previous section. Per Mitchell et al.'s approach [8] ($N = 174$), we used a list of key phrases developed in Ernala et al. [243] (Refer Chapter 3) to identify self-reports of schizophrenia on Twitter from 2014. Following manual filtering to remove noisy examples, without loss of generality, we collected the historic timeline data of all authors of these self-reports. We also collected the same metadata information and descriptive statistics are reported in the [186].

Clinically Appraised Self-report Data. Our third proxy dataset is inspired from the third body of work that used external expert appraisals on social media data to obtain diagnostic signals of mental illnesses. Following Birnbaum et al.'s approach [58] ($N = 146$), we combined machine learning with clinical appraisals to obtain data of 153 individuals whose self-reports were labelled by experts to be genuine. As before, we collected all metadata associated with their Twitter profiles, descriptive statistics of which are given in the [186].

Matched Control Data The predictive task of identifying individuals with schizophrenia necessitates comparisons to matched control individuals who do not provide an equivalent proxy diagnostic signal. Accordingly, we used the Twitter streaming API to obtain a random sample of public posts and extracted their authors($N = 640$). We filtered out any individuals who had mentions of schizophrenia in their posts.

Then, we adopted a statistical matching approach [244] to ensure that the control users and the individuals in each of our proxy datasets are comparable by trait attributes. Since social media behaviors are a reliable indicator of people's personality, psychological states, and even demographic attributes [245], we included the following covariates for the purpose of matching: total number of statuses, chosen language on Twitter, total number of followers and total number of followees. Through an iterative k-nearest-neighbor matching (k=1-15) based on the well validated Mahalanobis distance metric [246, 247], we com-

Table 5.1: Descriptive statistics for the proxy diagnostic signal datasets and their corresponding matched controls.

	Affiliation Data		Self-report Data		Appraised Self-report Data		Gold Standard Patient Data	
	Target Class	Control Class	Target Class	Control Class	Target Class	Control Class	Target Class	Control Class
Total #users	861	539	412	345	153	107	88	55
Total #posts	1,417,688	2,145,319	1,724,237	1,083,790	663,428	233,253	9,821,938	4,958,793
Avg #posts	1646.56	3980.18	4185.04	3141.42	4336.13	2179.93	111,612.93	90159.87
Median #posts	320	1113.0	1682	830	1376	737	28554.5	21178.0

pared the covariates of each individual's Twitter content in each proxy dataset (affiliation, self-report, appraised self-report) with that of each of the control users obtained above, and identified a set of most similar control users based on a heuristically chosen distance threshold. For the affiliation dataset, we obtained a matched control sample of 539 users. We obtained 345 and 107 matched controls for the self-report and the clinically appraised self-report datasets respectively. The descriptive statistics of these matched controls are given in Table 5.1.

Schizophrenia Patient Data and Healthy Controls As the fourth dataset, we include social media data of patients clinically diagnosed with schizophrenia and that of clinically verified healthy controls, based on a clinical examination or DSM-5 [202] criteria (Refer to Section 3.2.1 for more details). The consented participants included 88 patients who had been diagnosed with schizophrenia. Of these 88, 73 participants consented to provide their Facebook data, whereas 15 provided their Twitter data. Additionally, 55 healthy controls were recruited through the study, out of which 32 provided their Facebook data and 23 participants provided their Twitter data. We use all linguistic content from participants' Facebook and Twitter archives i.e. status updates and comments made on Facebook, and posts shared on Twitter. We conducted linguistic equivalence tests between the two data sources, a known approach in the transfer learning literature [248], to quantify the linguistic similarity and a high value indicated that the content in the two datasets (cosine similarity of 0.98 for schizophrenia patient population and 0.84 for healthy control population) was linguistically equivalent [249]. Thus, our final dataset comprised either Twitter or Facebook archives of 88 schizophrenia patients and 55 healthy controls. The descriptive statistics for

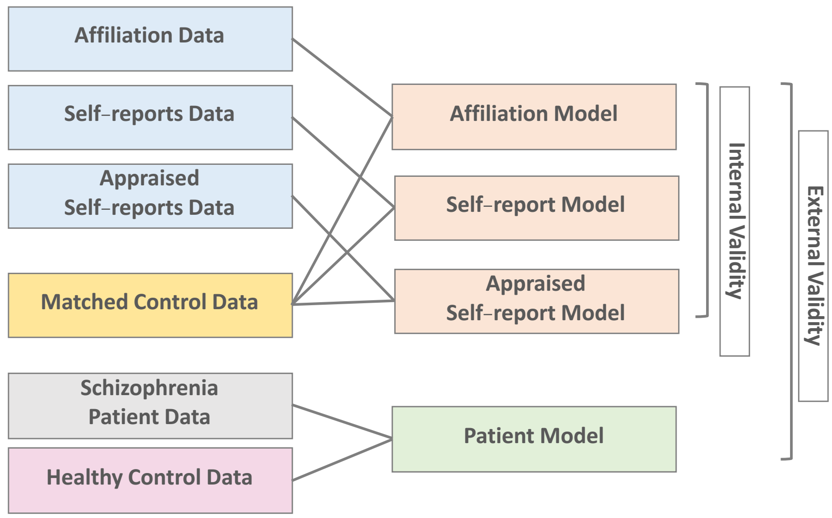


Figure 5.1: Schematic diagram of our proposed methodology.

the combined patient and healthy control dataset are reported in Table 5.1.

5.2 Methods

Rationale and Overview We adopt quantitative data triangulation as our methodological framework. Triangulation is an evaluation approach that uses multiple or heterogeneous methods, or data sources compiled via varied mechanisms, to develop a comprehensive understanding of a phenomenon, or to elucidate its complementary aspects [250]. Specifically, this approach is used to confirm the results of a research, and provide external validation to existing findings [251]. In essence, triangulation is an attempt to map out, or explain more fully, the richness and complexity of human behavior by studying it from more than one standpoint. Using this approach, we assess the efficacy of the three proxy diagnostic signals in identifying diagnoses of individuals with schizophrenia, both within their corresponding proxy datasets, as well in the data of schizophrenia patients. This way, we seek to establish their internal and external validity respectively. Figure 5.1 gives an overview of our approach.

Classification Framework We set up a binary classification task to distinguish between individuals with schizophrenia identified by each proxy dataset and its corresponding matched controls. We built four models: three based on the proxy datasets denoted as the *Affiliation*, *Self-report* and *Appraised Self-report Models* and one on the clinically validated patient data known as the *Patient Model*.

Preparing Training and Validation Data: We use the proxy datasets and their corresponding matched control data in their entirety for training and validating the above proxy classifiers. For the *Affiliation Model*, the positive examples (Class 1) comprised the Twitter data of the 861 users while the negative examples (Class 0) consisted of the 539 matched control users. The positive examples for the *Self-report* and *Appraised Self-report Models* spanned the data of 412 and 153 users respectively, while the corresponding negative examples included the Twitter data of 345 and 107 matched controls. For the *Patient Model*, we selected a random sample of 80% of the patient dataset for model training and validation, resulting in 68 patients with schizophrenia in the positive class, and 46 healthy control participants forming the negative class.

Preparing Unseen Test Data: We incorporated the held-out 20% patient data as an unseen test dataset, that could be consistently used across all models (*Affiliation, Self-report, Appraised Self-report* and *Patient*) for triangulation. This comprised 20 patients with schizophrenia and 9 healthy controls.

Features: Linguistic features from text data have been widely adopted and are known to be largely successful in predicting mental health states using social media data [194, 7]. We adopt two forms of linguistic content as features for classification. First, we build a term-frequency, inverse document-frequency based language model using the most frequent 500 n-grams (n=1-3) from the preprocessed data upon removal of stop words and URLs. Second, we use three categories of psycholinguistic measures: (1) Affective attributes, (2) Cognitive attributes and (3) Linguistic style attributes—from the well-validated psycholinguistic lexicon Linguistic Inquiry and Word Count (LIWC) [194]. Combining the two feature sets together, our overall feature space included 550 numeric features. Adopting the ANOVA F-test we reduced the feature space from 550 features to k-best features per classifier.

We experimented with non-linear and ensemble classification algorithms such as Support Vector Machines, Random Forest, and Logistic Regression [252]. For each classifier, we test its performance in two steps: First, for parameter tuning and assessing internal validity, we used stratified k-fold cross validation. We varied model parameters for all classification approaches during the validation step to find the best performing model. Second, choosing this best performing model from the validation step, we evaluated its performance on the unseen test data for external validity. Across the four classifiers, for relative comparison, we report model performance using a variety of metrics: Receiver Operating Characteristic Area Under Curve (ROC AUC), accuracy and F1 scores.

Table 5.2: Average model performance on the validation and unseen test datasets.

	Class 1	Class 0	**Cross validation**					**Testing**				
			P	R	f1	Acc	AUC	P	R	f1	Acc	AUC
Affiliation Model	861	539	0.89	0.94	**0.91**	**0.89**	**0.95**	0.28	0.1	0.15	0.21	0.20
Self-report Model	412	345	0.72	0.81	0.76	0.72	0.80	0.63	0.6	0.61	0.48	0.38
Appraised Self-report Model	153	107	0.81	0.88	0.84	0.80	0.85	0.65	0.75	0.70	0.55	0.51
Patient Model	68	46	0.76	0.80	0.77	0.72	0.76	0.93	0.7	**0.8**	**0.76**	**0.82**

5.3 Results

Internal Validity We present in Table 5.2 the cross validation performance of the four classifiers in distinguishing individuals with schizophrenia from matched controls. Overall, the *Affiliation Model* outperforms the other classifiers with the highest accuracy (Best: 0.94, Mean: 0.88, std: 0.02) and F1 (Best: 0.95, Mean: 0.91, std: 0.02) and a 27% improvement in accuracy over a ZeroR baseline (Accuracy: 0.61). The reported accuracy of this model is close to McManus et al. [20], demonstrating that the trained model can infer distinct patterns between the two classes.

Although both *Self-report* and *Appraised Self-report* models improve over their ZeroR baseline (accuracy: 0.54, 0.44 respectively), the *Appraised Self-report Model* performs better (Best: 0.88, Mean: 0.80, std: 0.03) than the *Self-report Model* (Mean: 0.72, Best:0.79,

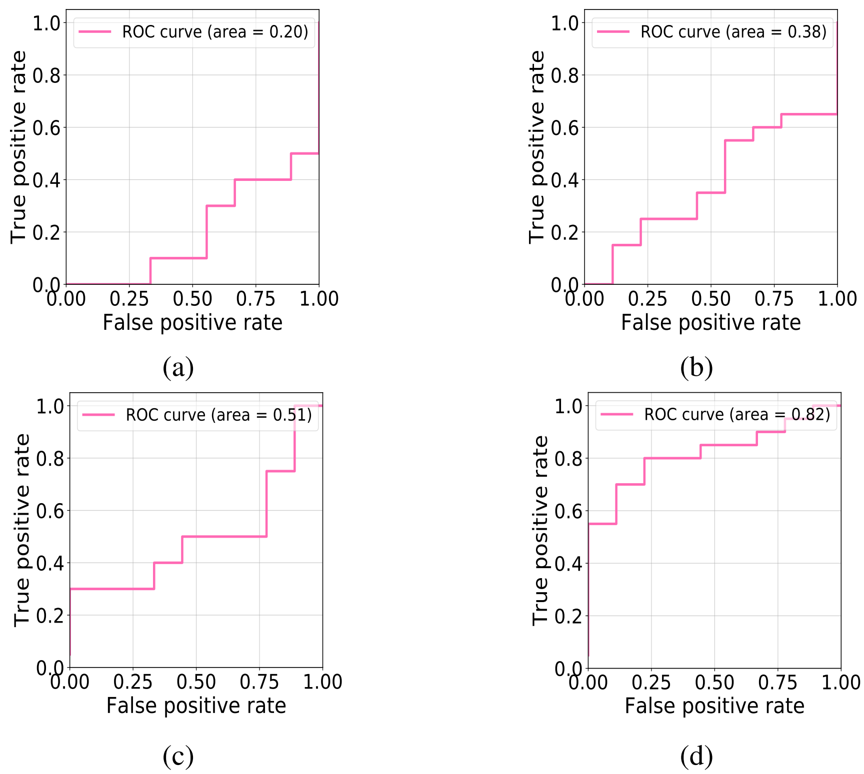


Figure 5.2: ROC (Receiver Operating Characteristic) curves per classifier (a) *Affiliation Model*, (b) *Self-report Model*, (c) *Appraised Self-report Model*, (d) *Patient Model*.

std: 0.02) across all metrics.

Comparing the performance of the proxy classifiers on their respective validation sets, we find that the *Appraised Self-report Model* has higher precision than the *Self-report Model*. This was also observed by Birnbaum et al. [243]; the clinician annotation task eliminated inauthentic noisy samples leading to a high precision sample of genuine self-reports.

Finally, the *Patient Model* trained on patient data performs modestly, although better, compared to the proxy classifiers, with average accuracy of 0.72 (best: 0.75) and average F1 score of 0.77 (best:0.79) across 5-fold cross validation[1].

[1]Given the relatively small sample sizes, to check for overfitting, we examined model stability through the standard deviation of evaluation metrics across folds. A low standard deviation of 0.02 indicated that despite low sample sizes, the model had stable performance.

External Validity Next, to examine their external validity on unseen patient test data, we present the performance of the proxy classifiers. Figure 5.2 (a-c) presents the ROC plots, per proxy classifier, showing the trade-off between true positive rate (sensitivity) against the false positive rate (1-specificity).

Among the three proxy classifiers, the *Affiliation Model* shows poor external validity with the lowest accuracy (0.21), the lowest F1 (0.14), and the lowest AUC (0.2) on the 20% sample of unseen patient data (refer Table 5.2). The next best performing model is the *Self-report Model* outperforming the *Affiliation Model* with a 27% improvement in the overall accuracy (0.48), 47% improvement in F1 score (0.61) and 18% improvement (0.38) in the ROC AUC. Although this indicates that self-reports might be a better diagnostic signal than affiliation, the performance of this classifier is still weak compared to its performance during the validation step (test of internal validity). Lastly, among the three proxy classifiers, we see the strongest external validity or best performance for the *Appraised Self-report Model*. This classifier shows a 9% and 55% improvement in F1 (0.70), and 7% and 34% improvement in accuracy (0.55) over the *Self-report* and *Affiliation Model* respectively. Although the *Appraised Self-report Model* demonstrates the strongest external validity so far, there is substantive decrease in its performance compared to the validation phase. Summarily, testing the proxy classifiers on unseen patient data revealed poor external validity and that relative performance between the validation and testing steps was not preserved when tested in a clinical setting.

Comparison of Classifiers on Unseen Patient Data Triangulating the three proxy datasets corresponding to their diagnostic signals, we compare their predictive performance with the *Patient Model*, again trained on the 20% sample of unseen patient test data. Through this, we establish an empirical estimate of the error incorporated by using the proxy classifiers, when applied on patient populations.

First, we report the performance of the *Patient Model*. From Table 5.2, we see that this model outperforms the proxy classifiers, in distinguishing healthy controls from schizophrenia patients, giving lower false positives and false negatives. We also find that this is a highly precise model (precision: 0.93), correctly predicting schizophrenia patients as the positive class. The performance, however, is affected by low recall, and we find lower precision for the negative class due to the false negatives (=6) wherein schizophrenia patients are wrongly predicted as healthy controls. We use the performance of the *Patient Model* as gold standard and examine the error incorporated by each of the proxy classifiers. We use F1 and ROC AUC to situate these differences.

We note the highest difference in performance exists between the *Patient Model* and the *Affiliation Model*. The *Patient Model* outperforms the *Affiliation Model* by 65% in F1 and 62% in AUC. Comparing the *Patient Model* with the *Self-report Model*, we observe a 19% and 44% gain in F1 and ROC AUC respectively. This indicates that the online behavior of self-reporting a mental illness diagnoses might be a better diagnostic signal than the affiliation behavior. Finally, the *Appraised Self-report Model* shows least difference in performance when compared to the *Patient Model* with 10% and 31% difference in F1 and AUC respectively. This indicates that when using self-reports as a diagnostic signal, clinical appraisal leads to better predictions. In short, the triangulation step reveals variability in predictive performances of the proxy diagnostic signals when tested on unseen patient data, demonstrating trade-offs when proxy signals are used for predicting clinical mental health states, versus when information is gathered directly from patients.

Deep Dive into Performance of Proxy Classifiers To evaluate beyond performance metrics and to reason about the poor external validity of the proxy classifiers, we present a deeper analysis of the proxy classifiers' performance.

Error Analysis. We begin by unpacking mismatches in predictions made by the proxy classifiers on unseen patient data, in terms of example false positives and false negatives. *Unpacking false positive classifications:* Consider an example X who is a healthy con-

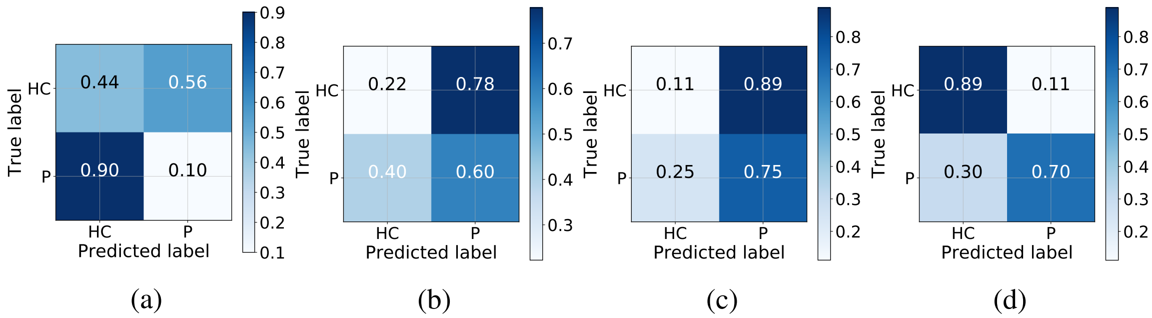


Figure 5.3: Confusion matrix per classifier (a) *Affiliation Model*, (b) *Self-report Model*, (c) *Appraised Self-report Model*, (d) *Patient Model*. Here HC: Healthy controls (Class 0); P: patients with schizophrenia (Class 1).

trol, per a clinically validated diagnostic assessment. But, the *Affiliation Model* wrongly predicted them as having schizophrenia. Examining their social media timeline, we find (paraphrased) posts including excerpts such as, "mental screenshot of notes", "are you bad for my mental health" and "use my phone in day mode because I am mentally ill". We note that terms like 'mental' ($\beta = 2.17$), 'health' ($\beta = 1.44$), 'illness' ($\beta = 1.45$) in these excerpts are highly predictive of the positive class in the *Affiliation Model*, leading to a misclassification of X as a schizophrenia patient. Moreover, because the *Affiliation Model* simply measures engagement, association with, or interest in mental health content and resources, it missed capturing the *context* in which these topics were discussed by X, leading to a misclassification of X as a schizophrenia patient. Now consider a healthy control participant Y's timeline. It includes prolific usage of terms such as 'creepy' (β = 0.241), 'hell' (β = 0.096), 'jesus' (β = 0.091), and 'help' (β = 0.401). These tokens are learned as highly predictive of the positive class by the *Appraised Self-report Model*, thereby leading to a misclassification of Y. Although these tokens reveal symptomatic expression, spirituality and support-seeking behaviors, notable in schizophrenia disclosures made on social media [185], the current example demonstrates varied usage of these tokens by healthy controls, in reference to pop-culture or in casual conversations.

We frame these observations as the following methodological gaps: that the outcomes yielded by the proxy classifiers are not valid indicators of a clinical diagnosis of schizophrenia (poor construct validity); and that the behaviors of individuals captured by the proxy signals might not be representative of the behaviors of schizophrenia patients (sampling bias).

Unpacking false negative classifications: Consider a different example A, a clinically diagnosed patient with schizophrenia. Their social media timeline data shows extensive usage of swear terms such as 'fuck' ($\beta = -0.94$), 'ass' ($\beta = -0.63$), 'bitch' ($\beta = -0.67$) that according to the *Affiliation Model* were highly predictive of the negative class, resulting in a false negative classification. Consider example B, a schizophrenia patient whose timeline largely consisted of travel and hobbies related posts with no evidence of schizophrenia experiences. The *Appraised Self-report Model* predicted B as a healthy control, due to lack of explicit disclosures of the illness, like symptomatic expressions and personal struggles (feature importance for LIWC categories: anger $(0; 0.03)$, body $(0; 0.06)$, swear $(0; 0.05)$ anxiety $(0; 0.03)$). These differences reveal that the proxy signals are not measuring what they intend to measure (poor construct validity). Further, that the social media language of patients might not be very different from control users (population bias).

Issues of Dataset Shift & Bias The population and sampling biases revealed by our error analysis goes on to show that the statistical data distributions might be drastically different between the proxy datasets and the actual patient dataset—a phenomenon referred to as "dataset shift" [253]. As a next step in our deep dive, we present the following analysis to systematically examine this dataset shift and assess its effects. Specifically, to quantify dataset shift, we adopt a measure of semantic distance computation between the linguistic content of proxy and patient datasets [248]. Our results bolster the findings of the error analysis, wherein we observe the farthest distance between the proxy and patient data in case of the affiliation dataset (similarity: 0.907, distance:0.092). The self-report dataset is at a closer semantic distance to the patient data distribution than the affiliation data, with a distance of 0.019 and similarity of 0.980. Finally, confirming the observations thus far, the appraised self-report dataset appears at the closest distance to the patient data with a distance of 0.017 and similarity of 0.982.

Table 5.3: Comparing the top features across the Affiliation, Appraised self-report and Patient Model. β weights (significant at the $p = 0.05$ level) denote feature importance. LIWC categories are presented in italics.

Affiliation	β	*Appraised*	β	*Patient*	β
i'm	-0.825	*NegAffect*	0.063	*cog mech*	-0.003
stigma	0.665	negation	0.074	*present*	-0.002
mhchat	0.696	*present*	0.40	*body*	-0.002
body	0.729	help	0.401	*verbs*	-0.002
bipolar	0.774	thought	0.41	*social*	-0.002
work	0.919	i'm	0.44	*aux verbs*	-0.002
self	0.961	die	0.45	help	0.0002
social	1.109	alone	0.45	feeling	0.001
care	1.111	hard	0.457	i'm	0.002
depression	1.116	cry	0.50	gonna	0.002
suicide	1.133	*body*	0.52	angel	0.002
thanks	1.445	feeling	0.523	burning	0.002
illness	1.447	*verbs*	0.58	pray	0.003
help	1.632	sorry	0.662	lifetime	0.005
mental health	1.866	gonna	0.63	attack	0.006

Issues of Construct Validity A second issue revealed by our error analysis was that the behavioral patterns learned by the proxy classifiers were absent in the schizophrenia patient population, raising concerns around construct validity. Therefore, next, we examine the features learned by the proxy classifiers in comparison to the features learned by the *Patient Model*. Table 5.3 shows the top features, and their feature weights for the worst and best proxy classifiers, and the *Patient Model*.

Overlap of features: Comparing the top features of the *Affiliation Model* with the *Patient Model*, we see little overlap between the two feature spaces, prominently, in terms of use of first person pronouns and LIWC category terms about 'social' and 'body'. We find that these features are predictive of one class in the *Affiliation Model*, whereas predictive of the opposite class in the *Patient Model*. Further comparing the top features of the *Appraised Self-report Model* with the *Patient Model*, we see a higher overlap than in the case of the *Affiliation Model*. Some of these features such as 'feeling', 'help' and use of first person pronouns are predictive of the positive class in both models, which explains the higher external validity of the *Appraised Self-report Model*. Although the *Appraised Self-report Model* is accurately learning certain patterns specific to the patient population, it misconstrues explicit mental illness disclosure behaviors (symptomatic expressions, combating stigma, and support seeking) as signals of a schizophrenia diagnosis.

Mismatch of features: Finally, we observe that the most predictive features (of the positive class) in the *Affiliation Model* are explicit signals of mental health care and support ('mental health', 'illness', 'depression', 'stigma', 'mhchat'), that have few occurrences in the patient data. Similarly, in the case of the *Appraised Self-report Model*, content related to schizophrenia experiences ('die', 'alone', 'sorry', 'creepy', LIWC categories of negative affect and negation) are either missing or not predictive of the positive class in the *Patient Model*. Therefore, we argue that what these proxy classifiers actually learn is the language use of individuals actively opening up about schizophrenia experiences, seeking informational and emotional support on Twitter. In comparison, our patient population does not exhibit such disclosure or support seeking behaviors on social media.

5.4 Discussion

In this study, we presented the first insights into some methodological gaps that exist in using social media derived diagnostic signals for predicting clinical mental health states. We found a lack of external validity when the prediction models developed using the proxy signals were tested on actual patient data. Our triangulation approach further surfaced issues of construct validity, limited theoretical underpinning, and population and sampling biases that permeate in the prediction task, through these diagnostic signals.

Uncertainty in Construct Validity. Drawing on the definition of this construct, we explore two methodological implications: 1) *Do these diagnostic signals measure what they claim to measure?* Our results show that the diagnostic signals are not measuring what they claim i.e. the clinical diagnosis of an individual's mental health (schizophrenia) state. This is revealed by the considerable mismatch we observed while comparing the top predictive features of the proxy classifiers and those of the *Patient Model*. Unpacking the context of these features in the actual social media posts, we found that they capture support seeking behaviors, interest in others' lived experiences of the illness, self-reported accounts of stigma and inhibition—patterns absent from the features extracted by the *Patient Model* from the clinical schizophrenia population.

2) *Is what is being measured by a diagnostic signal itself valid?* To the latter point about construct validity, we found a lack of clinical grounding in the diagnostic information (individual's clinical mental health state) that these signals intend to measure. Instead, what these signals presume as diagnostic information are essentially behavioral patterns associated with the appropriation of social media by a wide variety of stakeholders, not necessarily patients, in relation to the illness. These forms of appropriation include individuals posting resources for mental health awareness, individuals seeking therapeutics benefits, or individuals breaking free inhibitions and mental health stigma by disclosing their illness. Although these appropriation patterns can be a valuable resource to understand the experiences of schizophrenia [254], they do not provide clinically grounded information about an individual's diagnosis of a mental illness—thereby making them less suitable for the prediction tasks.

Theoretical Contextualization. Related to the above two issues lies another limitation, which is *a lack of theoretical underpinning* in the ways the diagnostic signals were identified. All of the scales and questionnaires used for clinical diagnosis, including the ones used in this study's patient population, draw upon theoretical frameworks, such as neurobiology, dimensional personality assessment, behavioral science, psychodynamic, and cognitive theories [255]. They undergo rigorous psychometric testing and are continually adjusted as the frameworks around mental illnesses evolve, or as the DSM [202], or more recently the National Institute of Mental Health introduced Research Domain Criteria (RDoC) framework [256] offer newer guidelines for mental health diagnostic and treatment. The proxy diagnostic signals are, however, not inspired by this theory. Instead they focus on online behaviors, which may or may not align with theoretical models, frameworks, or guidelines of mental illnesses.

The other methodological gap we identify in the use of the proxy diagnostic signals for predicting clinical diagnoses relates to *dataset shift* [253]. In the literature, datasets shifts in supervised learning are attributed to population or data sampling biases inherent in the data [70]. We therefore discuss the foundations of this phenomenon in two ways:

Population Biases. We observed that the datasets constructed using the proxy diagnostic signals include social media data of a unique set of individuals, who may not be representative of schizophrenia patients who are actually diagnosed with the illness and under treatment. Consequently, this population bias may manifest in several different ways: 1) The social media activities of an individual who follows online mental health resources, may be different from someone who publicly discloses their illness and experiences—and these, in turn, might be different from a clinically diagnosed patient's social media usage and behaviors [161]; 2) The diagnostic signals capture subpopulations who may not be truthfully reporting their illnesses or may be reporting about their self-derived assessments of a mental illness experience in an exaggerated fashion, that did not involve the feedback of a clinician; and 3) The diagnostic signals consist of subpopulations who may not be mental illness patients currently under treatment, and the social media activities of those who are under formal care and those who are not, might be considerably different.

Identifying and quantifying the biases between the populations targeted by the diagnostic signals, alongside examining their theoretical and construct validities is, therefore, crucial before the signals are deployed to make clinical predictions.

Clinical (Patient-Provider) Implications Alongside the methodological implications of making predictions of mental illness diagnoses with the proxy diagnostic signals, it is equally important to consider their impact on the key stakeholders such as clinicians and patients.

To the clinician community, whose primary source of diagnostic information comprises clinically validated questionnaires, scales, interviews, and symptoms reported by the patient [26], these new forms of proxy diagnostic signals derived from social media, despite the right intentions, add complexities to the conventional psychiatric assessment method. We highlight some of these complexities in the questions below. For instance, in the absence of supplementary and accessible details of their inner workings and biases, how can clinicians trust these new forms of diagnostic signals and their validity, and thereafter act upon them? How do these new signals complement or even contradict clinicians' mental models of reasoning, or how clinicians pursue diagnosis and treatment of their patients?

Importantly, decision-making by the clinicians (for diagnosis, treatment, or patient-provider interventions) involves both high stakes and high costs. Therefore, incorrect predictions made as a result of data with poor external, construct validity, or those suffering from population and sampling biases can be dangerous and have serious consequences for the patients' well-being, and social and professional life. While personalized patient care is touted as a strong motivation for adopting social media for clinical diagnosis and treatment [240], validity and bias issues may additionally adversely impact patients trust and attitudes towards mental healthcare. When outcomes of these proxy classifiers are incorporated into clinical decisions without the patients' awareness, poor validity can even negatively impact patients' perceived agency in treatment, or the therapeutic relationship they share with their clinicians. These issues may further conflict with patients' preferences, needs, and values in treatment [257]. Thus bridging these methodological gaps with interactions with and involvement of the patient and clinician stakeholders is key to translating the potential of social media to support clinical diagnosis and treatment.

5.4.1 Remedial Guidelines: A Proposal

In the light of the above discussion, we suggest some guidelines for researchers to bolster efforts in examining and establishing the efficacy of social media based signals for prediction of mental health states in clinical populations.

Improving Methodological Rigor and Adopting Alternative Research Designs. A first set of guidelines center around reducing or eliminating the issues noted above. We conjecture that combining multiple proxy diagnostic signals, especially those that are complementary to each other, could provide more rigor because of their potential to target more diverse social media populations. However, this warrants empirical investigation. Alternatively, given the stigma around experiences of mental illness [258], some of the proxy diagnostic signals can be leveraged in a respondent-driven sampling framework [259]. This can be a viable mechanism to reach and recruit individuals for clinical studies that seek to collect gold standard patient data. Implementing an online-offline framework [6], that combines social media data with pre-existing offline longitudinal information of comparable sub-populations, can also reduce the dataset shift challenges. Further, issues of dataset shift can be overcome by adopting recent approaches from the machine learning field, such as including importance weighting of training instances based on similarity to test set [260], and employing online learning of prediction models to identify and recover from incorrect predictions [261, 262]. Crowdsourcing based data analysis and replication efforts [263, 264] can also be used to make transparent the impact of proxy dataset biases on predictive models.

Building and Utilizing Shared Infrastructures for Data Collection, and Data Donation Efforts. Another set of guidelines center around building, contributing to, and leveraging shared infrastructures and data repositories for conducting this research. Our findings showed the value of using patient data in building predictive models of mental illness diagnosis. However, we recognize that researchers without access to patient populations within large healthcare systems, or without involved collaborations in the clinical field may be at an unfortunate disadvantage. Further, patient data collection can be complex, including technological and ethical dimensions, due to the need to engage with a vulnerable population and gather sensitive (largely non-public) information, that might include HIPAA [265] protected data. Open source, HIPAA compliant infrastructures with customizable data collection functionalities can be helpful to overcome some of these technical challenges. Participatory research efforts such as the Connected and Open Research Ethics (CORE) initiative [266] can be used to develop dynamic and relevant ethical practices to guide and navigate the social and ethical complexities of patient data collection. Initiatives focusing on voluntary data donation approaches, such as the notable OurDataHelps [267] program for suicide prevention research, can be utilized to gather high quality data about people's clinical mental health states, alongside their social media data.

Harnessing Partnerships Between Computational and Clinical Researchers, and Patients. Finally, this research area can benefit extensively from cross-disciplinary partnerships. Collecting patient data for building the predictive models involves human costs, and suffers from resource and logistical constraints. In working with sensitive population such as patients with mental illnesses, it is important to have appropriate clinical risk management protocols in place [268], especially when the source of data concerns social media activities of patients monitored in a near real-time fashion [269]. Computational researchers by themselves may not be best equipped to define or implement such protocols. Moreover, clinical expertise is needed to identify and navigate the right way and the right time to approach patients for informed consent regarding data sharing, and assess how it would impact their perceptions of clinical care. Partnership of computational researchers and clinicians throughout the research pipeline—e.g., right from establishing validity of measured online behaviors, providing appraisal of the data via qualitative coding tasks, to interpreting and situating large scale data analysis, can also improve rigor and eliminate issues of construct validity and improve theoretical grounding of the approach. Moreover, directly incorporating patients' feedback in the construction and acquisition of the clinical diagnostic signals will not only help represent their voices in the functioning of the predictive models and engage them as partners in treatment, but also support advancing the vision of participatory mental healthcare [270].

