Proceedings on Privacy Enhancing Technologies ; 2021 (3):373–393

Jooyoung Lee*, Sarah Rajtmajer, Eesha Srivatsavaya, and Shomir Wilson

# Digital Inequality Through the Lens of Self-Disclosure

**Abstract:** Recent work has brought to light disparities in privacy-related concerns based on socioeconomic status, race and ethnicity. This paper examines relationships between U.S. based Twitter users' socio-demographic characteristics and their privacy behaviors. Income, gender, age, race/ethnicity, education level and occupation are correlated with stated and observed privacy preferences of 110 active Twitter users. Contrary to our expectations, analyses suggest that neither socioeconomic status (*SES*) nor demographics is a significant predictor of the use of account security features. We do find that gender and education predict rate of self-disclosure, or voluntary sharing of personal information. We explore variability in the types of information disclosed amongst socio-demographic groups. Exploratory findings indicate that: 1) participants shared less personal information than they recall having shared in exit surveys; 2) there is no strong correlation between people's stated attitudes and their observed behaviors.

**Keywords:** Digital Inequality, Personal Data, Online Social Network, Socioeconomic Status, Online Self-Disclosure

DOI 10.2478/popets-2021-0052
Received 2020-11-30; revised 2021-03-15; accepted 2021-03-16.

## 1 Introduction

Prior work has suggested that people with lower income and education (*low SES*) lack sufficient understanding of privacy settings embedded within social media platforms to adequately protect their personal data, even if they are aware of online privacy threats [1]. Similar concerns about privacy self-management have been reported for women in developing countries [2], undocumented immigrants [3], and young adults [4]. This literature highlights so-called digital inequality, or the unbalanced distribution of risks and resources related to online activities [5]. This imbalance represents an obstacle to progress in information technologies, and raises important ethical concerns.

In addition to SES, gender [2, 6–8] and age [4, 9] have been identified as meaningful variables for user-centric studies of privacy online. A few studies have examined the relative contribution of race and ethnicity to privacy attitudes and behaviors [1, 6, 10, 11]. One study found members of Hispanic immigrant communities to be more reluctant to use platform-provided privacy settings, and more likely to share their location in posts on social media [1]. While another [11] compared personal information sharing practices between African Americans and European Americans but found no meaningful differences.

In parallel, a body of literature has emerged around so-called self-disclosure, or the voluntary sharing of personal information with others [12]. Prior work has explored motivations and contextual influences on self-disclosure [13–18], as well as associated privacy risks [19–21]. However, there has been relatively little empirical research about the relationship between self-disclosure and socio-demographic factors.

In this work, we explore correlations between income, gender, age, race/ethnicity, education level, occupation and self-disclosure on Twitter. Our work is based primarily on a study of privacy awareness and sharing behaviors of 110 Twitter users from the U.S. over a one month period. To explore how socio-demographic factors influence their privacy behavior, we formulated the following specific research questions: *Do socio-demographic factors have an impact on the usage of login verification? Do socio-demographic factors have an impact on self-disclosure? Do topics of self-disclosure vary across socio-demographic groups?* All public tweets posted by each user and their self-reported privacy setting usage are thoroughly examined. Inspired by prior work of Consedine et al. [11], we also collect socio-demographic information to explore variations in topics of disclosure amongst different groups.

***Corresponding Author: Jooyoung Lee:** Pennsylvania State University, E-mail: jfl5838@psu.edu
**Sarah Rajtmajer:** Pennsylvania State University, E-mail: smr48@psu.edu
**Eesha Srivatsavaya:** Pennsylvania State University, E-mail: efs5377@psu.edu
**Shomir Wilson:** Pennsylvania State University, E-mail: shomir@psu.edu

Based on findings in existing literature, we expected that certain SES and demographic groups (particularly, high-SES and female) would be more likely to adopt available privacy settings as compared to men and low-SES populations. On the contrary to our expectation, neither socioeconomic status (*SES*) nor demographics is identified as a key predictor of the use of advanced security features. We also predicted lower rates of observed self-disclosure in individuals with greater economic and educational resources. Similar to our prediction, we do find that gender and education predict rate of voluntary sharing of personal information. As far as we know, there exists only one study that has examined ethnicity, gender, and socioeconomic differences in online self-disclosure collectively [11], and their results suggest that income is a key predictor of self-disclosure, but ethnicity and gender are not. Lastly, our findings indicate that types of personal information revealed by particular groups vary significantly.

Following, we present the first holistic study of the relationship amongst multiple socio-demographic variables and both reported and observed privacy behaviors on Twitter. Our findings contribute to ongoing conversations around digital inequality and its impacts on privacy, and lay groundwork for future inquiry in this area.

## 2 Research Questions

We focus our inquiry around the following three research questions.

**RQ1. Do socio-demographic factors have an impact on the usage of login verification?**
Social networks (SNs) such as Facebook and Twitter have introduced various privacy and security features, including two-factor login authentication and tweet protection, to mitigate the risks of hacking [22]. Previous work has reported mixed findings on the differential usage of security features across populations [23]. For example, a study conducted by Nosko, Wood, et al. [24] found that women are more likely to utilize available privacy settings online, while another study by Oomen and Leenes [25] found that men are more likely to use these controls. Other works note age as a factor in user engagement with available privacy and security options [25]. Our first research question explores some of these inconsistencies in the specific case of two-factor authentication (equivalently, login verification) on Twitter. To our knowledge, this is the first study to explore the the role of socio-demographic predictors on the login verification feature on Twitter.

**RQ2. Do socio-demographic factors have an impact on self-disclosure?**
Previous work has illustrated social rewards for users sharing personal information in online communities: they can stay connected with existing ties [26–28], build new relationships [29], and receive emotional comfort and intimacy [30–32]. However, it has also been reported that not everyone has an equal chance to take advantage of these benefits due to different levels of privacy awareness and familiarity with relevant technologies [6, 33–35]. Low-SES individuals are particularly prone to challenges in managing their personal information in the cyberspace. Several authors have explored these challenges through the lens of occupation as a proxy for SES [36–38]. Related to demographics, a number of studies have explored age and gender differences in self-disclosure [2, 4, 6–9, 39]. Studies suggest that women are more concerned about the leakage of personal information and privacy loss than men [6, 8]. Yet, despite their awareness, women tend to be more active social media users and more likely to share profile information with strangers [6, 8, 40, 41]. With respect to age, studies with multiple age groups report that young adults may, in fact, be stricter in their online privacy behaviors than older adults [35, 42]. While previous work made an independent attempt to identify the relationship between privacy behaviors and one or two of the factors, we take a comprehensive approach studying how several of these variables affect information sharing behaviors individually and collectively.

