

# Privacy in Crisis: A Study of Self-Disclosure During the Coronavirus Pandemic

Taylor Blose, Prasanna Umar, Anna Squicciarini, Sarah Rajtmajer

College of Information Sciences and Technology, The Pennsylvania State University, USA  
 thb5018@psu.edu, pzu3@psu.edu, acs20@psu.edu, smr48@psu.edu

## Abstract

We study observed incidence of self-disclosure in a large dataset of Tweets representing user-led English-language conversation about the Coronavirus pandemic. Using an unsupervised approach to detect voluntary disclosure of personal information, we provide early evidence that situational factors surrounding the Coronavirus pandemic may impact individuals' privacy calculus. Text analyses reveal topical shift toward supportiveness and support-seeking in self-disclosing conversation on Twitter. We run a comparable analysis of Tweets from Hurricane Harvey to provide context for observed effects and suggest opportunities for further study.

## 1 Introduction

At the time of writing, one-third of the world's population is directly impacted by restrictions on movement and activity, aimed at slowing the Coronavirus pandemic. All fifty states of the U.S. have issued guidelines on permissible travel outside the home, limitations on public gatherings, and business closures. Experts suggest that some of these measures may continue for 18 months or more.

Unsurprisingly, there has been an unprecedented surge in online activity (James 2020). Much of the increased traffic extends beyond typical Internet surfing and video streaming, as people find ways to leverage online resources to stay connected with one another, personally and professionally. Video-conferencing and social media usage are soaring as people turn to these platforms to support routine social activities (Reuters 2020; Iyengar 2020).

It is to be expected that the expanded breadth and depth of online activity will magnify privacy risks for individual users. In addition to simply spending more time online, the emergence of virtual play dates and book clubs suggests that during this time of social distancing, people are looking for ways to stay close (e.g., "apart but together"). This may be particularly the case for the many individuals facing heightened anxiety, stress and depression due to social isolation, grief, financial insecurity, and of course health-related fears of the virus itself (Association 2020; Braverman 2020).

The literature on social communication suggests that interpersonal connectedness and relationship development is

fundamentally facilitated through iterative *self-disclosure* (Altman and Taylor 1973), that is, intentionally revealing personal information such as personal motives, desires, feelings, thoughts, and experiences to others (Derlaga and Berg 1987). In fact, there is a robust literature on (routine) self-disclosure in online social media outside the particular domain of crisis (Bak, Lin, and Oh 2014; Joinson and Paine 2007; Nguyen, Bin, and Campbell 2012; Houghton and Joinson 2012; Attrill and Jalil 2011). Research indicates that as users engage in discussion online, they leverage self-disclosure as a way to enhance immediate social rewards (Hallam and Zanella 2017), increase legitimacy and likeability (Bak, Kim, and Oh 2012), and derive social support (Tidwell and Walther 2002). Despite the "upsides", i.e., socially adaptive motivations for disclosure, we know that self-privacy violations can come at a cost, leaving users exposed to identity theft, cyber fraud and other crimes (Hasan et al. 2013), discrimination in job searches, credit and visa applications (McGregor, Murray, and Ng 2018), harassment and bullying (Peluchette et al. 2015). Indeed, studies have repeatedly shown that an overwhelming majority of users have privacy concerns about their online interactions (Smith, Dinev, and Xu 2011). These concerns evolve over time, tied to the day's events and the longer arc of shifting norms (Acquisti, Brandimarte, and Loewenstein 2015; Adjerid, Peer, and Acquisti 2016).

Little is known about the evolution of users' sharing practices during crisis. We know that victims of Hurricane Harvey sought assistance through social media, in some cases revealing their full names and addresses online (Seetharaman and Wells 2017). However, what we are witnessing in the case of the Coronavirus pandemic is distinct from previous crises in important ways. COVID-19 is a global, relatively protracted acute threat. Unlike natural disasters or military engagements, the pandemic has left communication infrastructure intact. Digital outlets have become lifelines. We posit that self-privacy violations during the Coronavirus pandemic can help individuals feel more socially connected during a time of anxiety and physical distance, even though the long term impact of these disclosures is unknown.

In this work, we carry out analysis of instances of self-disclosure in a dataset of 53,557,975 Tweets representing conversations on Coronavirus-related topics. We leverage an unsupervised method for identifying and labeling voluntar-

ily disclosed personal information, both subjective and objective in nature. Our main finding reveals a steep increase in instances of self-disclosure, particularly related to users' emotional state and personal experience of the crisis. To our knowledge, this work is the first to study self-disclosure on social media during the Coronavirus pandemic. We run comparative analyses on a dataset of Tweets representing conversations about Hurricane Harvey in late summer 2017. The Harvey study provides valuable context for these unprecedented times, suggesting similarities and differences that better inform our observations during the current crisis and the privacy and crisis literatures in general.

## 2 Related work

Research in the space of online interactions has sought to understand the actualization of self-disclosure in digitally-mediated social communication. Studies suggest that disclosure behaviors in online environments may be meaningfully different than their offline counterparts, e.g., anonymity and lack of nonverbal cues afforded by social media may encourage greater disclosure of sensitive information (Forest and Wood 2012; Johnson 2001). Similar findings are reported in (Ma, Hancock, and Naaman 2016), where authors explore the impact of content intimacy on self-disclosure. It is well-established for face-to-face communication that people disclose less as content intimacy increases, but this effect seems to be weakened in online interactions.

Recent work has positioned online self-disclosure as strategic behavior targeting social connectedness, self-expression, relationship development, identity clarification and social control (Abramova et al. 2017; Bazarova and Choi 2014; Choudhury and De 2014). Voluntary disclosure of personal information has been associated with improved well-being, meaningfully related to increased informational and emotional support (Huang 2016).