CHAPTER 6

INTERSECTION OF SOCIAL AND CLINICAL PATHWAYS TO CARE FOR MENTAL HEALTH ON SOCIAL MEDIA

Chapters 3 and 4 illustrate how social media supports both social and clinical pathways to care for people with schizophrenia. By disaggregating *patients* with schizophrenia (those clinically diagnosed) and *people* with schizophrenia (those whose clinical diagnosis is unknown), Chapter 5 provides a lens to understand the varying needs and differentiating uses of social media for mental health care. However, in reality, being a patient and not being a patient does not indicate two separate populations but different states that the same individual enters and leaves. While someone is admitted to the hospital, they enter an institutional role as a patient and receive clinical treatment. After discharge from the hospital, the same person is no longer under institutionalized care as a patient and has to manage their illness by themselves and seek support. Similarly, pathways to social care and clinical care are not dichotomous. Both social relationships and clinical care are required to improve the overall well-being of those with mental illnesses. Although prior studies have established the link between social relationships and health outcomes [271, 272], current delivery of mental health care and our understanding of interventions is heavily focused *either* on treatment of the illness *or* management of the illness and social support. In this chapter, I present two studies focusing on the intersection of clinical and social pathways of care to examine the role of social media in people's lives along the course of a mental illness.

From a clinical perspective, paradigms such as person centered-care [273, 107, 108] have advocated for the intersection of social and clinical aspects of health – putting the individual in center and emphasizing on social, mental, emotional and spiritual needs. Under this paradigm, events like hospitalizations are considered as part of *life-course experiences with health* [109]. Similarly, the recovery model looks at a person's journey as a "*deeply* *personal, unique process of changing one's attitudes, values, feelings, goals, skills and/or roles . . . a way of living a satisfying, hopeful and contributing life even with the limitations caused by illness*" [10, 105, 106, 274]. However, these frameworks that combine clinical and social care have not been applied to research in social media and mental health to advance our understanding of how needs, behaviors and uses of social media change along the course of illness.

From a Social Computing perspective, there is extensive research on technology use during major life events [118, 121, 127], and how social media acts as a transitionary machinery [119]. Literature shows that social media plays a key role in helping individuals establish a "*new normal*" in light of changing circumstances and life disruptions [118] and for social support and information seeking [124]. With the associated stigma and loss of access to technology, resources and social connections, events like psychiatric hospitalizations can also be considered major life events. If social media can support mental health interventions, how can we extend our understanding of social media and liminality to apply to events such as mental illness hospitalizations?

This chapter synthesizes these viewpoints from clinical and social computing literature to study how major life transitions around mental illnesses are exhibited on social media and how social and clinical care intersect around these transitionary periods? Anchoring on hospitalizations as transitionary periods where social and clinical care intersect, this chapter details two accompanying studies. The first study aims to understand health transitions as exhibited on social media and the second study unpacks one particular consequence of health transitions focusing on social re-integration.

6.1 A social media study on mental health status transitions surrounding psychiatric hospitalizations

Emerging from the deinstitutionalization movement in the late 20th century, the recovery model has been the guiding principle of mental health policy in many countries [105]. It views recovery as "*a personal journey rather than a set outcome, and one that may involve developing hope, a secure base and sense of self, supportive relationships, empowerment, social inclusion, coping skills, and meaning*" [275]. In essence, recovery from mental illness not only involves removal of symptoms and restoration of functioning but also involves recovery from the stigma and negative stereotypes, from the lack of opportunities and finding a way of living a hopeful and contributing life despite the limitations caused by the illness [276, 106, 274, 10, 277]. It provides a holistic perspective arguing that people with mental illness face the complementary experiences of *clinical recovery* (i.e., reduction of symptoms), and *social reintegration* (i.e., restoration of social lives and community inclusion) hand-in-hand [104, 278, 91]. Importantly, per this model, peoples' mental health states are temporally situated experiences that are part of their recovery and reintegration journeys themselves.

For mental health, work in the HCI and CSCW areas identify social media as an important tool serving as a mechanism to study peoples' mood, communication, social interactions, and psychological states. However, despite the guidelines from the recovery approach, this body of work has considered mutually disjoint, discrete conceptualizations of the individual experiencing mental illness. One line of work emphasizes the role of the individual as a patient, as someone with a validated diagnosis, receiving treatment, and on the road to clinical recovery from a mental illness. Work in this area leverages individuals' social media data to explore the efficacy of predictions in supporting early diagnosis, evidence-based treatment, and deploying timely patient-provider interventions [5, 43, 186, 63]. Another line of work emphasizes the individual's role as a support seeker who makes sensitive self-disclosures and participates in online health communities to maintain their mental health outside clinical care [2, 279, 1, 4, 185, 280].

However, as posited by the recovery approach, one's role as a patient and as a support seeker are not dichotomous, and peoples' experiences are often in transition along the course of illness. When we project narrow, oversimplified conceptualizations of the individual as either a patient or support seeker, and fail to understand the relationship between the two, we do not account for the contextualized, multiple, and heterogeneous experiences of people in reality. With emerging evidence about the potential of social media to support clinical diagnostic predictions and social support provisions [43, 5, 186, 2], it is timely to examine the intersection of these two perspectives.

In this work, we ask, *how can we study the different journeys of those with mental health conditions combining their experiences of clinical recovery and social reintegration?* Anchoring on psychiatric hospitalizations as a liminality [112], during which people enter and leave the role as a patient to self-manage their condition, we focus on three research questions:

RQ1: What self-presentation and behavioral signals on social media characterize individuals' mental health statuses around psychiatric hospitalizations?

RQ2: What trajectories on social media showcase transitions between these statuses surrounding the hospitalizations?

RQ3: What social media-based signals are indicative of social reintegration trajectories of individuals following hospitalizations?

To answer these questions, we combine data from medical records comprising clinical information related to diagnosis codes and hospitalization dates, with social media data from Facebook archives of 254 consented participants who have experienced at least one hospitalization for psychosis (N=142), mood disorders (N=106), or other mental health conditions (N=6). Across all participants, we compile over 980 thousand Facebook posts around 372 hospitalization events. Then towards answering RQ1, we adopt the Possible Selves framework [281] as a theoretical lens to capture and interpret peoples' conceptions of self-knowledge and alternate versions of themselves in the future. With this framework, we use Gaussian Mixture Models (GMM) [282] to enumerate common behavioral patterns or "possible selves statuses" (PSS) on Facebook seen around participants' hospitalizations: self-regulation, self-awareness, sociality, withdrawal, re-adjustment, and incorporation-focused. We validate the GMM components with qualitative interviews with clinical domain experts and define *a taxonomy of six possible selves exhibited on Facebook surrounding psychiatric hospitalizations*. Next, to address RQ2, we present a linear transition model between the derived PSS during the periods before and after hospitalizations to understand peoples' clinical recovery and social reintegration trajectories, *contributing an empirical framework of mental health transitions*. Finally, for RQ3, to demonstrate the utility of the derived taxonomy and the framework of mental health status transitions, we define a PSS-based operationalization of social reintegration. We conduct regression analyses to assess signals on Facebook that are associated with successful social reintegration post-hospitalization.

Through a theory-driven modeling approach based on the possible selves framework and insights clinically-grounded in the recovery model, this work presents a first step towards understanding personalized and heterogeneous behaviors and self-presentations of people as they experience mental health status transitions around psychiatric hospitalizations. We discuss the theoretical implications of combining peoples' clinical and social experiences in mental health care and the opportunities this intersection presents to post-discharge support, sensemaking in healthcare settings, and technology-based interventions for mental health. Finally, we put forth what it means to design social media platforms for online social reintegration after major life transitions.

6.1.1 Data

Participant Recruitment and Data Collection

For this study, we utilized Facebook data of consented participants with a medical history of at least one hospitalization due to a mental health condition. The data collection strategy (described in Chapter 3 and Chapter 5) aimed at identifying technology-based mental health information to provide early identification, intervention and treatment to patients with psychiatric disorders. The research protocol was approved by the Institutional

Table 6.1: Descriptive statistics of participants.

Variable	Statistics	Distribution
Demographics		
Age	Mean = 24.0, Median = 22.4, Range = (15.1, 60.7)	
Gender	Male — Female (55% Female)	
Race	Native American/Indian, African American, White, Other	
Ethnicity	Hispanic/Latino — Non-Hispanic/Latino (85% Non-Hispanic/Latino)	
Diagnosis	Psychosis, Mood disorders, Other disorders	
Facebook data		
#Posts	Mean = 2596.3, Median = 1471.0, Range = (0, 11018.0)	
Duration (days)	Mean = 1025.16, Median = 730.0, Range = (1, 3287)	
Medical data		
#Hospitalizations	Mean = 2.3, Median = 2.0, Std = 2.14, Range = (1, 17)	
Duration (days)	Mean = 15.5, Median = 11, Std = 13.7, Range = (1, 104)	
Gaps (days)	Mean = 365.2, Median = 175.5, Std = 513.9, Range = (2, 4179)	

Review Board (IRB) of the coordinating institution managing patient recruitment – a large healthcare organization in the north-east of the United States, as well as the local IRBs at collaborating sites.

Individuals over 15 years of age were recruited from various inpatient and outpatient psychiatric departments at the coordinating and partner institutions. Participants were eligible if they were diagnosed with a schizophrenia spectrum disorder, mood disorder with and without psychotic features, borderline personality disorder or anxiety disorder based on clinical assessment scales (e.g., the Psychiatric Diagnostic Screening Questionnaire or PDSQ [283]) and formal clinical examination conducted by a licensed clinical psychologist, and facilitated by the Structured Clinical Interview for DSM-5 (SCID) [284]. Participants also experienced at least one hospitalization for the mental health condition. Informed consent was obtained from participants after describing to them the research study, type of data to be collected, policies for storage and use, clinical risk mitigation protocols, and clarifying that their relationship with the medical institution would remain unaltered whether they chose to participate in the study or not. Consented participants included 142 psychiatric patients diagnosed with schizophrenia, 106 diagnosed with mood disorders, and 6 with other mental health conditions (Table 6.1). To answer the RQs, in downstream analyses, we combine participants with schizophrenia, mood disorders, and other mental health conditions into a single study population. While the objective outcomes of recovery, such as time taken for remission of symptoms and probability of relapse, vary across these conditions, clinical literature suggests commonalities in subjective recovery and reintegration experiences like a growing sense of agency and autonomy, quality of life, peer support, greater participation in normative activities, etc. [285]. As a formative investigation into transitions around hospitalizations, we focus on understanding the subjective experiences of recovery and reintegration transitions, thus, we combined analysis across conditions. All data collected from these participants were de-identified and stored in HIPAA compliant secure databases and servers, which were located at the coordinating institution with access privileges limited to only the core project personnel.

Facebook Data Upon informed consent, all participants were requested to extract and share their Facebook data archives. This Facebook data comprised activity traces on the platform, specifically data on friend requests, messaging, updates to profile fields, adding new photos/cover photos, sharing feelings via status updates, shares, likes, co-tagging, as well as the linguistic content of timeline posts (status updates) made by participants. Descriptive statistics of this data are shown in Table 6.1.

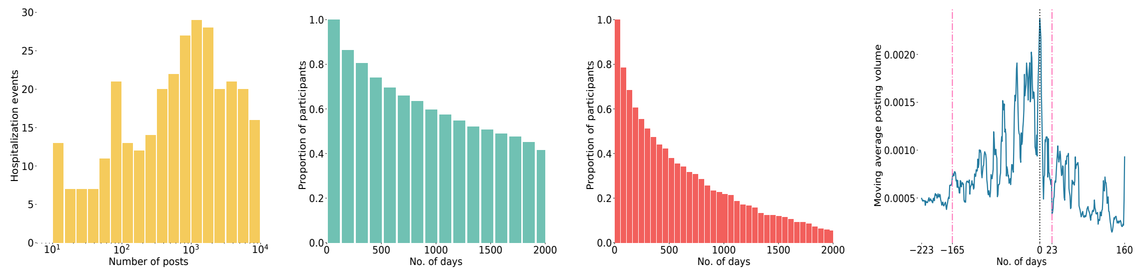
Medical Records and Hospitalization Data We also collected medical history for each participant (following consent and adoption of HIPAA compliant policies). This included primary and secondary diagnosis codes, the total number of hospitalizations and admission and discharge dates per each hospitalization event. Note that in this data a hospitalization typically indicated that the participant had spent at least one day (more typically up to 30 days) in an inpatient facility within the target healthcare system, because a licensed clinician had assessed a significant symptomatic exacerbation, or a risk of self-harm/suicide/homicide, that needed 24×7 medical care, and that was not addressable via adaptions to the patient's existing treatment plan if any. Across all patients, the medical records indicated 346 overall hospitalizations for schizophrenia patients, 230 hospitalizations for mood disorder patients and 9 hospitalizations for patients with other mental health conditions with 2.3 mean hospitalizations per participant (median = 2.0, std = 2.14).

Curating Facebook Data around Hospitalization Events

The goal of this study is to identify different journeys of people during life transitions around a psychiatric hospitalization, as observed and expressed on Facebook. In doing so, we note that not all hospitalization events will have similar experiences, even for the same individual. If psychiatric hospitalizations are considered major life transitions, the effects of each hospitalization as experienced by individuals might be different. Therefore, in this work, we consider each hospitalization event per consented participant as a different observation.

To gather Facebook data surrounding each hospitalization, we first collate the admission dates for all hospitalization events in each participant's medical history. Using the hospitalization admission dates as temporal markers, we center the participant's Facebook data such that the day of admission to the hospital is treated as *day 0*. Then we extend the temporal window before and after the hospitalization (*day 0*) and stop only when we reach another hospitalization admission for the same individual. Note that for a participant with only one hospitalization, we include their entire Facebook archive from the earliest to most recent date. We repeat this process for every hospitalization event recorded per participant. Across all 254 participants, we have a total of 584 recorded hospitalization events; however, only 372 hospitalizations had some digital traces from Facebook archives during that time. Corresponding to each of the 372 hospitalizations, we gather and center the participant's Facebook data around their admission date. The distribution of the number of posts surrounding each hospitalization is shown in Figure 6.1. The average number of posts surrounding each hospitalization event is 3162.44 (median = 670.0, std = 8503.81). Across all hospitalization events, the minimum and maximum number of Facebook posts are 1 and 86,465 respectively.

Next, to analyze data across all hospitalization events (which ranged from 2009-2019), we identify a fixed time period preceding and succeeding each hospitalization. We adopt



(a) (b) (c) (d)

Figure 6.1: (a) Distribution of number of Facebook posts over hospitalization events (b) CDF of post distribution preceeding the hospitalization event (c) CDF of post distribution suceeding the hospitalization event (d) Temporal phases identified around the hospitalization using a moving average model of posting volume. The central vertical line indicates the hospitalization admission date, while the vertical lines on its two sides indicate the boundaries of the before and after hospitalization phases.

an empirical approach by generating cumulative density functions (CDFs) of the number of posts[1] before and after each hospitalization event. These CDFs are shown in Fig. Figure 6.1. Based on these figures, we observe that most users (around 70-80%) have posts for at least 230 days before and 160 days after each hospitalization event. Therefore, we choose Facebook timeline data spanning 230 days before and 160 days after each hospitalization event as a fixed-length time period for downstream analyses.

6.1.2 Methods

Theoretical Framework: Possible Selves

Corresponding to RQ1, our goal is to identify and understand patterns on Facebook that characterize peoples' individualized mental health states around psychiatric hospitalizations. To do so, we adopt the Possible Selves framework [281], as introduced earlier. The possible selves concept is used in psychology to complement current conceptions of self-knowledge and to capture cognitive representations of alternative versions of the self in the future. In more detail, possible selves represent "*individuals' ideas of what they might become, what they would like to become, and what they are afraid of becoming, and thus provide a conceptual link between cognition and motivation*" [281]. Possible

[1] Overall number of posts includes status updates and activities such as check-ins on the timeline.

selves are largely based on past experiences, but their essence lies in clear references to the future [281] – they are cognitive representations of hopes, fears, and fantasies regarding oneself. Thus, per the Possible Selves framework "*an individual's collection of self-conceptions and self-images can include the good selves (the ones we remember fondly), the bad selves (the ones we would just as soon forget), the hoped-for selves, the feared selves, the not-me selves, the ideal selves, the ought selves. They can vary dramatically in their degree of affective, cognitive and behavioral elaboration*" [281].

As described in Section 2, we consider psychiatric hospitalizations to be major life transitions in an individual's life. Psychotherapy research has argued that surrounding this liminality, it is important to consider the patient's "possibilities" (i.e., the possible future states of the self) as an important instrument of effecting a change [286]. These possible selves can be understood as a kind of positive resource that the patient draws on when making desirable changes in their behavior or self-regulation to manage their underlying mental illness, around the hospitalizations. Prior clinical literature has studied individuals' possible selves in relationship to their diagnosed mental health condition itself [287, 288, 289]. For instance, Janis et al. found that participants with borderline personality disorder were less likely than controls to endorse positive possible selves as current, but more likely to endorse negative possible selves as "current, probable, desired, and important" [289]. In [287], Clarke found that being positive about achieving possible selves was positively related to functional outcomes in first episode psychosis. We hypothesize that people's language, behaviors and self-presentation signals surrounding hospitalization, as expressed on Facebook, may represent the various possible selves or "future-projected" aspects of self-knowledge, that they perceive as potentially possible.

Our rationale is grounded in the fact that existing behavior change theories (such as the Transtheoretical Model [136] or the Theory of Planned Behavior [290]), often used to capture health transitions, have been criticized for their lack of adaptive capabilities as well as their inability to take into account an individual's unique psychological state, social context, activity, and behavior patterns [291]. Furthermore, these models rarely match the reality of health transitions, due to differences between actual and perceived behaviors and assumptions related to homogeneity within behavioral change [292]. In contrast, our conceptualization of mental health statuses using the Possible Selves framework allows us to consider that the same individual can exhibit multiple possible selves on Facebook during a certain period before or after a hospitalization. Also, different individuals may express different possible selves on Facebook at any given time preceding or succeeding psychiatric hospitalizations because of their contrasting life situations. Finally, as one navigates the transition caused by the psychiatric hospitalization, these possible selves may evolve and change over time. In this study, we refer to an individual's diverse mental health statuses surrounding hospitalizations as a collection of their possible selves.

To operationalize and capture possible selves surrounding hospitalizations guided by this framework, we adopt the following empirical approach described in the remainder of this subsection.

Identifying Temporal Phases around Hospitalizations Studying life transitions surrounding psychiatric hospitalizations necessitates identifying data spanning pre- and post-hospitalization phases where the self-presentation and behavioral changes are most likely to be manifested. Prior work in CSCW and HCI on behavioral changes during major life transitions found abrupt declines in posting activity on social media and noted it as a sign of social withdrawal [185]. For instance, in individuals challenged with postpartum depression, changes in sociality, and behavior on social media manifested through patterns of posting volume [28]. Other work has also noted a sudden increase in posting volume referred to as a "rush" of excitement for the future in the case of life transitions like engagement, starting a new job, or having a child [7]. Also, per clinical literature, we expect people to show markers of social withdrawal, which is known to be a notable risk marker around hospitalizations [293]. Therefore, we use measures of changes in posting volume of an individual (normalized number of posts per day) to identify temporal phases around psychiatric hospitalizations. Our phase identification approach is adopted from Section 4.1 [185] and includes the following two steps:

1. First, we calculate the daily posting volumes on Facebook timeline around each of the 372 hospitalization events spanning 230 days before and 160 days after the hospitalization. We computed the rates of change throughout the before and after hospitalization periods by employing a weekly moving average model on this posting volume time series data. This would allow us to smooth out local fluctuations and seasonality while allowing comparison between the posting volume at day *t* and that during the seven days preceding it.

2. Next, we compute the medians of the weekly rates of changes during the periods before and after hospitalization and use a median split method to define phase boundaries. Specifically, the first time point (day) in the pre-hospitalization data when the rate of change of posting volume becomes higher than the pre-hospitalization median rate of change is taken as a cutoff. Similarly, the first time point (day) in post-hospitalization data when the rate of change of posting volume becomes lower than the post-hospitalization median rate of change is taken to indicate another cutoff.

As shown in Figure 6.1, the median rate of change in posting volume pre-hospitalization was 0.0007 and 165 days prior to hospitalization (*day 0*) is the first time the rate of change surpassed this median. Similarly, the median rate of change in posting volume post-hospitalization is 0.0005 and 23 days after the hospital admission is when this rate is lower than the post-hospitalization median. Since the rates of changes in posting volume are computed weekly, we adopt the following day demarcations to define four temporal phases around hospitalization:

- Bf_long or Long before hospitalization (N = 214): 223 to 165 days prior to hospitalization.

- Bf_hosp or Before hospitalization (N = 271): 165 days prior to the hospitalization day.
- Af_hosp or After hospitalization (N = 130): 23 days after the hospitalization admission.
- Af_long or Long after hospitalization (N = 166): 23 to 160 days after the hospitalization.

Thus, segmenting each of the 372 hospitalization events into four phases, we obtain 781 phases.

Modeling Possible Selves Around Psychiatric Hospitalizations

Next, to define and identify people's individualized mental health states, or their possible selves during the above identified four temporal phases surrounding psychiatric hospitalizations, we build a Gaussian Mixture Model [282]. This approach has been used in prior HCI research to capture the heterogeneity in people's social roles and their evolution [157].

Gaussian Mixture Modeling Approach Gaussian Mixture Model (GMM) is a probabilistic model that clusters heterogeneous, multimodal data into a fixed number of coherent components [282]. Unlike traditional clustering algorithms like *k*-means that perform hard-clustering where each data point is assigned a single cluster, GMMs perform soft-clustering where each data point can belong to multiple clusters with different weights. Using the model we assume that each temporal phase surrounding hospitalization can be represented as a feature vector x having d behavioral features and there exists K components $c_{i=1}^{K}$, one for each type of possible self status described by the features. Each of the K components c_i is modeled using a multi-variate Gaussian distribution with an associated vector μ_i of average values for each feature $x \in X$. Each temporal phase is then generated from a mixture of these K components and co-variance ($\sum_i$), which gives the likelihood of each pair of possible selves. Mathematically, each temporal phase x is represented as a linear combination of these K Gaussians, with the probability function as:

$$p(x) = \sum_{i=1}^{K} \pi_i * N(x|\mu_i, \sum_i), where \sum_{i=1}^{K} \pi_i = 1$$

Here, $\{\pi_{i=1}^{K}\}$ are called the mixing coefficients and denote the probability of each individual Gaussian. Learning a GMM involves learning the mean, co-variance and mixing coefficient $\{\mu_i, \sum_i, \pi_i\}_{i=1}^{K}$ of each Gaussian. We use a Gaussian Mixture Model to cluster the 781 temporal phases around hospitalizations into K components such that each component has its own single variance. Each component then describes a **possible self status** (PSS) representing common behavioral patterns seen across all observed patients and their Facebook data surrounding their hospitalization.

Operationalizing Linguistic and Behavioral Signals for the GMM Clusters

Next, we propose a set of linguistic and behavioral signals that characterize individuals' mental health status transitions. We operationalize these signals from Facebook data during each of the 718 temporal phases. Each temporal phase is represented by a d-dimensional feature ($d = 43$) vector x, consisting of features described in the following section. All feature values were converted to z-scores.

Psychological processes. *Affective measures* that reflect one's emotional response are expected to notably change around major life transitions [294]. We use affective words based on the Linguistic Enquiry and Word Count lexicon [195] (LIWC). We extract the normalized frequency of word occurrence in Facebook posts belonging to the following categories: positive affect, negative affect, sadness, anxiety, and anger. *Cognitive measures* also play an important role during transitions related to mental health by mediating the affective and attitudinal responses [295]. We calculate the normalized word frequency of the following LIWC categories in Facebook posts: insight, tentativeness, discrepancy, causation, certainty, differentiation.

Linguistic style. The use of function words is known to provide a non-reactive way to explore social and personality processes [194]. We use LIWC to define linguistic style: personal pronouns (first-person singular and plural, second-person and third-person), impersonal pronouns, adverbs, auxiliary verbs, conjunctions, article, preposition, and negation.

Mental health related. Beyond everyday activities shared on Facebook, individuals experiencing major life transitions such as a hospitalization are likely to share content specific to their experiences of symptoms of the condition. We adopt two sets of features to identify signals specific to the mental health symptoms and experiences. First, we use the LIWC measures related to health, sexual, body, and ingest categories to identify word usage around the specific health experiences. Second, we adopt validated machine learning classifiers of social media language indicative of depression, anxiety, stress, suicidal ideation, and psychosis from prior literature [296]. The classifiers have demonstrated linguistic equivalence across platforms and accuracy ranging from 0.82 to 0.92 (recall ranges 0.82 to 0.91 and precision between 0.85 and 0.92) on unseen test data in prior work. We ran these classifiers on Facebook timeline posts during each phase and calculated an aggregate proportion of posts that were predicted as indicative of different mental health concerns.

Temporal orientation. Alongside affective, cognitive, and behavioral variations between an individual's collection of self-conceptions, the theory of possible selves posits that they also vary in "tense" or "temporal sign" of the self, that is individuals holding notions of their past selves, present selves, and future selves [281, 297]. To capture the temporal signs in self-presentation on Facebook, we calculate normalized counts of word usage belonging to the following LIWC categories: focus on the past, focus on the present, and focus on the future.

Social and personal concerns. The disruption caused by life transitions like mental health hospitalizations is often accompanied by stress [298]. The stressors brought about by mental health experiences are not only related to peoples' psychological well-being but also their personal and social concerns such as work or employment [299], housing [300], social role changes and so on. To identify social and personal concerns during the phases surrounding hospitalization, we use the LIWC lexicon to calculate the proportion of words belonging to each of the following categories: home, religion, money, death, leisure, friend, and family.

Self-presentation on Facebook. Transitions such as experiencing a psychiatric hospitalization involve people "redefining their sense of self and redevelop self-agency in response to disruptive life events" [275]. Identity work [120] and signaling social role changes [301] are important aspects of navigating such transitions. Drawing from literature of identity and major life transitions [302], we expect our participants to show changes in self-presentational signals on Facebook during the transitional phases surrounding hospitalization. To capture these signals, we measure the following assessment and conventional signals [303] derived from Facebook: updating profile fields, adding photos, adding cover photos, sharing feelings with status updates, and broadcasting behaviors using shares and likes on Facebook.

Social interactions. Technologies like social media are known to play an important role during major life transitions by helping individuals establish a "new normal" [118], conduct identity work [120] and reach out to similar others [122]. Also, social functioning is a key marker for recovery in mental health conditions [304]. To capture aspects related to peoples' social interactions and functioning during the temporal phases surrounding hospitalization, we consider the following features from Facebook data: number of friend requests sent or accepted on Facebook, one-one interactions measured via the number of distinct people with whom the participant shared messages, number of messages exchanged, number of posts where the participant was co-tagged with others, and an overall measure of posts and activities on Facebook.

GMM Parameter Tuning: Determining the Number of Components Training the GMM involves selecting the parameter K to indicate the number of components. We experimented with K from [2, 10] to empirically determine the optimal number of components/possible self statuses. To prevent over-tuning, we used the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the optimal fit.

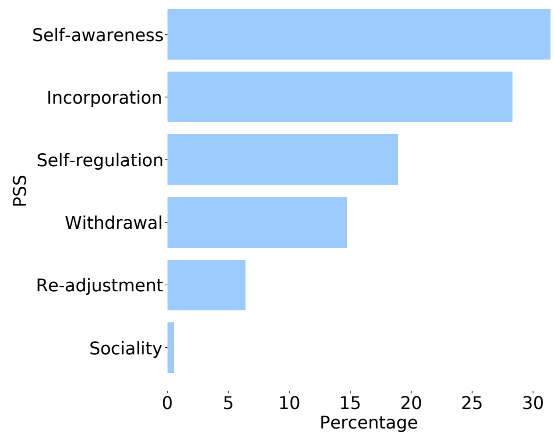


Figure 6.2: Frequency of the GMM-derived PSS

The lower the AIC and BIC values, the better the model is at predicting the underlying unknown distribution. Based on the values and the gradient of the AIC and BIC scores curve, we found that GMMs with $K \in [4, 6]$ were a good fit on the phases data.

Clinical Validation and Grounding of the Possible Self Statuses: A Taxonomy

Validating the output of generative models like Gaussian Mixture Model components or Latent Dirichlet Allocation based topic models is typically done by human coding tasks [185], goodness of fit or predictive likelihood measures, performance on external tasks or validation of coherence [305]. However, qualitatively interpreting the derived components is challenging due to lack of contextual knowledge, difficulty in understanding the operationalization of features and researcher bias. Recent work has reflected on the convergence and divergence between statistical machine learning methods (especially unsupervised approaches) and grounded theory method [306] and suggests hybrid, iterative approaches that combine the two [306, 307] as possible alternatives. We take inspiration from this literature and prior work interpreting GMM components [157] to finalize the final setting of number of GMM components and their labels.

Using the approach described above, we fit Gaussian Mixture Models for values of K that showed optimal fit based on the AIC, BIC scores. Then for each model's extracted components, we find the top behavioral features that are representative of the component.

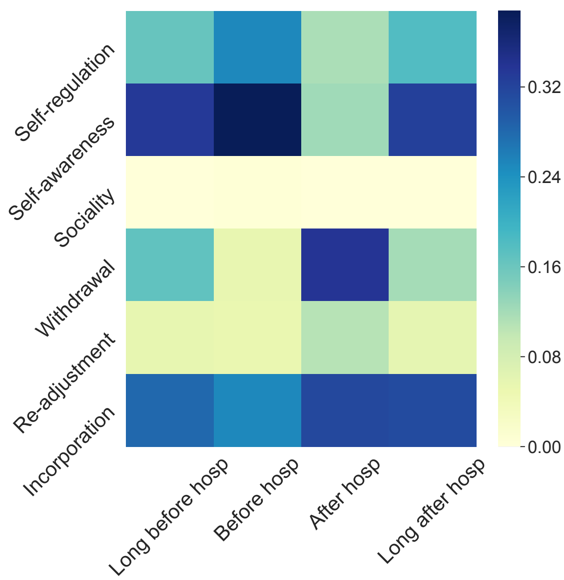


Figure 6.3: Proportion of each PSS in each temporal period surrounding the hospitalization events.

To get the top features per component, we use two measures: 1) We take the feature means per component, which in the form of z-scores, shows whether the behavior was performed more or less than its average value. Based on the magnitude of z-scores, we filter the top features and their mean values per component. 2) We build a linear regression model on the probability of a phase belonging to the component and extract features that are statistically significant and predictive of membership to the component.

We presented the top features per component for each of the GMM models to 4 annotators and gathered their input to help interpret and name the extracted components as possible self statuses. To provide additional context for interpretation, we also provided them with the temporal periods during which the component was most prominent (for instance, whether most phases predicted to belong to a component were right before hospitalization). Using this information, we conducted individual semi-structured interviews and follow-up sessions with annotators to iterate over the component labels and reach the most theoretically grounded and clinically informed GMM setting that would capture the PSS surrounding the hospitalizations. The four annotators included one psychiatrist and two computer science researchers who were domain experts in mental health and social media studies and a graduate student familiar with the Facebook data. The first author conducted the sessions to iteratively build a shared vocabulary combining the annotators' expertise in clinical care and behavioral analysis of social media data. Annotators compared different GMM model outputs based on the discernability of the components and whether the extracted components were comprehensive. Based on feedback from these sessions, we chose $K = 6$ as the final GMM configuration. Table 6.2 shows the names of the six PSS alongside the top representative features per PSS as well as example paraphrased Facebook posts, explaining the behaviors within the PSS. We used moderate levels of disguise [308] while paraphrasing posts in Table 6.2, i.e. identifying details (such as places) were changed and verbatim quotes were modified grammatically to safeguard privacy of participants. Figure 6.2 shows the frequency of occurrence of each PSS in our data.

1. **Self-regulation focused PSS:** involving lower engagement, reduced posting, and activity signaling a detachment from the online social network with no indications of disclosure about their mental health status (indicated by lower use of pronouns). This PSS also suggests boundary regulation strategies [309] while only sharing positive content on Facebook. This PSS is seen most frequently during Bf_hosp and is presented with lower frequency during other periods.

2. **Self-awareness focused PSS:** involving high self-referential thinking and pre-occupation indicated by the increased usage of pronouns [310] and posting of content related to mental health symptoms and conditions. The higher use of words related to anger, work, and money and showing a focus on the present and future in language demonstrate an understanding of the transition and subsequent consequences. This PSS also has reduced social interactions and posting about emotional content. This PSS is most salient during Bf_hosp but also persists during Bf_long and Af_long indicating continued self-reflection surrounding psychiatric hospitalization events [311].