**RQ3. Do topics of self-disclosure vary across socio-demographic groups?**
Although there have been numerous studies highlighting the distribution of topics associated with self-disclosure in online social media [13, 14, 40, 43], to our knowledge, a full study of topics of self-disclosure along socio-demographic lines has not been done with the exception of gender. Existing literature suggests that women tend to disclose more about intimate topics such as relationships with family and friends, feelings, and accomplishments at school or work, whereas men disclose more about cars, sports, work, and politics [40, 41]. Little work has been done on age-related topical differences. Hollenbaugh and Everett suggest that younger participants are more likely to disclose a larger amount of information on a greater variety of topics than older participants [44]. Additionally, Consedine et al. successfully incorporated ethnicity, gender and SES to study self-

disclosure levels and topics, but restricted their study to young adults [11]. In this work, we identify users based on 12 types of personal information to determine how topics of self-disclosure vary across demographics and SES.

## 3 Related Work

**Privacy at Risk** The worldwide expansion of SNs has drawn research attention to users' online privacy patterns. Madden et al. [1] have reported that 76% of Americans are worried about having their financial information lost or stolen online, and that 73% express a serious concern about social media companies' data collection and management processes. While, other work shows that users often post with minimal regard for negative consequences [45, 46], sometimes later feeling remorse for their actions [47–49]. Their regret tends to result from sharing sensitive topics or content with negative sentiments. The problem of unintended audience has been identified as another factor triggering sharing-related regret on Facebook [48]. Despite amplified anxieties towards potential or real privacy harms, users opt to sacrifice their privacy protections for minor conveniences and rewards [50, 51]. This paradox – willingness to expose personal information on social media despite privacy concerns [52] – can result in over-sharing of personal data and rare changes to default privacy settings [18, 48, 53–56]. Seeking explanations for observed discrepancies between privacy concerns and behaviors, authors have examined whether people read privacy policies or not [57, 58], while others have compared actual and intended disclosures [59, 60].

Several studies have explored the ways in which users are inclined to utilize privacy-enhancing strategies to manage privacy-related concerns [25, 61, 62]. Work of Oomen et al. [25] suggests that heightened perceptions of privacy risk leads to use of pseudonyms, cookie crunchers, anonymous email and providing false personal data. Some studies further observe a positive correlation between customized privacy settings and disclosing behaviors [61, 63]. That is, those who adopt customized security tools tend to disclose less personal information those using default settings.

While all SN users are subject to privacy risks, these risks are not evenly distributed. Underrepresented racial or education-level groups and women, for example, have been shown to be more vulnerable to privacy threats online [6, 33–35]. Although a number of studies have attempted to explain this phenomenon, some have provided contradictory and incomplete findings. Consedine et al. report no meaningful difference in levels of self-disclosure between men and women [11], although young men described a greater disclosure about some topics (e.g. sexual experiences). Similarly, among multiple SES factors, Redmiles et al. [64] found that education is the only factor affiliated with the respondents' likelihood of reporting a negative privacy experience. Unlike what Madden [1] found with survey methods, Redmiles et al. found that people with lower educational attainment tended to report equal or fewer incidents than people with a higher education level regardless of their feelings of vulnerability. We believe these conflicting results are attributable to different study populations, contexts, and platforms, or explainable by distinct interpretations of privacy concerns and practices.

**Online Self-Disclosure** Self-disclosure refers to "the process of making the self-known to other persons" [65]. Empirical studies illustrate that the frequency of self-disclosure increases significantly in computer-mediated communication interactions than in face-to-face discussions [66, 67]. Self-disclosure in online platforms is explainable by various reasons such as a desire for positive emotional motivations [30–32] and connectedness to society [26–28]. Prior work by Rajtmajer and colleagues [68, 69] suggests that peer behaviors play an important role in individual sharing decisions as users are willing to reach an agreement and share contents jointly. Similarly, anonymity is found to influence users' willingness to share personal information [13, 70, 71]. Consequences of public discourse may be negative as well – increased vulnerability in privacy protection [26, 72], an adverse impact on public self-image [73], or a loss of societal or emotional benefits [74, 75]. Previous work reported gender differences in online self-disclosing behavior, which revealed that women share their data on their social network sites more frequently compared to men [8, 39–41]. Barak and Gluck-Ofri, on the other hand, discovered no difference at all [6, 16]. A smaller number of prior works have examined online self-disclosure within different racial groups. Some studies exclusively focused on Caucasian samples [76, 77], and others compared particular ethnic groups (White vs. Black) [11]. Similar problems can be found in other socio-demographic indicators. Most studies about online self-disclosure investigated the self-disclosing behavior of youth in social media [11, 78, 79]. To our knowledge, no research has been conducted involving occupational variations in self-disclosure.

# 4 Methods

We describe our survey design, participant recruitment and data collection processes. Our approach included targeted recruitment of active Twitter users, tweet collection over a one month period, and pre- and post-surveys of participants. Users with private accounts were excluded from the study, as we were required to collect their postings on Twitter. The study described was approved by the The Pennsylvania State University's Institutional Review Board (IRB). Since IRB declared that the proposed study met the criteria for an exemption determination category, participants were not asked to sign the consent document. We instead informed participants in the advertisement and invitation emails that the research was about investigating public behavior on Twitter. To minimize the Hawthorn effect, which refers to participants' behavioral change during the study, participants were not specifically told that their public tweets would be collected.

## 4.1 Participant Recruitment and Screening

Study participants were recruited through Twitter's ad targeting service. Our recruitment ad highlighted the following participation requirements: (1) A participant should be at least 18 years old; and (2) a participant should be an active Twitter user. Included in this ad was a link to our screening survey. The screening survey (see Appendix A.1) asked users for their Twitter handles in addition to demographic information including gender, age, race/ethnicity, location, education level and income. We evaluated prospective participants based on the following hidden criteria, implemented to ensure we were engaging with active users tweeting primarily in English. The hidden criteria were: (1) the participant's Twitter account is at least a month old; (2) the participant has posted 10 tweets within the past month; (3) the participant has at least 10 followers and 10 following; (4) 90 percent of the participant's tweets within the past month are in English; and (5) 80 percent of the participant's tweets within the past month are not retweets. Participation recruitment took place over a two-month period, from February 6th through April 3rd, 2020.[1] Over the two month study recruitment period, 1,211 Twitter users participated in the screening survey. A much smaller number of users (n=207) were determined to be eligible to move forward given our hidden criteria.