There are, however, potential costs to users' privacy, including the unforeseen use and sharing of disclosed data. Early work by Acquisti and Gross (Acquisti and Gross 2006) suggested that social network users were neither fully aware nor responsive to privacy risks. Over time, studies have captured a shift toward increased privacy awareness (Johnson, Egelman, and Bellovin 2012; Vitak and Kim 2014), but there remains great variability in information sharing behaviors amongst individuals and across platforms (Zhao, Lampe, and Ellison 2016). It has been shown that culture plays a role in disclosure decisions (Zhao, Hinds, and Gao 2012; Krasnova, Veltre, and Günther 2012; Trepte et al. 2017), as does gender (Sun et al. 2015) and socioeconomic status (SES) (Marwick, Fontaine, and Boyd 2017). Overarchingly, the cost-benefit analyses underlying an individual's decision to share in the presence of privacy risk is postulated by social exchange theory (Emerson 1976) and re-framed in the context of online social networks as the so-called privacy calculus (Krasnova et al. 2010; Dienlin and Metzger 2016).

Critically, work in a number of domains suggests that contextual and situational factors, e.g., trust, anonymity, financial incentives, are embedded within the privacy calculus (Johnson and Paine 2012; Hann et al. 2007; Li, Sarathy, and Xu 2010). Amongst these factors, emotion has also

been suggested to play a meaningful role in privacy behaviors (Laufer and Wolfe 1977; Berendt, Günther, and Spiekermann 2005; Li et al. 2017). This finding is in keeping with the general theory of feeling-as-information (Petty, DeSteno, and Rucker 2001), whereby emotions serve as information cues directly invoking adaptive behaviors (Lazarus and Lazarus 1991). Studies linking emotion to disclosure online have thus far limited scope to considering the emotional impact of a particular website. The ways in which an individual's mood and general emotional state impact individual privacy calculus are unknown.

To our knowledge, there is no literature examining changes to patterns of self-disclosure in crisis through the lens of privacy risk. The crisis community is interested in the related but fundamentally distinct problem of mining self-disclosed information for the purposes of identifying and deploying assistance and relief to impacted individuals and communities (see (Muniz-Rodriguez et al. 2020) for review).

This work also dovetails with the literature on detection and tagging of self-disclosure in text, e.g., (Caliskan Islam, Walsh, and Greenstadt 2014; Wang, Burke, and Kraut 2016; Vasalou et al. 2011; Bak, Lin, and Oh 2014; Chow, Golle, and Staddon 2008; Choi et al. 2013). Chow et. al. (Chow, Golle, and Staddon 2008) developed an association rules-based inference model that identified sensitive keywords which could be used to infer a private topic. Similarly, multiple studies utilized pattern or rule based methods to detect specific types of disclosures (Vasalou et al. 2011; Umar, Squicciarini, and Rajtmajer 2019).

Past work has attempted to classify self-disclosure by levels, or degree of disclosure. Caliskan et.al (Caliskan Islam, Walsh, and Greenstadt 2014) used AdaBoost with Naive Bayes classifier to detect privacy scores for Twitter users' timelines. Bak et.al. (Bak, Lin, and Oh 2014) applied modified Latent Dirichlet Allocation (LDA) topic models for semi-supervised classification of Twitter conversations into three self-disclosure levels: general, medium and high. Wang et. al. (Wang, Burke, and Kraut 2016) used regression models with extensive feature sets to detect degree of self-disclosure. Because the notion of sensitive information is based on user perception and context, studies on detection of self-disclosure levels are often difficult to generalize beyond their original context.

## 3 Primary dataset

Our primary dataset is a repository of Tweet IDs corresponding to content posted on Twitter related to the Coronavirus pandemic (Chen, Lerman, and Ferrara 2020). At the time of writing, the repository contains 508,088,777 Tweet IDs for the period of activity from January 21, 2020 through August 28, 2020. Tweets were compiled utilizing a combination of Twitter's Search API (for activity January 21 through January 28) and Streaming API (for activity January 28 through July 31). The repository represents topically-relevant Tweets across the platform, canvassed based on designated keywords, as well as the full activity of selected accounts (See Tables 1, 2). Around June 6th, the repository

collection infrastructure transitioned to Amazon Web Services which generated a significant increase in Tweet ID volume. No search parameters were adjusted or data gaps presented because of the transition; therefore, we analyzed the entirety of the dataset in a consistent manner. In analyses that follow, we focused our scope to highlight an important transitional period in the pandemic for most users in the United States by considering two sub-periods – prior to and after March 11th, the date of the World Health Organization’s pandemic declaration and just two days preceding U.S. President Trump’s declared national emergency.

Text and metadata corresponding to these 508,088,777 Tweet IDs were obtained through rehydration using the Tward<sup>1</sup> Python library. Of the IDs passed for rehydration, 461,259,923 were successfully rehydrated. The 9.21% loss represents deleted content, therefore irretrievable through Twitter’s API.

For the purpose of measuring and studying self-disclosure, we filtered the corpus to capture Tweets that represent original content posted by individual users. Specifically, we removed quoted Tweets, retweets,<sup>2</sup> as well as all Tweets associated with verified accounts and the specific organizational accounts listed in Table 2. We narrowed our analysis to English-language content in order to reduce situational heterogeneity and maintain confidence in our labeling approach, which has been developed and validated on English-language text. The resulting corpus, which forms the basis of our analyses, consists of 53,557,975 unique Tweets.

#### 4 Automated detection of self-disclosure

We use an unsupervised method (Umar, Squicciarini, and Rajtmajer 2019) to detect instances of self-disclosure in our dataset. Consistent with the literature on detection of self-disclosure in text (see, e.g., Bak, Lin, and Oh 2014; Choudhury and De 2014; Houghton and Joinson 2012; Wang, Burke, and Kraut 2016), we consider the presence of first-person pronouns. Specifically, we consider sentences containing self-reference as the subject, a category-related verb and associated named entity. Consider the example of location self-disclosure shown in Figure 1. A first person pronoun “I” is the self-referent subject of the sentence. It is used with a location-related verb “live” in the vicinity of the associated location entity “Pennsylvania”. Notably, subjective categories of self-disclosure such as interests and feelings do not have associated named entities. These are differentiated through rule-based schemas based on subject-verb pairs. The approach described is implemented in three phases: 1) subject, verb and object triplet extraction with awareness to voice (active or passive) in the sentence; 2) named entity recognition; 3) rule-based matching to established dictionaries. Our dictionaries are adopted from (Umar, Squicciarini, and Rajtmajer 2019).