3. **Sociality focused PSS:** involving heightened sociality demonstrated by the greater

Table 6.2: Derived possible self states (PSS) on Facebook surrounding psychiatric hospitalizations. The increasing behaviors indicate actions that are performed more as part of the PSS compared to its average level, while those decreasing indicate actions that are performed less as part of the PSS compared to its average level. Example increasing/decreasing behaviors for the same individual are indicated in blue. All Facebook posts are paraphrased to protect participants' privacy. *** indicate p-values for behaviors that are predictive of membership to the PSS component. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

PSS	***Increase* in behaviors**	***Decrease* in behaviors**	**Example posts and behaviors**
Self-regulation	Use of words indicative of positive emotions***. *"I just wanted to say that everyone danced so well tonight. I really enjoyed the performance."*	All other actions. E.g., use of function words, pronouns,first-person singular pronouns, show focus on the present in language, post content indicative of mental health symptoms and experiences, use of impersonal pronouns, show focus on the past in language*, third-person plural pronouns, words related to anger, body, work, sadness, death, leisure*, words indicative of negative emotions**, words about friends*. *"My roommate was watching OUAT, and I remember how quick Ruby was to condemn Regina."*	*"It is world humanitarian day. I'm doing something good, somewhere for someone else. Join me. #WHD2012 #IWASHERE"*
Self-awareness	Use of function words, personal pronouns, first-person singular pronouns, words related to cognitive processes, first-person plural pronouns, words related to anger**, words related to money*, show focus on the present in language, show focus on the future in language*, posting content indicative of mental health symptoms or experiences. *"I just can't sleep, I watched American Horror Story for the whole day. I promised myself I would wake up early and clean my room."*	Posts indicative of positive and negative emotions, sending messages to friends on FB, posting photos on FB, words related to leisure, one-on-one interactions on FB, words indicative of anxiety*, sharing feelings with status updates on FB**, shares on FB*. *Number of FB shares relatively dropped by 100%.*	*"It's just a slap in the face when you are your only sole motivation and advice giver... you have no one saying 'keep going', 'i'm proud of you"', "You work hard on your mental health to the point your new psychiatrist doesn't want you on meds anymore.", "Feeling accomplished and great.", "Hungry and bored again. blah!"*
Sociality	Overall posts and activities on FB, uploading photos and cover photos on FB, sharing feelings via posts on FB, likes on FB, one-on-one interactions on FB, sending messages on FB, co-tagged with others on FB, use of informal words. *Number of FB posts relatively increased by 341%.*	Post content related to mental health symptoms and experiences, use of second-person pronouns, first-person plural pronouns, words indicative of negative emotions, showing focus on the past in language, use of words related to leisure, adding new friends on FB. *"These winds are blowing down everything except the Trump tower."*	*"I have a lot of best friends lol so happy... national best friend day to everyone who are my best friend."*
Withdrawal	No action.	Use of function words, pronouns, posting content indicative of mental health symptoms and conditions, posting content indicative of positive emotions**, use of person pronouns, showing focus on the present in language, use of first-person singular pronouns, anger, words related to body. *"omg! You've got great hair styling skills sister."*	*"[The user] went to [a certain music festival]", "[The user] added education to his timeline", "[The user] added [a city] to his current city."*
Re-adjustment	Use of words related to leisure, sexual words, words related to work, ingestion***, pronouns, function words , informal words, co-tagging with others on FB, words related to anger, death, adding new friends on FB, use of words related to health***, showing focus on the past in language***, sending messages to friends on FB***, shares on FB***. *"I'm on the verge of a manic episode. WHAT DO I DO?"*	Only use of third-person singular pronouns. *"Woot! commemorating my 7th good hair day in a row"*	*"Friends always: fight for you, include you, respect you... stand by you. People believe your actions more than your words.", "I am not my hair. I am not this skin. I am not your expectations. I am a soul that lives within", "#oldclassmates reunion", "#fuckedup", "#amen".*
Incoporation	Showing focus on the present in language*, words indicative of negative emotions***, posting content indicative of mental health symptoms and experiences, us of informal words**, function words, pronouns, sending messages to friends on FB, use of personal pronouns, first-person singular pronouns, words related to religion**, body one-on-one interactions on FB, use of words about friends***, health**, sharing feelings with status updates on FB***. *"my girl..I need your help pls get back to me as soon as possible"*	Use of words indicating positive emotions, co-tagging with others on FB, overall posting and activities on FB, use of words related to anxiety*. *"Aww! Thanks for the feel better card!"*	*"Anyone with a TV watching the movie Avengers I might be able to join?", "Anyone coming from [a location] that might be able to give me a ride to [another location]?", "Sorry I missed your show last night. Make sure you keep me posted with everything going on."*

volumes in posting status updates and photos, sending messages, one-on-one interactions, and other activities on Facebook. This PSS only appears during Bf_hosp.

4. **Withdrawal focused PSS:** involving an overall reduction in posting and activities on Facebook. As part of this PSS, people use significantly fewer function words and personal pronouns and reduce posting content indicative of mental health symptoms and experiences. The overall lack of activity and engagement shows that people are possibly withdrawing from the active use of social media. This PSS is most salient during Af_hosp.

5. **Re-adjustment focused PSS:** involving increased sociality, such as adding new friends, sending messages, and co-tagging others. But this PSS also reveals more self-attentional focus and expressive behaviors, shown by usage of words about anger, health, and death without inhibition. This PSS is most salient during Af_hosp and is less visible during other phases.

6. **Incorporation focused PSS:** involving the inclusion of experiences and narratives related to mental health transition into peoples' online social lives. The increased use of emotional content and personal pronouns, showing a focus on the present and sharing content about their mental health status indicates self-focus, awareness and disclosure of experiences on Facebook. On the other hand, this PSS also shows signs of increased sociality demonstrated by messages shared and one-on-one interactions with others. This PSS is most salient during Af_hosp and persists with lower frequency before.

Figure 6.3 shows a heatmap of membership probabilities of each PSS around the four periods surrounding the hospitalization. In essence, annotators found that the derived PSS captured the heterogeneity in peoples' experiences along two main themes: symptomatic expression before hospitalization and social reintegration after hospitalization. Self-regulation, self-awareness, and sociality focused PSS that are most salient before hospitalization reveal different manifestations of peoples' social media use and disclosure levels regarding their mental health condition online. While some people might choose to share their experiences related to mental health on Facebook (as in the self-awareness focused PSS), others might regulate and censor content regarding their mental health status transition (self-regulation focused PSS). This variability in social media use is also noted in prior work in people managing depression [312] and disclosures about schizophrenia diagnosis [186]. Our expert annotators also pointed out that the various PSS that are salient before hospitalization show varying levels of awareness and insights that people might have while experiencing mental health symptoms. For instance, during psychotic and manic episodes, that are most commonly experienced by individuals diagnosed with schizophrenia and mood disorders respectively (the majority of our study population), individuals are known to have distinctive levels of awareness, pre-occupation, and emotion regulation[313]. Annotators also identified the withdrawal, re-adjustment, and incorporation focused PSS that are most salient after hospitalization periods as different trajectories into social reintegration. While some people might withdraw themselves away from social technologies after discharge from the hospital to cope with the stigma and repercussions [276], others might increase activity on these platforms to get back to their online social lives and create a "new normal" [118].

Although five of the six PSS occur with sizable frequencies, the sociality PSS occurs very infrequently. Examining the model, we found that the behaviors in this component were outliers and always captured as a separate cluster is all GMM models. We omit this PSS in subsequent analysis due to the infrequent occurrence and lack of discernable behaviors grounded in literature.

6.1.3 Results

Per RQ2, we now study the different trajectories of transitions in PSS experienced by people surrounding psychiatric hospitalizations; for this, we use the above derived taxonomy

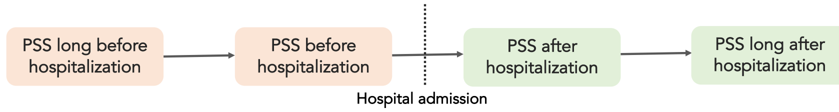


Figure 6.4: Temporal chains used to generate consecutive possible selves statuses for modeling transitions.

of PSS. Prior work has noted that such transitions can be "complex, nonlinear, sometimes cyclical and potentially recurring" [113]. As an initial step, we focus on individual hospitalization events in a person's entire journey with mental illness. We propose a framework that models linear transitions surrounding hospitalizations as a Markov process [314] with temporal chains generated to connect consecutive PSS (Figure 6.4). Based on the temporal chains, we compute the transition probabilities, i.e., given a PSS is presented at time period t, the probability that the person would present any of the six PSS at time period $t + 1$. Here t indicates the four temporal periods surrounding hospitalization: long before (Bf_long), before (Bf_hosp), after (Af_hosp), and long after (Af_long). Each temporal phase is assigned a PSS if that PSS has the highest probability compared to others.

The top-10 likely PSS transitions across all four periods surrounding the hospitalization are shown in Figure 6.5. First, we notice stability in the self-awareness (cond. probability = 0.48), self-regulation (cond. probability = 0.26) and incorporation focused PSS (cond. probability = 0.36), that are carried over from one temporal period to the next; those who present these PSS are more likely to maintain it in the future. The constancy in these PSS is also supported by the fact that self-awareness and incorporation focused PSS are the most common ones that people transition into – 37.5% of re-adjustment, 23.3% of self-awareness and 27% of self-regulation focused PSS transition into incorporation focused PSS, and 27.9% of incorporation, 23.9% of withdrawal and 28.1% of re-adjustment focused PSS transition into the self-awareness PSS.

Notably, those who exhibit the re-adjustment PSS have a conditional probability of 0.37, transition into the incorporation PSS. This transition shows individuals moving from a focus on the past to focusing on the present which is consistent with literature on major

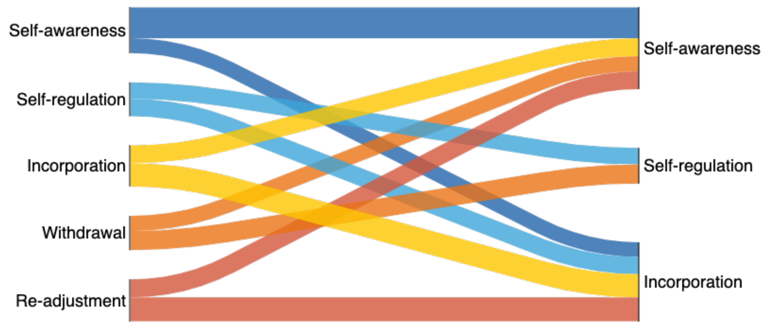
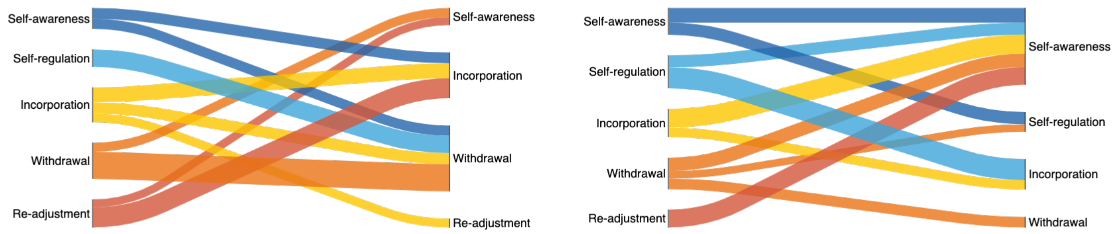


Figure 6.5: Top 10 transitions between PSS that were likely to be seen around hospitalization periods.

life transitions: reintegration is a process that first involves confronting life with the illness and then reconstructing life with the illness [111]. This transition also reflects a shift in boundary regulation practices [309]. While the re-adjustment focused PSS has heightened posting and activities on Facebook overall (every behavior is performed more than the average amount), people move past this PSS into the incorporation focused PSS which has moderate activity and posting on Facebook. Similar boundary regulation is also seen in people transitioning from the withdrawal focused PSS to self-regulation. Here, those who have completely disengaged from the platform (evident from the overall lower activity and posting), transition back into actively using the platform by posting positive content.

Next, we present a more in-depth analysis of transitions focusing on two important junctures around the psychiatric hospitalization. 1) The first involves recovery trajectories captured by the transitions from before to after the hospitalization (Bf_hosp to Af_hosp) revealing experiences related to clinical inpatient care. 2) The second centers around reintegration trajectories captured by those transitions that are from after to long after hospitalization (Af_hosp to Af_long) revealing experiences related to post-hospitalization social care and getting back to "normal" life outside of their clinical status. As noted before, a holistic understanding of clinical recovery and social reintegration is an essential aspect of understanding the mental health experience.



(a) $Bf\_hosp \rightarrow Af\_hosp$ (b) $Af\_hosp \rightarrow Af\_long$

Figure 6.6: (a) Top 10 transitions between PSS likely to be seen around recovery i.e. transitions from Bf_hosp to Af_hosp. (b) Top 10 transitions between PSS likely to be seen around reintegration i.e. transitions between Af_hosp and Af_long. Nodes indicate PSS and width/thickness of the flow is proportionate to transition probability.

Using Possible Selves to Understand Recovery Trajectories Figure 6.6(a) shows the 10 most likely transitions seen around recovery trajectories i.e. the PSS exhibited before a hospitalization to the PSS exhibited after the hospitalization. First, we see stability in the presentation of withdrawal and incorporation-focused PSS; those who exhibit these PSS Bf_hosp are likely to maintain it Af_hosp. The stability of this PSS is also confirmed by the fact that withdrawal and incorporation focused PSS are the most common trajectories we see Af_hosp – 48.3% of self-regulation, 31% of incoporation and 28% of self-awareness focused PSS Bf_hosp transition into withdrawal focused PSS, and 56% of re-adjustment and 30% of self-awareness focused PSS transition into the incorporation focused PSS Af_hosp.

However, there is a difference between transitions that end in withdrawal focused PSS and those that end in the incorporation focused PSS. This depends on the whether the event was the first hospitalization experienced by the individual. Considering only those who have portrayed the incorporation focused PSS Bf_hosp, 66.6% of the transitions from incorporation to withdrawal focused PSS are seen during the first hospitalization events experienced by people. On the other hand, the majority of the cases where individuals remain in the incorporation focused PSS before and after the hospitalization (58.3%) are observed during subsequent hospitalizations (like the 2nd, 4th, or 5th hospitalization recorded for the individual). This reveals that while those who have been re-hospitalized can maintain an incorporation focused PSS, the first hospitalization experience leaves most people transitioning into the withdrawal focused PSS. This could be attributed to the stigma, isolation, or "other"-ing experience related to first time psychiatric hospitalizations [315]. This finding is also consistent with prior literature on mental health care that suggests that negative experiences associated with the first psychiatric hospitalisation remains significant even after many years have passed [100]. In a focus group study with participants from 6 countries, [100] note that psychiatric rehospitalizations encompassed some amount of familiarity; however, participants described the first hospitalization as "something shocking, intolerable, and terrible" [100]. For example, consider participant `A` whose recovery transition showed to shift from a self-awareness focused PSS to a withdrawal focused PSS after their first hospitalization. Participant `A` posted 15 status updates on Facebook during the two weeks prior to their hospitalization, such as the following paraphrased posts: *"Live as if you were to die tomorrow and learn as if were to live forever; The only competition you'll ever face is with your own ignorance."* However, since their hospitalization `A` made no posts on Facebook, except accepting new friends requests. In contrast, consider participant `B`. Around their fifth hospitalization `B` showed a recovery transition from re-adjustment focused PSS to reintegration focused PSS with paraphrased posts such as *"Don't lend people money because that person might be a *** and not pay you back."* before hospitalization to posts such as *"my brother..the love we share is eternal. You live on in your loved ones memories and hearts."* after hospitalization.

Using Possible Selves to Understand Reintegration Trajectories As a second deeper dive into recovery trajectories, Figure 6.6(b) shows the top 10 transitions that begin after the hospitalization and end long after the hospitalization. The post-discharge period after hospitalization involves a shift in focus from institutionalized treatment, to coping and management of illness. This transition is accompanied by mechanisms to break inhibitions and stigma, open up about their experiences with intimate others, reach out for social support and improve overall well-being [91]. The two most common trajectories we observe during the post-hospitalization period end in the self-awareness or incorporation focused PSS. While both of these PSS involve self-referential thinking and self-reflection, only the incorporation focused PSS shows signals of social functioning (such as messaging friends, one-on-one interactions, co-tagging, etc.) that are considered positive signs of reintegration. Among people who transition into the incorporation focused PSS, 53.8% transition from the self-awareness focused PSS, and 24% are those who have maintained the incorporation focused PSS. Studies from the psychiatry literature report that the stigma related to hospitalization could be transitory in cases where people resume occupancy of normal societal roles [316] . This suggests that the transition from self-awareness focused PSS to incorporation PSS is a positive reintegration trajectory where people move past the uncertainty post-hospitalization and re-establish their online social connections as part of the incorporation-focused PSS.

Next, we find that every PSS exhibited immediately after the hospitalization (Af_hosp) has a transition ending in the "self-awareness" focused PSS long after the hospitalization (Af_long). While self-focus demonstrated by the awareness focused PSS is beneficial, excessive self-reflection is considered a central feature in mood and anxiety disorders [317], and greater focus on self-concept is also related to internalized stigma in mental health [318, 319]. In contrast to the social stigma about mental illness that entails discrimination, negative stereotypes, loss of opportunities, etc., self-stigma relates to awareness, agreement, and application [320] – a person with mental illness must first be aware of corresponding stereotypes before agreeing with them and then apply self-stigma to one's self. A behavioral consequence of self-stigma is social avoidance [320, 321], that is the person may avoid situations where they might feel publicly disrespected because of self-stigma and low self-esteem. Awareness, reflection and social avoidance that are related to self-stigma are representative features of the self-awareness focused PSS, that most individuals are found to transition into long after hospitalization (Af_long). Based on this literature, the majority of reintegration trajectories we see ending in the self-awareness focused PSS might be reflective of the long term self-stigma, that happens as a consequence of psychiatric hospitalizations. For instance, participant C who transitioned into a self-awareness focused PSS long after their hospitalization posted on Facebook about the stigma they have been experiencing, saying (paraphrased): *"I think the stigma about mental health really needs to be broken. I'm tired of every single person treating me like the plague. For the last year, I have been dealing with bi polar disorder. It was the scariest thing I ever dealt with but I learned that I'm not alone. People automatically treat you like your a different person, but don't show me your sympathy. Treat me the same. I'm stronger than I've ever been."*

Predictive framework to access reintegration Given the importance of reintegration from both clinical and social perspectives, in this section, per RQ3 we set up a prediction framework to assess how individuals reintegrate after psychiatric hospitalizations. By doing so, we demonstrate the utility of the taxonomy of PSS (RQ1) and the empirical framework capturing their transition trajectories on Facebook (RQ2).

We consider the incorporation focused PSS, which involves narratives related to mental health and people re-establishing social interactions after hospitalization, as a PSS-based operationalization of social reintegration. To examine the relationship between PSS trajectories and reintegration post-hospitalization, we consider models regressing on the probability of the phase long after hospitalization (Af_long) being predicted as the incorporation focused PSS by the GMM model. We choose Af_long as the focus of prediction, as we expect reintegration to be the final stage of mental health transitions experienced by the person. As covariates, we include the behavioral signals from the preceding phases (ref. Section 4.2.2), such that the model reveals how PSS-based behavioral signals in the past (before and after the hospitalization[2]) predict the probability of reintegration in the future

[2]As information about the past PSS, we consider a weighted average of features before and after hospitalization.

Table 6.3: Summary of regression models. Null is the intercept-only model, the first baseline. M_hosp is the second baseline deriving reintegration probabilities based on evidence of a future re-hospitalization. All comparisons are made with the Null model. p-values reported at $p < 0.01$.

	Ridge	Lasso	D Trees	M_hosp	Null
Pearson's r	0.46	0.29	0.38	0.05	¡ 0.01
MSE	0.08	0.12	0.09	0.16	0.12
R^2	0.35	0.19	-	0.20	-
χ^2	82.80***	34.13	6.77	20.03	-

(long after hospitalization or Af_long). We standardize all covariates and use the variance inflation factor (VIF) to eliminate multicollinearity. The dependent variables i.e., probability of belonging to the incorporation-focused PSS are log-transformed.

We consider several non-linear (decision tree regressor) and linear regression models with regularization (Lasso, Ridge and Elastic Net) and use grid search for parameter tuning. We use k-fold ($k = 10$) cross-validation approach to iteratively train the model and predict on held-out data. We collate the predictions, and obtain the pooled model performance measures – including Pearson's correlation r and Mean Square Error (MSE) to evaluate predictive accuracy and R^2 value to evaluate the model fit. For model performance comparison, we consider two baselines: first, an intercept only model (Null), which assumes a constant probability of reintegration irrespective of the covariates. As a second baseline (M_hosp), we consider the probability of reintegration based on the likelihood that a person has exactly one hospitalization. This model assumes that a lack of re-hospitalization indicates successful reintegration. We regress covariates on probability = 1, for those with no re-hospitalizations based on medical records (and probability = 0 for those with re-hospitalizations).

Table 6.3 shows summaries of the best performing models. We find that the Ridge regression model has the best performance compared to other models. Compared to the Null model, the Ridge model provides considerable predictive power (shown statistically significant based on χ^2 tests on model performance on held-out data). To reject the possi-

Table 6.4: Summary of covariates of best performing Ridge regression model on reintegration. p-values use Bonferroni correction (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

Covariate	Estimate
LIWC Category: Negative affect	0.06*
LIWC Category: 1st P Sing Pronoun	0.23**
LIWC Category: 2nd Person Pronoun	0.18***
LIWC Category: 3rd Person Pronoun	0.16**
LIWC Category: Temporal Focus, Past	0.05*
Facebook: Photos Shared	-0.06*
Facebook: Number of Shares	1.35**
LIWC Category: Certainty	0.12*
LIWC Category: Differ	-0.1*

bility that the performance of the Ridge model is by chance, we run permutation tests [322] to reject the null hypothesis that a randomly generated vector of reintegration probabilities will perform better than the Ridge model. We run 10,000 permutations of randomly generated dependent variables, and find that the probability (p-value) of improvement by a randomly generated vector is 0.003. This rejects the null hypothesis and reveals statistical significance in the observed improvement by the Ridge model.

Next, we present findings about the most predictive covariates from a Ridge regression model,[3] Table 6.4. First, sharing content that is emotional (*negative emotion, estimate = 0.06***) and personal, including personal pronouns like 'i' (*1st person singular, estimate = 0.24***), 'you' (*2nd person, estimate = 0.18****), and 'they' (*3rd person, estimate =0.16***) during the periods before and after hospitalization is predictive of reintegration long after hospitalization (Af_long). This finding is reflective of literature from the expressive writing paradigm that observed long term benefits of expressive writing in the form of both health outcomes (less stress related visits to the doctor, improved mood, lowered blood pressure, fewer post-traumatic intrusion and avoidance symptoms [323]) and social/behavioral outcomes [218]. Recent work has also revealed how social media is increasingly appropriated for sensitive disclosures related to mental health to obtain support and therapeutic benefits [185]. Although the level of disclosure might vary in our

[3] Note that regression coefficients should be interpreted conditional on the penalty.

Table 6.5: Differences in features between those who show high vs. low likelihood of reintegration. Reported measures are mean feature values per group, Kruskal Wallis test statistic H, Mann Whitney test statistic U, and effect size (Cohen's d). p-values use Bonferroni correction (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

Feature	Low	High	H	U	d
LIWC Category: Negative Affect	-0.13	0.05	7.93***	2226.5***	-0.32
LIWC Category: Home	0.14	-0.02	3.98*	2449*	0.10
LIWC Category: Religion	-0.06	0.29	4.60*	2405.5*	-0.38
Facebook: Updates to Profile Fields	-0.05	0.02	6.67**	2337.5**	-0.33
Facebook: Number of Messages Shared	-0.16	0.05	3.94*	2452*	-0.31
Facebook: Number of Likes Received	-0.04	-0.02	2.55	2576.5*	-0.05
Facebook: Number of Shares	-0.06	-0.05	3.77*	2516*	-0.31
LIWC Category: Negation	-0.07	0.02	2.48	2568*	-0.16

patient participants' social media data compared to explicit, public disclosures of mental illness [186], we find that expressive behaviors on the platform are associated with reintegration. Next, we find that showing a focus on the past in posting language (*estimate = 0.05**) during the hospitalization, is predictive of reintegration long after hospitalization (Af_long). Drawing from the Possible Selves framework [281], an emphasis on the past-selves could indicate a focus on getting back to what was "normal" in the past and creating a "new normal" after the transition due to hospitalization. Finally, certainty word usage (*estimate = 0.12**) around the hospitalization, demonstrates heightened emotional stability and is predictive of reintegration after hospitalization.

We finally ask the question: what are the differences between those who show high reintegration signals on Facebook and those who do not? For this, we conduct post-hoc Kruskal Wallis and Mann Whitney tests to examine feature means between the two groups. We split observations into low and high reintegration based on the probability of belonging to the incorporation focused PSS being greater than chance (probability by chance is one in six PSS = 0.16. High reintegration = $p > 0.16$ and low reintegration = $p < 0.16$). Table 6.5 shows the statistically significant differences between the two groups. Those who have a higher likelihood of reintegration display more negative emotion and share content related to religion more than those who have a lower likelihood. Disclosure, emotional support, and faith are known to act as buffers against negative effects of stressful life events, which likely help with coping and reintegration after the event [324]. We also find that those who show reintegration signals stay in touch with friends on Facebook via messages. In contrast, those who reduce interactions via messages are less likely to show reintegration signals. The implications of these findings are discussed in Section 6.3.

6.2 The reintegration journey following a psychiatric hospitalization: Examining the role of social technologies

> *"I joined a whole bunch of new groups [after psychiatric hospitalization] to try to make sure that my Facebook feed was nourishing me, and not strangling me."* [P1]

For mental health, both clinical recovery and social reintegration need to go hand in hand for the overall well-being of individuals [91, 278, 104]. Particularly, for individuals who have experienced a psychiatric hospitalization, recovery is viewed as "*a personal journey rather than a set outcome*" [275], and involves not only removal of symptoms and restoration of functioning but also social reintegration, referring to "*the degree to which an individual's social network reflects adequate size and multiple social roles (e.g., as friend, family member, coworker) and the extent to which an individual engages in mutual exchange, or reciprocity, in social relationships*" [325]. Reintegration is therefore viewed as a journey, during which the focus shifts from institutionalized treatment and medication to self-management of illness. It is accompanied by mechanisms to break inhibitions and stigma around mental health, open up about an individual's experiences with intimate others, reach out for social support, re-enter educational or occupational roles, and improve overall well-being. Successful reintegration across these dimensions is a crucial marker for recovery and an important goal for mental health policy, more broadly [326, 327]. However, stigma, negative consequences, and lack of access to resources for support and care present challenges to individuals who experience a psychiatric hospitalization to get back to their social lives.

Despite the central role of social reintegration, the topic has received relatively less attention in mental health research and practice. From a clinical perspective, after discharge from the hospital, clinicians often lose timely contact with their patients which present challenges for continued care and support [328]. Even when patients adhere to clinical follow-up appointments, the emphasis is on reduction of symptoms via continued therapy or medication management, rather than on improvement of other aspects such as social relationships, employment, education, and leisure [329]. This emphasis on clinical recovery alone has received criticism from scholars, as it tends to objectify the person with mental illness [330], disregards their sufferings and identity outside of the clinical definition of the illness [331], reduces patients' participation in recovery [332], and fails to consider the challenges people need to overcome beyond symptom management to get back to life following hospitalizations [110]. Furthermore, recovery journeys are known to have a high likelihood of re-hospitalization if reintegrating back to social life and roles becomes challenging. Thus, the guiding principle of mental health policy in many countries – the recovery model [105] and patient-centered model of healthcare [333] – posit keeping the patient at the center of decision-making in recovery. Along these lines, we argue that adopting such a "whole-person perspective" and developing a holistic understanding of clinical recovery and social reintegration can provide a person more agency and control in the management of their condition, following a psychiatric hospitalization.

A significant portion of people's social lives is now technology-mediated through the use of social networking sites, online communities, messaging applications, and so on. Especially for young adults, a demographic most susceptible to onset of mental health conditions [334], technology and online presence form an integral part of their identity and social lives [161]. Furthermore, social technologies have emerged to play an important role in mental health – as spaces for disclosure [2], social support [335], raising awareness [336], fighting stigma [1], and enabling clinical interventions [7]. A growing body of literature in human-computer interaction (HCI) and computer-supported cooperative work and social computing (CSCW) identifies the role of social technologies in recovery and management of health conditions [155, 337, 338]. In parallel, HCI/CSCW researchers have studied how social technology play an important role during major life transitions and help individuals establish a "new normal" [118, 86, 339], conduct identity work [121, 120] and reach out to similar others [122]. However, health transitions like psychiatric hospitalizations are less explored under this lens, and little work has been done to understand how technology-mediated social interactions impact reintegration and recovery journeys in mental health. As noted in P1's quote above, like other major life transitions [340, 339, 341], people often turn to social technologies during reintegration after hospitalization to navigate shifting mental health goals and social networks and to access support resources. Understanding the role of social technology in this particular life transition, mental health reintegration, can improve our understanding on how platforms support or hinder people's social lives and health management after events like psychiatric hospitalizations.

In this study, *we examine the role of social technology as people get back to their social lives and negotiate the transition they experience due to the psychiatric hospitalization.* Specifically, we ask:

RQ1. How do people get back to their social lives after experiencing a psychiatric hospitalization?

RQ2. What is the role of social technologies in people's reintegration journeys after experiencing a psychiatric hospitalization?

To answer these questions, we conducted semi-structured interviews, spanning over nine months, with 19 adults who had experienced a psychiatric hospitalization for schizophrenia spectrum disorder, mood disorder, bipolar, borderline personality or anxiety disorder in the recent past. Participants actively used at least one social technology platform like Facebook, Twitter Reddit, Snapchat, Whatsapp, Tumblr, etc. We examined the data adopting a hybrid inductive and deductive approach to thematic analysis. In doing so, we contribute an understanding of people's offline and online social lives after psychiatric hospitalization in the context of managing the illness.

We found that participant's social lives after hospitalization were deeply intertwined with factors linked to self-management of the mental health condition, such as stigma, inhibitions, and over reliance on others after the psychiatric hospitalization [258]. We identified different approaches participants adopted to re-establish social connections immediately after discharge from the hospital, often driven by a sense of urgency, obligation and stigma. Social technology platforms mediated people's interactions after the hospitalization, providing spaces for disclosure, social support, and sources for positive behavioral change. While participants drew several social benefits from social technology use, some felt that their use of these platforms hindered their path to reintegration, due to feelings of social comparison, negative interactions, and emotional triggers to their mental health symptoms. We discuss the theoretical implications of social technology use/non-use for reintegration in an individual's recovery journey and highlight the clinical implications for post-discharge care and design suggestions for social technology to support reintegration following a major life transition like a psychiatric hospitalization.

Privacy and Ethics. This research was conducted with approval from the Institutional Review Board. Further information on approaches taken to protect participant privacy, safety, and accurately represent the lived experiences of participants without compromising anonymity can be found in Section 6.2.1. Since the topic of this study concerns mental health and psychiatric hospitalizations, some quotes and descriptions of participants' lived experiences may be triggering to readers. We suggest caution while reading, printing, or disseminating these findings.

6.2.1 Methods

We designed an interview study to investigate how people who experienced a psychiatric hospitalization got back to their social lives post-hospitalization (RQ1) and the role of so-

Table 6.6: Participant demographics, including self-reported diagnosis of mental health condition, duration and place of their last psychiatric hospitalization.

ID	Gender	Age	Education	Race	Self-reported diagnosis	Time	Place
P1	W	46	some college	White	Bipolar Disorder	9 d	SC
P2	M	30	bachelor's	Black or African American	Anxiety Disorder	1 wk	CA
P3	W	24	some college	Middle Eastern or North African	Anxiety Disorder	2 mo	NY
P4	W	24	some college	Black or African American	Anxiety Disorder	1 wk	NY
P5	M	27	bachelor's	Black or African American	Borderline Personality Disorder	2 mo	CA
P6	W	25	some college	Black or African American	Anxiety Disorder	3 wk	GA
P7	W	25	bachelor's	Black or African American	Anxiety Disorder	2 wk	TX
P8	M	25	bachelor's	Native Hawaiian or Other Pacific Islander	Anxiety Disorder	3 mo	TX
P9	W	37	Master's	Some other race, ethnicity or origin	Depression and Anxiety Disorder	3 wk	TX
P10	W	25	bachelor's	Black or African American	Depression and Anxiety Disorder	4 mo	CA
P11	M	25	bachelor's	Black or African American	Mood Disorder	6 mo	TX
P12	W	27	bachelor's	White	Depression and Anxiety Disorder	4 wk	NY
P13	M	30	bachelor's	Black or African American	Borderline Personality Disorder	1 wk	TX
P14	W	28	bachelor's	Black or African American	Anxiety Disorder	1 d	VA
P15	W	21	some college	Black or African American	Schizophrenia	1 mo	NY
P16	W	22	bachelor's	White	Schizophrenia	15 d	NY
P17	M	38	bachelor's	Black or African American	Depression and Anxiety Disorder	2 wk	AL
P18	W	34	bachelor's	Black or African American	Postpartum Psychosis	1 mo	NC
P19	M	37	master's	Some other race, ethnicity or origin	Schizophrenia	1 wk	GA

cial technologies in this process of reintegration (RQ2). We conducted semi-structured interviews with 19 adults (ages 21-46 years; M = 28.9 years, 63% women) who have experienced a psychiatric hospitalization between 2009 and 2020. In this section, we describe our recruitment methodology and analysis process and discuss the ethics of our work.