As a majority of selected participants were Caucasians, additional efforts were made for one week to recruit more people of color. We employed Twitter's *follower look-alikes* targeting strategy[2], and specified popular celebrities or TV programs in underrepresented groups. A primary goal of this feature is to target people with interests similar to the followers of particular accounts. For instance, if advertisers enter *"@TelemundoNews"*, the post is more likely to be exposed to the Latino community. Targeted users are determined based on a variety of signals, including what they retweet, click on, tweet, and more. In addition, we set keywords (e.g., blacktwitter, hispanics, asianamerican)[3] in an effort to reach the relevant populations. Of 207 participants, 89 were recruited with a *follower look-alikes* targeting method. Among the 89, 16 participants (17.97%) self-reported as people of color. As the percentage of underrepresented populations before and after applying the strategy (18.64% vs. 17.97% respectively) is not significantly different, we concluded that the strategy was not effective.

The self-reported income, gender, age and education level breakdown of the 89 study participants recruited with a *follower look-alikes* targeting method are as follows: 1) the average income is 109,930 dollars; 2) there are 50 female participants, 35 male participants. and 4 non-binary participants. 3) participants are skewed younger with 66.29% (59) of participants being younger than 50; 4) 20 participants (22.47%) were attending college, with more than 73% (65) reporting having a Bachelor's or Associate degree, or higher.

## 4.2 Participant Surveys

Entry surveys were sent by email to 207 users who completed the screening survey and met participation criteria (see Appendix A.2). As occupation information is frequently used to determine individuals' SES status [36–38], we designed the entry survey to collect participants' occupation status in depth. They were ini-

**1** We recognize the possibility that the COVID-19 pandemic affected study results, and we explore this possibility later in the manuscript.

**2** https://business.twitter.com/en/help/campaign-setup/campaign-targeting/interest-and-follower-targeting.html

**3** https://business.twitter.com/en/help/campaign-setup/campaign-targeting/keyword-targeting.html tially asked to choose a particular category that best described their current occupation. Following their selection, the survey also asked them to provide a short description of their job. Six socio-demographic variables (income, gender, age, education level, race, occupation) were collected to be used as independent variables for the proposed research questions. Of the 207 qualified participants, 125 completed the entry survey and moved forward to the next step. They were compensated with a $3 Amazon gift card for completing the entry survey.

Prior to sending out a final survey, we collected their real-time tweets of 125 participants for a month. Exit surveys were sent by email to participants at the time of study completion, one month after receipt of entry survey (see Appendix A.3). The exit survey asked participants about their use of Twitter's login verification setting and their recollection of sharing specific information. Of the 125 participants, 115 successfully completed the final survey and were sent a $7 Amazon gift card. Of these 115 participants, 5 were excluded in our final analyses because they submitted an incorrect Twitter handle or did not post during the study period. This resulted in 110 users in total.

## 4.3 Tweet Collection

Using the Twitter API, we collected all tweets over a one month period for the 125 participants who met all inclusion criteria and completed the entry survey. Tweets were collected at a single time point, exactly one month after entry survey completion. We did not periodically refetch their tweets. That being said, deleted tweets were excluded in the study. Since users were not aware of their tweets being collected, we reckon the potential impact of users' self-censorship is minimal.

We received all public tweets posted by each user as well as retweets, replies to their tweets, and retweets of their tweets. Additionally, we collected metadata for each tweet including profile description, hashtags, favorite counts, and retweet counts. In total, we collected 23,880 public tweets, including retweets and replies, from 125 participants. 2,812 tweets collected from 15 participants were ultimately discarded because they failed to complete our exit survey or provided invalid Twitter handles. Table 1 reports basic statistics of the included users and their tweets. 74.64% of tweets were identified as "original", meaning they were not retweets. For the purposes of measuring self-disclosure, we excluded retweets as not representative of user-generated content [14]. The resulting total number of tweets considered for self-disclosure tagging was reduced from 21,068 to 15,727.

**Table 1.** Descriptive statistics for our dataset

| | |
|---|---|
| Participants, total | 110 |
| Posted tweets, total | 21,068 |
| Original tweets, total | 15,727 |
| Original tweets with first-person pronouns, total | 6,962 |
| Posted tweets, mean per user | 192 |
| Tweets with first-person pronouns, mean per user | 65 |
| Self-disclosing tweets, mean per user | 52 |

## 4.4 Annotation for Self-Disclosure

An automated labeling of self-disclosure in user-generated comments is an active area of ongoing research [13, 14, 40]. State of the art approaches focus on presence or absence of personal information, or broad category of disclosure (e.g., informational vs. emotional) [80], but do not afford fine-grained categorical labels. Accordingly, we manually annotated collected tweets for 12 categories of personal information derived from prior work [13], expanding them to more closely examine the dissemination of personally-identifiable information (specifically, phone numbers, email addresses and social security numbers), which is closely linked to multiple privacy threats [81–83]. The categories of self-disclosure are shown in Table 2. Each tweet was labeled as representing one or more category of personal information, or was labeled "No Disclosure" if none of the other categories were present.

To reduce the size of the labeling task, we narrowed our focus to include only those tweets containing explicit self-reference. That is, we included only tweets containing one or more first-person pronouns such as "I", "my", "me", and "mine". Previous work highlighted first-person pronouns as an important indicator of self-disclosure [13–16, 43, 84]. Our resulting corpus contained 6,962 tweets. To be clear, users' profile description fields were not part of the annotation corpus for analysis.

## 4.5 Annotation Evaluation

A complete labeling of the corpus was performed by the first author of the paper. Three additional researchers re-annotated a random subset of annotated tweets (200)

**Table 2.** Descriptions of personal information categories in the annotation scheme

| Categories | Sub-Categories | Description |
|---|---|---|
| Demographic | Birthday/Age | Sharing a tweet that directly implies information about one's own date of birth or age |
| | Race/Ethnicity | Sharing a tweet that directly implies information about one's own race or ethnicity such as being Black, White, Hispanic, etc. |
| | Gender | Sharing a tweet that directly implies information about one's own gender |
| | Marital Status | Sharing a tweet that directly implies information about one's own marital status such as being single, married, separated, divorced or widowed |
| | Education | Sharing a tweet that directly implies information about one's own education level |
| | Employment Status | Sharing a tweet that directly implies information about one's current employment status |
| Personal Identifier | Phone Number | Sharing a tweet that directly implies information about one's own phone number |
| | Email Address | Sharing a tweet that directly implies information about one's own email address |
| | SSN | Sharing a tweet that directly implies information about one's own social security number |
| Location | Location | Sharing a tweet that directly implies information about one's own location such as postal code, street address, city, and state |
| Subjective | Emotion/Opinion | Sharing a tweet that directly implies information about one's own personal feelings, opinions, thoughts, statements, attitudes, or beliefs |
| | Interest | Sharing a tweet that directly implies information about one's own hobbies and interests such as pastimes, favorites, tastes in music, movies, TV programs, or books |
| None | No Disclosure | Sharing no information about the above 12 personal information types |

to check the reliability and consistency of the annotation task. To ensure the integrity of annotation for every category, the number of instances in each category of the subset was equally distributed except for "Phone Number", "Email Address" and "SSN" category. These categories contained less than 5 instances. All four annotators received a complete set of detailed instructions describing the task and criteria for labeling.