<sup>1</sup><https://github.com/DocNow/twarc>

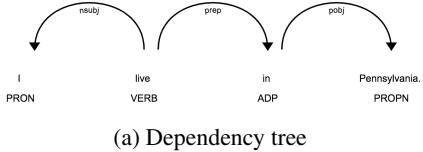
<sup>2</sup>Retweets were identified through the existence of the “retweetedstatus” field in the Tweet object returned by the API. Tweets beginning with the string ‘RT @’ were also treated as retweeted records.

Keyword Followed	Start Date
Coronavirus, Koronavirus, Corona, CDC, Wuhancoronavirus, Wuhanlockdown, Ncov, Wuhan, N95, Kungflu, Epidemic, outbreak, Sinophobia, China	1/28/2020
covid-19	2/16/2020
corona virus	3/2/2020
covid, covid19, sars-cov-2	3/6/2020
COVID-19	3/8/2020
COVD, pandemic	3/12/2020
coronapocalypse, cancel everything, Coronials, SocialDistancingNow, Social Distancing, SocialDistancing	3/13/2020
panicbuy, panic buy, panicbuying, panic buying, 14DayQuarantine, DuringMy14DayQuarantine, panic shop, panic shopping, panicshop, InMyQuarantineSurvivalKit, panic-buy, panic-shop	3/14/2020
coronakindness	3/15/2020
quarantinelife, chinese virus, chinesevirus, stayhomechallenge, stay home challenge, sflockdown, DontBeASpreader, lockdown, lock down	3/16/2020
shelteringinplace, sheltering in place, staysafestayhome, stay safe stay home, trumppandemic, trump pandemic, flattenthecurve, flatten the curve, china virus, chinavirus	3/18/2020
quarantinelife, PPEshortage, saferathome, stayathome, stay at home, stay home, stayhome	3/19/2020
GetMePPE	3/21/2020
covidiot	3/26/2020
epitwitter	3/28/2020
pandemie	3/31/2020
wear a mask, wearamas, kung flu, covididiot	6/28/2020
COVID_19	7/9/2020

Table 1: Keywords followed, by start date

Account Followed	Start Date
Pneumonia, Wuhan, CoronaVirusInfo, V2019N, CDCEmergency, CDCgov, WHO, HHSgov, NIAIDNews	1/28/2020
drtdros	3/15/2020

Table 2: Accounts followed, by start date



I live in **Pennsylvania GPE** .

(b) Named entity recognition

Figure 1: Illustration of phases in self-disclosure categorization scheme (Umar, Squicciarini, and Rajtmajer 2019)

As the proposed approach is based on sentence structure and syntactic resources (subject, verb, object and entities), it can be applied to any textual content. However, we acknowledge that Tweets present unique characteristics. Due to character limits and consequent emerging norms of the platform, users more frequently engage acronyms and abbreviations (Han and Baldwin 2011). Relatedly, sentence structure and syntax are noisier when compared to more verbose platforms (Boot et al. 2019). User mentions, hashtags, and graphic symbols are embedded within text. Considering these differences, we pre-processed Tweets as follows. All Unicode encoding errors were corrected. We removed markers associated with retweets (e.g., “RT”) and filtered user mentions and hashtags. Additionally, symbols like “&” and “\$” were replaced with their respective word representations. Email addresses and phone numbers were replaced with placeholders “emailid” and “phonenumbers”, while URLs were filtered. We also replaced contractions in the Tweets like “I’m” to “I am” and corrected incorrect use of spacing between words. These pre-processing steps enabled cleaner input to the detection algorithm.

We validated our approach on a baseline dataset of 3,708 annotated Tweets (Caliskan Islam, Walsh, and Greenstadt 2014). The dataset was manually labelled for presence or absence of self-disclosed personal information such as location, medical information, demographic information and so on. For our purposes, category-specific labels of self-disclosure were binarized, resulting in a total of 3,188 self-disclosing and 520 non-self-disclosing Tweets. Using the unsupervised method described, we classified the baseline data and compared against the binarized manual labels. Our approach demonstrated acceptable performance yielding precision, recall and F scores of 92.5%, 50.3% and 65.1% respectively. Recall is lower than reported in (Umar, Squicciarini, and Rajtmajer 2019), attributable to differences between the taxonomy used to manually label the baseline Twitter dataset and the taxonomy used to create the unsupervised detection method (e.g., our unsupervised approach was not developed to detect categories like drug use and personal attacks, whereas these categories were explicitly offered to manual labelers).

## 5 Topic modeling

Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) was used for topic modeling of Tweets in our dataset. In this approach, each document (a single Tweet) in the corpus (set of Tweets) is considered to be generated as a mixture of latent topics and each topic is a distribution over words. For a document, each word is assigned topics according to a

Dirichlet distribution. Iteratively cycling through each word in each document and all documents in the corpus, topic assignments are updated based on the prevalence of words across topics and the prevalence of topics in the document. Based on this process, final topic distributions for documents and word distributions for topics are generated.

In addition to the cleaning steps described in Section 4, we pre-processed Tweets using tokenization, conversion to lower case, lemmatization and removal of punctuation and stopwords. Topics were generated from the resulting corpora using the LDA model within the Gensim Python library.<sup>3</sup>

We ran topic analyses over subsets of interest within the complete Coronavirus dataset. Namely, we explored unique topic models for Tweets in two time windows pre- and post-March 11 (January 21 - March 11; March 12 - May 15), and compared topical themes between Tweets with presence or absence of detected instances of self-disclosure over the entirety of this subset. To better understand disclosure trends presented post-March 11, we also generated topic models over one month windows for the remainder of the Coronavirus dataset: May 16 through June 15, June 16 through July 15, and July 16 through August 15.

## 6 Findings

Of the 53,557,975 Coronavirus-related Tweets analyzed, approximately 19.07% (10,215,752 Tweets) contain elements of self-disclosure. Looking more closely at daily variance from January until June, we identified a significant transition point in activity around March 11, 2020, as shown in Figure 2. The significance of the behavioral change is supported by overlaying 7-day and 30-day simple moving averages which smooth day-to-day variance by taking the average disclosure percentage value over the given time window.