Recruitment and Participants We used four channels to recruit participants for the interview study: 1) clinician referrals within a large health care system, 2) local Craigslist ads, 3) online social media platforms, 4) online mental health support communities. We collaborated with clinician researchers and practitioners in a large healthcare organization in the north-east of the United States who posted recruitment flyers around their centers and contacted potential participants regarding the research study. We also posted recruitment ads on Craigslist, shared the call for participation on Twitter and online mental health support communities on Reddit with moderator approval. We recruited from all four channels in parallel until we had reached a point of theoretical saturation [342]. Participants were eligible if they were adults between the ages 18 and 65 who experienced a psychiatric hospitalization for a diagnosis of schizophrenia spectrum disorder, mood disorder, bipolar, borderline personality or anxiety disorder and who had an active account on at least one social technology platform (e.g., Facebook, Twitter, Reddit, Snapchat, Tumblr.) We chose these mental health conditions because of the significant importance of social functioning and reintegration for clinical recovery, the lifelong management of the condition and the high likelihood of relapse.

We sent out a brief screening survey with the recruitment call for eligible participants to sign up for participation in the study. Participants self-reported their hospitalization, diagnosis of mental health condition and social technology use via the screening survey. Participants were also required to provide an email address so that they could be contacted for scheduling and compensation. We offered participants a $25 Amazon gift card as a token of appreciation.

The screening survey was active September to November 2020 and we received a total of 138 responses. Among survey respondents, 42 were eligible for participation and they were contacted via email with study information and an online consent form. I conducted remote interviews with 19 consented adults within the U.S who experienced at least one psychiatric hospitalization. The average age of participants was 28.9 and 63% identified as women, 11 participants reported a diagnosis of anxiety disorders, 2 reported borderline personality disorders, one participant reported a diagnosis of bipolar disorder and 4 participants reported schizophrenia form disorders. Participants had experienced psychiatric hospitalization (in-patient or emergency room facilities) for time periods ranging from 1 week to 6 months (average = 38 days, std = 46 days, median = 21 days) between 2014 to 2020 across the United States including Alabama (1), California (3), Georgia (2), New York (5), North Carolina (1), South Carolina (1), Texas (5), and Virginia (1) (Refer Table 6.6.) Participants' reported reasons for the psychiatric hospitalization included escalation of symptoms related to their mental health condition, management of medication, as well as high risk adverse experiences related to self-harm, suicidal ideation, trauma and postpartum psychosis. All participants reported using at least one social platform, with Facebook, Instagram and Whatsapp being the most commonly used ones.

Participant Safety and Risk Mitigation Measures The study was conducted following Institutional Review Board (IRB) approval and informed consent from the participants. To ensure participant safety, as a part of the consent process, participants were clearly told that they are free to end the interview at any time, and to let the interviewer know if there are parts of their disclosure that felt too sensitive or deanonymizing for publication. Additionally, following Draucker et al. [343], after particularly overwhelming questions (e.g., those on past suicidal ideation or self-harm), participants were briefly asked after answering if they felt okay and wanted to continue the interview. Further, our consent form included links to prominent mental health resources like 7 Cups of Tea[4], Crisis Hotline[4], Crisis Text Line[4], and National Suicide Prevention Lifeline[4], which we encouraged our participants to use if the interview left them emotionally overwhelmed. To protect privacy, all personally identifiable information have been deidentified or obfuscated in our reporting of the findings.

Data Collection Following our approved IRB protocol, we collected data for this study from October-November 2020. We conducted semi-structured, remote interviews via video or phone call based on the participant's preference. We incorporated remote interviews to extend the reach of the study and for safety concerns during the COVID-19 pandemic. We developed guiding interview questions by drawing from literature on clinical recovery and social reintegration and input from the mental health clinician collaborators. I conducted the interviews using a video-conferencing software approved by the IRB and the author's academic institution, including for the local participants, due to the ongoing COVID-19 pandemic at the time. I began the session by informing participants the study goals, the risks and benefits of participation and asking for their permission to record the session. Then participants were asked to walk us through the day they were discharged from psychiatric hospitalization and what followed next. In cases where participants experienced more than one psychiatric hospitalization, we asked them to pick a hospitalization expe-

[4] www.7cupsoftea.com, www.imalive.com, www.crisistextline.org/, suicidepreventionlifeline.org

rience that they considered prominent and to answer all subsequent questions around that hospitalization. Follow up questions focused on getting back to social lives, disclosure of experiences related to mental illness, social support, general social technology use and changes in use surrounding hospitalization. When possible, we asked for specific examples and probed participants to understand the role of social technologies during their reintegration journeys after hospitalization. Each interview lasted approximately 60 minutes. Only audio was recorded and then transcribed for analysis using Otter.ai software for transcription services.

Qualitative Data Analysis To analyze the interview data, we followed an integrated inductive and deductive approach to thematic analysis to combine data-driven codes with theory-driven ones [344]. The analysis began with the inductive part – open coding of the transcripts independently by three researchers to identify patterns in data and establish a thematic framework. The themes were then organized into an initial codebook. The team met frequently to resolve disagreements, discuss emerging concepts, and refine the themes. This coding process resulted in the formation of 10 themes such as "transitioning from hospital to home," and "online social support." We consolidated and organized these themes to highlight the social lives of people after psychiatric hospitalization and how management of a mental illness and technology use intertwines with people's social lives during the reintegration process.

Limitations As with other qualitative work with similar research goals and methodological orientation, our findings are limited in their generalizability. While, we sought for a diverse pool of participants with four different recruitment channels and involving people with different mental health conditions, our study sample is not representative of individuals in the U.S. who have experienced psychiatric hospitalization. Similarly, we only included participants who have been diagnosed with schizophrenia form disorders, anxiety, mood, bipolar, and borderline personality disorders, and did not identify experiential differences based on diagnostic type. Future work can evaluate and extend our findings with other populations, including people with different mental health conditions, people who have experienced very long hospitalization periods, and people in different countries and cultures. All of our participants mentioned using at least one social technology platform actively. Therefore, our results on the role of social technology in reintegration after psychiatric hospitalization are not representative of how these technologies impact reintegration journeys of people who do not use them actively or at all. Despite these limitations, our work presents first insights into how social technologies support and hinder people's social lives after psychiatric hospitalization.

Positionality This research has been conducted by a team diverse in many ways. In terms of academic, disciplinary, and professional backgrounds, our team includes social computing and human-computer interaction (HCI) researchers, as well as clinical psychologists and psychiatrists. Our team is also demographically diverse, including people of color, those holding LGBTQ+ identities, and immigrants. Notably, the team includes members with lived experience of mental illness as well as those who interact with such individuals on an everyday basis as part of their (clinical) profession. Together, our team holds a profound commitment to mental health research and practice, critically considering the potential offered by technology and computational artifacts in mental health, whether from the perspective of benefits or from that of harms. Therefore, we collectively recognize the emotional labor it takes on the part of a researcher to conduct research that involves a marginalized population, that centers around a highly sensitive topic, and whose social interpretations are shaped by demographics and culture. These personal and professional experiences have influenced both the questions we ask and the analytic lens we have adopted in this work.

6.2.2 Results

Social lives after psychiatric hospitalization

Transitioning from hospital to home. Psychiatric hospitalizations are considered life-altering experiences, as the admission often implies that individuals are unequipped to manage their psychiatric needs and require removal from their existing environment to receive appropriate, urgent care [95]. While admitted in the hospital, an individual's role as a patient is often defined by the many rules and rituals set by the hospital that they have to follow, including initially being in a locked ward that they cannot leave at will, and following a schedule for their meals, treatments, and activities. Importantly, in many cases, there is a lack of access to technology, social support, and offline connections [95, 98]. Participants' transition from the restrictions of the hospital and their role as a patient to their own home was often described as a significant social re-adjustment.

Some participants expressed feeling a sense of freedom as they transitioned from the hospital back to their home. This was indicating everyday mundane activities like '*sleeping in my own bed*' (P1, P2, P16) or '*having a cigarette,*' (P1) but also a shift in power and control over other activities and interactions. For example, P1 notes that after leaving the hospital, she was not being told what to do by others anymore and this was a sudden transition back to social relationships where she had equal power. Similarly, P2 shared feeling an immediate 'feeling of freedom' and being 'back in control' as soon as he was discharged. P2 said he felt this way because he disliked hospitals and the in-patient experience and was waiting to get back home.

> "*The first thing that came through my mind even just before the tests and after I was discharged. I don't know, that feeling of freedom.*" [P2]

This sense of freedom was reflected in not only how participants could act, but also in their social interactions. P11 also echoed disliking the hospitalization experience and feeling relieved to see family and having the freedom in choosing who they interacted with.

However, this sense of increased control and freedom was not shared by all participants. The transition from the hospital to home was drastic for some participants. P12 shared how factors at home contributed to her hospitalization. She expressed fear that interacting with people at home would make it challenging for her to manage and cope with the mental health condition. P12 said,

> "*The first thing that came to my mind was, how do I cope? How do I get back to the reality that I left at home? Because there's no escaping the reality that was at home. There was no escaping the people. I was scared that I'd get depressed, all over again. But somehow, somehow, sticking it out a minute at a time, a day at a time, I was able to cope.*" [P12]

P18 shared that she did not have anyone to pick her up after the discharge, so she reached home in a hospital van. Upon reaching home, she learnt that her family kept her from living in the same house and she found herself homeless for a bit after the hospitalization. She said,

> *"And I was homeless for a little bit because my own husband has kicked me out. And so, you know, it was just really frustrating. And, yes, it's this frustrating because it's a lot going on, and happening, kind of fast. Things that I wasn't really prepared for."* [P18]

Some of our participants experienced the hospitalization during the COVID-19 pandemic [345], which made their transition to home and their reintegration journey further challenging. P16 noted *"I went from like having a very busy schedule of work, school, volunteering, seeing friends, having a social life, going out all those things to having a psychotic episode and then coming out of that going to a wedding the week after and then being in quarantine [due to the COVID-19 pandemic.]"*

Assessing patients' capacity for management of the illness and self care, their clinical needs and their socioeconomic and cultural needs including where the patient would stay after the discharge, the levels of support available and needed, the wishes and decisions of the patient and the family, etc., are an integral part of discharge planning for patients hospitalized for a mental health condition [103, 346]. Our findings suggest the preparedness that participants might feel at the time of discharge might change as they immediately transition to home due to unforeseen conditions that happen while they were in the hospital.

Re-gaining access to social connections. We found two distinct approaches participants took to re-establish their social connections immediately after discharge from the hospital.

For some participants, after the hospitalization, we found that a sense of urgency and prioritization determined how they interacted with others in their lives – either because they were missing out on social interactions while they were in the hospital, or because they felt obligated to get back to people who couldn't reach them while they were hospitalized. One of the immediate social interactions participants had after discharge from the hospital was letting people in their lives know that they were back home. This communication happened most often via technology-mediated channels and rarely in person. Some participants shared feeling a sense of overwhelm and urgency in reaching out to people and described how they prioritized whom to contact in their lives. For instance, P1 mentioned creating a list to identify the order in which she should reach out to her family and friends to let them know she was back home and doing well. We found that being able to re-establish social connections and draw upon benefits from social circles was dependent on whether people in participant's lives knew about their psychiatric hospitalization. P1 highlighted the importance of disclosing hospitalization experiences to people in her life, sharing that important people in her life already knew about her condition and they were the first people she reached out to after discharge from the hospital. In contrast, P16 did not have such a social circle who knew about her hospitalization experiences to immediately draw upon their support after the hospitalization. She shared that she experienced restrictions on who they can reach out to. P16 said that since she had not seen their friends in a while, one of the first things on her mind was getting back to them. However, she said, *"My parents made sure not to tell people, not because they're ashamed, but because they didn't want me to feel like I had to go and explain myself to other people."* P16 mentioned how it was only after a period of time that she opened up to her close friends about her hospitalization experiences.

Other participants shared that they did not immediately re-establish their social connections because they were worried or skeptical about how their family/friends might react to the news about their psychiatric hospitalization or because they considered slowing down the reintegration process as a coping mechanism to manage their condition. As P8 shared,

> *" I don't know what I'm expecting and [how] people may react to the news that I'm back, and then, bearing in mind that it was difficult for them to understand me."* [P8]

P13 echoed a similar slowed down approach to re-gaining access to their social network. He said,

> *" I mean, you know, just trying to stay away from many things, just like keeping a low profile...I meant to just relax [...] It makes me have a good plan. Yeah, just like no pressure at all."* [P13]

Prior work studying the social lives of people following traumatic brain injury also found withdrawal from social interactions during recovery [142]. Clinical literature suggests that social functioning, i.e. re-establishing social connections and interactions after hospitalization is an essential marker for recovery in mental health [347]. Importantly, problems with social functioning are known to lead to long-term social difficulties, such as withdrawal, isolation, and lack of integration into the community [348]. In relation to our finding, this body of work suggests that while people might not feel immediately ready for social interactions, waiting too long to re-establish social connections is not beneficial for recovery and reintegration.

Re-gaining access to technology and re-establishing online presence A common experience during psychiatric hospitalization is loss of access to personal devices like mobile phones and limited use of technology and the internet [95, 98]. Participants expressed a wide range of emotions including feelings of anxiety, overwhelm and the fear of missing out, when they re-gained access to their phones at the time of discharge. For instance, P17 shared that he felt anxious about missing work-related emails while he was in the hospital. P13 mentioned how he missed his phone and online interactions while he was in the hospital. He mentioned that the first thing he did after discharge was catching up with online interactions and notifications.

> *"I wasn't allowed to have my phone. So, I felt like I missed a lot of things online. When I got back home, it was so hard to catch up with everything on like like on Instagram, Facebook, the forums that I'm always in. I'm in [different] groups, so I just felt left out. I mean, everything was behind me. Yeah, so I had to catch up, to get to get updated."* [P13]

However, not all participants expressed positive feelings towards re-gaining access to their devices. One of the major reason noted by participants was the overwhelming volume of mobile push notifications they received when they turned on their phones after the discharge. P1 shared that the volume of mobile notifications "*ruined [her] experience of returning home*", and she would strongly advise others to not immediately turn to their phones and calls or read the notifications they have missed right after discharge from the hospital.

> *" The first thing you want to do is turn your phone on, because gosh you missed your phone so much, and it won't stop making noise at you because you've missed, you know, a week's worth of texts, and emails and any other*

notification. And, you know, so you're trying really hard not to be overwhelmed by anything. So eventually, you know, I just, I just turned it on and put it down, because, I knew that there were people I wanted to let them know that I was home and all that. But, that was too much all at once, so I try to just, you know, put the phone down, enjoy the car ride with my mom." [P1]

P2 also echoed this feeling of overwhelm, saying, *"there's a lot of notifications...people are chatting chatting always. When you go offline and come back online or or leave your phone. So many notifications."* In Social Computing and HCI literature, prior work has studied various forms of technology non-use [349, 350, 351] in the context of people who actively choose to stop their online presence and interactions or people who were never able to access a technology [350]. In contrast, in the context of our participants who have experienced a psychiatric hospitalization, the institutional rules and guidelines do not permit them to use technology while admitted in the hospital. Situating this finding in technology non-use literature, we find that our participants feelings about their break with technology extend beyond existing categories and conceptualizations of non-use [350].

Social lives intertwined with management of the illness after hospitalization

Self-reliance. The deinstitutionalization movement shifted the role of the psychiatric hospital from a place of long-term stay and treatment to emphasizing reducing feelings of dependence and supporting community integration [105]. This model encouraged rapid discharge from the hospital once patients' symptoms stabilized, so they may continue care in outpatient settings. We found that not all participants felt like they were ready to get back to their normal routine and social lives after discharge from the hospital. P16 mentioned difficulty with focusing and paying attention and fine motor skills after the hospitalization. She also spoke about how receiving accommodations at school was helpful to navigate life after the hospitalization.

"*I was still like frazzled coming out of it like you're not like I was recovered* *enough to go home, but not enough to like be back to normal routine... [I] was having a lot of my gross and fine motor skills were very like not, I'm not gonna say depleted but I'm gonna say like, not as refined as what they typically were. like I was having like I was getting dizzy from walking in the hospital so even just like light exercise was difficult for me because I found myself getting dizzy from walking.*" [P16]

Participants shared how their recovery journey outside of the hospital affected their self-perception and self-reliance. P19 experienced problems with his memory after discharge from the hospital and needed people in his life to help him remember things. One of the first social interactions after discharge that P19 mentioned involved friends and family members showing photos and videos to recall past memories. P12 mentioned being heavily dependent on her mother for coping skill and managing her symptoms because her mother also experienced the same condition. The reliance on others for everyday activities was particularly challenging for those who experienced their first psychiatric hospitalization. P16 who was diagnosed and hospitalized for schizophrenia form disorder for the first time noted, *"Prior to that like I had been such like an able person. Like even coming out of the hospital I wasn't allowed to drive, because I was still gonna adjust to medication so I wasn't allowed to drive so I had to Uber everywhere."* This reliance on others during the recovery period and a change in self-perception impacted how participants described their social interactions immediately following the psychiatric hospitalization.

Stigma. Extensive research establishes the challenges associated with stigma around mental health conditions globally [277, 276, 293, 352]. We found that the post-discharge period is particularly challenging due to stigma as participants re-established their social connections and opened up about their psychiatric hospitalization experiences with others. Several participants shared feelings of stigma they experienced from their family members and friends. On the one hand, the stigma manifested as a hindrance to self-disclosure, obtaining social support and reintegrating back to social lives. As P16 notes, "*I think* *there's a huge disconnect. In between what people know about... what they think they know about mental illness and what it actually looks like, and what leads people to have it.*" P2 also expressed the difficulty in feeling they could not tell friends they had been in the hospital because they wouldn't be accepting. P18 shared,

> "*It's something that I guess from an outsider's perspective, people don't really understand and they might say, you know, 'you're just being lazy,' you know, when you're, you know, really depressed, you know, and you legit can't function, you know, and them not understanding that or empathizing with that.*" [P18]

On the other hand, participants also noted the societal level stigma associated with mental illness that led them to unwillingly lie about their condition and hospitalization. P1 shared that she had to lie about the hospitalization to people at her workplace due to the stigma and negative consequences she might face. She noted:

> "*if I'm not telling them that I'm going into the hospital that I tell them that, Oh, my, my aunt is having surgery, and I'm going to stay with her for a week...and, the cell..you know, reception is, you know, not always great out there. But yeah, unfortunately, we're still at a point in this country in the world that if you say that you are in a behavioral health center, you know, to fancy it up.*" [P1]

We found that one of the most significant negative effects of stigma was it obstructed people's path to social reintegration after psychiatric hospitalization. Participants noted cutting family members and friends off their life because they were '*scared*' (P11) or could not handle what the person was going through with management of the mental illness. P11 says "*Some people think maybe you are not okay. So they'll be a bit scared of how you're going to react.*" These obstacles that are presented as a result of stigma highlight a re-assessment of pathways to social reintegration (due to loss of previous social connections) and self-management of the illness (due to lack of social support resources, perceived or actual.)

Shift in goals. The period after discharge from psychiatric hospitalization is characterized by a shift from institutional, clinical treatment to self-management of the mental health condition by individuals. We found that several participants identified this shift and created new goals to manage their condition outside the hospital, often noting that the hospital was only one step in their journey with mental illness.

> *"The hospital is a temporary thing, you know I'm saying, whatever problem or situation is going on with you. That is still going on with you. You know, none of that changes. You change, your environment changes, your perspective changes, but, they don't change anything. So yeah, most of my managing definitely came from when I was outside."* [P18]

Participants noted several strategies to cope with their symptoms and manage their condition after the hospitalization, most commonly identifying these mechanisms as part of self-care. Some participants shared how their stopped adhering to their prescription medication or choosing alternative forms of medication to manage their condition. For instance, P19 mentioned that he started taking natural medications such as activated charcoal to "*clear [his] gut of all those medications*" [P19]. P18 also shared her negative perceptions towards medication she received at the hospital and how she chose to stop it and focus on changes to her lifestyle. She described her approach as, "*taking care of myself is just being more aware of what's going on with me.*"

> *"To this day, and even in the hospital they forced me to take pills, I wouldn't, I don't take medication now. So, to me it was just a lot of self care and making sure that I pay attention to me, make sure I eat and sleep, which again, somehow [I] wasn't getting because I just had a baby and making sure that I just pay attention to, you know what's going on with me...if I'm feeling, you know, emotional or whatever, I just choose not to do certain activities that day."* [P18]

Other participants revealed how they made changes to their lifestyle to maintain their health outside the hospital and how this was a significant change to their lives post-hospitalization. P19 spoke about how he started eating healthy, exercising, meditating, and noticed a significant improvement in managing his symptoms: *"Instead of just saying, you know, go to your therapist and all that... no, we need to, people need to be teaching people about healthy food...Suicidal thoughts, or like voices, or whatever, it's so low now. It's like it almost doesn't exist because when I started changing a diet."* Similarly, P2 spoke about making major decisions in his life after the hospitalization such as quitting a stressful job and ending bad friendships to maintain his health and well-being.

> *"I had to quit my job. I was under a lot of pressure. I don't know...my eating habits changed. Even friends, I stopped talking to some of my friends, putting a little pressure and all that...so I had to drop my friends and the job."* [P2]

P16 mentioned consciously making the effort and allowing time for herself and checking in with how she was feeling.

> *"And I'm making the effort to allow myself time for myself and, like, giving myself time for self-care and making sure that I'm checking in with therapy and, like, making sure that I'm not overwhelmed or like, my time is being cut so short that I'm not like sleeping enough hours and stuff like that...I'd say, there has to be a time period where I put [other things] away. Let me go take a walk outside, listen to some music, relax a bit and just kind of be away from my phone, or do something else that's not related. To just kind of connect with myself a little bit better."* [P16]

Lastly, P19 shared how he changed his self-presentation to symbolize the new beginning he was marking after the hospitalization.

> *"You got to create, like a whole new person, you know, you can go back to the old ways or you'll probably, you'll probably go back to to bad mental health*

> *issues, so basically try to create a new person...and even dressing different, you know, people see me now and I'd have hoop earrings on, like, I never did that. Just creating a whole new human being, that's what I'm about."* [P19]

A common theme across these goals that people set for themselves was that they were actively identifying and addressing daily stressors in their lives and making positive behavioral changes to better manage their condition. P18 summarizes this accurately speaking about how reintegration is more than just treating symptoms, particularly, focusing on the everyday stressors with social relationship, finances, jobs, etc., that people face in their lives outside the hospital.

> *"A lot of people just have a lot of issues that are not being addressed. They're, again, so focused on, you know, the pills and things. But like I say, maybe there was something that was stressing them out, like, hey, I don't have enough money to feed myself, you know, I don't have a license. which means I can't get services, you know, like food stamps or housing or whatever. I think that [clinicians and hospital staff] don't really help people in their everyday lives, to kind of cope with that stress a little bit to help what's inside. which I really feel like puts people in a place where they're just constantly in this fight or flight response and survival mode."* [P18]

It takes a village. Beyond the individual's role in managing their mental health condition outside the hospital, we found that family members, friends and online social connections together played an important role in supporting participants' reintegration after psychiatric hospitalization. We found that the role of others is important not only after discharge but also during in-patient hospitalization experience. Particularly among participants who were hospitalized for a longer duration ranging from weeks to months, we found the importance of close friends visited them during the hospitalization or sending cards thinking of them (P5, P19).

After discharge, participants spoke about how their family made them '*feel welcome*' (P5) to be back to their home. Others shared how feelings of acceptance after the hospitalization experience and normalizing mental health challenges played a crucial role in their recovery and reintegration journeys. P12 spoke about how her family treated her the same after hospitalization due to lack of stigma. She shared her mother being especially understanding, due to her own experiences with depression and anxiety, and helped her get through the recovery and reintegration period.

> *"They didn't change the way that they view me for having a mental illness. That meant a lot to me."* [P16]

Participants' friends and family also helped with managing their symptoms and cope with their mental health condition after the hospitalization. P11 mentioned receiving a lot of support from both family and friends. He mentioned how they would visit him frequently after discharge and help him get back to a normal routine. P19, who faced challenges with his memory after hospitalization, spoke about how his sister showed him pictures on his phone to recollect past memories. Similarly, he spoke about watching videos with his friends and this helped him to remember a lot of his friends he had lost memory of. P19 also mentioned how his family recorded moments during which he faced suicidal thoughts and walked him through those videos as a way to give structure to his experience with schizophrenia. This form of confronting inhibited thoughts and giving an experience structure and meaning is known to help self-management of mental health conditions and facilitate a sense of resolution [353]. P12 shared how her mother, who also experienced mental health challenges, helped her with coping mechanisms during the reintegration period.

Another role that participants' social connections played was as a source of informational support. P2 mentioned how his girlfriend shared links related to mental health support with him, even when he was not seeking for such informational support. He expressed the important role this played in his recovery after the hospitalization.

> *" Actually my girlfriend used to give me links. So, personally, I wasn't like*

looking for such content, but my girlfriend sent me links to read about being in the hospital." [P2]

P2 shared how having at least one supportive person during the reintegration journey plays a significant role.

> *" I don't have many friends. I think I spoke to one of my friends, and it is very different. And so when we went to hospital because of my issues. They might end up, I don't know, may be they couldn't take me for who I am. So, I would just keep it myself apart from one friend or didn't even tell them exactly where I was...in the hospital. But for my girlfriend she knew everything. Because to her it doesn't matter. With my girlfriend on my side I think things were not bad...she was really there for me."* [P2]

Apart from close friends and family members, participants spoke about the role other social connections and support groups played in their lives during reintegration. For instance, P12 spoke about how her boss at work was extremely supportive and understanding of her hospitalization experiences and allowed her to take time off and slowly get back to a normal work routine.

> *"my boss give me a break, and allowed me to work at home more because part of the pressure at work was triggering my depression. but he cut work time for me and eventually went back to working full time. but he helped me gradually, gradually, ease into work. that helped a lot."* [P12]

The important role of family and friend was also emphasized by participants who felt like they did not have such a community supporting their recovery and reintegration. P18 expressed feeling a lack of support from family and friends. She came back home after discharge from the hospital by herself and had negative experiences of rejection from her family. She shared that she made new friends while she was admitted in the hospital and how joining a support group, albeit unenthusiastically, helped with recovery after hospitalization.

> *"I don't feel like I got a lot of support, I was kind of forced into a support group, and actually really enjoyed that. Because of people that I felt like it kind of surprised with what I was going through. And then I made some friends in the hospital so I think that really kind of helped my overall recovery talking to them and you know getting their perspective on what happened."* [P18]

Prior literature on mental health supports that the mechanisms we identified relating to emotional and informational support, coping and management of symptoms are associated with successful recovery and overall well-being [10]. We found that the social context of the individual, either a single person or a community of people, involving family members, friends, colleagues, 'sympathetic others', together shared the labor involved in navigating reintegration after psychiatric hospitalization. Furthermore, we found that both pre-existing social relationships as well as newly established connections have a role in supporting reintegration journeys.

Online social lives intertwined with self-management of the mental illness

Online spaces for self-disclosure Confirming findings in prior literature, our participants also appropriated social technology platforms for mental health disclosures [185, 2]. We identified three different approaches to how participants considered disclosing about their mental health challenges on social technologies like Facebook, Instagram, Youtube, and others. Some participants noted being very private on online spaces and they mentioned that they would never consider posting about their mental health experiences on online social platforms (P2, P6, P13). The stigma around mental health and the uncertainty in how others might perceive them inhibited participants from disclosing to those outside their close friends and family circles (P11). For example, P12 shared that she only disclosed to people who needed to know about her condition, but that she would also tell others if they asked. The second approach to online disclosures (of psychiatric hospitalizations) involved participants who felt comfortable to post about their mental health experiences on personal social media platforms or online support groups. Among reasons that made participants feel comfortable to post on their personal social media profiles about their mental health condition, one frequently noted reason was other people's prior knowledge of the participant's condition (P1, P11.) P11 said, *"Most of my friends knew about my situation. So, I did not have any challenge posting it because they are aware I was in hospital."* In contrast to face to face disclosures, P18 expressed that it was easier for her to disclose in online support groups because she felt it was easier to say everything she wanted to say without any interruptions. P18 also mentioned that with online support groups, there is no fear of therapist mandatory reports. She said, on the benefit of disclosing in online support groups compared to offline groups:

> *"I think online, you're a little bit more receptive to open up and you are not under like a time constraint, like in group in person, you may have a certain amount of time. And because again there's other people there, you can't talk over people or whatever. when you're by yourself online, you can really say everything you have to say and get it all out. Some people's posts are very very long I don't know if everybody reads through the entire thing or not. But you do feel like, again I'm being heard. I'm not being interrupted, I'm not under constraint, I can really just say everything that I have to say. Whereas in person, you may not necessarily get to do that and again because they can put a face, you know with the name. we may be a little bit more hesitant to say certain things, especially since it's facilitated by, you know, therapists and they're mandatory reporters. So there are certain things, you're probably not going to say to them."* [P18]

The third approach participants adopted for online self-disclosures involved public broadcasting disclosures to those outside their social circles. P19 used YouTube and Facebook live streaming to tell his story with mental illness to followers and discuss his recovery process, especially following a hospitalization. P5 shared that he posted on Facebook and Instagram about his experiences with mental illness and gained a significant number of followers because of their stories. P15 also had a YouTube channel where she posted videos on Christianity and music. On opening up about her hospitalization experiences, she said:

> *"I'll probably open with my hospitalization experienced probably on YouTube because I do gain a lot of views on YouTube. So I might put it on YouTube. I might not really say it by posting on Instagram or posting on Tumblr."* [P15]

When asked about what she would share on YouTube, P15 said she wanted people to know that, *"life will get better. things can get better like right now you might be going through something, but it takes time for, for you to get better."* P15 also noted she would feel more comfortable opening up to her followers on YouTube than people in her life because she did not like waiting for people's response and the lack of reciprocation from her friends. She said by posting on YouTube, she would be able to help a larger group of people:

> *"I think sharing it on YouTube is meaningful because of your experience, what you've gone through and then, you can probably help someone. And that's how I feel, I feel like it's more meaningful than talking to your friend about it, because that's just one person, and compare this with many other people."* [P15]

While most participants experienced positive feedback and support when they opened up about their condition online, one participant (P17), who had a public Facebook profile for both his personal and business use, shared experiencing negative interactions and harassment that made him skeptical to share more about their illness.

> *"People harassed me over the phone, I have like my business phone number up on my website. But I have been harassed and I blocked people on Facebook,*

> *yes. I have been attacked emotionally by people on Facebook. They [people who attacked] were people from high school, that, you know, people will add you on Facebook based on connections from school or church or whatever. And that doesn't mean I was personally connected with them."* [P17]

Online spaces for support. Most commonly, we found that participants adopted social technologies for reaching out and accessing social support. Majority of our participants shared about finding and accessing support via technology only after their first psychiatric hospitalization. Two mechanisms were predominant in the ways participants found social support via social technology platforms after hospitalization.

First, majority of the participants mentioned belonging to an online support group, most commonly on Facebook or on Whatsapp. On Facebook, participants mentioned large groups formed by organizations like Mental Health America (MHA), National Alliance on Mental Illness (NAMI), as well as smaller, local mental health awareness and support groups. Some participants mentioned being part of WhatsApp groups, with several hundred people, that facilitating sharing about mental health experiences, coping mechanisms and supportive resources.

When we asked about how participants learned about these groups, we found several points of entry through which participants entered online support groups. Most commonly, participants shared that someone else in their lives introduced them to the groups (P2, P7, P8) or pointed to them as a support resource. For instance, P2 spoke about how his girlfriend shared a link with him through which we could join a local mental health support group on WhatsApp. Other participants mentioned actively seeking out for support groups after their hospitalization using features such as Facebook search. For instance, P12 spoke about seeking out mental health groups on Facebook after hospitalization because she was more aware of her problems and how these groups could benefit her. Once participants joined a Facebook group, they revealed how they subsequently found more support groups through the auto-generated suggested groups on Facebook (P1).