**Table 3.** A distribution of tweets and Kappa values for each category

| Category | n (%) | Fleiss' Kappa | Cohen's Kappa |
|---|---|---|---|
| Birthday/Age | 74 (1.06%) | 0.582 | 0.363 |
| Race/Ethnicity | 8(0.11%) | 0.706 | 0.745 |
| Gender | 67 (0.96%) | 0.513 | 0.314 |
| Marital Status | 77 (1.11%) | 0.868 | 0.906 |
| Education | 95 (1.36%) | 0.568 | 0.655 |
| Employment Status | 168 (2.41%) | 0.614 | 0.723 |
| Phone Number | 0 (0%) | 1 | 1 |
| Email Address | 1 (0.01%) | 0.332 | 0 |
| SSN | 0(0%) | 1 | 1 |
| Location | 45 (0.65%) | 0.723 | 0.692 |
| Emotion / Opinion | 5,380 (77.28%) | 0.229 | 0.174 |
| Interest | 384 (5.52%) | 0.425 | 0.476 |
| No Disclosure | 1,415 (20.32%) | 0.447 | 0.397 |
| Total / Mean Score | 6,962 | 0.616 | 0.573 |

To measure agreement amongst annotators, two Kappa values were calculated: 1) Fleiss' Kappa amongst 4 annotators; 2) Cohen's Kappa between the primary annotator and a "majority voter" of the three additional annotators [85]. Table 3 reports the calculated Kappa scores for each category. Lower values for some categories including "Emotion/Opinion" and "Interest" were expected, taking into account their subjectivity [13, 86]. Low agreement for "Gender" was surprising, but retrospectively attributable to a lack of inclusive examples in the provided guidelines. For instance, tweets with gender-specific emojis were excluded in the examples for the "Gender" category. Similarly, low agreement rates on "Birthday/Age" and "Education" categories could be explainable by the diversity of ways that people describe their education level, limiting the effectiveness of the annotation instructions. Some people may annotate "I graduated from college 20 years ago" as a tweet that implies age information since one can infer the user's current age, but others may think that it is too vague to categorize it as age. For the sake of simplicity, all the sub-categories in "Demographic", "Personal Identifier", and "Location" will be referred to as *objective* Information. *Subjective* information include "Emotion/Opinion" and "Interest" categories. Overall, although annotators tended to show less consensus on subjective categories, there was greater agreement on objective categories.

### 4.6 Potential Impact of COVID-19

Since we collected Twitter user data during the COVID-19 pandemic, we briefly discuss how it may have affected participants' self-disclosing behaviors in online environments. Recent research showed that social media platforms have seen a 61% increase in usage during the pandemic [87]. Topics have shifted as well. Users are more involved in sharing mental health and support expressions [88]. At the same time, they self-censor their posts by evaluating whether they will hold negative impacts on others [87]. We examined the percentages of self-disclosing tweets across the three subsets of our data, collected over the following periods: (March, 4th - April, 1st), (March, 22nd - April, 19th) and (April, 4th-May, 2nd) respectively. The calculated percentages did not vary much (25%/24%/26%), but we acknowledge a possibility of certain types of personal information being discussed more frequently during the COVID-19.

**Table 4.** Demographic characteristics of study participants

| Variable | | N (% ) |
|---|---|---|
| Income | Mean | 97,727 |
| | Median | 75,000 |
| | Standard deviation | 88,011 |
| Gender | Female | 57 (51.82) |
| | Male | 50 (45.45) |
| | Non-binary | 3 (2.73) |
| Age | 18-20 | 13 (11.82) |
| | 21-29 | 24 (21.82) |
| | 30-39 | 25 (22.73) |
| | 40-49 | 15 (13.63) |
| | 50-59 | 22 (20) |
| | 60 or older | 11 (10) |
| Education | Less than high school degree or equivalent | 7 (6.36) |
| | Attending college | 22 (20) |
| | Bachelor / Associate degree | 43 (39.09) |
| | Graduate degree | 38 (34.55) |
| Race | White / Caucasian | 87 (79.09) |
| | Hispanic American | 6 (5.45) |
| | Multiple ethnicity | 14 (12.73) |
| | Black or African American | 3 (2.73) |
| Occupation | Unemployed | 16 (14.55) |
| | Natural Resources, construction, maintenance | 5 (4.55) |
| | Sales and office | 7 (6.35) |
| | Service | 11 (10) |
| | Professional and related | 59 (53.64) |
| | Management, business, financial operations | 12 (10.91) |

## 5 General Analysis

### 5.1 Participant Demographics

Table 4 gives the self-reported income, gender, age, education level, race/ethnicity and occupation breakdown of the 110 study participants. Our study participants are skewed younger with 70% (77) of participants being younger than 50. Most participants have some post-secondary education, with more than 70% (81) reporting having a Bachelor's or Associate degree, or higher. It is important to note that Twitter is not entirely representative of the overall U.S. adult population. According to the Pew Research Center, Twitter users are reported to be much younger than the average U.S adult and are also more likely than the general public to have a college degree [89], which seems to justify the skewness of our dataset. However, we acknowledge an uneven distribution of a racial makeup in our dataset, and we explore this more thoroughly later in the "Limitations" section.

For simplicity's sake, the occupation categories were reduced from 18 different categories to 5 high-level categories utilized in the Current Population Survey[4] along a category for "Unemployed" as well. From these new categories, 64.55% (71) reported to work in management, professional, and related occupations, which are the highest paying of the major occupational categories, and 14.55% (16) participants reported to be unemployed.

### 5.2 Analysis of Self-Disclosure Tweets

As described above, 21,068 total collected tweets were narrowed to 6,962 for self-disclosure labeling. Of these, 79.24% (5,517) of tweets with first-person pronouns contained elements of self-disclosure. The resulting distribution of tweets across personal information labels are included in a Table 3. All participants disclosed at least one type of personal information during the study period. 79 participants shared objective information at least once. Overall, 7.22% (503) of tweets revealed objective information and 77.39% (5,388) disclosed subjective information. In particular, 77.28% of labeled tweets contain sentiments or opinions. Only one tweet explicitly contained the author's email address, personally identifiable information. Of the 6,962, 1,415 tweets were labeled as "None", having not been identified as containing information about the 12 predefined categories in the annotation scheme.