In the period January 21 through March 11, the average daily percentage of self-disclosing Tweets is 14.63%; from March 12 through May 15, the average daily percentage is 18.89%. This change in activity coincides with an escalation in severity and increased global awareness of the crisis, with the World Health Organization (WHO) officially classifying Coronavirus as a pandemic (March 11). Current events coincident with observed changes in the rate of self-disclosure are noted on March 13 and March 19 when the United States officially declared a national state of emergency, and when the governor of California issued the first statewide ‘stay-at-home’ order, respectively. Self-disclosure activity remains high for the remainder of the dataset with an average daily percentage of 19.79% from May 16 through

<sup>3</sup><https://radimrehurek.com/gensim/models/ldamodel.html>

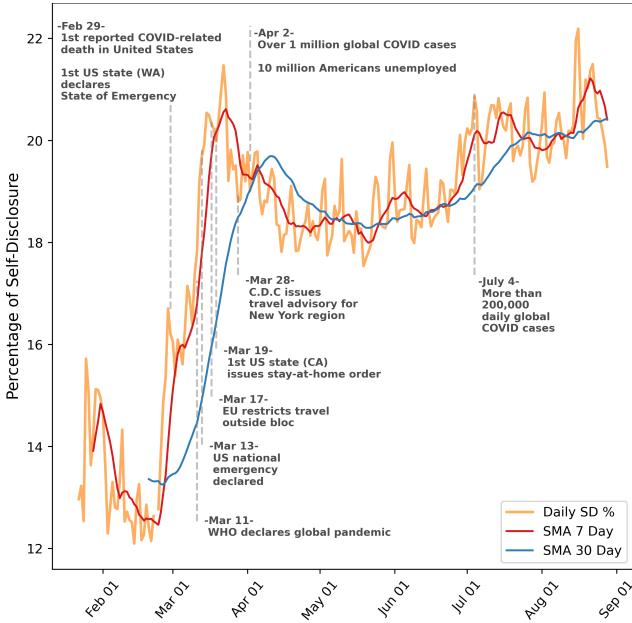


Figure 2: Percentage of Tweets containing self-disclosure, assessed daily with 7- and 30-day simple moving averages

August 28. These observations suggest that situational context, in particular during crisis, may meaningfully influence short- and long-term disclosure behaviors.

We further examined messaging around the the Coronavirus pandemic through topic modeling, as outlined in Section 5. As discussed, self-disclosing behaviors increased steeply around March 11th, and interesting distinctions are noticeable in the topical breakdown comparing the periods just before (January 21 - March 11) and after (March 12 - May 15) this date of interest, as reported in Figure 3. Leading up to March 11th (Figure 3a), self-disclosing conversation focused on general information (Topic 1) and sentimental impacts (Topic 3) represented 20% and 24% of all Tweets, respectively. After March 11 (Figure 3b), terms related to global and sentimental aspects of the crisis remain present (Topic 3, Topic 2) but prominent conversations also shift to managing the spread of the virus (Topic 1, Topic 4) and discussions about needs, help, thanks and support (Topic 5, Topic 6). In fact, Topics 2 and 4 together make up the majority of the Tweets in that period, representing 23% and 24% of activity, respectively. This represents a shift from outward-looking to self-centric messaging as well as early evidence of emotional support and support seeking through disclosure.

Centered on the mid-March increase in disclosure behaviors, we also compare topical variance between disclosing and non-disclosing Coronavirus-related Tweets for entirety of the aforementioned subsetted time window (January 21 - May 15; see Figure 4). Generally, extracted topics reflect terminology pervasive in mainstream media at the onset of the crisis, including but not limited to expected impacts of COVID-19 (health, market, lockdown, quarantine), recommendations (stay home, mask, cancel, school), and political

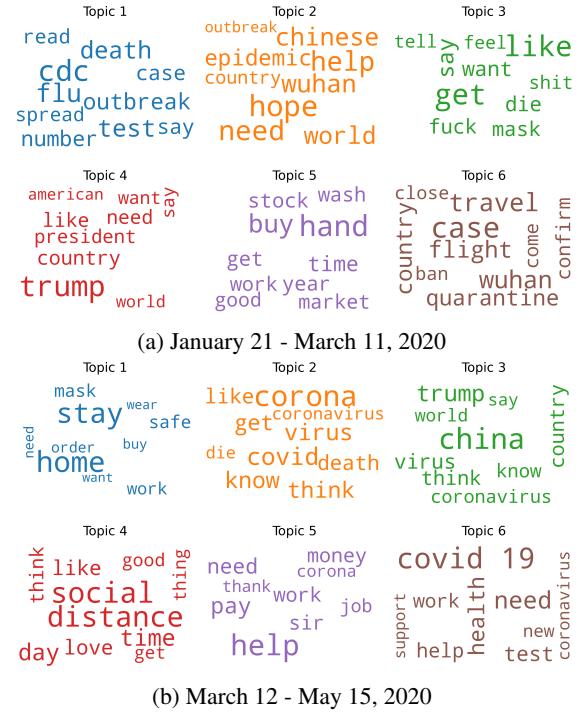


Figure 3: Topical Comparison of Self-Disclosing Coronavirus Tweets Pre- & Post-March 11, 2020

impact (Trump, administration, democrat). There is no noticeable distinction between topics extracted from the subset of Tweets containing self-disclosure and those without.

We also analyzed monthly subsets within the remaining timeline of our dataset, May 15 through August 15 (Figure 5), where the proportion of self-disclosing conversations remains elevated, and in fact, continues to gradually increase higher over time. The observed sustained higher rate of self-disclosure through the summer is particularly interesting. We speculate that users may have recalibrated their own sharing practices during the pandemic, and that this recalibration may represent a longer-term effect. Consistent with preceding time windows were themes of support seeking and political opinion. We observe the introduction of new themes related to social movements and education, while, discussions surrounding daily life (working/staying home) become the most prominent topic representing 38% and 41% of all self-disclosing Tweets in the final two months of the collection. This again suggests a possible transition from discussions of anxiety and need to potentially coping with lifestyle changes and establishing a new normal.