We probed into the structure of these online support groups and how participants participated and found meaningful support during their reintegration journeys. On Facebook, most groups that participants belonged to were private Facebook groups that required moderator approval to join. Participants shared how this enabled them to feel like these groups were safe spaces and made them feel comfortable speaking about mental illness (P1, P18). In contrast to the structure and moderation of Facebook groups, the WhatsApp groups that participants belonged to were largely not moderated. P2 spoke about how the chatting medium on Whatsapp and push notifications made it difficult for him to follow content on the group.

> *"There's a lot of notifications...people are chatting. Chatting always. When you go offline and come back online or or leave your phone. So many notifications."* [P2]

One participant (P2) shared that he thought one member of his WhatsApp group might be a licensed therapist because they frequently answered others' questions. But, the lack of affordances to archive roles and norms on WhatsApp groups made P2 unsure about recommendations on the group. Across Facebook and WhatsApp support groups, participants described how members introduced themselves, shared stories about their mental health condition and hospitalization experiences, as well as coping mechanisms. As also evidenced in prior work [1, 354, 355], reciprocity, informational support, reducing inhibitions and stigma and normalizing mental health experiences contributed to the value our participants drew from these groups. P5 spoke about how there were so many people with many different experiences on these groups that they were always able to relate to some content or person on the group.

> *"I'm in there, talking about you know my experience being chronically in and out. And recently, and for the extended period of time. And they all say the*

same thing you know it's like it's like coming into a new world, like being born again." [P19]

P10 shared about feeling inspired by others' stories on the Facebook group she was participated. She started sharing stories of people's recovery and reintegration journeys from the Facebook support group on her own timeline (with the individual's consent), to raise mental health awareness among her own social circle.

While not all participants actively participated in these online support groups by posting or commenting, they noting benefiting from reading advice on coping with mental illness and making positive life changes based on information shared in online groups (P2). The second predominant mechanism through which participants drew benefits via social technologies was by passively consuming supportive content on these platforms. Sometimes this included posts made by other individuals in an online support group. But, more commonly participants mentioned following content on inspirational quotes, positive life changes, positive behavioral changes, spirituality, etc. For instance, P4 spoke about following a Twitter account that posted motivation quotes and how this content helped her.

> *"There's a page on Twitter that usually shares motivational quotes. I go through the articles they post. you know, get inspired and that really help me to deal with my with my inner self. So I was able to be happy again."* [P4]

P8 also spoke about following Facebook pages that post inspirational quotes and spiritual videos. He said, *"[I get] the ups and the inspiration and that one would let me recover, very quickly. Yes. And, yeah, most of them did they have the Facebook pages so just follow the pages and get the posts and inspiration on a daily basis. And with that, it has helped me to get the mental illness in control."* Lastly, some participants also mentioned finding online meetings (P1, P3) and webinars (P6) related to mental health via social technologies. Participants spoke about how these interactive meetings helped them during the COVID-19 pandemic when they could not access resources and care in-person.

Negative aspects of social technology use that hinder management of mental illness.

So far, we discussed how participants found that social technologies supported their reintegration journeys after the hospitalization. Alongside these benefits, participants also identified aspects of social technology use that they found harmful or not beneficial to their recovery and reintegration.

Some participants found that spending too much time online on social technology platforms was replacing the time they spent on social interactions in-person (P16).

> *"I felt as if I was spending too much time on my phone, to the point where I was not physically present in the conversation or like I just needed time that I wasn't being bombarded by, you know, advertisements, friends from high school doing this, friends and colleagues doing this, comparing yourself to other people." [P16]*

The most commonly noted negative aspect of social technology use was related to feelings of social comparison [356]. Participants described how seeing other people's posts about doing well in life made them fell less accomplished (P12) or bad about themselves (P1). As P16 explained, *"'Well, look, there's a fitness Instagram model and like I just feel like it creates so much of distress in a lot of ways that it's kind of unnecessary. like as much as it is entertaining it causes a lot of like, low key distress. to put it."* We found that feelings of social comparison were particularly significant when they acted as triggers to people's mental health conditions. P12 also echoed feeling left behind, seeing other people's posts on Facebook. She said,

> *"I've been trying to avoid Facebook ever since because I will admit that some of my triggers come from seeing how other people are doing so well and I feel like I'm stuck."* [P12]

While P12 also participated in online support groups on Facebook as we described above, she noted how these support groups have been helpful in managing her condition, but, she viewed them as short-term solutions. Specifically, P12 spoke about the content on these support groups reminded her of her old experiences with mental illness that she wished to move away from. From P12 (on negative aspect of support groups):*"I wouldn't want to keep replaying my experience over and over and over again. I'd like to move on from that."* Another participant spoke about the negative effects of consuming information that was not helping their recovery and reintegration. P17, who was hospitalized in 2020, shared that he deactivated his social media accounts after the hospitalization because he did not wish to see posts about the California wildfires, and COVID-19 related deaths as that affected his moods and mental state.

> *"That was kind of intense I didn't want to say too much about that [mental health], or the wildfires and COVID, which is kind of sad. I didn't want to. I didn't want to see much about that. Yeah. Even though I wanted to be mindful of people that are suffering I just did. It was kind of sad to hear about it."* [P17]

Prior experiences on the platform also affected P17's decision to deactivate his Facebook account after the hospitalization. He mentioned being cyberbullied, where someone used profanity and hateful language on his posts. P17 also shared that a family member posted hurtful comments about being hospitalized for mental health. Due to these past experiences, P17 said he commonly deactivated his account when he decided to take a break and focus on his mental well-being.

Transformation in digital habits and routines. Finally, we asked participants if their use of social technologies changed after the hospitalization, compared to how they used these platforms prior to the hospitalization. For many participants, we did not find any active changes to time spent, digital habits and posting behaviors on social technologies. They were already using these platforms in a specific, limited manner and their use did not change due to the hospitalization (P2). For instance, P18 shared about always being private and that she prefers to directly call or send messages to her friends as opposed to interactions on social media. She mentioned that she did not like the idea of having information about her available to anyone other than people who already knew her background.

Other participants actively made changes to their online social lives, digital routines and posting behaviors on social technologies after the psychiatric hospitalization. For example, P16 shared about spending less time on Facebook and Instagram after the hospitalization because she felt it was beneficial to limit screen time for her mental health. She mentioned using a logging app on her phone to track her screen time. In contrast, some participants increased their screen time and social technology use because they believed these platforms mediated their offline social reintegration, i.e., they helped re-establish social connections that they missed while they were hospitalized. P11, who had been in the hospital for four months, shared that after discharge he got back on social media to re-establish his online presence. He also felt comfortable posting again because that was the only way he could share with friends his experiences and plan social interactions since his discharge. Similarly, P5 spoke about always being active on WhatsApp and using it more frequently after the hospitalization as he was not seeing his friends in person as much due to the COVID-19 pandemic.

A few participants spoke about re-configuring their digital habits and social technology use to better facilitate their reintegration journeys after the hospitalization. P1, who only had a few high school friends on her Facebook friends list, spoke about actively joining five Facebook groups for mental health support. She also mentioned restructuring her Facebook feed by unfriending both toxic acquaintances and strangers, following health-related pages, and joining mental health support groups: *"I joined a whole bunch of new groups to try to make sure that my Facebook feed was nourishing me and not, and not strangling me."* [P1]

> *"[I recommend] anybody struggling with any of these issues, to try to find some groups on Facebook, especially if they use Facebook, that help increase your support chances, the support system, your chances of staying mentally healthy, because you've got all these extra people to interact with. Find online meetings, or, you know, NAMI or any of the other mental health support places*

> *out there. Because, they're out there! your people are out there, no matter who you are, especially if you're struggling with mental health. And you can be in these groups, and you're going to get affirmation statements, you're going to get really good quotes. You're going to have people talking about their issues. And you can be like, oh, wow, those are my issues here. So I helped with that. I would definitely that would be the first thing I would encourage people to do. Like, drop all of your friends from high school, pare your list down to people you actually care about and are interested in, and definitely find a group or two or five. I mean, you can make technology and social media, you can make that work for you."* [P1]

P17 whose past experiences on Facebook included negative interactions and frequent de-activation, spoke about how he blocked people who were hurtful, and started following funny videos on Facebook Watch. He said he felt relaxed watching such videos and noted that it was helpful for him after the hospitalization.

> *"Yeah, some of the videos that people post. I guess they are like bloopers or just pranks, without hurting someone, but just kind of quirky, funny videos, those help me...just seeing people make this video so I guess even like dance videos and videos of people making different food from across the world. That kind of helps."* [P17]

P17 also highlighted how social technology platforms differ from one another, and how one might be better for him after the hospitalization experience. He mentioned that he continued using Instagram (and not Facebook as much), because he does not have to see other people's status updates and interact with them: *"You can just follow, post, and like pictures, and you don't have all these status updates all the time, and you can just scroll down, keep scrolling."*

P19 who did not share personal details on Facebook prior to the hospitalization shared how he transformed his profile to a public-facing account. He continued using Facebook and YouTube to stream a fitness series he had started before hospitalization. After the hospitalization, he used these platforms to start a mental health series that benefited both him and his subscribers. From P19 (on how getting back to making their Facebook series helped them): *"You know, regain your mental health, you know just doing stuff that helps other people, versus focusing on yourself often."*

Interpretation of findings By investigating the social lives of individuals after psychiatric hospitalization, this study presents people's shifting goals, priorities and challenges during the process of reintegration and unpacks how social lives are intertwined with management of illness. Our findings on people's reintegration journeys after psychiatric hospitalization corroborate that both clinical recovery (i.e. reduction in symptoms) and social reintegration (resuming social roles) need to go hand-in-hand for overall well-being of those with mental illness [105, 329]. Further, we lay out details on the intersection between clinical and social factors affecting people's lives after psychiatric hospitalization. During the period after discharge, we identified participants' goals and responsibilities such as re-establishing social connections, resuming social roles, and management of the illness outside the hospital. However, stigma related to mental illness, over-reliance on others, lack of support and change in living circumstances presented challenges to achieving these post-hospitalization goals. The interplay between people's goals and challenges due to mental health impacted both clinical recovery and social reintegration. On one hand, we identified how social factors impacted management of illness. We found that perceptions of stigma, at both an individual and societal level, affected how participants viewed themselves and accessed pathways to care. For instance, participants most affected by stigma perceptions mentioned cutting social ties with family members and friends, online and offline, because they were unsure how they would react to mental health experiences, presenting obstacles to successful recovery [103]. On the other hand, clinical aspects like aberrations in mental health symptoms during the recovery period also impacted people's social lives. Some participants distanced themselves from social interactions either because they were still recovering from mental health symptoms and did not feel comfortable being around others, or because reducing interactions was perceived as beneficial for their recovery.

The intersection of clinical recovery and social reintegration especially impacted people's online social lives after the hospitalization. Participants mentioned that it was mostly after the psychiatric hospitalization that they considered disclosure and support resources on online social technology platforms. We found that social technologies supported participants recovery and reintegration journeys by mediating social interactions, providing spaces for disclosure and support and sources for positive health changes. However, negative feelings related to social comparison, emotional triggers from content seen online and negative online interactions caused distress to some participants and presented hindrances to their efforts towards recovery and reintegration journeys.

Examining these factors together calls attention to people's shifting goals and priorities and the transformations in their social lives during the periods after hospitalization. In the following subsections, we reflect on how these findings inform researchers, clinicians and designers of social technologies invested in improving mental health care.

6.3 Discussion

Theoretical Implications Anchoring on psychiatric hospitalizations as a liminality, in the first study, we combined *clinical perspectives* – around symptomatic expression and recovery, with *social perspectives* – around stigma and reintegration. Guided by the Possible Selves framework [281], we provided an empirically derived taxonomy to represent and then understand individuals' mental health status transitions. Our approach thus enabled incorporating people's heterogeneous experiences around recovery and reintegration in the computational modeling of mental health. By combining the two perspectives, we presented an exploration of the intersection and the significance of considering social and clinical perspectives on mental health. This illuminates important theoretical implications.

As noted above, computational approaches to study mental health within HCI, digital psychiatry, and machine learning frame the individual in two broad, distinct ways: as a *patient* who is receiving *clinical care* and institutionalized treatment for a validated diagnosis [7, 357], or as a *vulnerable individual* who is seeking *social care* and support for management of a mental health condition, such as from their networks of loved ones and peers, offline and online [335, 1, 2]. The word patient comes from the Latin “patiens,” from “patior,” meaning to suffer or bear. Scholars have expressed criticism of this conceptualization. It tends to objectify the person with mental illness [330], disregards their sufferings and identity outside of the clinical definition of the illness [331], and fails to consider the challenges people need to overcome beyond symptom management to get back to life following hospitalizations [110]. In contrast, when the clinical facet of one’s mental illness trajectory is ignored only to consider their vulnerability and self- and peer-supported management efforts, this latter conceptualization fails to account for the obstacles an individual faces in navigating a formalized treatment plan in concert with their own attempts. Research has shown that in some cases this lack of consideration of the relationship of social and clinical care can interfere with evidence-based treatment, create tensions in the patient-provider alliance, and even lead to detrimental mental health outcomes [101].

The repercussions of this disjoint consideration have been noted in recent social computing literature [358]. Chancellor, Baumer, and De Choudhury highlighted that the representation of people’s experiences and data in machine learning work on mental health – as a patient, disorder, data point or a person – may inadvertently risk dehumanization, poor and incomplete characterization of the mental health experience, and present serious consequences to scientific rigor [359]. Relatedly, findings from Chapter 5 surfaced the differentiating uses of social media for mental health, showing that representations of people’s experiences who make disclosures on social media and seek social support for mental health, poorly generalize to those of patients receiving clinical care [186]. At the crux of these observations is the fact that peoples' mental health statuses and needs are not isolated, discrete attributes, but are temporally situated experiences of the *same person* who is in transition, given the dynamic nature of their illness.

Framed in terms of Van Gennep's liminality [112], using our derived taxonomy of PSS, we could examine the transitions of individuals as they enter and leave the role of a patient (Section 5). For example, an individual could move from the self-regulation focused PSS they expressed as a non-patient before hospitalization, to the withdrawal-focused PSS post-hospitalization, that is reminiscent of them being a patient at the hospital facility. Our taxonomy could further reveal the transitions that are indicative of reintegration after psychiatric hospitalizations (Section 6), such as those shifting from the self-awareness PSS immediately after they were a patient, to the incorporatation-focused PSS in the longer-term, as a way to embrace the non-patient identity again. This way, we showcased how a holistic view of an individual's mental health state, as indicated in their *clinical* recovery and *social* reintegration can be derived by understanding their various possible selves surrounding hospitalizations. We did so by demonstrating a theoretically- and clinically-grounded computational approach that amalgamates insights from people's medical histories as well as their social media data.

Drawing upon our work, future research leveraging digital traces for prediction of mental health states can benefit from noting that people's mental health statuses and their exhibition on social media are heterogeneous, and contextualized within the individuals' specific needs for clinical treatment and social support. Multimodal approaches, like Gaussian Mixture Models [282], that we adopt in RQ1, when formulated with theory (such as the Possible Selves framework [281]) and punctuated with clinical insights (Section 4), can reveal differentiating patterns of behaviors and use of social media within the same population. This way our work contrasts the theoretical models used to understand health transitions in prior CSCW research [135], which consider that the experiences of all individuals across the illness journey are the same, and that these transitions are largely uni-directional and permanent. Adopting a personalized approach, our taxonomy gives three distinct sets of behaviors that are dominant before psychiatric hospitalization – self-regulation, self-awareness, and sociality focused PSS. These characterizations of mental health statuses can enrich personalized predictive models for early interventions that consider an individual's clinical as well as social pathways of care. Similarly, research primarily focusing on social support mechanisms for mental health could leverage how clinical aspects to care such as hospitalizations, medication adherence etc. impact support outcomes. For instance, in this study our proposed empirical framework of transitions revealed that first hospitalization experiences lead most people to transition into the withdrawal focused PSS, whereas those re-hospitalized are able to maintain the incorporation focused PSS. The support seeking goals and outcomes for the former group might be significantly distinct from the latter – insights that could contextualize how different individuals use of social media to find help and advice around their mental distress.

Digital breaks in the context of mental health experiences Research in CSCW and HCI has focused on articulating and classifying different types of technology use and non-use including framings such as digital divide [360], and dimensions like volitionality [351], disenchantment or disinterest [350], resistance [351], among others. In the context of psychiatric hospitalization and social technology use, the period of digital break is due to institutionalized mandates that do not permit individuals to access their devices and online sources during the period of hospitalization. As we discussed in section 6.2.2, participants' feelings about their digital breaks extended beyond existing categories and conceptualizations of non-use. Initially, the relationship between technology and use in our participant sample might relate to the "limiting use" category of non-use – people who systematically limited their use of a platform due to social, professional or institutional pressures [349]. However, in contrast to institutional pressures at, say, a workplace, individuals hospitalized at a psychiatric facility do not have the power to negotiate technology access and use until they are discharged. How can we understand participant's feelings of anxiety, overwhelm and the fear of missing out in relation to digital breaks and re-gaining access to technology? What, then, are the nuances in digital breaks in the context of reintegration for mental health – circumstances when digital breaks can both be nourishing as well as alienating?

Our findings help to triangulate and confirm results from prior work on people returning to social media after a break. We find that un-friending practices and updating friends lists is a common practice across both contexts [361, 362]. Consistent with prior work, we also note concerns about boundary regulation and privacy as people get back online after periods of digital breaks, observed in our findings about disclosure and the segregation of public and private online profiles [361].

However, there are a few characteristics of mental health experiences and psychiatric hospitalizations as periods of digital break that separate our results from other work. First, participants had no agency or choice in taking digital breaks during hospitalization. Also, the period of digital break in this context is unknown, because it is often unclear when one is ready to be discharged from the hospital. The periods of hospitalization in our study ranged from one week to several months. While our findings do not let us disentangle the effects of hospitalization duration on how participants re-established online presence, we anticipate this duration to be an important variable affecting getting back online after digital breaks. Second, more so than in other contexts [363, 349], social surveillance [364] (i.e. by people gathering information about other people) and identity management during re-entry periods online can play a bigger role in populations experiencing psychiatric hospitalization due to the societal stigma around mental illness. Lastly, reversion to online social platforms in the context of our participants can be perceived as a need or necessity due to the post-hospitalization goals of reintegration and re-gaining social connections. Future work on mental health and technology use can pay attention to such nuances, idiosyncrasies, and uniqueness in different people's digital breaks in the context of mental illness hospitalization.

Social technology affordances and their role in mental health support Our study on people's reintegration journeys (Section 6.2) shows that social technology affordances such as visibility [365], pseudo-/anonymity [366], broadcasting communication [367, 368], one-on-one interactions [238, 369], etc., have enabled participants to use these platforms towards their mental health recovery and reintegration, by participating in online support groups [1], making disclosures about their experiences [355], seeking informational support, sources for positive behavioral changes and re-connecting with people in their lives. Outside of "active" [370] interactions with people and content online, participants also mentioned the benefits they drew from scrolling, viewing videos, and passive consumption of online content. The benefits people draw from the latter practices are often invisible to researchers when we focus on archived digital trace data or "active" use of social media.

Recent work has highlighted the importance of understanding and incorporating these invisible practices such as passive browsing of social media postings, or non-clicking into experiences on social media platforms [371]. Situating our findings in this body of work draws attention to the concept of engaged lurking, a "strategic and idiosyncratic activity" that lets users meet their needs online while avoiding other concerns [371, 372, 373]. Scrolling and non-clicking practices are also viewed by this literature as privacy-protecting activities [374, 375]. This is particularly important in the case of mental health experiences in the aftermath of a socially stigmizing experience like a hospitalization, as users may fear revealing their health status to online audience by leaving behind visible digital traces. While we did not further investigate participants' motivations for passively consuming online content in this study, it is likely that they do so as for privacy reasons or to circumvent online advertisements and algorithmic content curation.

Furthermore, these invisible practices might also differentiate the social benefits that people draw from social technology use for reintegration. While our participants mentioned drawing certain benefits from passively consuming content online, how to do the social benefits and negative feelings of social comparison play out when people adopt these invisible practices of viewing content? How does the role of social technology in reintegration for mental health vary among groups that "actively" engage with people and content online vs. groups that adopt invisible engagement practices online. Future work can examine these differences and the differentiating role of social technology use for mental health.

Clinical Implications Reintegration and self-management of mental health symptoms outside the hospital and institutionalized clinical treatment strongly impact future clinical outcomes and overall well-being of individuals [91, 329]. Clinical literature highlights the importance of social reintegration and community participation in avoiding re-hospitalizations and supporting the clinical goals of recovery [103, 117] . However, reintegration is a complicated phenomenon to identify and existing instruments and validated measurements are limited; often they are specific to the domain are such as incarceration [376], refugee migration [377] etc.). Our empirical findings can inform clinical practices along the following directions.

Post-discharge care and support How one manages their illness outside of institutionalized clinical treatment strongly impacts both future clinical outcomes and overall well-being [329, 91]. But after discharge from the hospital, clinicians often lose timely contact with their patients which present challenges for continued care and support [328]. Even when patients adhere to clinical followup appointments, the emphasis is on treatment of symptoms and managing medication use and adherence, rather than on improvement in social well-being of the individual [103]. Clinical studies have highlighted the importance of social reintegration and community participation in avoiding re-hospitalizations and supporting the clinical goals of recovery [103]. However, reintegration is a complicated phenomenon to identify and existing instruments and validated measurements are limited [378]. Importantly, reintegration in mental illness is poorly understood for the reasons above. The predictive assessment of reintegration from Section 6.1 revealing likelihood of reintegration after hospitalization can help clinicians better understand the outside-hospital, post-discharge experiences of people. Especially for young adolescent populations, who are one of the most affected demographic with mental disorders [379], social media is a significant part of their identity and social lives [380]. How people re-establish their online presence and social connections after stigmatizing hospitalization events can act as collateral information [381] for follow-up during clinical sessions, discharge practices and community care. Findings from the interview study further highlight the implication of social technology use in people's reintegration journeys.

Our findings from Section 6.2 showed that psychiatric hospitalization removes individuals from their social lives, both online and offline. On top of that, participants felt overwhelmed by the sudden transition back into their normal lives, as well as triggered by life circumstances that contributed to their initial hospitalization. Accordingly, we suggest that prior to discharge, clinicians (including clinical psychologists, psychiatrists, and social workers) can discuss the process of social reintegration with patients. They can assist patients in developing a plan that allows them to slowly return to their everyday routine. As a patient reintegrates, they are leaving an environment centered around their mental illness for one that generally stigmatizes it. Our participants said that they may also exit the hospital with additional anxiety regarding the stigma that surrounds psychiatric hospitalization, as well as personal and professional complexities brought to the fore due to the hospitalization, whether around getting back to an abusive partner or dealing with homelessness. Clinicians could alleviate some of this anxiety by periodically assessing patient's preparedness towards reintegration and getting back to their normal lives. Clinicians can further help patients learn how to disclose their conditions to trusted others, whether online or offline. However, patients must also feel prepared in deciding whether or not to disclose, as social stigma may result in negative reactions from others. Many participants noted the crucial role that friends and family played in helping them readjust to their lifestyles. We discussed how several participants ended unhealthy friendships following hospitalization, both in their personal lives and on social media. It is important that clinicians ensure their patients have some form of a support system to assist them in the reintegration process. In doing so, they can also discuss the negative aspects of their patients' social lives, as some patients may have relationships or life circumstances that contribute to their struggles with mental health.

Sensemaking in health care settings Making sense of health status transitions is of critical importance for people who are dealing with life disruptions, such as health-related challenges [382]. Clinical research suggests that individuals involve in a "laborious sense-making activity" to "create a new link between past, present, and future" [383] in their search for normality and self-regulation [383]. Godbold et al. [384, 385] found that people with compromised health bridged information gaps by orienting themselves to repeated themes and to notions of what is "normal" among a group of peers. Accordingly, the taxonomy and findings from the first study demonstrate how people's embodiment in a combination of possible selves statuses surrounding psychiatric hospitalizations can act as new information in discursive therapy sessions and help in sensemaking of hospitalization experiences. Drawing from work in personal informatics and self-reflection [386, 18], the extracted PSS can act as questions that people pursue about their data to maintain awareness of their status relative to a goal (like social functioning or reintegration), or even for the purpose of self-experimentation to examine the efficacy of current treatment strategies. The Possible Selves framework [281] is also known to elicit behavioral change and has been adopted in psychotherapy [387], because the very conceptualization process of possible selves may lead the way toward more planned and intentional interventions. Supported by our taxonomy of PSS transitions, clinicians and therapists can encourage patients to discuss their recovery and reintegration trajectories. For instance, if a person was transitioning into a withdrawal-focused PSS in the long term (indicating an overall reduction in activity and engagement with others) clinicians can encourage them to discuss and address any challenges to social reintegration they might be facing after hospitalization. The derived transitions between PSS from RQ2 in Section 6.1 can further act as a component through which self-motivation and self-knowledge are influenced [286] in clinical care settings – after all, the essence of sensemaking for an individual with mental distress is often to embrace a positive change.

Holistic understanding of mental health resources available post-hospitalization Beyond mediating social interactions after the hospitalization, we found that social technologies played a significant role in participant's self management of their condition. In this work we found a plethora on social technology platforms such as Facebook, Whatsapp, Instagram, Youtube, Reddit, including video-conferencing tools and webinars, that played a role in care pathways after hospitalization. Our findings suggest that better understanding of the sources available to people via social technologies for management of mental illness can inform clinicians in their provision of post-discharge care. We found that after first psychiatric hospitalization experiences, participants appropriated platforms like Facebook, Twitter, Instagram, Youtube and Reddit to disclose about their illness, reach out and provide social support to similar others and share coping mechanisms and challenges to their reintegration online. Online disclosure and social support were particularly helpful for participants who did not feel comfortable disclosing to many people in their life. Majority of our participants shared that they had no prior knowledge about the types of online support groups available to them – most commonly they became aware of such groups' existence after being informed by another person in their life. Patients might benefit from having a compiled set of online resources available to them on different social platforms, and considerations that go into the use of these platforms for overall well-being. While clinicians themselves might not always be aware of all available resources, given that they are aware of their patients' needs and histories, they might be able to use such a resource to recommend which specific ones to use, and which ones patients might want to avoid. Furthermore, the ways participants appropriated social technologies presented positive benefits as well as challenges to their mental health journeys. In collaboration with their clinicians, patients could consider the aspects of their social technology use that may be damaging to their mental health and come up with a plan collaboratively to develop a more positive practices. Such a plan could even be adapted over time and persistently, as an individual pursues their recovery and reintegration journey. Lastly, we note a caveat that the inclusion of social technology platforms as resources for post-discharge care should not exclude opportunities for care for those who do not use such platforms.

Design implications ***Designing for online social reintegration*** In our work, we found evidence suggesting that Facebook can act as space for online social reintegration for individuals transitioning around psychiatric hospitalizations. Findings from the first study showed that emotional and personal sharing around hospitalization is predictive of reintegration in the future. Those who show a high likelihood of reintegration stay in touch with friends via messages more so than those who do not show these signals. So how can we design social media to support peoples' reintegration journeys? Extensive literature shows that social media use is associated with social capital gains – both bonding capital in the form of emotional support received from strong ties [90], and bridging capital in the form of new information received from weak ties [88]. Social capital benefits are immensely helpful to individuals during post-hospitalization periods. Strong ties can provide relationship maintenance benefits [388] and emotional support [389] after the psychiatric hospitalization. Weak ties can similarly be helpful. Regaining employment, access to new information and resources, support for health-related stigma and community support are all important factors that support reintegration journeys that weak ties are known to provide [125, 390]. Consistent with prior work on identity transitions [119], we found that social media enables people to embody multiple, heterogeneous possible selves surrounding hospitalization. Prior work suggests that individuals attempt to manage the link between their previous and current identity by editing self-presentational data, and the configuration of the network itself [339]. In the case of transitions surrounding hospitalization, individuals may wish to draw on social capital benefits discussed above from a selective set of Facebook friends, while not disclosing to others on their network. They might embody multiple PSS, say both incorporation-focused and withdrawal-focused PSS after hospitalization. How can we design social media to support selective presentation of the PSS to different audiences? To support this multiplicity, it will be helpful if platforms provided better controls for audience segmentation and selection [391], so that the multiple possible selves can be presented to appropriate audiences without inhibition and negative repercussions like context collapse or compromised privacy.

Personalized technology-based interventions for mental health Technology-based behavioral and psychological intervention strategies are increasingly applied to mental health for self-assessment and self-monitoring, psychoeducation, goal setting, and skill building [392]. To obtain desired outcomes, sustained client engagement and participation are crucial; however, technology-based interventions are known to have challenges with limited participation and high attrition rates [393]. Moreover, existing interventions consider different symptoms as distinct entities without taking into account the person's social and ecological context. However, the clinical literature says that the boundaries between the presentation of different disorders are often not so strict [394]. One way to improve engagement and create room for representing individualized heterogeneity of mental health states and experiences could be factoring in the specific individual's experiences into the intervention design and tailor support and care according to their context. The Possible Selves framework [281] and methodology we adopt in study 1 can be applied to engagement data from intervention tools to uncover patterns of behaviors that can inform personalized delivery of support and care [395]. From RQ2 in Section 6.1, we found that after hospitalization two distinct prominent trajectories emerge: those who incorporate the illness and get back to social lives, and those who withdraw from active use of platforms. The support needs of these two groups might vary drastically and the intervention tools can account for these differences to maintain engagement from clients with varied experiences. In essence, we posit that such an approach to describe a person's clinical and social care needs, as well as their mental health state aligns with the vision of NIMH's Research Domain Criteria (RDoC) [394], that defines five 'domains' each reflecting a psychophysiological system in which a person's functioning is impaired, to different degrees. The PSS-based recovery and reintegration trajectories can also inform out-patient programs or community integration programs that are traditionally referred to patients by social workers or staff during the time of discharge from the hospital. Specifically, this information might help contextualize how clinical treatment (as provided by the hospital), and social care (outside the hospital), including technology-mediated support, might contribute to a collaborative recovery and reintegration plan.

Designing for digital breaks Drawing on the nuances of digital breaks in the context of psychiatric hospitalization and mental health experiences, we identify design opportunities for social technology platforms to support people getting back online. Platform designers can pay attention to facilitating taking breaks from social technology. In the context of reintegration after psychiatric hospitalization, facilitating breaks would involve a set of closely connected, trusted considerations. First, platforms can begin to consider the design of push notifications after digital breaks or periods of inactivity to reduce feelings of overwhelm and anxiety due to information overload. Second, designs and affordances that provide safe spaces and support self-disclosure and selective sharing, i.e. supporting users to share personal, sensitive content in a way they feel comfortable about the privacy of their posts, could better facilitate mechanisms for reintegration. Importantly, as has been noted in recent HCI research, platforms need to be sensitive and respectful to people's life circumstances around major events and transitions [341], including psychiatric hospitalization. For instance, special consideration could be given to algorithmic ranking of information feeds, personalized features such as Facebook's "Year in Review" and advertisement recommendations for users who return after long breaks from the platform, including thinking through when and for who are these features appropriate at all.

Controls to manage information feeds While several participants in the second study highlighted the benefits of consuming inspirational content about positive life changes on social technology platforms, others mentioned how the information they consume online was stress-inducing, triggering their mental health symptoms or eliciting feelings of social comparison. However, only a few participants adopted practices like blocking, unfriending, re-configuring Facebook's News Feed settings or deactivating their accounts to counteract the negative effects. Platform designers can better account for these effects by providing users more agency and control in configuring their information feeds and recommendations. This could include features for controlling how much they see a specific type of content on their feeds. For instance, platforms can design "algorithmic marketplaces" providing a suite of content ranking and recommendation algorithms to users, who can select the desired ones based on their life circumstances. A feature like this would support several participants in our study, including who wished to only see funny videos to relax their moods, those who wished to stop seeing content triggering social comparison or mental health symptoms, and those who wished to only see positive, inspirational, or reaffirming quotes. Similarly, existing features for blocking, taking a break from other people's content, un-following, etc., can be made more apparent and easily accessible to users by providing "feature guides", for instance, to recommend their use.

We note an important caveat in these design suggestions. For psychiatric hospitalization, relapse is exceptionally common and people undergo multiple transitions and periods of lack of access to resources and technology non-use throughout their illness trajectory. Several participants in our study noted how one hospitalization was more prominent in their experience than others. People's reintegration journeys with and without social technologies may vary from one hospitalization experience to another. These design suggestions, therefore, cannot be uncritically followed without additional context about the specific person's needs, demands, or the broader life situation. Further research is needed to understand digital breaks in the context of psychiatric hospitalizations under the lens of technology non-use.