4 https://www.bls.gov/cps/cpsaat11.htm

# 6 Statistical Tests

Six socio-demographic factors (income, gender, age, education level, race/ethnicity, occupation) were used as independent variables for the proposed research questions. Research questions are as follows: do socio-demographic factors have an impact on the usage of login verification? (RQ1); do socio-demographic factors have an impact on self-disclosure? (RQ2); do topics of self-disclosure vary across socio-demographic groups? (RQ3). Income is a continuous variable and the remaining variables are categorical. In support of the first research question (RQ1), we explored the relationship between these independent variables and login verification usage. We performed the Pearson's Chi-Square test or Fisher's Exact test as appropriate (continuous vs. categorical). For the second question (RQ2), we examined users' self-disclosing behaviors as evidenced in our set of annotated tweets. Specifically, the dependent variable for RQ2 was calculated by dividing the number of self-disclosing tweets with the total number of tweets for each user. This percentage was inspected by an analysis of variance (ANOVA). Additionally, multiple regression models were performed for both RQ1 and RQ2. Our final analysis in support of the third research question (RQ3) compared the distribution of personal information categories across the independent variables. Data were analyzed using R, an open-source statistical software package, and were screened for missing values, multicollinearity, homoscedasticity, and normality assumptions. Based on a Shapiro-Wilk's test, income was determined not to be normally distributed and was therefore log-transformed. The results show that: 1) none of the socio-demographic variables have significant effects on reported login verification usage; 2) gender and education predictors affect the percentage of self-disclosing tweets; 3) self-disclosing behaviors with respect to the "Emotion/Opinion" category vary within the gender groups; 4) self-disclosures of "Birthday/Age" were associated with the level of education; 5) the percentage of self-disclosing tweets in "Race/Ethnicity", "Gender" and "Location" categories within different occupations was statistically different. Following, we discuss the results of statistical testing in a greater detail.

## 6.1 Research Question 1: Login Verification Setting

In the screening and exit survey, subjects were asked about two privacy settings on Twitter: login verification and tweet protection. The tweet protection feature allows only the user's followers to view and interact with the user's tweets. We considered participants who had public accounts for our study because an access to their real-time tweets was required. Since the data contains small sample sizes, Fisher's Exact test was conducted to determine if the use of login verification is dependent on each socio-demographic variable. We converted income to a categorical variable by thresholding according to divisions suggested by the U.S. Census Bureau.[5]. The results of the Fisher's Exact test are shown in Table 5.

As shown, none of the socio-demographic variables are found to have statistically significant effects on reported login verification usage. We also ran a logistic regression test, using reported login verification usage ("Yes" or "No") as the response variable and omitting any responses of "I don't know". The quantitative income variable was included in this test and reference levels were created for the categorical variables. The reference levels were selected randomly to prevent any bias: "White / Caucasian" for race/ethnicity, "18-20" for age, "Unemployed" for occupation, "Less than high school degree or equivalent" for education, and "Female" for gender. Results of the logistic regression are given in Table 6 as Model 1.

To mitigate Type I Error, a Bonferroni correction was performed to correct for multiple tests ($\alpha = 0.05/6 = 0.0083$). No statistically significant results were found; p-values with respect to the age categories "21-29", "30-39", and "40-49" were below 0.05, but they failed to the adjusted p-value threshold (0.0083). A box plot of the income variable and the survey responses is shown in Figure 1.

## 6.2 Research Question 2: Self-Disclosure by Quantity

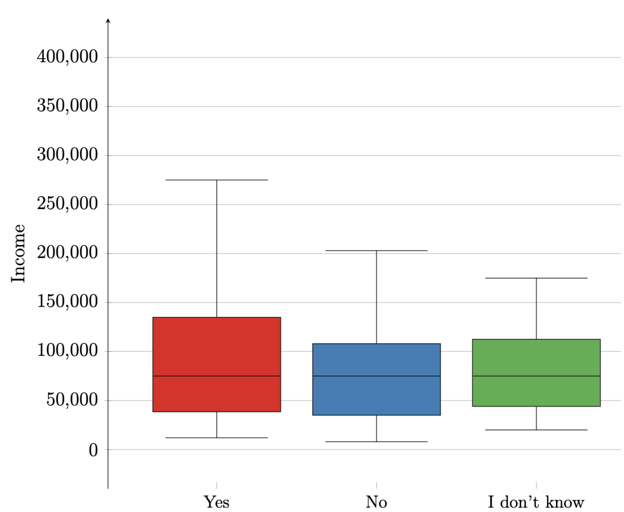
The results of one-way ANOVA on each independent variable are summarized in Table 7. Income variable was converted to a categorical variable by thresholding. Table 7 shows that only gender was statistically significant

5 https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.html

**Table 5.** Results of Fisher's Exact test for login verification setting (RQ1)

| Variable | | Yes N | Yes % | No N | No % | I don't know N | I don't know % | P |
|---|---|---|---|---|---|---|---|---|
| Income | Low (less than $45,000) | 14 | 46.67 | 10 | 33.33 | 6 | 20 | 0.526 |
| | Middle ($45,000-149,999) | 22 | 36.67 | 20 | 33.33 | 18 | 30 | |
| | High (more than $150,000) | 11 | 55 | 5 | 25 | 4 | 20 | |
| Gender | Male | 18 | 36 | 17 | 34 | 15 | 30 | 0.455 |
| | Female | 28 | 49.12 | 16 | 28.07 | 13 | 22.80 | |
| | Non-binary | 1 | 33.33 | 2 | 66.67 | 0 | 0 | |
| Age | 18-20 | 6 | 46.15 | 6 | 46.15 | 1 | 7.69 | 0.292 |
| | 21-29 | 10 | 41.67 | 9 | 37.50 | 5 | 20.83 | |
| | 30-39 | 12 | 48 | 7 | 28 | 6 | 24 | |
| | 40-49 | 8 | 53.33 | 5 | 33.33 | 2 | 13.33 | |
| | 50-59 | 9 | 40.91 | 6 | 27.27 | 7 | 31.82 | |
| | 60 or older | 2 | 18.18 | 2 | 18.18 | 7 | 63.64 | |
| Education | Less than high school or equivalent | 4 | 57.10 | 2 | 28.57 | 1 | 14.29 | 0.814 |
| | Attending college | 11 | 50 | 8 | 36.36 | 3 | 13.63 | |
| | Bachelor's degree / Associate degree | 17 | 39.53 | 13 | 30.23 | 13 | 30.23 | |
| | Graduate degree | 15 | 39.47 | 12 | 31.58 | 11 | 28.95 | |
| Race/Ethnicity | White/Caucasian | 33 | 37.93 | 28 | 32.18 | 26 | 29.89 | 0.331 |
| | Hispanic American | 5 | 83.30 | 1 | 16.67 | 0 | 0 | |
| | Multiple Ethnicity | 7 | 50 | 5 | 35.71 | 2 | 14.28 | |
| | Black or African American | 2 | 66.67 | 1 | 33.30 | 0 | 0 | |
| Occupation | Unemployed | 8 | 50 | 5 | 31.25 | 3 | 18.75 | 0.59 |
| | Natural resources, construction, maintenance | 1 | 20 | 2 | 40 | 2 | 40 | |
| | Sales and office | 4 | 57.14 | 2 | 28.57 | 1 | 14.29 | |
| | Service | 6 | 54.55 | 4 | 36.36 | 1 | 9.09 | |
| | Professional and related | 26 | 47.46 | 16 | 27.12 | 17 | 25.42 | |
| | Management, business, financial operations | 2 | 16.67 | 6 | 50 | 4 | 33.33 | |