## 7 Comparative analysis

For context, we compared observed self-disclosure during the Coronavirus pandemic to observed self-disclosure during Hurricane Harvey (2017). Although hurricanes are an annual expectation, the landfall duration and subsequent impact of Hurricane Harvey created a crisis throughout communities in the South Central region of the United States. This overwhelmed traditional emergency response infras-

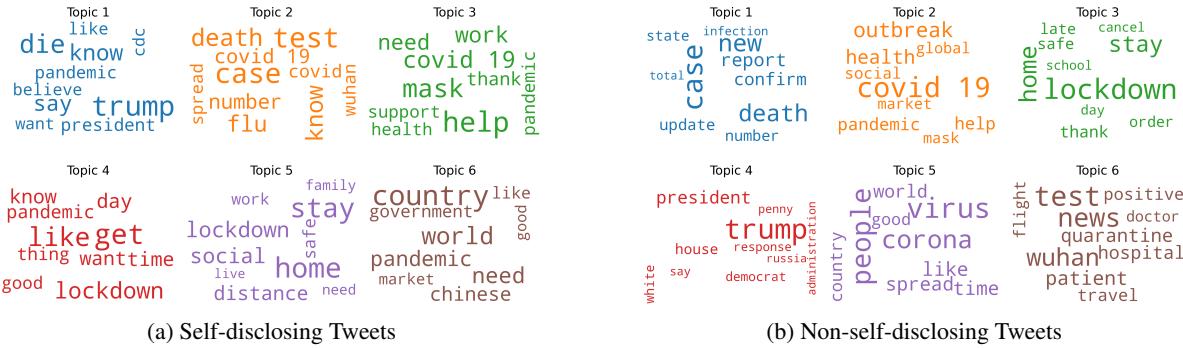


Figure 4: Topical Comparison of Coronavirus Tweets through May 15, 2020

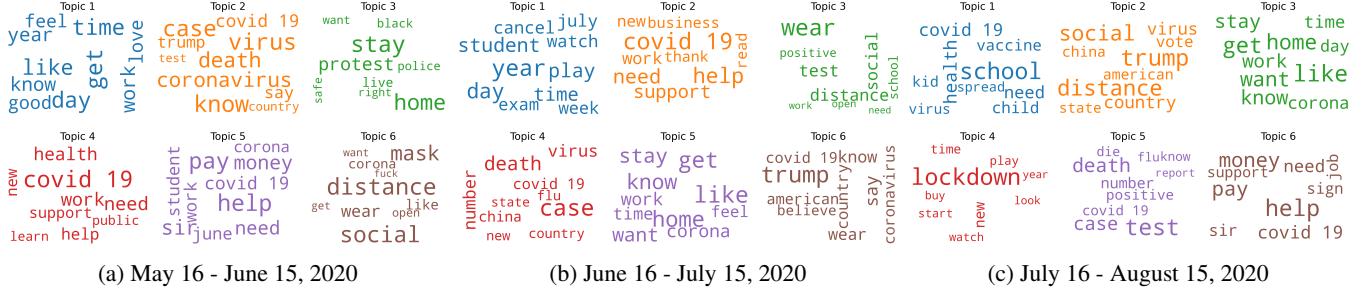


Figure 5: Topical Comparison of Self-Disclosing Coronavirus Tweets: May 16 - August 15, 2020

structure and affected citizens took to social media to seek emergency assistance (Sebastian et al. 2017; Smith et al. 2018).

We consider a collection of 6,732,546 Tweet IDs representing posted content inclusive of keywords “Hurricane Harvey”, “Harvey”, and/or “HurricaneHarvey” during the 12-day period August 25, when Harvey first made landfall, through September 5, 2017 (Alam et al. 2018). Similar to the Coronavirus dataset, we passed the set of Hurricane Harvey-related Tweet IDs through the Hydrator Tweet Retrieval Tool (v2.0)<sup>4</sup>, a sister desktop application to Twarc. We experienced a 33% loss during data reconstitution (compare to 9.21% loss in the Coronavirus-related dataset), attributable to Tweet and account deletion during the nearly 3 years which have passed. We filtered the resulting 4,379,462 Tweets for original content in the same fashion as we handled the Coronavirus data – removing all quoted Tweets, retweets, content from verified accounts, and non-English content as identified by Twitter. The resulting 551,061 Tweets were passed through pre-processing and unsupervised labeling to detect instances of self-disclosure, and through topic analysis as detailed in Sections 4 and 5.

Across all 551,061 Tweets in this subset, we observed an average 9% self-disclosure rate (49,595 Tweets) over the 12-day collection period – substantially lower than the 19.07% observed in the Coronavirus-related dataset. Several factors might account for the difference. The current pandemic has been marked by efforts to maintain social distance and, we have proposed, increased levels of disclosure may be related

to relationship-building with online cohorts. But what we may also be seeing is a reflection of a general trend toward greater self-disclosure in online social media over the course of nearly three years separating the two events.

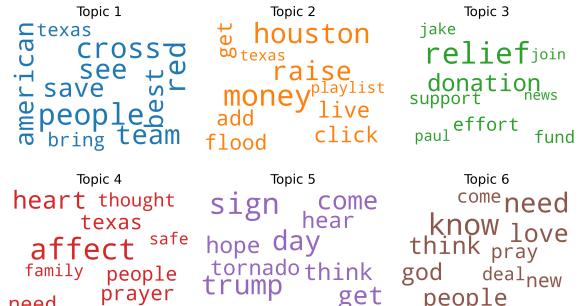
With respect to topical focus, we see important parallels between the two crises. As illustrated in Figure 6a, self-disclosing Tweets revealed emotional messaging centered on seeking immediate spiritual, physical, and monetary support (Topics 1, 2, 3, 4, 6) with top terms including “red cross”, “raise money”, “donation”, “relief”, and “prayer”. While present, these themes are less prominent in the non-self-disclosing data. This finding across both datasets suggests that support-seeking during crisis might be a driver of self-disclosure and play a meaningful role in users’ sharing practices.

Also mirroring the Coronavirus dataset, we observe one self-disclosing topic representing politically-motivated conversation (Topic 5). While there is topical overlap between the two classes of Harvey-related Tweets, non-disclosing Tweets (Figure 6b) presented a focus on contextual information related to the crisis with relevant keywords “flood”, “Texas”, and “Houston” (Topic 4).