CHAPTER 7
CONCLUSION

The central hypothesis in this thesis is: social media, and algorithmic approaches informed by clinical and patient stakeholder perspectives, can support clinical and social pathways to care for mental health in the form of patient-provider interventions and social support provisions. My approach to developing a comprehensive understanding of the efficacy of social media for mental health care in this thesis has been primarily computational and empirical: I presented how mental health attributes can be characterized based on social media data, I developed computational models and evaluation techniques to examine whether social media data can assist in predicting clinical outcomes, and I discussed both qualitative and quantitative observational studies to improve understanding of mental health experiences and social media use. This approach necessitated close collaborations with domain stakeholders like clinical researchers and practitioners (psychiatrists and clinical psychologists) and people with lived experiences. Similarly, in characterizing and understanding mental health attributes from digital traces, this work heavily draws from social psychology, health sciences and psycholinguistics literature. In this chapter, I summarize the contributions of this thesis to theory and practice and discuss broader challenges, limitations and future directions that are possible due to this work.

7.1 Contributions

7.1.1 Theoretical contributions

Differential uses of social technologies for mental health This thesis presents a first empirical study to assess the quality of different social media-derived signals in predicting clinical diagnoses of mental illness, for treatment and patient-provider interventions. In doing so, this work unpacks the differentiating ways in which people appropriate social technologies for mental health and the efficacy of characterizing these behaviors as clinical outcomes. Through a series of works, we demonstrate that working with data volunteered and contributed by clinically diagnosed patient populations is imperative to realize the true potential of social media data in assisting clinical decision making. Furthermore, these findings surface methodological gaps in prior work employing social media data for predicting mental health states ranging from the uncertainties in the construct validity of the proxy signals, and poor theoretical grounding, to a variety of population and data sampling biases. As a remedial proposal, this thesis presents guidelines and research practices for participatory algorithmic development to address these methodological challenges.

Therapeutic outcomes of online broadcasting disclosures By analyzing the linguistic content shared around schizophrenia disclosures on Twitter, this work found that therapeutic benefits, are apparent even in broadcasting disclosures shared via a public social media platform like Twitter. For instance, long term trends indicative of reduction in the negative syndromes of the condition, such as decreasing negative affect, increasing positive affect, greater future orientation and reduced self preoccupation are observed. Consistent with prior work, these findings support the "venting out" phenomenon on social media; by disclosing one's deepest thoughts and feelings on social media, one can suppress and inhibit dysfunctional negative thoughts.

Impact of reciprocity and social support from online social networks on mental health disclosures This work also sheds light on the role of an "invisible audience" on online social platforms and the disclosure benefits people draw in connecting and opening up to such an audience. As such the online disclosure of mental health concerns and experiences may be framed as an interpersonal process, in which people regulate their disclosures based on what the invisible audience chooses to disclose about them – this supported by the observation that after the disclosure events, individuals tend to discuss more frequently about stigma related issues and consistently about mental illness topics, both of which are known indicators of reduced inhibition or self-restraint. Chapter 4 further provides evidence of reciprocity, both topically and temporally, in the interactions between the audience and disclosers. Although the nature of audience providing these social capital resources is nebulous, i.e. the disclosers may not necessarily know who this audience is, even if they have an imagined mental conception of who it might be, the reciprocal engagement that the audience provides over time confirms prior observations about online social platforms facilitating formation and maintenance of social capital and social support.

An empirically-derived taxonomy of heterogeneous behavioral patterns characterizing people's health transitions around psychiatric hospitalizations People's mental health statuses and their exhibition on social media are heterogeneous, and contextualized within the individuals' specific needs for clinical treatment and social support. In contrast to theoretical models used to understand health transitions in prior CSCW research that assume homogeneity, the taxonomy presented in Chapter 6 identifies heterogeneous behavioral patterns exhibited online around psychiatric hospitalization. The insights derived from the taxonomy, combining medical records and social media data, such as distinct behaviors exhibited prior and post hospitalization, can inform and enrich personalized predictive models for early interventions as well as tailored social support provisions along the course of the illness.

Recovery and Reintegration trajectories in mental health Insights from this thesis add to understanding of recovery and reintegration journeys in mental health. Based on the taxonomy of status exhibited by people with mental illness surrounding psychiatric hospitalization, this work identified common recovery and reintegration trajectories. These trajectories further unpacked the differences in mental health experiences across study population – while 66.6% of the transitions from incorporation to withdrawal focused status are seen during the first hospitalization events experienced by people, the majority of the cases where individuals remain in the incorporation focused status before and after the hospitalization (58.3%) were observed during subsequent hospitalizations (like the 2nd, 4th, or 5th hospitalization recorded for the individual). Consistent with literature from psychiatry, nursing and social work, these findings provide contextualized details about people's mental health status along the course of the illness, informing the design and development of technology-based interventions for mental health.

Role of social technologies in reintegration after psychiatric hospitalizations Findings from this thesis reveal the different approaches people adopted to re-establish social connections immediately after discharge from the hospital, demonstrating the role of online social platforms as spaces supporting reintegration in mental health. In relation to their reintegration journeys, social technology use supported as well as hindered participants' illness trajectories. While participants drew several social benefits from disclosure and social support via technology, some felt that their use of these platforms hindered their path to reintegration, due to feelings of social comparison, negative interactions, and emotional triggers to their mental health symptoms.

7.1.2 Practical contributions

Demonstration of efficacy of social media data in predicting clinical outcomes This thesis provides evidence that social media activity captures objective linguistic and behavioral markers of psychotic relapse in young individuals with recent onset psychosis. The machine learning models described in Chapter 3 demonstrate that it is possible to make personalized predictions of imminent relapse hospitalizations at the patient-specific level. These models alongside the participatory research approach involving clinician and patient perspectives serve as critical building blocks for the development of social-media based algorithmic systems for clinical decision making. The specific details on features that are predictive of schizophrenia disclosures or relapse hospitalizations, such as increased use of swear and anger related words, first person pronoun use, co-tagging behaviors and late night social media use, also inform clinicians new risk markers indicative of exacerbated mental health conditions.

Model of online social reintegration Reintegration is a complicated phenomenon to identify and existing instruments and validated measurements are limited; often they are specific to the domain being investigated (such as incarceration, refugee migration etc.). The predictive assessment of reintegration from Chapter 6 revealing likelihood of reintegration after hospitalization helps clinicians better understand the outside-hospital, post-discharge experiences of people. Especially for young adolescent populations, who are one of the most affected demographic with mental disorders, social media is a significant part of their identity and social lives. How people re-establish their online presence and social connections after stigmatizing hospitalization events can act as collateral information for follow-up during clinical sessions, discharge practices and community care.

Details on social media features that support disclosure and experiences related to mental health This thesis presents evidence on social media features that support people in making sensitive disclosures, reaching out and providing social support and managing self-presentation and connections during major life transitions like psychiatric hospitalization. The presence of the observed therapeutic benefits on Twitter, despite the stark affordances and norms of use of the platform, extends existing discussions in recent research: that the dichotomy between online and online expression, and its role in enabling candid self-disclosures, whether in the dyadic (private) or the broadcasting (public) form, might be blurring after all. Or that, the disclosers' are creating new opportunities to derive therapeutic benefits from short-form, spontaneous blurbs shared on public social media platforms, going beyond the structure of online dyadic therapist-client settings. Features such as Twitter mentions and the ability to reciprocate with personal, sensitive stories is found to impact disclosers' sharing of mental health experiences. Chapter 4 details how online platforms can be designed to support safe spaces for sharing mental health experiences. Similarly, findings from Chapter 7 highlight that while consumption and participation in online communities supported people's reintegration journeys after psychiatric hospitalization, emotional triggers to feelings of social comparison and negative interactions on these platforms presented hindrances to successful recovery. Understanding these features and their impact on people's well-being and mental health outcomes is critical in designing safe, healthy, supportive spaces on social media.

Informing the design of technology-assisted therapy tools and online mental health communities A crucial aspect of these technology-assisted therapy tools is providing the volunteers or the AI agents adequate resources, so they can successfully engage in conversations with help seekers. To do so, there is a need to capture timely feedback, in terms of the nature and quality of engagement (of the volunteer or AI agent), and their impact on future disclosure behavior of the help seekers. With the forecasting methodology described in Chapter 4, interactive systems can be built to enable the volunteers/agents/algorithms act on the help seekers/disclosers feedback on engagement in a timely manner. Similarly, the framework for studying patterns in audience engagement with respect to what the disclosers reveal about themselves can be adopted to identify specific engagement patterns signaling reciprocity. Upon identification, the usage of these markers can by promoted — either manually as guidelines to volunteers and support providers or algorithmically in the case of conversational agents. Finally, moderation efforts in online support communities and social media platforms can adopt our methodologies to similarly motivate audiences engage meaningfully with vulnerable self-disclosing individuals and to thereby create positively beneficial online therapeutic spaces.

7.2 Ethics

Participant Safety The introduction of insights from social media data into clinical settings, as discussed above, requires careful consideration of ethical implications concerning privacy, ethics, consent and clinical responsibility. First, consent procedures would not only need to include privacy and confidentiality with respect to data collection and data use, but also the clinical settings (in-patient, out-patient, therapy sessions) in which social media data might be incorporated. To elucidate this information, more research is needed from interdisciplinary teams of clinicians, researchers and patients to understand the efficacy and potential outcomes of introducing social media data into clinical contexts. When to seek consent from participants managing mental illness is also important. In the studies descried in this thesis, clinicians assessed potential participants' symptomatic conditions before reaching out regarding study participation. Participants should also be given an option to opt-out of research programs without any consequences to their ongoing treatment and care at the facility. Lastly, the data used in this study were obtained from consenting participants who were fully informed of the risks and benefits of participation. For future work, the potential for this information to reveal sensitive clinical insights may motivate other parties to collect and analyze it without consent. Thus, it is important for clinicians and researchers to develop standards to protect the confidentiality and the rights of this sensitive population to avoid misuse of personal information and to maintain individual autonomy.

Social Media Data Collection and Use The data used for studies described in Chapter 4 are publicly available and we do not interact with the users; therefore it did not qualify for approval from our respective Institutional Review Boards. However, without the users' consent, knowledge, or awareness, we are cognizant of the ethical limitations that occur in the absence of consent and feedback from the study population. To reduce risk of users' identity and data being revealed inadvertently, we paraphrased quotes in the paper, obfuscated any personally identifiable information, a method that has been used in other similar social computing work [147]. We acknowledge that these sensitive predictions of people's mental health state require ethical guidance beyond the purview of traditional ethics board. To that end, collaboratively developing dynamic and relevant ethical practices like participatory research efforts by the Connected and Open Research Ethics Initiative (CORE), to guide and navigate the social and ethical complexities of this research is incredibly important.

Negative Implications Recent literature [1] calls for researchers to pro-actively consider the negative implications and harms that might be caused by research practices and artifacts. In light of these discussion, I would like to note some unintended consequences in using data-driven, machine learning approaches for predicting mental health states and behaviors. The products of this research such as the relapse prediction model, are designed in a participatory manner with two stakeholders in mind, clinicians and their patients. One risk of this work is the misuse of such models by nefarious actors who do not have a role in supporting people's pathways to care. Identifying these actors and their intentions is complex and challenging. As the field of machine learning and health moves forward, it becomes imperative to develop standards, approvals and protections (similar to HIPAA compliance) for the use of computational, predictive models in clinical settings.

Caveats Regarding Design Implications Disclosing about stigmatized concerns like schizophrenia might call upon negative impacts such as social discrimination and rejection, which are detrimental to well-being. Therefore, the support recommendations discussed under design implications (in Chapters 3 and 6) need to be cognizant of the boundary regulation choices of the disclosing individuals, e.g., restricting recommendations to a chosen audience of the discloser, to prevent unintended negative consequences. Ideally, they also need to adapt to the responses that disclosures may elicit from an individual's social net-

[1] https://acm-fca.org/2018/03/29/negativeimpacts/

work, so that the amount of disclosure information revealed is adequate to gather support, however does not divulge excessive details about the user. Similarly, we indicated the possibility of sharing of social media archives with a therapist. These design approaches need to factor in boundary regulation issues in the patient-therapist interpersonal relationship, and need to develop adequate data and informational abstractions and curation methods. This would allow balancing the disclosers' clinical needs and their privacy expectations, attending to their privacy concerns at the forefront.

Another important caveat in these design suggestions is related to health transitions, reintegration and social media use (Chapter 6). Unlike other life transitions such as parenthood [124] or gender identity transitions [86], transitions of mental health statuses could be cyclical or non-linear, rarely unidirectional, or permanent. Relapse and re-hospitalizations are exceptionally common for the conditions we study in this work, like schizophrenia and mood disorders [176]. Therefore people undergo multiple transitions and possibly return to past selves several times along the course of the illness; this is unlikely the case with other life transitions explored in prior literature [119]. People may also choose to not use social media (or use it differently) during these major life transitions. Therefore, the above design suggestions cannot be uncritically followed without accounting for the complexities of the context. Further research is needed to uncover the specific ways and points in time that social media could support reintegration.

7.3 Limitations

This body of work has some notable limitations that I acknowledge in this section.

First, our examination of social media for mental health is limited in terms of the demographics and characteristics of the study population and platform of study.

Study population. This thesis focuses on one mental health condition, schizophrenia, to understand the pathways to care via social media for mental health. While social media has the potential to support interventions for other mental health conditions, as evidenced by prior work, the findings from this thesis (specifically, from chapters 3-5) may not generalize directly to other mental illnesses such as depression, mood disorder, bipolar and borderline personality disorders, etc. The studies on the intersection of clinical and social pathways to care included study populations with other diagnosis such as mood, anxiety and borderline personality disorders. While symptomatic experiences and objective outcomes of recovery, such as time taken for remission of symptoms, or probability of relapse might vary across conditions, clinical literature suggests commonalities in reintegration and recovery experiences across mental illnesses [396]. The commonalities and differences in mental health experiences and transitions experienced by people with different illnesses and co-morbid conditions needs to be further investigated. Future work can evaluate and extend our findings with other populations, including people with different mental health conditions, people who have experienced very long hospitalization periods, and people in different countries and cultures.

Demographics. The participants who consented and shared their medical records and social media data for the research described in Chapter 3 and Chapter 6 sought treatment within a single healthcare system in a specific geography, albeit one of the largest in the United States, comprising English-speaking, Western populations. The eligibility criteria for patient participant recruitment ranged from 15 to 35 years to reflect the inclusion criteria of the Early Treatment Program at the hospital, however, adolescents may engage with social media in a distinct manner compared to young adults. Similarly, participants in the interview study (Section 6.2) experienced psychiatric hospitalization only in the United States. The demographics therefore, are skewed and we caution against sweeping generalizations. Clinical literature suggests that schizophrenia manifests uniformly across demographic groups (gender, ethnicity, race) and geography [397] so we conjecture the demographic biases to be minimal. More research is needed across different demographic groups and different healthcare systems to extend the insights in this work.

Platform of study. Overall, this thesis includes the study of multiple social media platforms like Facebook, Twitter, Reddit, Instagram, etc., to understand the role of these platforms for mental health care interventions. The individual research studies, however, have been largely limited to a single platform of choice. For instance, the feasibility of social media data in predicting clinical outcomes like relapse is demonstrated based on Facebook data. The nature of self disclosures examined in Chapter 4 focuses on a micro-blogging platform like Twitter, which is likely to be very different from disclosures made on online communities platforms like Reddit, social networks like Facebook and so on. In Chapter 5, we have considered only three proxy diagnostic signals, although they are amongst the most widely used in the community. Additional investigations are required on alternative proxy signals and the potential of employing the use of multiple proxy signals in a concerted fashion. In chapter 6, we are unable to delineate the benefits/harms and the role of individual social platforms in reintegration for mental health. Prior work highlights that while research studies on social media focus on a single platform, people's lived experiences suggest that they incorporate multiple social media platforms into their communication practices and social needs [398] (consistent with our findings in Chapter 6.2.) Therefore, future work can seek to unravel the nuances of pathways of care for mental health across the social media ecology.

Second, the methodological choices, operationalization of constructs and decisions made during the research process have an impact on the findings in this work.

Social pathways to care. To operationalize self-disclosure on Twitter, we relied on a set of hand curated key phrases to assist in data collection. Although these phrases are clinician validated, they do not include all possible ways in which Twitter users disclose their diagnosis of schizophrenia. Relatedly, in our operationalization of intimacy of disclosures, we limit our focus to studying the impact of active, incoming audience engagement. Stemming from our interest in the invisible audience, we focused our attention on finding evidence for a general form of social benefits received by disclosure. These choices limit us from making claims about specific motivations, goals, and social benefits of online disclosures of mental health. For instance, disclosers might pursue goals other than social benefits, such as trust, impression management, and social validation that we do not disentangle in our analysis. We have also not probed into the nature of the audience and questions surrounding their own social media. Studying the alignment between discovered patterns of audience engagement and specific disclosure goals, and how non-responsive or non-supportive audience impacts future disclosure behaviors constitutes an interesting direction for future research. Further, the social benefits that we identify in our study (such as therapeutic outcomes, reciprocity) need further causal evidence and validation using self-reported data. Causal inference studies and qualitative data such as interviews can be powerful in complementing this line of work.

Clinical pathways to care. The performance of our relapse prediction algorithm in Chapter 3 was likely impacted by our definition of relapse, which was defined as a hospitalization due to psychotic symptoms. Relapse, however is a complicated phenomenon, and has other definitions, including symptomatic exacerbations that do not result in hospitalization. Furthermore, the decision to hospitalize is often multifactorial and may not always be a reliable indicator of psychotic symptoms. Our error analysis, described in Chapter 3 suggested that several periods believed to be incorrectly identified as periods of relapse did in fact have documented evidence for the presence of psychotic symptoms, although they did not necessarily result in a hospitalization.

Additionally, our approach was limited by our characterization of monthly periods of relative health and relative illness. First, illness trajectory for many individuals with psychotic disorders does not neatly fall into distinct segments of "health" and "illness", rather symptoms fluctuate over time. Furthermore, the recording of inpatient hospitalization dates were obtained via medical records, and it is possible that some hospitalizations were missing from the record and, therefore, not included in our analyses. In order to address these limitations and to improve our ability to find associations between social media activity and psychotic symptom exacerbations, future studies need to monitor participants prospectively and utilize frequent symptom rating scales to more accurately assess symptom severity. Second, while all participants included in our analyses experienced at least one relapse hospitalization, the specific symptoms that define an exacerbation for each individual with psychotic disorders are often unique, and although symptom heterogeneity was addressed in our analyses, generalizability may be limited. Third, some participants were more active on Facebook than others, providing varying degrees of extractable data. An important question for future research will be how much social media data is necessary in order to make a reliable clinical predictions. Lastly, the Facebook archives used for our analyses were collected retrospectively. While retrospective collection eliminates the possibility of altering behavior as a result of being monitored [399], to achieve the goal of early relapse identification, prospective monitoring will be necessary in future work.

Intersection of clinical and social care Along the third theme of this thesis, we were motivated by the recovery model to combine medical records with social media data, as they provide complementary insights into mental health transition experiences. Medical records indicate the clinical aspect of hospitalizations and provide temporal markers of health and illness – a person is hospitalized when their symptoms exacerbate and they are discharged from the hospital only after receiving appropriate treatment and medication. Social media data provide insights into people's social interactions with others (through messaging, check-ins, etc.), their emotions, and self-presentation aspects related to mental health status. However, we acknowledge that these data do not comprise an exhaustive set of representative signals for clinical and social care. Among the clinical attributes, we are looking at summary attributes: diagnosis code and hospitalization dates that provide valid clinical information, often condensed from richer medical records capturing patient's status, such as symptomatic expression, self-reports and collateral information gathered by the clinician from during appointments, responses on any clinical instruments/assessments employed, or medication adherence and history of therapy. Similarly, not all aspects of social care might be visible on social media. For instance, caregivers and offline connections, community integration programs and support groups both offline and online also contribute to social aspects of care in mental health. Future work can explore these offline facets and other frameworks to investigate the intersection of clinical and social aspects of mental health care.

Another factor that needs further investigation in these works is the role of device non-use during the period of psychiatric hospitalization. People are not allowed to use their phone/computer during in-patient hospitalization. They might occasionally be permitted to access a computer within the hospital but they do not have access to social media sites during this time. As people are not allowed access to technology during in-patient hospitalization digital traces on their Facebook data during this time are mostly empty. This aspect is uniform across all participants, therefore, we do not expect changes in the empirical results from 6.1. However, the role of device non-use and digital breaks on social media posting behavior and its association with withdrawal-focused status after hospitalization needs further delineation. Finally, contextualizing the PSS-based trajectories in people's experiences and the use of this information in clinical contexts requires additional rigorous validation in future work.

Across these works, we are unable to present causal evidence between social media based online behaviors and mental health attributes (both clinical outcomes such as relapse and social outcomes like social support). Where it is possible, ethically and logistically, we employed methods such as comparison with a matched control group to establish weak causation. However, further work is needed to examine to what extent social media based behavioral data, causally relate to mental health attributes and individuals' well-being.

7.4 Future Work

Health Equity. The application of data-driven algorithmic techniques to problems in mental health, for instance, clinical decision making, raises critical questions about who are the individuals who benefit from the ML-driven clinical interventions? And, who are the individuals who might be harmed or left out by these technologies? In my work, I've studied multiple groups of individuals to understand mental health and social media use; people who publicly disclose about their illness online, people who participate in awareness campaigns online, people who follow online support forums for mental health, clinically diagnosed patients, and people who experienced a psychiatric hospitalization. However, this excludes all those who do not use social media platforms, those who are not formally diagnosed, those who do not reach out to care and support due to the stigma surrounding mental health, and those who do not have access to resources for care. Findings from this thesis show that the envisioned ML/AI approaches on social media data for mental health do not generalize well across these groups of individuals who are included or excluded by research practices. How can we develop AI/ML systems for mental health to be more equitable and support people's differential health needs? In operationalizing health outcomes for computational approaches, how can we adopt a human-centered perspective that considers the complete individual and their context? Having AI systems for mental health work fairly across subgroups while avoiding harm, and being accountable is a challenge. A fruitful direction for future work is to systematically investigate how we can develop standards to document the individuals and their context in machine learning/artificial intelligence systems for mental health and innovate machine learning approaches that might successfully scale across subgroups.

Human AI Interaction. The effectiveness of social media based interventions in conjunction with machine learning approaches depends on whether processes exist for domain experts to trust, understand and interact with ML models and incorporate it into their decision making. What consequences do these predictions have on the therapeutic alliance between clinicians and patients? What recourse processes can we design when the ML/AI algorithm makes prediction errors? One approach to this challenge is developing less sophisticated models that can be easily communicated to clinicians collaborators by innovating around post-hoc explanations for model's predictions. Another perspective is understanding from stakeholders' perspectives the gaps in knowledge or vocabulary about computational techniques applied to predict health outcomes. Particularly, building shared vocabularies for computational concepts and practices and bridging knowledge gaps can be an impactful future direction sustain inter-disciplinary collaborations between computer scientists clinician and health care experts. Thus, what comes next in realizing the potential of social media in supporting interventions for mental health is understanding how domain experts interact with data-driven, machine learning model predictions of clinical outcomes and, how these new decision support systems transform existing clinical practices, care pathways and trust between patients and providers.

Harm reduction. While the focus of this dissertation is on the potential of social media in supporting mental health care, literature also points to harmful health content online that can have dangerous contagion effects [400]. And not all the ways people appropriate social media is supportive for mental health outcomes. Thus, pro-active identification and mitigation of harmful behaviors and effects of harmful content related to mental health is a critical next step. More importantly, developing techniques to identify such behaviors (that might work in clandestine ways) at scale and designing fair, appropriate moderation mechanisms to manage the content spread is crucial.

REFERENCES

[1] N. Andalibi, O. L. Haimson, M. De Choudhury, and A. Forte, "Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2016, pp. 3906–3918.

[2] N. Andalibi, P. Ozturk, and A. Forte, "Sensitive self-disclosures, responses, and social support on instagram: The case of #depression.," in *CSCW*, 2017, pp. 1485–1500.

[3] M. De Choudhury and E. Kiciman, "The language of social support in social media and its effect on suicidal ideation risk," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[4] E. Sharma and M. De Choudhury, "Mental health support and its relationship to linguistic accommodation in online communities," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 2018, p. 641.

[5] S. Chancellor, Z. Lin, E. L. Goodman, S. Zerwas, and M. De Choudhury, "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, 2016, pp. 1171–1184.

[6] B. Inkster, D. Stillwell, M. Kosinski, and P. Jones, "A decade into facebook: Where is psychiatry in the digital age?" *The Lancet Psychiatry*, vol. 3, no. 11, pp. 1087–1090, 2016.

[7] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[8] M. Mitchell, K. Hollingshead, and G. Coppersmith, "Quantifying the language of schizophrenia in social media.," in *CLPsych@ HLT-NAACL*, 2015, pp. 11–20.

[9] G. Coppersmith, C. Harman, and M. Dredze, "Measuring post traumatic stress disorder in twitter," in *International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[10] M. Leamy, V. Bird, C. Le Boutillier, J. Williams, and M. Slade, "Conceptual framework for personal recovery in mental health: Systematic review and narrative synthesis," *The British Journal of Psychiatry*, vol. 199, no. 6, pp. 445–452, 2011.

[11] J. Mezzich, J. Snaedal, C. van Weel, and I. Heath, "The international network for person-centered medicine: Background and first steps," *World Medical Journal*, vol. 55, no. 3, pp. 104–107, 2009.

[12] C. Sandvig and E. Hargittai, "How to think about digital research," *Digital confidential: The secrets of studying online behavior*, pp. 1–25, 2015.

[13] A. Van Gennep, *The rites of passage*. Routledge, 2013.

[14] L. Harrington, R. Siegert, and J. McClure, "Theory of mind in schizophrenia: A critical review," *Cognitive neuropsychiatry*, vol. 10, no. 4, pp. 249–286, 2005.

[15] R. S. Keefe, P. D. Harvey, M. F. Lenzenweger, M. Davidson, S. H. Apter, J. Schmeidler, R. C. Mohs, and K. L. Davis, "Empirical assessment of the factorial structure of clinical symptoms in schizophrenia: Negative symptoms," *Psychiatry Research*, vol. 44, no. 2, pp. 153–165, 1992.

[16] R. P. Liberman and A. Kopelowicz, "Recovery from schizophrenia: A challenge for the 21st century," *International review of psychiatry*, vol. 14, no. 4, pp. 245–255, 2002.

[17] G. Bateson, D. D. Jackson, J. Haley, and J. Weakland, "Toward a theory of schizophrenia," *Behavioral science*, vol. 1, no. 4, pp. 251–264, 1956.

[18] E. L. Murnane, D. Cosley, P. Chang, S. Guha, E. Frank, G. Gay, and M. Matthews, "Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: Implications for the design of technologies to manage mental health," *Journal of the American Medical Informatics Association*, vol. 23, no. 3, pp. 477–484, 2016.

[19] M. Matthews, E. Murnane, J. Snyder, S. Guha, P. Chang, G. Doherty, and G. Gay, "The double-edged sword: A mixed methods study of the interplay between bipolar disorder and technology use," *Computers in Human Behavior*, vol. 75, pp. 288–300, 2017.

[20] K. McManus, E. K. Mallory, R. L. Goldfeder, W. A. Haynes, and J. D. Tatum, "Mining twitter data to improve detection of schizophrenia," *AMIA Summits on Translational Science Proceedings*, vol. 2015, p. 122, 2015.

[21] M. Välimäki, C. Athanasopoulou, M. Lahti, and C. E. Adams, "Effectiveness of social media interventions for people with schizophrenia: A systematic review and meta-analysis," *Journal of medical Internet research*, vol. 18, no. 4, e92, 2016.

[22] R. Emsley, B. Chiliza, L. Asmal, and B. H. Harvey, "The nature of relapse in schizophrenia," *BMC psychiatry*, vol. 13, no. 1, p. 50, 2013.

[23] J. A. Lieberman, D. Perkins, A. Belger, M. Chakos, F. Jarskog, K. Boteva, and J. Gilmore, “The early stages of schizophrenia: Speculations on pathogenesis, pathophysiology, and therapeutic approaches,” *Biological psychiatry*, vol. 50, no. 11, pp. 884–897, 2001.

[24] F. J. Frese, “Advocacy, recovery, and the challenges of consumerism for schizophrenia,” *Psychiatric Clinics*, vol. 21, no. 1, pp. 233–249, 1998.

[25] G. R. Kuperberg, “Language in schizophrenia part 1: An introduction,” *Language and linguistics compass*, vol. 4, no. 8, pp. 576–589, 2010.

[26] C. R. Altable, “Logic structure of clinical judgment and its relation to medical and psychiatric semiology,” *Psychopathology*, vol. 45, no. 6, pp. 344–351, 2012.

[27] M. Park, D. W. McDonald, and M. Cha, “Perception differences between the depressed and non-depressed users in twitter.,” in *ICWSM*, 2013.

[28] M. De Choudhury, S. Counts, and E. Horvitz, “Major life changes and behavioral markers in social media: Case of childbirth,” in *CSCW*, ACM, 2013, pp. 1431–1442.

[29] S. Chancellor, J. A. Pater, T. Clear, E. Gilbert, and M. De Choudhury, “# thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities,” in *CSCW*, ACM, 2016, pp. 1201–1213.

[30] 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,” *NAACL HLT 2015*, p. 1, 2015.

[31] L. Manikonda and M. De Choudhury, “Modeling and understanding visual attributes of mental health disclosures in social media,” 2017.

[32] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, “Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines.,” *American Psychologist*, vol. 70, no. 6, p. 543, 2015.

[33] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, *et al.*, “Life in the network: The coming age of computational social science,” *Science (New York, NY)*, vol. 323, no. 5915, p. 721, 2009.

[34] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, “The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care in-

terventions: Explanation and elaboration," *PLoS medicine*, vol. 6, no. 7, e1000100, 2009.

[35] Z. Jamil, "Monitoring tweets for depression to detect at-risk users," Ph.D. dissertation, Université d'Ottawa/University of Ottawa, 2017.

[36] Y. Zhou, J. Zhan, and J. Luo, "Predicting multiple risky behaviors via multimedia content," in *Social Informatics*, G. L. Ciampaglia, A. Mashhadi, and T. Yasseri, Eds., Cham: Springer International Publishing, 2017, pp. 65–73.

[37] G. Gkotsis, A. Oellrich, S. Velupillai, M. Liakata, T. J. Hubbard, R. J. Dobson, and R. Dutta, "Characterisation of mental health conditions in social media using informed deep learning," *Scientific reports*, vol. 7, p. 45 141, 2017.

[38] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 217–226, 2014.

[39] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu, "Depression detection via harvesting social media: A multimodal dictionary learning solution."

[40] J. H. Shen and F. Rudzicz, "Detecting anxiety through reddit," in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, 2017, pp. 58–65.

[41] A. Benton, M. Mitchell, and D. Hovy, "Multi-task learning for mental health using social media text," *arXiv preprint arXiv:1712.03538*, 2017.

[42] P. Burnap, W. Colombo, and J. Scourfield, "Machine classification and analysis of suicide-related communication on twitter," in *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, ACM, 2015, pp. 75–84.

[43] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *ACL Workshop on Computational Linguistics and Clinical Psychology*, 2014.

[44] G. Coppersmith, R. Leary, E. Whyne, and T. Wood, "Quantifying suicidal ideation via language usage on social media," in *Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM*, 2015.

[45] M. De Choudhury, M. Gamon, A. Hoff, and A. Roseway, "''moon phrases": A social media faciliated tool for emotional reflection and wellness," in *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '13, ICST (Institute for Computer Sciences,

Social-Informatics and Telecommunications Engineering), 2013, pp. 41–44, ISBN: 978-1-936968-80-0.

[46] X. Huang, X. Li, T. Liu, D. Chiu, T. Zhu, and L. Zhang, "Topic model for identifying suicidal ideation in chinese microblog," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, 2015, pp. 553–562.

[47] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li, and T. Zhu, "Detecting suicidal ideation in chinese microblogs with psychological lexicons," in *Ubiquitous Intelligence and Computing, 2014 IEEE 11th Intl Conf on and IEEE 11th Intl Conf on and Autonomic and Trusted Computing, and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom)*, IEEE, 2014, pp. 844–849.

[48] H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J. Huang, L. Cai, and L. Feng, "User-level psychological stress detection from social media using deep neural network," in *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, 2014, pp. 507–516.

[49] H. Lin, J. Jia, L. Nie, G. Shen, and T.-S. Chua, "What does social media say about your stress?.."

[50] K. Loveys, P. Crutchley, E. Wyatt, and G. Coppersmith, "Small but mighty: Affective micropatterns for quantifying mental health from social media language," in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, 2017, pp. 85–95.

[51] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on twitter," *Internet Interventions*, vol. 2, no. 2, pp. 183–188, 2015.

[52] V. M. Prieto, S. Matos, M. Alvarez, F. Cacheda, and J. L. Oliveira, "Twitter: A good place to detect health conditions," *PloS one*, vol. 9, no. 1, e86191, 2014.

[53] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, "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*, 2015, pp. 99–107.

[54] T. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez, and C. Giraud-Carrier, "Detecting cognitive distortions through machine learning text analytics," in *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*, IEEE, 2017, pp. 508–512.