**Fig. 1.** Box plot of survey responses to the use of login verification and income




($p = 0.002$) while there were no significant differences in self-disclosing behaviors amongst the remaining 4 variables (age, education, race/ethnicity, occupation). As the ANOVA test results suggested a significant difference in disclosure rate based on gender, post hoc analyses were conducted with a Tukey HSD test. This test concluded that female and male groups are statistically different ($p < 0.01$) in levels of self-disclosure. Specifically, women are more likely to disclose their personal information than men (28.7% vs. 19.2%). This finding is confirmatory of previous studies [2, 6–8]. Interactions between these variables were inspected using a five-way ANOVA (gender X age X race/ethnicity X education X occupation). Results revealed that only gender ($p < 0.01$) has a statistically significant impact on self-disclosing behaviors, and that the interactions between factors are not significant.

As a last step, we fit a multiple linear regression model to identify the size of the effect the variables have on online self-disclosure. This analysis assumes that the factors do not interact with each other in their effect on the dependent variable. Categorical indicators including gender, age, education level, race/ethnicity and occupation were dummy-encoded. Results of the linear regression model are shown in Table 6. To protect from Type I Error, a Bonferroni correction was performed to adjust for multiple tests ($\alpha = 0.05/6 = 0.0083$). Findings

**Table 6.** Regression analysis for login verification setting (Model 1) and self-disclosure by quantity (Model 2)

| | | | Model1 (login verification setting) | | Model2 (self-disclosure) | |
|---|---|---|---|---|---|---|
| Variable | | N | $\beta$ Coefficient | P | $\beta$ Coefficient | P |
| Income | (continuous variable) | 110 | 0.609 | 0.6091 | -0.001 | 0.952 |
| Gender | Female | 57 | | ref | | |
| | Male | 50 | -0.458 | 0.459 | -0.115 | 0.0000049*** |
| | Non-binary | 3 | -2.394 | 0.139 | 0.016 | 0.822 |
| Age | 18-20 | 13 | | ref | | |
| | 21-29 | 24 | 2.483 | 0.04* | 0.05 | 0.32 |
| | 30-39 | 25 | 2.741 | 0.03* | 0.113 | 0.034* |
| | 40-49 | 15 | 2.838 | 0.03* | 0.095 | 0.087 |
| | 50-59 | 22 | 2.744 | 0.052 | 0.05 | 0.38 |
| | 60 or older | 11 | 2.807 | 0.103 | -0.02 | 0.712 |
| Education | Less than high school degree or equivalent | 7 | | ref | | |
| | Attending college | 22 | 1.537 | 0.277 | -0.091 | 0.103 |
| | Bachelor or Associate degree | 43 | -0.76 | 0.549 | -0.144 | 0.004** |
| | Graduate degree | 38 | -0.85 | 0.532 | -0.144 | 0.008** |
| Race/Ethnicity | White / Caucasian | 87 | | ref | | |
| | Hispanic American | 6 | 1.973 | 0.164 | -0.077 | 0.158 |
| | Multiple Ethnicity | 14 | -0.12 | 0.871 | -0.026 | 0.422 |
| | Black or African American | 3 | 1.075 | 0.534 | -0.067 | 0.327 |
| Occupation | Unemployed | 16 | | ref | | |
| | Natural resources, construction, maintenance | 5 | -0.29 | 0.864 | -0.038 | 0.525 |
| | Sales and office | 7 | 0.212 | 0.862 | -0.042 | 0.431 |
| | Service | 11 | -0.766 | 0.505 | 0.065 | 0.187 |
| | Professional and related | 59 | 0.154 | 0.861 | 0.016 | 0.661 |
| | Management, business, financial operations | 12 | -1.87 | 0.129 | 0.107 | 0.028* |

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

of Model 2 (Table 6) demonstrate that gender and education predictors affect the percentage of self-disclosing tweets. Consistent with previous analyses, male participants were 11.5 % less associated with self-disclosure than female participants (coefficient: -0.115, $p < 0.001$). Moreover, participants holding a college diploma or a graduate degree were more reluctant to share their personal information in online environments than participants with less than a high school degree or equivalent (coefficient: -0.144, $p < 0.01$). Participants in their thirties were more likely to disclose personal information online than those of ages 18-20 (coefficient: 0.113, $p < 0.05$). Participants engaged in management, business, and finance-related operations were 10.6 % more associated with self-disclosure than participants with no job (coefficient: 0.106, $p < 0.05$). However, these two findings became statistically insignificant after applying the Bonferroni correction which set the threshold for significance at $p=0.0083$.

A multivariate stepwise regression model was used to select a subset of factors that yielded the best prediction for rates of self-disclosure. The significant parameters (gender, age, education and occupation) were selected. Findings from the previous analysis with respect to age and occupation has become significant after fitting a new regression model with the selected features ($p < 0.0125$).

## 6.3 Research Question 3: Self-Disclosure by Topics

To study the types of personal information disclosed by different groups of participants, we calculated the mean percentage of self-disclosing tweets in each category using labels acquired from the human annotator. Appendix B contains the distribution of tweets with personal information categories across independent variables. Categories with fewer than two instances were excluded from the table. We converted income to a categorical variable by thresholding according to divisions suggested by the U.S. Census Bureau.[6] Results indicate that emotions and opinions are dominant topics of conversation regardless of socio-demographic characteristics.