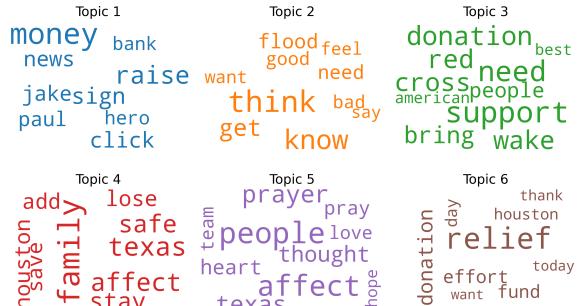
## 8 Discussion

Perhaps the most striking observation in the analyses we have described is early evidence of heightened and sustained levels of self-disclosure during the ongoing Coronavirus pandemic, as compared to observed disclosure during Hurricane Harvey. This global crisis is unprecedented in a number of ways, one being the scale and scope of human in-

<sup>4</sup><https://github.com/DocNow/hydrator>



(a) Self-disclosing Tweets



(b) Non-self-disclosing Tweets

Figure 6: Topical Comparison of Hurricane Harvey Tweets

teraction through social media. Concerns about privacy have been at the center of discussion in popular press (see, e.g., (Servick 2020; Cellan-Jones 2020; Lin and Martin 2020)), but most of this conversation has been about privacy trade-offs related to cellphone tracking and similar approaches to location surveillance and individual health monitoring in service to public health. Many are willing to sacrifice some privacy in hopes of stemming the spread of the disease and helping to accelerate the return to normalcy; others are not. Scholarly work has begun to propose “privacy first” decentralized approaches for COVID-related tracking and notification (see, e.g., (PAC 2020; Cov 2020; DP3 2020)).

We aim, with this work, to engage the research community in the work of better understanding more subtle, voluntary self-privacy violations emergent in the Coronavirus pandemic and in crisis more generally. We know that, during Hurricane Harvey, individuals took to social media in search of immediate aid. Today, individuals are leaning on their online social communities for ongoing engagement and support. Isolation, economic uncertainty, and health-related anxiety pose serious threat to mental health and well-being (Holmes et al. 2020; of Medical Sciences UK 2020), but the potential manifestations of psychological impact in the domain of voluntary self-disclosure are unknown. The existing literature hints at the role of mood and emotion in the privacy calculus, but these relationships have not be well-established. Our analyses suggest that the current pandemic and its effects may impact self-disclosure behaviors. An open question is whether these heightened oversharing practices will become a new norm, or whether they fade over time and we will ultimately return to pre-COVID baselines. If heightened self-disclosure sustains beyond this crisis, we might look to social theory for explanations. Threshold models of collective behavior (see, e.g., (Granovetter 1978; Centola et al. 2018; Wiedermann et al. 2020)) may offer insights. As the crisis continues to unfold in the next several months, we suggest that additional work should aim to further develop, refine and test these hypotheses.

## References

2020. Covid Watch. URL <https://www.covid-watch.org>

2020. DP3T: Decentralized Privacy-Preserving Proximity Tracing. URL <https://github.com/DP-3T>

2020. PACT: Private Automated Contact Tracing. URL <https://pact.mit.edu>

Abramova, O.; Wagner, A.; Krasnova, H.; and Buxmann, P. 2017. Understanding Self-Disclosure on Social Networking Sites-A Literature Review .

Acquisti, A.; Brandimarte, L.; and Loewenstein, G. 2015. Privacy and human behavior in the age of information. *Science* 347(6221): 509–514.

Acquisti, A.; and Gross, R. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *International workshop on privacy enhancing technologies*, 36–58. Springer.

Adjerid, I.; Peer, E.; and Acquisti, A. 2016. Beyond the privacy paradox: Objective versus relative risk in privacy decision making. Available at SSRN 2765097 .

Alam, F.; Ofli, F.; Imran, M.; and Aupetit, M. 2018. A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria. *Proc. of ISCRAM, Rochester, USA* .

Altman, I.; and Taylor, D. A. 1973. *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.

Association, A. P. 2020. Finding local mental health resources during the COVID-19 crisis. URL <https://www.apa.org/topics/covid-19/local-mental-health>

Attrill, A.; and Jalil, R. 2011. Revealing only the superficial me: Exploring categorical self-disclosure online. In *Computers in Human Behavior*, volume 27, 1634–1642. ISSN 07475632. doi:10.1016/j.chb.2011.02.001.

Bak, J.; Lin, C.-Y.; and Oh, A. 2014. Self-disclosure topic model for classifying and analyzing Twitter conversations. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1986–1996. Doha, Qatar: Association for Computational Linguistics. doi:10.3115/v1/D14-1213. URL <https://www.aclweb.org/anthology/D14-1213>

Bak, J. Y.; Kim, S.; and Oh, A. 2012. Self-disclosure and relationship strength in Twitter conversations. In *Proceedings*

*of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, 60–64. Association for Computational Linguistics.

Bazarova, N. N.; and Choi, Y. H. 2014. Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication* 64(4): 635–657.

Berendt, B.; Günther, O.; and Spiekermann, S. 2005. Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM* 48(4): 101–106.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan): 993–1022.

Boot, A. B.; Sang, E. T. K.; Dijkstra, K.; and Zwaan, R. A. 2019. How character limit affects language usage in tweets. *Palgrave Communications* 5(1): 1–13.

Braverman, B. 2020. The coronavirus is taking a huge toll on workers' mental health across America. *CNBC Workforce Wire* URL <https://www.cnbc.com/2020/04/06/coronavirus-is-taking-a-toll-on-workers-mental-health-across-america.html>.

Caliskan Islam, A.; Walsh, J.; and Greenstadt, R. 2014. Privacy Detective: Detecting Private Information and Collective Privacy Behavior in a Large Social Network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, WPES '14, 35–46. New York, NY, USA: ACM. ISBN 978-1-4503-3148-7. doi:10.1145/2665943.2665958. URL <http://doi.acm.org/10.1145/2665943.2665958>

Cellan-Jones, R. 2020. Coronavirus: Privacy in a pandemic. *BBC News* URL <https://www.bbc.com/news/technology-52135916>.