[55] N. Vedula and S. Parthasarathy, "Emotional and linguistic cues of depression from social media," in *Proceedings of the 2017 International Conference on Digital Health*, ACM, 2017, pp. 127–136.

[56] 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.

[57] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2013, pp. 201–213.

[58] M. L. Birnbaum, S. K. Ernala, A. F. Rizvi, M. De Choudhury, and J. M. Kane, "A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals," *J. Med. Internet Res*, 2017.

[59] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating machine learning algorithms for twitter data against established measures of suicidality," *JMIR mental health*, vol. 3, no. 2, 2016.

[60] S. Chancellor, Y. Kalantidis, J. A. Pater, M. De Choudhury, and D. A. Shamma, "Multimodal classification of moderated online pro-eating disorder content," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 2017, pp. 3213–3226.

[61] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," *EPJ Data Science*, vol. 6, no. 1, p. 15, 2017.

[62] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, p. 13 006, 2017.

[63] P. Resnik, A. Garron, and R. Resnik, "Using topic modeling to improve prediction of neuroticism and depression," in *Proceedings of the 2013 Conference on Empirical Methods in Natural*, Association for Computational Linguistics, 2013, pp. 1348–1353.

[64] K. Saha, L. Chan, K. De Barbaro, G. D. Abowd, and M. De Choudhury, "Inferring mood instability on social media by leveraging ecological momentary assessments," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 95, 2017.

[65] H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar, "Towards assessing changes in degree of depression through facebook," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 118–125.

[66] S. Tsugawa, Y. Mogi, Y. Kikuchi, F. Kishino, K. Fujita, Y. Itoh, and H. Ohsaki, "On estimating depressive tendencies of twitter users utilizing their tweet data," in *Virtual Reality (VR), 2013 IEEE*, IEEE, 2013, pp. 1–4.

[67] d. boyd danah and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, communication & society*, vol. 15, no. 5, pp. 662–679, 2012.

[68] R. Kitchin, "Big data, new epistemologies and paradigm shifts," *Big Data & Society*, vol. 1, no. 1, p. 2 053 951 714 528 481, 2014.

[69] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The parable of google flu: Traps in big data analysis," *Science*, vol. 343, no. 6176, pp. 1203–1205, 2014.

[70] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman, "Social data: Biases, methodological pitfalls, and ethical boundaries," 2016.

[71] A. F. Lehman, J. A. Lieberman, L. B. Dixon, T. H. McGlashan, A. L. Miller, D. O. Perkins, J. Kreyenbuhl, J. S. McIntyre, S. C. Charles, K. Altshuler, *et al.*, "Practice guideline for the treatment of partients with schizophrenia," *American Journal of psychiatry*, vol. 161, no. 2 SUPPL. 2004.

[72] E. Goffman, *Stigma: Notes on the management of spoiled identity*. Simon and Schuster, 2009.

[73] S. M. Jourard, "Self-disclosure: An experimental analysis of the transparent self.," 1971.

[74] D. M. Quinn and S. R. Chaudoir, "Living with a concealable stigmatized identity: The impact of anticipated stigma, centrality, salience, and cultural stigma on psychological distress and health.," *Journal of personality and social psychology*, vol. 97, no. 4, p. 634, 2009.

[75] R. R. Rodriguez and A. E. Kelly, "Health effects of disclosing secrets to imagined accepting versus nonaccepting confidants," *Journal of Social and Clinical Psychology*, vol. 25, no. 9, pp. 1023–1047, 2006.

[76] J. W. Pennebaker, "Writing about emotional experiences as a therapeutic process," *Psychological science*, vol. 8, no. 3, pp. 162–166, 1997.

[77] S. E. Ullman and H. H. Filipas, "Predictors of ptsd symptom severity and social reactions in sexual assault victims," *Journal of traumatic stress*, vol. 14, no. 2, pp. 369–389, 2001.

[78] A. N. Joinson, "Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity," *European journal of social psychology*, vol. 31, no. 2, pp. 177–192, 2001.

[79] A. Joinson, "Causes and implications of disinhibited behavior on the internet.," 1998.

[80] A. N. Joinson and C. B. Paine, "Self-disclosure, privacy and the internet," *The Oxford handbook of Internet psychology*, pp. 237–252, 2007.

[81] P. C. Cozby, "Self-disclosure: A literature review.," *Psychological bulletin*, vol. 79, no. 2, p. 73, 1973.

[82] S. Balani and M. De Choudhury, "Detecting and characterizing mental health related self-disclosure in social media," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 2015, pp. 1373–1378.

[83] M. De Choudhury and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, 2014.

[84] M. De Choudhury, S. Sharma, T. Logar, W. Eekhout, and R. Nielsen, "Quantifying and understanding gender and cross-cultural differences in mental health expression via social media," in *CSCW*, 2017.

[85] Y.-C. Wang, M. Burke, and R. Kraut, "Modeling self-disclosure in social networking sites," in *CSCW*, ACM, 2016, pp. 74–85.

[86] O. L. Haimson, J. R. Brubaker, L. Dombrowski, and G. R. Hayes, "Disclosure, stress, and support during gender transition on facebook," in *CSCW*, ACM, 2015, pp. 1176–1190.

[87] J. S. Coleman, "Social capital in the creation of human capital," 1988.

[88] R. D. Putnam *et al.*, *Bowling alone: The collapse and revival of American community*. Simon and schuster, 2000.

[89] S. Szreter and M. Woolcock, "Health by association? social capital, social theory, and the political economy of public health," *International journal of epidemiology*, vol. 33, no. 4, pp. 650–667, 2004.

[90] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook "friends:" social capital and college students' use of online social network sites," *JCMC*, 2007.

[91] J. F. Helliwell and R. D. Putnam, "The social context of well-being.," *Philos. Trans. Royal Soc. B*, 2004.

[92] R. Zhang, "The stress-buffering effect of self-disclosure on facebook: An examination of stressful life events, social support, and mental health among college students," *Comput. Hum. Behav.*, 2017.

[93] L. L. Bachrach, *Deinstitutionalization-an Analytical Review and Sociological Perspective*, 4. US Department of Health, Education, and Welfare, Public Health Service . . ., 1975.

[94] S. S. Sharfstein, "Goals of inpatient treatment for psychiatric disorders," *Annual Review of Medicine*, vol. 60, pp. 393–403, 2009.

[95] L. J. Cohen, "Psychiatric hospitalization as an experience of trauma," *Archives of Psychiatric Nursing*, vol. 8, no. 2, pp. 78–81, 1994.

[96] S. Kent and P. Yellowlees, "Psychiatric and social reasons for frequent rehospitalization," *Psychiatric Services*, vol. 45, no. 4, pp. 347–350, 1994.

[97] M. C. Zanarini, F. R. Frankenburg, G. S. Khera, and J. Bleichmar, "Treatment histories of borderline inpatients," *Comprehensive psychiatry*, vol. 42, no. 2, pp. 144–150, 2001.

[98] D. Paksarian, R. Mojtabai, R. Kotov, B. Cullen, K. L. Nugent, and E. J. Bromet, "Perceived trauma during hospitalization and treatment participation among individuals with psychotic disorders," *Psychiatric Services*, vol. 65, no. 2, pp. 266–269, 2014.

[99] A. Tulloch, A. David, and G. Thornicroft, "Exploring the predictors of early readmission to psychiatric hospital," *Epidemiology and psychiatric sciences*, vol. 25, no. 2, pp. 181–193, 2016.

[100] M. Ådnanes, L. Melby, J. Cresswell-Smith, H. Westerlund, L. Rabbi, M. Dernovšek, L. Šprah, R. Sfetcu, C. Straßmayr, and V. Donisi, "Mental health service users' experiences of psychiatric re-hospitalisation-an explorative focus group study in six european countries," *BMC health services research*, vol. 18, no. 1, pp. 1–8, 2018.

[101] K. Newman-Taylor, C. Garner, E. Vernon-Wilson, K. H. Paas, L. Herbert, and S. K. Au-Yeung, "Psychometric evaluation of the hope, agency and opportunity (hao); a

brief measure of mental health recovery," *Journal of Mental Health*, vol. 26, no. 6, pp. 562–568, 2017.

[102] M. P. Dijkers, G. Whiteneck, and R. El-Jaroudi, "Measures of social outcomes in disability research," *Archives of physical medicine and rehabilitation*, vol. 81, S63–S80, 2000.

[103] M. Ådnanes, J. Cresswell-Smith, L. Melby, H. Westerlund, L. Šprah, R. Sfetcu, C. Straßmayr, and V. Donisi, "Discharge planning, self-management, and community support: Strategies to avoid psychiatric rehospitalisation from a service user perspective," *Patient Education and Counseling*, 2019.

[104] T. Viggiano, H. A. Pincus, and S. Crystal, "Care transition interventions in mental health," *Current opinion in psychiatry*, vol. 25, no. 6, pp. 551–558, 2012.

[105] W. A. Anthony, "Recovery from mental illness: The guiding vision of the mental health service system in the 1990s.," *Psychosocial rehabilitation journal*, vol. 16, no. 4, p. 11, 1993.

[106] S. Ramon, B. Healy, and N. Renouf, "Recovery from mental illness as an emergent concept and practice in australia and the uk," *International Journal of Social Psychiatry*, vol. 53, no. 2, pp. 108–122, 2007.

[107] L. M. Lines, M. Lepore, and J. M. Wiener, "Patient-centered, person-centered, and person-directed care: They are not the same," *Medical care*, vol. 53, no. 7, pp. 561–563, 2015.

[108] V. A. Entwistle and I. S. Watt, "Treating patients as persons: A capabilities approach to support delivery of person-centered care," *The American Journal of Bioethics*, vol. 13, no. 8, pp. 29–39, 2013.

[109] B. Starfield, "Is patient-centered care the same as person-focused care?" *The Permanente Journal*, vol. 15, no. 2, p. 63, 2011.

[110] H. L. Provencher and C. L. Keyes, "Recovery: A complete mental health perspective," in *Mental well-being*, Springer, 2013, pp. 277–297.

[111] D. Kralik, K. Visentin, and A. Van Loon, "Transition: A literature review," *Journal of advanced nursing*, vol. 55, no. 3, pp. 320–329, 2006.

[112] A. Van Gennep, *The rites of passage*, 1909.

[113] D. Kralik, "The quest for ordinariness: Transition experienced by midlife women living with chronic illness," *Journal of advanced nursing*, vol. 39, no. 2, pp. 146–154, 2002.

[114] M. Glacken, G. Kernohan, and V. Coates, "Diagnosed with hepatitis c: A descriptive exploratory study," *International journal of nursing studies*, vol. 38, no. 1, pp. 107–116, 2001.

[115] E. Kaziunas, A. G. Buyuktur, J. Jones, S. W. Choi, D. A. Hanauer, and M. S. Ackerman, "Transition and reflection in the use of health information: The case of pediatric bone marrow transplant caregivers," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 2015, pp. 1763–1774.

[116] G. A. Masters, R. J. Baldessarini, D. Öngür, and F. Centorrino, "Factors associated with length of psychiatric hospitalization," *Comprehensive psychiatry*, vol. 55, no. 3, pp. 681–687, 2014.

[117] N. C. Silva, D. G. Bassani, and L. S. Palazzo, "A case-control study of factors associated with multiple psychiatric readmissions," *Psychiatric Services*, vol. 60, no. 6, pp. 786–791, 2009.

[118] M. Massimi, J. P. Dimond, and C. A. Le Dantec, "Finding a new normal: The role of technology in life disruptions," in *Proceedings of the acm 2012 conference on computer supported cooperative work*, ACM, 2012, pp. 719–728.

[119] O. Haimson, "Social media as social transition machinery," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, p. 63, 2018.

[120] T. Morioka, N. B. Ellison, and M. Brown, "Identity work on social media sites: Disadvantaged students' college transition processes," in *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, ACM, 2016, pp. 848–859.

[121] M. De Choudhury and M. Massimi, """ she said yes"–liminality and engagement announcements on twitter," *iConference 2015 Proceedings*, 2015.

[122] B. C. Semaan, L. M. Britton, and B. Dosono, "Transition resilience with icts:'identity awareness' in veteran re-integration," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2016, pp. 2882–2894.

[123] M. Massimi, R. Harper, and A. J. Sellen, "Real, but glossy: Technology and the practical pursuit of magic in modern weddings," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM, 2014, pp. 854–865.

[124] L. Gibson and V. L. Hanson, "Digital motherhood: How does technology help new mothers?" In *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 2013, pp. 313–322.

[125] 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*, 2013, pp. 1419–1430.

[126] S. Yarosh, Y. Chieh, G. D. Abowd, *et al.*, "Supporting parent–child communication in divorced families," *International Journal of Human-Computer Studies*, vol. 67, no. 2, pp. 192–203, 2009.

[127] M. Massimi and R. M. Baecker, "A death in the family: Opportunities for designing technologies for the bereaved," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, 2010, pp. 1821–1830.

[128] W. R. Hobbs and M. K. Burke, "Connective recovery in social networks after the death of a friend," *Nature Human Behaviour*, vol. 1, no. 5, p. 0092, 2017.

[129] E. Getty, J. Cobb, M. Gabeler, C. Nelson, E. Weng, and J. Hancock, "I said your name in an empty room: Grieving and continuing bonds on facebook," in *Proceedings of the SIGCHI Conference on human factors in computing systems*, 2011, pp. 997–1000.

[130] J. R. Brubaker, F. Kivran-Swaine, L. Taber, and G. R. Hayes, "Grief-stricken in a crowd: The language of bereavement and distress in social media," in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[131] M. E. Smith, D. T. Nguyen, C. Lai, G. Leshed, and E. P. Baumer, "Going to college and staying connected: Communication between college freshmen and their parents," in *Proceedings of the ACM 2012 conference on computer supported cooperative work*, ACM, 2012, pp. 789–798.

[132] M. J. Paul, R. W. White, and E. Horvitz, "Diagnoses, decisions, and outcomes: Web search as decision support for cancer," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 831–841.

[133] J. Liu, E. R. Weitzman, and R. Chunara, "Assessing behavior stage progression from social media data," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 1320–1333.

[134] T. Huang, A. Elghafari, K. Relia, and R. Chunara, "High-resolution temporal representations of alcohol and tobacco behaviors from social media data," *Proceedings of the ACM on human-computer interaction*, vol. 1, no. CSCW, pp. 1–26, 2017.

[135] D. MacLean, S. Gupta, A. Lembke, C. Manning, and J. Heer, "Forum77: An analysis of an online health forum dedicated to addiction recovery," in *Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2015.

[136] J. O. Prochaska and W. F. Velicer, "The transtheoretical model of health behavior change," *American journal of health promotion*, vol. 12, no. 1, pp. 38–48, 1997.

[137] G. R. Hayes, G. D. Abowd, J. S. Davis, M. L. Blount, M. Ebling, and E. D. Mynatt, "Opportunities for pervasive computing in chronic cancer care," in *International Conference on Pervasive Computing*, Springer, 2008, pp. 262–279.

[138] M. Jacobs, J. Clawson, and E. D. Mynatt, "A cancer journey framework: Guiding the design of holistic health technology," in *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2016, pp. 114–121.

[139] Z. Levonian, D. R. Erikson, W. Luo, S. Narayanan, S. Rubya, P. Vachher, L. Terveen, and S. Yarosh, "Bridging qualitative and quantitative methods for user modeling: Tracing cancer patient behavior in an online health community," in *ICWSM*, 2020.

[140] J. Eschler and W. Pratt, "" i'm so glad i met you" designing dynamic collaborative support for young adult cancer survivors," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 1763–1774.

[141] M. Wen and C. P. Rosé, "Understanding participant behavior trajectories in online health support groups using automatic extraction methods," in *Proceedings of the 17th ACM international conference on Supporting group work*, 2012, pp. 179–188.

[142] J. L. Feuston, C. G. Marshall-Fricker, and A. M. Piper, "The social lives of individuals with traumatic brain injury," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 182–194.

[143] A. L. Strauss, *Social organization of medical work*. University of Chicago Press, 1985, ISBN: 9780226777078.

[144] J. M. Corbin and A. Strauss, "A nursing model for chronic illness management based upon the trajectory framework," *Scholarly inquiry for nursing practice*, vol. 5, no. 3, pp. 155–174, 1991.

[145] Y. Chen, "Health information use in chronic care cycles," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011, pp. 485–488.

[146] E. R. Burgess, M. C. Reddy, A. Davenport, P. Laboi, and A. Blandford, "" tricky to get your head around" information work of people managing chronic kidney disease in the uk," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–17.

[147] J. A. Pater, O. L. Haimson, N. Andalibi, and E. D. Mynatt, "“‘hunger hurts but starving works”: Characterizing the presentation of eating disorders online," in *CSCW*, ACM, 2016, pp. 1185–1200.

[148] M. W. Newman, D. Lauterbach, S. A. Munson, P. Resnick, and M. E. Morris, "It's not that i don't have problems, i'm just not putting them on facebook: Challenges and opportunities in using online social networks for health," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, ACM, 2011, pp. 341–350.

[149] X. Ding, Y. Chen, Z. Ding, and Y. Xu, "Boundary negotiation for patient-provider communication via wechat in china," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–24, 2019.

[150] C.-F. Chung, E. Agapie, J. Schroeder, S. Mishra, J. Fogarty, and S. A. Munson, "When personal tracking becomes social: Examining the use of instagram for healthy eating," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1674–1687.

[151] J. A. Pater, B. Farrington, A. Brown, L. E. Reining, T. Toscos, and E. D. Mynatt, "Exploring indicators of digital self-harm with eating disorder patients: A case study," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, 2019.

[152] D. Patel, A. Blandford, M. Warner, J. Shawe, and J. Stephenson, "”I Feel like Only Half a Man”: Online Forums as a Resource for Finding a ”New Normal” for Men Experiencing Fertility Issues," vol. 3, no. CSCW, Nov. 2019.

[153] A. L. Young and A. D. Miller, "”This Girl is on Fire”: Sensemaking in an Online Health Community for Vulvodynia," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, New York, NY, USA: Association for Computing Machinery, 2019.

[154] K. O. Leary, S. A. Munson, A. Bhattacharya, S. A. Munson, J. O. Wobbrock, and W. Pratt, "Design Opportunities for Mental Health Peer Support Technologies," *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, no. February, pp. 1470–1484, 2017.

[155] J. Huh and M. S. Ackerman, "Collaborative help in chronic disease management: Supporting individualized problems," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, 2012, pp. 853–862.

[156] D. Yang, R. Kraut, and J. M. Levine, "Commitment of newcomers and old-timers to online health support communities," in *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 6363–6375.

[157] D. Yang, R. E. Kraut, T. Smith, E. Mayfield, and D. Jurafsky, “Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.

[158] E. R. Burgess, K. E. Ringland, J. Nicholas, A. A. Knapp, J. Eschler, D. C. Mohr, and M. C. Reddy, “” i think people are powerful” the sociality of individuals managing depression,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–29, 2019.

[159] A. L. Hartzler, B. Weis, C. Cahill, W. Pratt, A. Park, U. Backonja, and D. W. McDonald, “Design and usability of interactive user profiles for online health communities,” *ACM Transactions on Computer-Human Interaction*, vol. 23, no. 3, 2016.

[160] 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,” *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2018-April, pp. 1–14, 2018.

[161] M. L. Birnbaum, A. F. Rizvi, C. U. Correll, J. M. Kane, and J. Confino, “Role of social media and the i nternet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders,” *Early intervention in psychiatry*, vol. 11, no. 4, pp. 290–295, 2017.

[162] B. Buck, K. S. Minor, and P. H. Lysaker, “Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives,” *Comprehensive psychiatry*, vol. 58, pp. 138–145, 2015.

[163] B. Buck and D. L. Penn, “Lexical characteristics of emotional narratives in schizophrenia: Relationships with symptoms, functioning, and social cognition,” *The Journal of nervous and mental disease*, vol. 203, no. 9, p. 702, 2015.

[164] K. Hong, A. Nenkova, M. E. March, A. P. Parker, R. Verma, and C. G. Kohler, “Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls,” *Psychiatry research*, vol. 225, no. 1, pp. 40–49, 2015.

[165] K. S. Minor, K. A. Bonfils, L. Luther, R. L. Firmin, M. Kukla, V. R. MacLain, B. Buck, P. H. Lysaker, and M. P. Salyers, “Lexical analysis in schizophrenia: How emotion and social word use informs our understanding of clinical presentation,” *Journal of psychiatric research*, vol. 64, pp. 74–78, 2015.

[166] R. D. Strous, M. Koppel, J. Fine, S. Nachliel, G. Shaked, and A. Z. Zivotofsky, “Automated characterization and identification of schizophrenia in writing,” *The Journal of nervous and mental disease*, vol. 197, no. 8, pp. 585–588, 2009.

[167] S. Fineberg, J. Leavitt, S. Deutsch-Link, S. Dealy, C. Landry, K. Pirruccio, S. Shea, S. Trent, G. Cecchi, and P. Corlett, "Self-reference in psychosis and depression: A language marker of illness," *Psychological medicine*, vol. 46, no. 12, pp. 2605–2615, 2016.

[168] G. Bedi, F. Carrillo, G. A. Cecchi, D. F. Slezak, M. Sigman, N. B. Mota, S. Ribeiro, D. C. Javitt, M. Copelli, and C. M. Corcoran, "Automated analysis of free speech predicts psychosis onset in high-risk youths," *npj Schizophrenia*, vol. 1, no. 1, pp. 1–7, 2015.

[169] M. L. Birnbaum, K. Candan, I. Libby, O. Pascucci, and J. Kane, "Impact of online resources and social media on help-seeking behaviour in youth with psychotic symptoms," *Early intervention in psychiatry*, vol. 10, no. 5, pp. 397–403, 2016.

[170] M. De Choudhury, S. Counts, and E. Horvitz, "Social media as a measurement tool of depression in populations," in *Proceedings of the 5th Annual ACM Web Science Conference*, ACM, 2013, pp. 47–56.

[171] C. M. Bishop, *Pattern recognition and machine learning*. Springer Science+ Business Media, 2006.

[172] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.

[173] J. M. Kane, D. G. Robinson, N. R. Schooler, K. T. Mueser, D. L. Penn, R. A. Rosenheck, J. Addington, M. F. Brunette, C. U. Correll, S. E. Estroff, *et al.*, "Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the nimh raise early treatment program," *American Journal of Psychiatry*, vol. 173, no. 4, pp. 362–372, 2015.

[174] J. D'Angelo, B. Kerr, and M. A. Moreno, "Facebook displays as predictors of binge drinking: From the virtual to the visceral," *Bulletin of science, technology & society*, vol. 34, no. 5-6, pp. 159–169, 2014.

[175] J. M. Kane and C. U. Correll, "Past and present progress in the pharmacologic treatment of schizophrenia," *The Journal of clinical psychiatry*, vol. 71, no. 9, p. 1115, 2010.

[176] D. Robinson, M. G. Woerner, J. M. J. Alvir, R. Bilder, R. Goldman, S. Geisler, A. Koreen, B. Sheitman, M. Chakos, D. Mayerhoff, *et al.*, "Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder," *Archives of general psychiatry*, vol. 56, no. 3, pp. 241–247, 1999.

[177] H. Ascher-Svanum, B. Zhu, D. E. Faries, D. Salkever, E. P. Slade, X. Peng, and R. R. Conley, “The cost of relapse and the predictors of relapse in the treatment of schizophrenia,” *BMC psychiatry*, vol. 10, no. 1, p. 2, 2010.

[178] J. W. Swanson, M. S. Swartz, R. A. Van Dorn, E. B. Elbogen, H. R. Wagner, R. A. Rosenheck, T. S. Stroup, J. P. McEvoy, and J. A. Lieberman, “A national study of violent behavior in persons with schizophrenia,” *Archives of general psychiatry*, vol. 63, no. 5, pp. 490–499, 2006.

[179] M. Birchwood, E. Spencer, and D. McGovern, “Schizophrenia: Early warning signs,” *Advances in Psychiatric Treatment*, vol. 6, no. 2, pp. 93–101, 2000.

[180] J. F. Gleeson, D. Rawlings, H. J. Jackson, and P. D. McGorry, “Early warning signs of relapse following a first episode of psychosis,” *Schizophrenia research*, vol. 80, no. 1, pp. 107–111, 2005.

[181] M. Alvarez-Jimenez, A. Priede, S. Hetrick, S. Bendall, E. Killackey, A. Parker, P. McGorry, and J. Gleeson, “Risk factors for relapse following treatment for first episode psychosis: A systematic review and meta-analysis of longitudinal studies,” *Schizophrenia Research*, vol. 139, no. 1-3, pp. 116–128, 2012.

[182] D. Kimhy, I. Myin-Germeys, J. Palmier-Claus, and J. Swendsen, “Mobile assessment guide for research in schizophrenia and severe mental disorders,” *Schizophrenia bulletin*, sbr186, 2012.

[183] D. Ben-Zeev, R. Brian, R. Wang, W. Wang, A. T. Campbell, M. S. Aung, M. Merrill, V. W. Tseng, T. Choudhury, M. Hauser, *et al.*, “Crosscheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse.,” *Psychiatric rehabilitation journal*, vol. 40, no. 3, p. 266, 2017.

[184] 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*, ACM, 2013, pp. 3267–3276.

[185] S. K. Ernala, A. F. Rizvi, M. L. Birnbaum, J. M. Kane, and M. De Choudhury, “Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, pp. 1–27, 2017.

[186] S. K. Ernala, M. L. Birnbaum, K. A. Candan, A. F. Rizvi, W. A. Sterling, J. M. Kane, and M. De Choudhury, “Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, 2019, p. 134.

[187] F. S. Collins and H. Varmus, "A new initiative on precision medicine," *New England Journal of Medicine*, vol. 372, no. 9, pp. 793–795, 2015.

[188] W. T. Carpenter Jr and B. Kirkpatrick, "The heterogeneity of the long-term course of schizophrenia," *Schizophrenia Bulletin*, vol. 14, no. 4, pp. 645–652, 1988.

[189] M. T. Tsuang, M. J. Lyons, and S. V. Faraone, "Heterogeneity of schizophrenia: Conceptual models and analytic strategies," *The British Journal of Psychiatry*, vol. 156, no. 1, pp. 17–26, 1990.

[190] L. M. Manevitz and M. Yousef, "One-class svms for document classification," *Journal of machine Learning research*, vol. 2, no. Dec, pp. 139–154, 2001.

[191] D. M. J. Tax, "One-class classification: Concept learning in the absence of counter-examples.," 2002.

[192] M. Hauskrecht, M. Valko, B. Kveton, S. Visweswaran, and G. F. Cooper, "Evidence-based anomaly detection in clinical domains," in *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, vol. 2007, 2007, p. 319.

[193] M. L. Birnbaum, S. K. Ernala, A. F. Rizvi, E. Arenare, A. R. Van Meter, M. De Choudhury, and J. M. Kane, "Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook," *npj Schizophrenia*, vol. 5, no. 1, p. 17, Dec. 2019.

[194] C. Chung and J. W. Pennebaker, "The psychological functions of function words," *Social communication*, pp. 343–359, 2007.

[195] 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.

[196] R. Kohrding *et al.*, "A test of equality of two normal population means assuming homogeneous coefficients of variation," *The Annals of Mathematical Statistics*, vol. 40, no. 4, pp. 1374–1385, 1969.

[197] M. Birchwood, J. Smith, F. Macmillan, B. Hogg, R. Prasad, C. Harvey, and S. Bering, "Predicting relapse in schizophrenia: The development and implementation of an early signs monitoring system using patients and families as observers, a preliminary investigation," *Psychological Medicine*, vol. 19, no. 3, pp. 649–656, 1989.

[198] Y. Henmi, "Prodromal symptoms of relapse in schizophrenic outpatients: Retrospective and prospective study," *Psychiatry and clinical Neurosciences*, vol. 47, no. 4, pp. 753–775, 1993.

[199] 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," *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11 203–11 208, 2018.

[200] F. Spaniel, E. Bakstein, J. Anyz, J. Hlinka, T. Sieger, J. Hrdlicka, N. Görnerová, and C. Höschl, "Relapse in schizophrenia: Definitively not a bolt from the blue," *Neuroscience letters*, vol. 669, pp. 68–74, 2018.

[201] N. Rezaii, E. Walker, and P. Wolff, "A machine learning approach to predicting psychosis using semantic density and latent content analysis," *NPJ schizophrenia*, vol. 5, 2019.

[202] A. P. Association, *Diagnostic and statistical manual of mental disorders, (DSM-5)*. American Psychiatric Pub, 2013.

[203] M. Birchwood, P. Mcgorry, and H. Jackson, "Early intervention in schizophrenia.," *The British Journal of Psychiatry*, vol. 170, no. 1, pp. 2–5, 1997.

[204] N. B. Mota, M. Copelli, and S. Ribeiro, "Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance," *npj Schizophrenia*, vol. 3, no. 1, p. 18, 2017.

[205] E. Eisner, S. Bucci, N. Berry, R. Emsley, C. Barrowclough, and R. J. Drake, "Feasibility of using a smartphone app to assess early signs, basic symptoms and psychotic symptoms over six months: A preliminary report," *Schizophrenia research*, vol. 208, pp. 105–113, 2019.

[206] J. M. Olivares, J. Sermon, M. Hemels, and A. Schreiner, "Definitions and drivers of relapse in patients with schizophrenia: A systematic literature review," *Annals of general psychiatry*, vol. 12, no. 1, p. 32, 2013.

[207] V. J. Derlaga and J. H. Berg, *Self-disclosure: Theory, research, and therapy*. Springer Science & Business Media, 2013.

[208] A. E. Marwick and D. Boyd, "I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience," *New media & society*, vol. 13, no. 1, pp. 114–133, 2011.

[209] N. N. Bazarova and Y. H. Choi, "Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites," *Journal of Communication*, vol. 64, no. 4, pp. 635–657, 2014.

[210] L. C. Miller, J. H. Berg, and R. L. Archer, "Openers: Individuals who elicit intimate self-disclosure.," *Journal of Personality and Social Psychology*, vol. 44, no. 6, p. 1234, 1983.

[211] E. A. Stuart, "Matching methods for causal inference: A review and a look forward," *Statistical science: a review journal of the Institute of Mathematical Statistics*, vol. 25, no. 1, p. 1, 2010.

[212] A. Jablensky, N. Sartorius, G. Ernberg, M. Anker, A. Korten, J. E. Cooper, R. Day, and A. Bertelsen, "Schizophrenia: Manifestations, incidence and course in different cultures a world health organization ten-country study," *Psychological Medicine Monograph Supplement*, vol. 20, pp. 1–97, 1992.

[213] J. A. Naslund, K. A. Aschbrenner, S. J. Kim, G. J. Mchugo, J. Unützer, S. J. Bartels, and L. A. Marsch, "Health behavior models for informing digital technology interventions for individuals with mental illness.," *Psychiatric rehabilitation journal*, 2017.

[214] W. Hinzen and J. Rosselló, "The linguistics of schizophrenia: Thought disturbance as language pathology across positive symptoms," *Frontiers in psychology*, vol. 6, p. 971, 2015.

[215] E. Pitler and A. Nenkova, "Revisiting readability: A unified framework for predicting text quality," in *Proceedings of the conference on empirical methods in natural language processing*, Association for Computational Linguistics, 2008, pp. 186–195.

[216] N. C. Andreasen, "Negative symptoms in schizophrenia: Definition and reliability," *Archives of general psychiatry*, vol. 39, no. 7, pp. 784–788, 1982.

[217] 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.

[218] K. A. Baikie and K. Wilhelm, "Emotional and physical health benefits of expressive writing," *Advances in psychiatric treatment*, vol. 11, no. 5, pp. 338–346, 2005.

[219] J. W. Pennebaker, "Theories, therapies, and taxpayers: On the complexities of the expressive writing paradigm," *Clinical Psychology: Science and Practice*, vol. 11, no. 2, pp. 138–142, 2004.

[220] R. G. Tedeschi and L. G. Calhoun, "" posttraumatic growth: Conceptual foundations and empirical evidence"," *Psychological inquiry*, vol. 15, no. 1, pp. 1–18, 2004.

[221] D. G. Robinson, M. G. Woerner, M. McMeniman, A. Mendelowitz, and R. M. Bilder, “Symptomatic and functional recovery from a first episode of schizophrenia or schizoaffective disorder,” *American Journal of Psychiatry*, vol. 161, no. 3, pp. 473–479, 2004.

[222] R. Vauth, B. Kleim, M. Wirtz, and P. W. Corrigan, “Self-efficacy and empowerment as outcomes of self-stigmatizing and coping in schizophrenia,” *Psychiatry research*, vol. 150, no. 1, pp. 71–80, 2007.

[223] J. W. Pennebaker and J. D. Seagal, “Forming a story: The health benefits of narrative,” *Journal of clinical psychology*, vol. 55, no. 10, pp. 1243–1254, 1999.