**6** https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.html

**Table 7.** One-way ANOVA within groups for each independent variable (RQ2)

| | variable | n (%) | mean % of disclosure | p |
|---|---|---|---|---|
| Income | low (less than $45,000) | 30 (27.27) | 25.7 | 0.514 |
| | middle ($45,000-149,999) | 60 (54.54) | 25.3 | |
| | high (more than $150,000) | 20 (18.19) | 25.6 | |
| Gender | Female | 57 (51.82) | 28.7 | 0.002** |
| | Male | 50 (45.45) | 19.2 | |
| | Non-binary | 3 (2.73) | 33.3 | |
| Age | 18-20 | 13 (11.82) | 22.6 | 0.482 |
| | 21-29 | 24 ( 21.82) | 25.1 | |
| | 30-39 | 25 (22.73) | 28.9 | |
| | 40-49 | 15 (13.64) | 26.7 | |
| | 50-59 | 22 (20.0) | 24.4 | |
| | 60 or older | 11 (10.0) | 2.1 | |
| Education | Less than high school degree or equivalent | 7 (6.36) | 0.311 | 0.547 |
| | Attending college | 22 (20.0) | 25.4 | |
| | Bachelor's degree / Associate degree | 43 (39.0) | 24 | |
| | Graduate degree | 38 (34.55) | 25.7 | |
| Race/Ethnicity | Black or African American | 3 (2.73) | 19.6 | 0.78 |
| | Hispanic American | 6 (5.45) | 24 | |
| | Multiple Ethnicity | 14 (12.73) | 23.9 | |
| | White / Caucasian | 87 (79.09) | 25.9 | |
| Occupation | Unemployed | 16 (14.55) | 25.3 | 0.298 |
| | Natural resources, construction, maintenance | 5 (4.55) | 17.7 | |
| | Sales and office | 7 (6.36) | 20.5 | |
| | Service | 11 (10.0) | 26.7 | |
| | Professional and related | 59 (53.64) | 25.1 | |
| | Management, business, financial operations | 59 (10.91) | 31.2 | |

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

A one-way ANOVA was performed on each personal information category to measure statistical differences within groups of each independent variable. Any statistical findings regarding non-binary or Black participants are disregarded due to its small sample size (n = 3). The results indicate that there is no significant difference among income, age and race/ethnicity predictors. However, self-disclosing behaviors with respect to the "Emotion/Opinion" ($p < 0.001$) category differ within the gender groups. In particular, women tended to share their emotions or thoughts more frequently than men ($p < 0.001$), which is consistent with the literature [40, 41]. Disclosures of "Birthday/Age" were significantly correlated with level of education. Post hoc analyses showed that those with a college degree were less likely to share information regarding their date of birth or current age than those who are attending college ($p < 0.05$). Differences in "Race/Ethnicity", "Gender" and "Location" categories within different occupations was significant as well. Specifically, unemployed respondents disclosed their racial, ethnic or gender information more frequently than those working in professional occupations ($p < 0.05$). Moreover, those people who are engaged in professional occupations or without current jobs were negatively associated with location-related information sharing than those in sales and office occupations ($p < 0.05$).

# 7 Recollection and Actual Disclosure

Previous research has demonstrated a consistency between users' stated information sharing and their actual disclosure on social media [53]. Specifically, prior work has found that more than 70% of Facebook users were accurate about their perceived sharing behaviors, and 11% revealed less than they claimed they do [53]. To test this claim, we asked participants to report their recollection of sharing particular types of information during the study period in the exit survey. Categories include 5 types of personal information: demographic characteristics (e.g. age, sex, marital status, education, employment status, income), lifestyle characteristics (including media habits), shopping/purchasing habits, financial data, and personal identifiers (e.g., names, addresses, social security numbers) [90]. In order to make a fair comparison between observed disclosures and stated recollections, we re-annotated several of the information categories and generated sub-categories that matched the survey categories. A difference between users' reported recollection of sharing and their actual sharing behaviors is shown in Figure 2. We visually observe that participants in fact shared less than they recalled for all of the categories except for marital status and education. Also, Welch's t-test ($\alpha = 0.05$) was performed to test if the means of two variables (observed disclosures vs. stated recollections) are significantly different. The result confirmed that participants statistically shared their information less frequently than they remembered ($p < 0.01$). To compare the mean differences of two variables (recollection vs. actual disclosure) with respect to types of personal information, we ran Welch's t-test ($\alpha = 0.05$) on each category respectively. As shown in Figure 2, mean differences associated with "Gender" ($p = 0.0002$), "Race/Ethnicity" ($p = 0.0004$), "Phone Number" ($p = 0.025$), "Email Address" ($p = 0.0003$), "State" ($p = 1.009e\text{-}06$), "City" ($p = 1.414e\text{-}07$), "Postal Code" ($p = 0.002$), "Workplace Location" ($p = 0.0004$), "Hobbies" ($p < 2.2e\text{-}16$), "TV programs" ($p = 0.0002$), and "Leisure Activities" ( $p < 2.2e\text{-}16$) categories were tested to be statistically significant. Particularly, we find a larger disparity in "Hobbies", "TV program", and "Leisure Activities" categories, possibly because one may not necessarily include first-person pronouns when discussing its favorites. A comment such as "The Walking Dead is the best" would not have been captured with the approach adopted in our study. This discussion is beyond the scope of this work but could be the subject of further research.