Centola, D.; Becker, J.; Brackbill, D.; and Baronchelli, A. 2018. Experimental evidence for tipping points in social convention. *Science* 360(6393): 1116–1119.

Chen, E.; Lerman, K.; and Ferrara, E. 2020. COVID-19: The First Public Coronavirus Twitter Dataset. *arXiv e-prints* arXiv:2003.07372.

Choi, D.; Kim, J.; Piao, X.; and Kim, P. 2013. Text Analysis for Monitoring Personal Information Leakage on Twitter. *Journal of Universal Computer Science* 19(16): 2472–2485. URL [http://www.jucs.org/jucs\\_19\\_16/text\\_analysis\\_for\\_monitoring](http://www.jucs.org/jucs_19_16/text_analysis_for_monitoring)

Choudhury, M. D.; and De, S. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity URL <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/8075>

Chow, R.; Golle, P.; and Staddon, J. 2008. Detecting Privacy Leaks Using Corpus-based Association Rules. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, 893–901. New York, NY, USA: ACM. ISBN 978-1-60558-193-4. doi:10.1145/1401890.1401997. URL <http://doi.acm.org/10.1145/1401890.1401997>

Derlaga, V. J.; and Berg, J. H. 1987. *Self-disclosure: Theory, research, and therapy*. Springer Science & Business Media.

Dienlin, T.; and Metzger, M. J. 2016. An extended privacy calculus model for SNSs: Analyzing self-disclosure and self-withdrawal in a representative US sample. *Journal of Computer-Mediated Communication* 21(5): 368–383.

Emerson, R. M. 1976. Social exchange theory. *Annual review of sociology* 2(1): 335–362.

Forest, A. L.; and Wood, J. V. 2012. When social networking is not working: Individuals with low self-esteem recognize but do not reap the benefits of self-disclosure on Facebook. *Psychological science* 23(3): 295–302.

Granovetter, M. 1978. Threshold models of collective behavior. *American journal of sociology* 83(6): 1420–1443.

Hallam, C.; and Zanella, G. 2017. Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards. *Computers in Human Behavior* 68: 217–227.

Han, B.; and Baldwin, T. 2011. Lexical normalisation of short text messages: Makn sens a# twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 368–378. Association for Computational Linguistics.

Hann, I.-H.; Hui, K.-L.; Lee, S.-Y. T.; and Png, I. P. 2007. Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems* 24(2): 13–42.

Hasan, O.; Habegger, B.; Brunie, L.; Bennani, N.; and Damiani, E. 2013. A discussion of privacy challenges in user profiling with big data techniques: The eexcess use case. In *Big Data (BigData Congress), 2013 IEEE International Congress on*, 25–30. IEEE.

Holmes, E. A.; O'Connor, R. C.; Perry, V. H.; Tracey, I.; Wessely, S.; Arseneault, L.; Ballard, C.; Christensen, H.; Silver, R. C.; Everall, I.; et al. 2020. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry* .

Houghton, D. J.; and Joinson, A. N. 2012. Linguistic markers of secrets and sensitive self-disclosure in twitter. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, 3480–3489. IEEE.

Huang, H.-Y. 2016. Examining the Beneficial Effects of Individual's Self-Disclosure on the Social Network Site. *Comput. Hum. Behav.* 57(C): 122–132. ISSN 0747-5632. doi:10.1016/j.chb.2015.12.030. URL <https://doi.org/10.1016/j.chb.2015.12.030>

Iyengar, R. 2020. The coronavirus is stretching Facebook to its limits. *CNN Business* URL <https://www.cnn.com/2020/03/18/tech/zuckerberg-facebook-coronavirus-response/index.html>

James, M. 2020. FCC OKs Verizon Request for More Capacity as Network Use Spikes. *Government and Technology* URL <https://www.govtech.com/news/FCC-OKs-Verizon-Request-for-More-Capacity-as-Network-Use-Spikes.html>

Johnson, M.; Egelman, S.; and Bellovin, S. M. 2012. Facebook and privacy: it's complicated. In *Proceedings of the eighth symposium on usable privacy and security*, 1–15.

Joinson, A. N. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European journal of social psychology* 31(2): 177–192.

Joinson, A. N.; and Paine, C. B. 2007. Self-disclosure, privacy and the Internet. *The Oxford handbook of Internet psychology* 2374252.

Joinson, A. N.; and Paine, C. B. 2012. Self-disclosure, Privacy and the Internet. In *Oxford Handbook of Internet Psychology*. ISBN 9780191743771. doi: 10.1093/oxfordhb/9780199561803.013.0016. URL <http://ekarapanos.com/courses/socialweb/Social{ }Web{ }-{ }fall{ }2011/Reading{ }Material{ }files/Joinson{ }sd.pdf>.

Krasnova, H.; Spiekermann, S.; Koroleva, K.; and Hildebrand, T. 2010. Online social networks: Why we disclose. *Journal of information technology* 25(2): 109–125.

Krasnova, H.; Veltre, N. F.; and Günther, O. 2012. Self-disclosure and privacy calculus on social networking sites: The role of culture. *Business & Information Systems Engineering* 4(3): 127–135.

Laufer, R. S.; and Wolfe, M. 1977. Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of social Issues* 33(3): 22–42.

Lazarus, R. S.; and Lazarus, R. S. 1991. *Emotion and adaptation*. Oxford University Press on Demand.

Li, H.; Luo, X. R.; Zhang, J.; and Xu, H. 2017. Resolving the privacy paradox: Toward a cognitive appraisal and emotion approach to online privacy behaviors. *Information & management* 54(8): 1012–1022.

Li, H.; Sarathy, R.; and Xu, H. 2010. Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems* 51(1): 62–71.