[224] S. R. Kay, L. A. Opler, and J.-P. Lindenmayer, “Reliability and validity of the positive and negative syndrome scale for schizophrenics,” *Psychiatry research*, vol. 23, no. 1, pp. 99–110, 1988.

[225] D. Turkington, D. Kingdon, and P. J. Weiden, “Cognitive behavior therapy for schizophrenia,” *American Journal of Psychiatry*, vol. 163, no. 3, pp. 365–373, 2006.

[226] A. Gruzd, B. Wellman, and Y. Takhteyev, “Imagining twitter as an imagined community,” *Am. Behav. Sci.*, 2011.

[227] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?” In *Proceedings of the 19th international conference on World wide web*, ACM, 2010, pp. 591–600.

[228] M. De Choudhury, S. S. Sharma, T. Logar, W. Eekhout, and R. C. Nielsen, “Gender and cross-cultural differences in social media disclosures of mental illness.,” in *CSCW*, 2017.

[229] I. Altman and D. Taylor, “
1973, social penetration: The development of interpersonal relationships,” *Holt, New York*,

[230] D. A. Taylor and I. Altman, “Self-disclosure as a function of reward-cost outcomes,” *Sociometry*, 1975.

[231] D. A. Dickey and W. A. Fuller, “Likelihood ratio statistics for autoregressive time series with a unit root,” *Econometrica*, 1981.

[232] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, “Characterizing and predicting postpartum depression from shared facebook data,” in *CSCW*, ACM, 2014, pp. 626–638.

[233] S. Sprecher, S. Treger, J. D. Wondra, N. Hilaire, and K. Wallpe, “Taking turns: Reciprocal self-disclosure promotes liking in initial interactions,” *J. Exp. Soc. Psychol.*, 2013.

[234] L. C. Jiang, N. N. Bazarova, and J. T. Hancock, “From perception to behavior: Disclosure reciprocity and the intensification of intimacy in computer-mediated communication,” *Commun. Res.*, 2013.

[235] T. A. Vlahovic, Y.-C. Wang, R. E. Kraut, and J. M. Levine, “Support matching and satisfaction in an online breast cancer support community,” in *CHI*, 2014.

[236] S. E. Hobfoll, A. Nadler, and J. Leiberman, “Satisfaction with social support during crisis: Intimacy and self-esteem as critical determinants.,” *J. Pers. Soc. Psychol.*, 1986.

[237] J. Durbin, “Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables,” *Econometrica*, 1970.

[238] M. Burke, C. Marlow, and T. Lento, “Social network activity and social well-being,” in *CHI*, 2010.

[239] C. Steinfield, N. B. Ellison, and C. Lampe, “Social capital, self-esteem, and use of online social network sites: A longitudinal analysis,” *Journal of Applied Developmental Psychology*, 2008.

[240] D. C. Mohr, M. Zhang, and S. Schueller, “Personal sensing: Understanding mental health using ubiquitous sensors and machine learning,” *Annual Review of Clinical Psychology*, vol. 13, no. 1, 2017.

[241] A. R. Daughton, M. J. Paul, and R. Chunara, “What do people tweet when they’re sick? a preliminary comparison of symptom reports and twitter timelines,” 2018.

[242] R. Zafarani and H. Liu, “Evaluation without ground truth in social media research,” *Communications of the ACM*, vol. 58, no. 6, pp. 54–60, 2015.

[243] M. L. Birnbaum, S. K. Ernala, A. Rizvi, M. De Choudhury, and J. Kane, “A clinician-machine collaborative approach to identifying social media markers of schizophrenia,” *Journal of medical Internet research*, 2017, To appear.

[244] D. B. Rubin, “Statistical matching using file concatenation with adjusted weights and multiple imputations,” *Journal of Business & Economic Statistics*, vol. 4, no. 1, pp. 87–94, 1986.

[245] H. A. Schwartz, M. Sap, M. L. Kern, J. C. Eichstaedt, A. Kapelner, M. Agrawal, E. Blanco, L. Dziurzynski, G. Park, D. Stillwell, *et al.*, “Predicting individual well-

being through the language of social media," in *Pac Symp Biocomput*, vol. 21, 2016, pp. 516–527.

[246] P. R. Rosenbaum and D. B. Rubin, "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, vol. 39, no. 1, pp. 33–38, 1985.

[247] K. Saha, I. Weber, and M. De Choudhury, "A social media based examination of the effects of counseling recommendations after student deaths on college campuses," 2018.

[248] E. H. Huang, R. Socher, C. D. Manning, and A. Y. Ng, "Improving word representations via global context and multiple word prototypes," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, Association for Computational Linguistics, 2012, pp. 873–882.

[249] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[250] M. Q. Patton, "Enhancing the quality and credibility of qualitative analysis.," *Health services research*, vol. 34, no. 5 Pt 2, p. 1189, 1999.

[251] N. K. Denzin, "Triangulation 2.0," *Journal of mixed methods research*, vol. 6, no. 2, pp. 80–88, 2012.

[252] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[253] M. Sugiyama, N. D. Lawrence, A. Schwaighofer, *et al.*, *Dataset shift in machine learning*. The MIT Press, 2017.

[254] K. Saha, I. Weber, M. L. Birnbaum, and M. De Choudhury, "Characterizing awareness of schizophrenia among facebook users by leveraging facebook advertisement estimates," *Journal of medical Internet research*, vol. 19, no. 5, 2017.

[255] F. C. Redlich and D. X. Freedman, "The theory and practice of psychiatry," 1966.

[256] T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow, and P. Wang, *Research domain criteria (rdoc): Toward a new classification framework for research on mental disorders*, 2010.

[257] J. Torous, M. Keshavan, and T. Gutheil, "Promise and perils of digital psychiatry," *Asian journal of psychiatry*, vol. 10, pp. 120–122, 2014.

[258] P. Corrigan, "How stigma interferes with mental health care.," *American Psychologist*, vol. 59, no. 7, p. 614, 2004.

[259] D. D. Heckathorn, "Respondent-driven sampling: A new approach to the study of hidden populations," *Social problems*, vol. 44, no. 2, pp. 174–199, 1997.

[260] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of statistical planning and inference*, vol. 90, no. 2, pp. 227–244, 2000.

[261] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge university press, 2006.

[262] H. Lakkaraju, E. Kamar, R. Caruana, and E. Horvitz, "Identifying unknown unknowns in the open world: Representations and policies for guided exploration.," in *AAAI*, vol. 1, 2017, p. 2.

[263] R. Silberzahn, E. L. Uhlmann, D. Martin, P. Anselmi, F. Aust, E. C. Awtrey, Š. Bahnık, F. Bai, C. Bannard, E. Bonnier, *et al.*, "Many analysts, one dataset: Making transparent how variations in analytical choices affect results," 2017.

[264] T. Emmens and A. Phippen, "Evaluating online safety programs," *Harvard Berkman Center for Internet and Society.[23 July 2011]*, 2010.

[265] R. B. Ness, J. P. Committee, *et al.*, "Influence of the hipaa privacy rule on health research," *Jama*, vol. 298, no. 18, pp. 2164–2170, 2007.

[266] J. Torous and C. Nebeker, "Navigating ethics in the digital age: Introducing connected and open research ethics (core), a tool for researchers and institutional review boards," *Journal of medical Internet research*, vol. 19, no. 2, 2017.

[267] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, "Natural language processing of social media as screening for suicide risk," *Biomedical informatics insights*, vol. 10, p. 1 178 222 618 792 860, 2018.

[268] E. J. Emanuel, D. Wendler, and C. Grady, "What makes clinical research ethical?" *Jama*, vol. 283, no. 20, pp. 2701–2711, 2000.

[269] S. D. Young and R. Garett, "Ethical issues in addressing social media posts about suicidal intentions during an online study among youth: Case study," *JMIR mental health*, vol. 5, no. 2, 2018.

[270] D. Schuler and A. Namioka, *Participatory design: Principles and practices*. CRC Press, 1993.

[271] D. Umberson and J. Karas Montez, “Social relationships and health: A flashpoint for health policy,” *Journal of health and social behavior*, vol. 51, no. 1_suppl, S54–S66, 2010.

[272] K. Wickrama, R. D. Conger, F. O. Lorenz, and L. Matthews, “Role identity, role satisfaction, and perceived physical health,” *Social Psychology Quarterly*, pp. 270–283, 1995.

[273] T. S. J. Alharbi, E. Carlström, I. Ekman, A. Jarneborn, and L.-E. Olsson, “Experiences of person-centred care-patients’ perceptions: Qualitative study,” *BMC nursing*, vol. 13, no. 1, p. 28, 2014.

[274] S. Tse and R. M. Ng, “Applying a mental health recovery approach for people from diverse backgrounds: The case of collectivism and individualism paradigms,” *Journal of Psychosocial Rehabilitation and Mental Health*, vol. 1, no. 1, pp. 7–13, 2014.

[275] J. H. Elm, J. P. Lewis, K. L. Walters, and J. M. Self, “
i’m in this world for a reason”: Resilience and recovery among american indian and alaska native two-spirit women,” *Journal of lesbian studies*, vol. 20, no. 3-4, pp. 352–371, 2016.

[276] W. R. Gove and T. Fain, “The stigma of mental hospitalization: An attempt to evaluate its consequences,” *Archives of General Psychiatry*, vol. 28, no. 4, pp. 494–500, 1973.

[277] M. Farkas, “Recovery, rehabilitation, reintegration: Words vs. meaning,” *World Association of Psychosocial Rehabilitation Bulletin*, vol. 8, no. 4, pp. 6–8, 1996.

[278] E. W. Sterling, A. Silke, S. Tucker, L. Fricks, and B. G. Druss, “Integrating wellness, recovery, and self-management for mental health consumers,” *Community mental health journal*, vol. 46, no. 2, pp. 130–138, 2010.

[279] 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*, ACM, 2016, pp. 2098–2110.

[280] U. Pavalanathan and M. De Choudhury, “Identity management and mental health discourse in social media,” in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 315–321.

[281] H. Markus and P. Nurius, “Possible selves.,” *American psychologist*, vol. 41, no. 9, p. 954, 1986.

[282] G. J. McLachlan and K. E. Basford, *Mixture models: Inference and applications to clustering*. 1988, vol. 38.

[283] M. Zimmerman and J. I. Mattia, "The psychiatric diagnostic screening questionnaire: Development, reliability and validity.," *Comprehensive psychiatry*, 2001.

[284] R. L. Spitzer, J. B. Williams, M. Gibbon, and M. B. First, "The structured clinical interview for dsm-iii-r (scid): I: History, rationale, and description," *Archives of general psychiatry*, vol. 49, no. 8, pp. 624–629, 1992.

[285] L. Davidson and W. White, "The concept of recovery as an organizing principle for integrating mental health and addiction services," *The journal of behavioral health services & research*, vol. 34, no. 2, pp. 109–120, 2007.

[286] R. H. Hoyle and M. R. Sherrill, "Future orientation in the self-system: Possible selves, self-regulation, and behavior," *Journal of personality*, vol. 74, no. 6, pp. 1673–1696, 2006.

[287] R. Clarke, "Possible selves in first episode psychosis. a mixed methods study," Ph.D. dissertation, University of East Anglia, 2016.

[288] Z. Iqbal, M. Birchwood, P. Chadwick, and P. Trower, "Cognitive approach to depression and suicidal thinking in psychosis: 2. testing the validity of a social ranking model," *The British Journal of Psychiatry*, vol. 177, no. 6, pp. 522–528, 2000.

[289] I. B. Janis, H. B. Veague, and E. Driver-Linn, "Possible selves and borderline personality disorder," *Journal of clinical psychology*, vol. 62, no. 3, pp. 387–394, 2006.

[290] I. Ajzen *et al.*, "The theory of planned behavior," *Organizational behavior and human decision processes*, vol. 50, no. 2, pp. 179–211, 1991.

[291] E. B. Hekler, S. Michie, M. Pavel, D. E. Rivera, L. M. Collins, H. B. Jimison, C. Garnett, S. Parral, and D. Spruijt-Metz, "Advancing models and theories for digital behavior change interventions," *American journal of preventive medicine*, vol. 51, no. 5, pp. 825–832, 2016.

[292] R. Povey, M. Conner, P. Sparks, R. James, and R. Shepherd, "A critical examination of the application of the transtheoretical model's stages of change to dietary behaviours," *Health Education Research*, vol. 14, no. 5, pp. 641–651, 1999.

[293] B. G. Link, E. L. Struening, S. Neese-Todd, S. Asmussen, and J. C. Phelan, "Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses," *Psychiatric services*, vol. 52, no. 12, pp. 1621–1626, 2001.

[294] R. H. Moos, "Coping with acute health crises," in *Handbook of clinical health psychology*, Springer, 1982, pp. 129–151.

[295] S. J. Breckler, "Empirical validation of affect, behavior, and cognition as distinct components of attitude," *Journal of Personality and Social Psychology*, vol. 47, no. 6, pp. 1191–1205, 1984.

[296] K. Saha, B. Sugar, J. Torous, B. Abrahao, E. Kıcıman, and M. De Choudhury, "A social media study on the effects of psychiatric medication use," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, 2019, pp. 440–451.

[297] J. Simons, M. Vansteenkiste, W. Lens, and M. Lacante, "Placing motivation and future time perspective theory in a temporal perspective," *Educational psychology review*, vol. 16, no. 2, pp. 121–139, 2004.

[298] B. P. Dohrenwend, K. G. Raphael, S. Schwartz, A. Stueve, and A. Skodol, "The structured event probe and narrative rating method for measuring stressful life events.," 1993.

[299] R. C. Mannell, S. E. Iso-Ahola, *et al.*, "Work constraints on leisure: A social psychological analysis.," *Work constraints on leisure: a social psychological analysis.*, pp. 155–187, 1985.

[300] J. M. Alisky and K. A. Iczkowski, "Barriers to housing for deinstitutionalized psychiatric patients," *Psychiatric Services*, vol. 41, no. 1, pp. 93–95, Jan. 1990.

[301] B. Wheaton, "Life transitions, role histories, and mental health," *American sociological review*, pp. 209–223, 1990.

[302] E. Sevin and R. Ladwein, "To start being. . . .the anticipation of a social role through consumption in life transition: The case of the first-time pregnancy," *ACR North American Advances*, vol. NA-35, 2008.

[303] J. S. Donath, "Identity and deception in the virtual community," in *Communities in cyberspace*, Routledge, 2002.

[304] M. Birchwood, J. Smith, R. Cochrane, S. Wetton, and S. Copestake, "The social functioning scale the development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients," *The British Journal of Psychiatry*, vol. 157, no. 6, pp. 853–859, Dec. 1990.

[305] J. Grimmer and B. M. Stewart, "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political analysis*, vol. 21, no. 3, pp. 267–297, 2013.

[306] E. P. Baumer, D. Mimno, S. Guha, E. Quan, and G. K. Gay, "Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence?" *Journal of the Association for Information Science and Technology*, vol. 68, no. 6, pp. 1397–1410, 2017.

[307] M. Muller, S. Guha, E. P. Baumer, D. Mimno, and N. S. Shami, "Machine learning and grounded theory method: Convergence, divergence, and combination," in *Proceedings of the 19th International Conference on Supporting Group Work*, 2016, pp. 3–8.

[308] A. Bruckman, "Studying the amateur artist: A perspective on disguising data collected in human subjects research on the internet," *Ethics and Information Technology*, vol. 4, no. 3, pp. 217–231, 2002.

[309] F. Stutzman and W. Hartzog, "Boundary regulation in social media," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ser. CSCW '12, Association for Computing Machinery, Feb. 2012, pp. 769–778, ISBN: 9781450310864.

[310] J. R. Z. Abela and D. U. D'Alessandro, "Beck's cognitive theory of depression: A test of the diathesis-stress and causal mediation components," *British Journal of Clinical Psychology*, vol. 41, no. 2, pp. 111–128, 2002.

[311] C. L. Philippi and M. Koenigs, "The neuropsychology of self-reflection in psychiatric illness," *Journal of psychiatric research*, vol. 0, pp. 55–63, Jul. 2014.

[312] J. Eschler, E. R. Burgess, M. Reddy, and D. C. Mohr, "Emergent self-regulation practices in technology and social media use of individuals living with depression," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, Association for Computing Machinery, Apr. 2020, pp. 1–13, ISBN: 9781450367080.

[313] J. R. Carl, D. P. Soskin, C. Kerns, and D. H. Barlow, "Positive emotion regulation in emotional disorders: A theoretical review," *Clinical Psychology Review*, vol. 33, no. 3, pp. 343–360, Apr. 2013.

[314] Y. Ephraim and N. Merhav, "Hidden markov processes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, Jun. 2002.

[315] T. Moses, "Determinants of mental illness stigma for adolescents discharged from psychiatric hospitalization," *Social Science & Medicine*, vol. 109, pp. 26–34, 2014.

[316] W. R. Gove, "The Stigma of Mental Hospitalization: An Attempt to Evaluate Its Consequences," *Archives of General Psychiatry*, vol. 28, no. 4, p. 494, Apr. 1973.

[317] S. Nolen-Hoeksema, “The role of rumination in depressive disorders and mixed anxiety/depressive symptoms.,” *Journal of abnormal psychology*, vol. 109, no. 3, p. 504, 2000.

[318] A. Lucksted and A. L. Drapalski, “Self-stigma regarding mental illness: Definition, impact, and relationship to societal stigma.,” 2015.

[319] A. Staring, M. Van der Gaag, M. Van den Berge, H. Duivenvoorden, and C. Mulder, “Stigma moderates the associations of insight with depressed mood, low self-esteem, and low quality of life in patients with schizophrenia spectrum disorders,” *Schizophrenia research*, vol. 115, no. 2-3, pp. 363–369, 2009.

[320] P. W. Corrigan, J. E. Larson, and N. Rüsch, “Self-stigma and the “why try” effect: Impact on life goals and evidence-based practices,” *World Psychiatry*, vol. 8, no. 2, pp. 75–81, Jun. 2009.

[321] A. Lucksted and A. L. Drapalski, “Self-stigma regarding mental illness: Definition, impact, and relationship to societal stigma,” *Psychiatric Rehabilitation Journal*, vol. 38, no. 2, pp. 99–102, 2015, Place: US Publisher: Educational Publishing Foundation.

[322] A. Anagnostopoulos, R. Kumar, and M. Mahdian, “Influence and correlation in social networks,” in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2008, pp. 7–15.

[323] J. W. Pennebaker and A. Graybeal, “Patterns of natural language use: Disclosure, personality, and social integration,” *Current Directions in Psychological Science*, vol. 10, no. 3, pp. 90–93, 2001.

[324] S. Cohen and T. A. Wills, “Stress, social support, and the buffering hypothesis.,” *Psychological bulletin*, vol. 98, no. 2, p. 310, 1985.

[325] Y.-L. I. Wong and P. L. Solomon, “Community integration of persons with psychiatric disabilities in supportive independent housing: A conceptual model and methodological considerations,” *Mental health services research*, vol. 4, no. 1, pp. 13–28, 2002.

[326] P. Carling, “Emerging approaches to housing and support for people with psychiatric disabilities,” *Handbook of mental health economics and health policy*, vol. 1, pp. 239–259, 1996.

[327] R. J. Flynn and R. Lemay, *A quarter-century of normalization and social role valorization: Evolution and impact*. University of Ottawa Press, 1999.

[328] W. Kim, S.-Y. Jang, T.-H. Lee, J. E. Lee, and E.-C. Park, “Association between continuity of care and subsequent hospitalization and mortality in patients with mood disorders: Results from the korea national health insurance cohort,” *PloS one*, vol. 13, no. 11, e0207740, 2018.

[329] C. L. Keyes, “Mental illness and/or mental health? investigating axioms of the complete state model of health.,” *Journal of consulting and clinical psychology*, vol. 73, no. 3, p. 539, 2005.

[330] J. Neuberger and R. Tallis, “We do need a new word for patients?” *British Medical Journal*, vol. 318, no. 7200, pp. 1756–1756, 1999.

[331] K. Jacob, “Recovery model of mental illness: A complementary approach to psychiatric care,” *Indian journal of psychological medicine*, vol. 37, no. 2, p. 117, 2015.

[332] D. Glenister, “Patient participation in psychiatric services: A literature review and proposal for a research strategy,” *Journal of Advanced Nursing*, vol. 19, no. 4, pp. 802–811, 1994.

[333] R. M. Epstein and R. L. Street, *The values and value of patient-centered care*, 2011.

[334] R. C. Kessler, G. P. Amminger, S. Aguilar-Gaxiola, J. Alonso, S. Lee, and T. B. Ustun, “Age of onset of mental disorders: A review of recent literature,” *Current opinion in psychiatry*, vol. 20, no. 4, p. 359, 2007.

[335] H. J. Oh, C. Lauckner, J. Boehmer, R. Fewins-Bliss, and K. Li, “Facebooking for health: An examination into the solicitation and effects of health-related social support on social networking sites,” *Computers in human behavior*, vol. 29, no. 5, pp. 2072–2080, 2013.

[336] K. Saha, J. Torous, S. K. Ernala, C. Rizuto, A. Stafford, and M. De Choudhury, “A computational study of mental health awareness campaigns on social media,” *Translational behavioral medicine*, vol. 9, no. 6, pp. 1197–1207, 2019.

[337] L. S. Liu, J. Huh, T. Neogi, K. Inkpen, and W. Pratt, “Health vlogger-viewer interaction in chronic illness management,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 2013, pp. 49–58.

[338] H. MacLeod, K. Oakes, D. Geisler, K. Connelly, and K. Siek, “Rare world: Towards technology for rare diseases,” in *Proceedings of the 33rd Annual ACM Conference on human factors in computing systems*, 2015, pp. 1145–1154.

[339] O. L. Haimson, J. R. Brubaker, L. Dombrowski, and G. R. Hayes, “Digital footprints and changing networks during online identity transitions,” in *Proceedings of*

the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 2895–2907.

[340] J. L. Bevan, R. Gomez, and L. Sparks, “Disclosures about important life events on facebook: Relationships with stress and quality of life,” *Computers in Human Behavior*, vol. 39, pp. 246–253, 2014.

[341] K. Saha, J. Seybolt, S. M. Mattingly, T. Aledavood, C. Konjeti, G. J. Martinez, T. Grover, G. Mark, and M. De Choudhury, “What life events are disclosed on social media, how, when, and by whom?” In *Proc. CHI*, 2021.

[342] T. Rowlands, N. Waddell, and B. McKenna, “Are we there yet? a technique to determine theoretical saturation,” *Journal of Computer Information Systems*, vol. 56, no. 1, pp. 40–47, 2016.

[343] C. B. Draucker, D. S. Martsolf, and C. Poole, “Developing distress protocols for research on sensitive topics,” *Archives of psychiatric nursing*, vol. 23, no. 5, pp. 343–350, 2009.

[344] J. Fereday and E. Muir-Cochrane, “Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development,” *International journal of qualitative methods*, vol. 5, no. 1, pp. 80–92, 2006.

[345] S. B. Omer, P. Malani, and C. Del Rio, “The covid-19 pandemic in the us: A clinical update,” *Jama*, vol. 323, no. 18, pp. 1767–1768, 2020.

[346] H. Altman, “A collaborative approach to discharge planning for chronic mental patients.,” *Hospital & community psychiatry*, 1983.

[347] M. Birchwood, J. Smith, R. Cochrane, S. Wetton, and S. Copestake, “The social functioning scale the development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients,” *The British Journal of Psychiatry*, vol. 157, no. 6, pp. 853–859, 1990.

[348] M. Webber and M. Fendt-Newlin, “A review of social participation interventions for people with mental health problems,” *Social psychiatry and psychiatric epidemiology*, vol. 52, no. 4, pp. 369–380, 2017.

[349] E. P. Baumer, P. Adams, V. D. Khovanskaya, T. C. Liao, M. E. Smith, V. Schwanda Sosik, and K. Williams, “Limiting, leaving, and (re) lapsing: An exploration of facebook non-use practices and experiences,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2013, pp. 3257–3266.

[350] C. Satchell and P. Dourish, "Beyond the user: Use and non-use in hci," in *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, 2009, pp. 9–16.

[351] S. M. Wyatt, "Non-users also matter: The construction of users and non-users of the internet," *Now users matter: The co-construction of users and technology*, pp. 67–79, 2003.

[352] A. Lucksted and A. L. Drapalski, "Self-stigma regarding mental illness: Definition, impact, and relationship to societal stigma," *Psychiatric Rehabilitation Journal*, vol. 38, no. 2, pp. 99–102, 2015.

[353] J. W. Pennebaker, "Putting stress into words: Health, linguistic, and therapeutic implications," *Behaviour research and therapy*, vol. 31, no. 6, pp. 539–548, 1993.

[354] N. Andalibi, P. Ozturk, and A. Forte, "Depression-related imagery on instagram," in *CSCW*, ACM, 2015, pp. 231–234.

[355] S. K. Ernala, T. Labetoulle, F. Bane, M. L. Birnbaum, A. F. Rizvi, J. M. Kane, and M. De Choudhury, "Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses," in *International Conference on Web and Social Media*, AAAI, 2018.

[356] M. Burke, J. Cheng, and B. de Gant, "Social comparison and facebook: Feedback, positivity, and opportunities for comparison," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[357] M. L. Birnbaum, S. K. Ernala, A. Rizvi, E. Arenare, A. Van Meter, M. De Choudhury, and J. M. Kane, "Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from facebook," *NPJ schizophrenia*, vol. 5, no. 1, pp. 1–9, 2019.

[358] J. L. Feuston and A. M. Piper, "Everyday experiences: Small stories and mental illness on instagram," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.

[359] S. Chancellor, E. P. Baumer, and M. De Choudhury, "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*, vol. 3, no. CSCW, pp. 1–32, 2019.

[360] A. Lenhart and J. B. Horrigan, "Re-visualizing the digital divide as a digital spectrum," *IT & society*, vol. 1, no. 5, pp. 23–39, 2003.

[361] E. P. Baumer, S. Guha, E. Quan, D. Mimno, and G. K. Gay, “Missing photos, suffering withdrawal, or finding freedom? how experiences of social media non-use influence the likelihood of reversion,” *Social Media+ Society*, vol. 1, no. 2, p. 2 056 305 115 614 851, 2015.

[362] S. Y. Schoenebeck, “Giving up twitter for lent: How and why we take breaks from social media,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 773–782.

[363] S. Guha and S. B. Wicker, “Do birds of a feather watch each other? homophily and social surveillance in location based social networks,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 2015, pp. 1010–1020.

[364] A. Marwick, “The public domain: Surveillance in everyday life,” *Surveillance & Society*, vol. 9, no. 4, pp. 378–393, 2012.

[365] J. W. Treem and P. M. Leonardi, “Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association,” *Annals of the International Communication Association*, vol. 36, no. 1, pp. 143–189, 2013.

[366] X. Ma, J. Hancock, and M. Naaman, “Anonymity, intimacy and self-disclosure in social media,” in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 3857–3869.

[367] Y. H. Choi and N. N. Bazarova, “Self-disclosure characteristics and motivations in social media: Extending the functional model to multiple social network sites,” *Human Communication Research*, vol. 41, no. 4, pp. 480–500, 2015.

[368] S. A. Rains and S. R. Brunner, “The outcomes of broadcasting self-disclosure using new communication technologies: Responses to disclosure vary across one’s social network,” *Communication Research*, vol. 45, no. 5, pp. 659–687, 2018.

[369] M. Burke, R. Kraut, and C. Marlow, “Social capital on facebook: Differentiating uses and users,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 2011, pp. 571–580.

[370] 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.

[371] N. B. Ellison, P. Triu, S. Schoenebeck, R. Brewer, and A. Israni, “Why we don’t click: Interrogating the relationship between viewing and clicking in social media

contexts by exploring the "non-click"," *Journal of Computer-Mediated Communication*, vol. 25, no. 6, pp. 402–426, 2020.

[372] B. Nonnecke and J. Preece, "Lurker demographics: Counting the silent," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2000, pp. 73–80.

[373] ——, "Why lurkers lurk," 2001.

[374] P. F. Chang, J. Whitlock, and N. N. Bazarova, "
to respond or not to respond, that is the question": The decision-making process of providing social support to distressed posters on facebook," *Social Media+ Society*, vol. 4, no. 1, p. 2 056 305 118 759 290, 2018.

[375] T.-Y. Wu, A. Oeldorf-Hirsch, and D. Atkin, "A click is worth a thousand words: Probing the predictors of using click speech for online opinion expression," *International Journal of Communication*, vol. 14, p. 20, 2020.

[376] C. T. Griffiths, Y. Dandurand, and D. Murdoch, *The social reintegration of offenders and crime prevention*. National Crime Prevention Centre Ottawa, Ontario, Canada, 2007, vol. 4.

[377] H. Carr, "Returning 'home': Experiences of reintegration for asylum seekers and refugees," *The British Journal of Social Work*, vol. 44, no. suppl_1, pp. i140–i156, 2014.

[378] P. Huxley, S. Evans, S. Madge, M. Webber, T. Burchardt, D. McDaid, and M. Knapp, "Social and community opportunities profile (scope) short version," 2012.

[379] R. U. Florenzano, "Chronic mental illness in adolescence: A global overview," *Pediatrician*, vol. 18, no. 2, 1991.

[380] A. Perrin and M. Anderson, "Share of us adults using social media, including facebook, is mostly unchanged since 2018," *Pew Research Center*, vol. 10, 2019.

[381] H. H. Perlman, "The caseworker's use of collateral information," *Social Casework*, vol. 32, no. 8, pp. 325–333, 1951.

[382] S. K. Genuis and J. Bronstein, "Looking for "normal": Sense making in the context of health disruption," *Journal of the Association for Information Science and Technology*, vol. 68, no. 3, pp. 750–761, 2017.

[383] M. Cardano, "Mental distress: Strategies of sense-making," *Health*, vol. 14, no. 3, pp. 253–271, May 2010.

[384] N. Godbold, "Beyond information seeking: Towards a general model of information behaviour," *Information Research: An International Electronic Journal*, vol. 11, no. 4, Jul. 2006.

[385] ——, "An evolving normal: Making sense of renal dialysis via online discussion boards," *Relational Concepts in Medicine*, pp. 205–214, May 2011.

[386] I. Li, A. K. Dey, and J. Forlizzi, "Understanding my data, myself: Supporting self-reflection with ubicomp technologies," in *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, ACM Press, 2011, ISBN: 9781450306300.

[387] W. Bak, "Possible selves: Implications for psychotherapy," *International journal of mental health and addiction*, vol. 13, no. 5, pp. 650–658, 2015.

[388] N. B. Ellison, J. Vitak, R. Gray, and C. Lampe, "Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes," *Journal of Computer-Mediated Communication*, vol. 19, no. 4, pp. 855–870, Jul. 2014.

[389] B. Wellman and S. Wortley, "Different strokes from different folks: Community ties and social support," *American journal of Sociology*, vol. 96, no. 3, pp. 558–588, 1990.

[390] C. E. Cutrona and D. W. Russell, "Type of social support and specific stress: Toward a theory of optimal matching.," 1990.

[391] M. M. Skeels and J. Grudin, "When social networks cross boundaries: A case study of workplace use of facebook and linkedin," in *Proceedings of the ACM 2009 international conference on Supporting group work*, 2009, pp. 95–104.

[392] D. C. Mohr, M. N. Burns, S. M. Schueller, G. Clarke, and M. Klinkman, "Behavioral intervention technologies: Evidence review and recommendations for future research in mental health," *General Hospital Psychiatry*, vol. 35, no. 4, pp. 332–338, Jul. 2013.

[393] C. M. Yeager and C. C. Benight, "If we build it, will they come? issues of engagement with digital health interventions for trauma recovery," *mHealth*, vol. 4, Sep. 2018.

[394] T. R. Insel, "The nimh research domain criteria (rdoc) project: Precision medicine for psychiatry," *American Journal of Psychiatry*, vol. 171, no. 4, pp. 395–397, 2014.

[395] P. Chikersal, D. Belgrave, G. Doherty, A. Enrique, J. E. Palacios, D. Richards, and A. Thieme, "Understanding client support strategies to improve clinical outcomes

in an online mental health intervention," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–16.

[396] R. E. Drake and R. Whitley, "Recovery and severe mental illness: Description and analysis," *The Canadian Journal of Psychiatry*, vol. 59, no. 5, pp. 236–242, 2014.

[397] D. Bhugra, "The global prevalence of schizophrenia," *PLoS medicine*, vol. 2, no. 5, e151, 2005.

[398] X. Zhao, C. Lampe, and N. B. Ellison, "The social media ecology: User perceptions, strategies and challenges," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 89–100.

[399] J. G. Adair, "The hawthorne effect: A reconsideration of the methodological artifact.," *Journal of applied psychology*, vol. 69, no. 2, p. 334, 1984.

[400] D. L. Borzekowski, S. Schenk, J. L. Wilson, and R. Peebles, "E-ana and e-mia: A content analysis of pro–eating disorder web sites," *American journal of public health*, vol. 100, no. 8, p. 1526, 2010.