Lin, L.; and Martin, T. 2020. How Coronavirus Is Eroding Privacy. *The Wall Street Journal* URL <https://www.wsj.com/articles/coronavirus-paves-way-for-new-age-of-digital-surveillance-11586963028>

Ma, X.; Hancock, J.; and Naaman, M. 2016. Anonymity, Intimacy and Self-Disclosure in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 3857–3869. New York, New York, USA: ACM Press. ISBN 9781450333627. doi:10.1145/2858036.2858414. URL <https://s.tech.cornell.edu/assets/papers/anonymity-intimacy-disclosure.pdf> <http://dl.acm.org/citation.cfm?doid=2858036.2858414>

Marwick, A.; Fontaine, C.; and Boyd, D. 2017. “Nobody sees it, nobody gets mad”: Social media, privacy, and personal responsibility among low-SES youth. *Social Media+ Society* 3(2): 2056305117710455.

McGregor, L.; Murray, D.; and Ng, V. 2018. Four ways your Google searches and social media affect your opportunities in life. *The Conversation* URL <https://theconversation.com/four-ways-your-google-searches-and-social-media-affect-your-opportunities-in-life-96809>

Muniz-Rodriguez, K.; Ofori, S. K.; Bayliss, L. C.; Schwind, J. S.; Diallo, K.; Liu, M.; Yin, J.; Chowell, G.; and Fung, I. C.-H. 2020. Social Media Use in Emergency Response to Natural Disasters: A Systematic Review With a Public Health Perspective. *Disaster Medicine and Public Health Preparedness* 14(1): 139–149. doi:10.1017/dmp.2020.3.

Nguyen, M.; Bin, Y. S.; and Campbell, A. 2012. Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking* 15(2): 103–111.

of Medical Sciences UK, A. 2020. Survey results: Understanding people’s concerns about the mental health impacts of the COVID-19 pandemic. URL <https://acmedsci.ac.uk/file-download/99436893>

Peluchette, J. V.; Karl, K.; Wood, C.; and Williams, J. 2015. Cyberbullying victimization: Do victims’ personality and risky social network behaviors contribute to the problem? *Computers in Human Behavior* 52: 424–435.

Petty, R. E.; DeSteno, D.; and Rucker, D. D. 2001. The role of affect in attitude change..

Reuters. 2020. Zoom’s daily active users jumped from 10 million to over 200 million in 3 months. *Venture Beat* URL <https://venturebeat.com/2020/04/02/zooms-daily-active-users-jumped-from-10-million-to-over-200-million-in-3-months/>

Sebastian, A.; Lendering, K.; Kothuis, B.; Brand, A.; Jonkman, S. N.; van Gelder, P.; Godfroij, M.; Kolen, B.; Comes, M.; Lhermitte, S.; et al. 2017. Hurricane Harvey Report: A fact-finding effort in the direct aftermath of Hurricane Harvey in the Greater Houston Region .

Seetharaman, D.; and Wells, G. 2017. Hurricane Harvey Victims Turn to Social Media for Assistance. *The Wall Street Journal* URL <https://www.wsj.com/articles/hurricane-harvey-victims-turn-to-social-media-for-assistance-1503999001>

Servick, K. 2020. Cellphone tracking could help stem the spread of coronavirus. Is privacy the price? *Science* URL <https://www.sciencemag.org/news/2020/03/cellphone-tracking-could-help-stem-spread-coronavirus-privacy-price>

Smith, H. J.; Dinev, T.; and Xu, H. 2011. Information privacy research: an interdisciplinary review. *MIS quarterly* 35(4): 989–1016.

Smith, W. R.; Stephens, K. K.; Robertson, B.; Li, J.; and Murthy, D. 2018. Social media in citizen-led disaster response: Rescuer roles, coordination challenges, and untapped potential. In *Proceedings of the International Conference on Information Systems for Crisis Response and Management (ISCRAM)*.

Sun, Y.; Wang, N.; Shen, X.-L.; and Zhang, J. X. 2015. Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior* 52: 278–292.

Tidwell, L. C.; and Walther, J. B. 2002. Computer-mediated communication effects on disclosure, impressions, and in-

terpersonal evaluations: Getting to know one another a bit at a time. *Human communication research* 28(3): 317–348.

Trepte, S.; Reinecke, L.; Ellison, N. B.; Quiring, O.; Yao, M. Z.; and Ziegele, M. 2017. A cross-cultural perspective on the privacy calculus. *Social Media+ Society* 3(1): 2056305116688035.

Umar, P.; Squicciarini, A.; and Rajtmajer, S. 2019. Detection and Analysis of Self-Disclosure in Online News Commentaries. In *The World Wide Web Conference on - WWW '19*, 3272–3278. New York, New York, USA: ACM Press. ISBN 9781450366748. doi:10.1145/3308558.3313669. URL <http://dl.acm.org/citation.cfm?doid=3308558.3313669>

Vasalou, A.; Gill, A. J.; Mazanderani, F.; Papoutsis, C.; and Joinson, A. 2011. Privacy dictionary: A new resource for the automated content analysis of privacy. *Journal of the American Society for Information Science and Technology* 62(11): 2095–2105. doi:10.1002/asi.21610. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21610>

Vitak, J.; and Kim, J. 2014. “You Can’t Block People Offline”: Examining How Facebook’s Affordances Shape the Disclosure Process. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW ’14, 461–474. New York, NY, USA: Association for Computing Machinery. ISBN 9781450325400. doi:10.1145/2531602.2531672. URL <https://doi.org/10.1145/2531602.2531672>

Wang, Y.-C.; Burke, M.; and Kraut, R. 2016. Modeling Self-Disclosure in Social Networking Sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW ’16, 74–85. New York, NY, USA: ACM. ISBN 978-1-4503-3592-8. doi:10.1145/2818048.2820010. URL <http://doi.acm.org/10.1145/2818048.2820010>

Wiedermann, M.; Smith, E. K.; Heitzig, J.; and Donges, J. F. 2020. A network-based microfoundation of Granovetter’s threshold model for social tipping. *Scientific Reports* 10(1): 1–10.

Zhao, C.; Hinds, P.; and Gao, G. 2012. How and to Whom People Share: The Role of Culture in Self-Disclosure in Online Communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW ’12, 67–76. New York, NY, USA: Association for Computing Machinery. ISBN 9781450310864. doi:10.1145/2145204.2145219. URL <https://doi.org/10.1145/2145204.2145219>

Zhao, X.; Lampe, C.; and Ellison, N. B. 2016. The social media ecology: User perceptions, strategies and challenges. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 89–100.