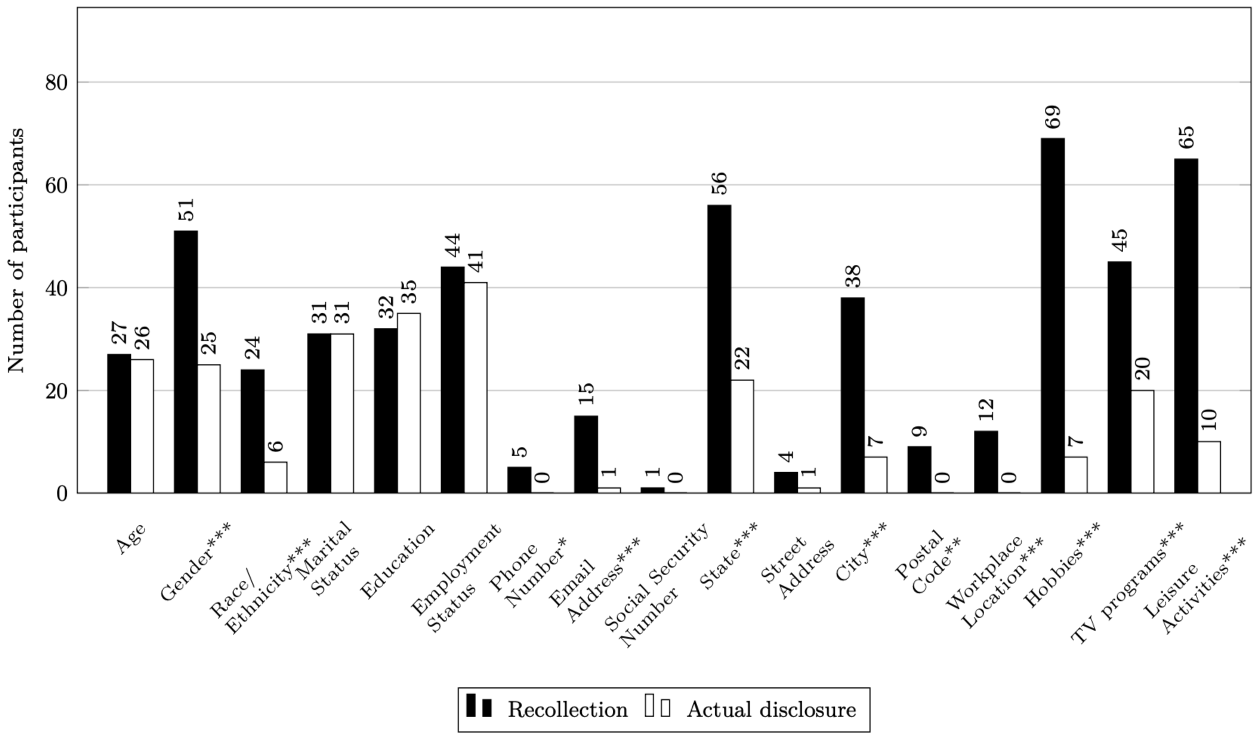
## 8 Topics and Self-Disclosure

Following the quantitative analysis described, we explored the topic distribution of both non-disclosing tweets and disclosing tweets using topic modeling. Topic modeling is a text mining tools that helps to discover underlying themes in a text. Topic models are regularly engaged to study users' online behavior [13, 14, 40, 43]. This approach allows us to examine differences between tweets with and without annotated self-disclosure. To this end, we constructed bi-grams and ran two separate topic models. We used two topic models, not one unified topic model, because it was the easiest way to clearly distinguish the topical differences of two subtext groups. A unified topic model trained on all texts, regardless of the element of self-disclosure, may create ambiguities when interpreting the outcomes. Text preprocessing included converting to lower case, removing punctuation, stopwords and URLs, tokenizing, and lemmatizing. We used NLTK for text pre-processing.[7] The top five most frequently used bigrams in self-disclosing tweets were found to be the following: (gon,na), (feel, like), (sorry, loss), (look, like), (would, like). These bigrams can be found in texts that share the authors' opinions or sentiments, which was the dominant category in our dataset. The top five most frequently used bigrams from tweets without self-disclosure were: (look, like), (happy, birthday), (gon, na), (donald, trump), (united,state), (social, distance). We note that (gon,na) and (look, like) are overlapping in both subsets, but the latter covers a broader set of topics including politics and the COVID-19 pandemic. Topic modeling proceeded as follows. After text pre-processing, topics were generated by the LDA Mallet from the Python wrapper.[8] To obtain an optimal number of topics, multiple topic models with different number of topics varying from 5 to 30 were created. Each model was evaluated by coherence score which measures quality and interpretability of the topic models. Subsequent analyses revealed the best topic model with 30 topics and coherence score of 0.563 for tweets with no disclosure, and 30 topics and coherence score of 0.539 for tweets with disclosure. Table 8 contains the extracted topics from each LDA model with relevant keywords. Overall, the most frequently discussed topics did not differ much when we compared disclosing tweets and non-disclosing tweets. During the study period, participants actively shared information about their daily routines, politics, and the COVID-19 pandemic. Considering a large number of self-disclosing tweets that were labeled as the "Emotion/Opinion" category, it is reasonable to witness multiple emotion-related keywords associated with the "Daily life / Feeling" topic. Moreover, topics related to school may be explainable due to the "Occupation" category, if there were many school faculty members or staffs amongst the participants.

**7** https://www.nltk.org

**8** https://radimrehurek.com/gensim/models/wrappers/ldamallet.html

**Fig. 2.** Recollection and actual disclosure




Signif. codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05

# 9 Comparing Self-Reported Privacy Awareness and Actual Self-Disclosure

We compared users' stated privacy and security risk awareness and their actual sharing behaviors. Prior studies have examined whether privacy risk perception explains the use of certain privacy protection strategies [25, 48, 53–55, 55, 61, 62]. Their work has yielded mixed findings; several studies have argued that there is no relationship between users' risk perceptions and privacy behaviors [48, 53–55, 55], while others have reported a positive correlation between these two factors [25, 61, 62]. We first examined the prevalence of and stated reasons for participants' utilization of privacy and security settings. Specifically, Twitter has two privacy enhancing features: *two-factor authentication* and *protect your tweets*. The two-factor authentication feature requires users to go through additional steps with either a code, a login confirmation via an app, or a physical security key when they log in to their accounts. At the same time, users may keep their accounts private with the tweet protection setting, which means only their current followers are able to see their tweets. The self-reported use of login verification settings among our study participants was as follows: 47 (42.73%) participants reported having activated the login verification feature, while 63 (57.27%) people answered “No” or “I don't know”. Among those who used this feature, more than 75 % of them (n = 37, 78.72%) responded they used it due to concerns over account security. This finding supports the assumption that people utilize privacy and security settings because they acknowledge and are concerned about potential privacy risks. In terms of the adoption of “protect your tweets” feature, 5.45% of subjects responded “Yes”, 85.45% responded “No”, and 9.09% responded “I don't know”. Considering the fact that all of the participants' accounts were actually public, their responses indicate that for some users, there is little understanding of the feature, which could be a result of a suboptimal user interface for the privacy setting or other reasons [91]. There was, however, no strong correlation between users' self-disclosure and their use of privacy/security enhancing options, particularly for the login verification setting, when it was analyzed independently. This claim was tested with a one-way ANOVA and results were not found to be statistically significant. In order to discover a set of in-

**Table 8.** Top five topics from topic model with relevant keywords

| | Topic | Keywords | Number of Tweets |
|---|---|---|---|
| Tweets with self-disclosure | Daily life / Feeling | make, time, love, feel, good, hope, today | 5,518 |
| | School | work, read, kid, school, book, thing, home, student | 24 |
| | Wearing mask | find, back, call, play, wear, mask | 3 |
| | Politics | trump, vote, biden, hard, lose, put, party | 2 |
| Tweets without self-disclosure | Daily life | time, make, live, home, world, work | 11,061 |
| | Current event | day, today, watch, thing, president, read, make, stop | 3,341 |
| | Covid19 | people, state, die, coronavirus, pandemic, case | 786 |
| | Trump | good, trump, back, bad, guy, man, great | 288 |
| | Politics | trump, vote, biden, bernie, people, american | 39 |

dependent variables (income, gender, age, education, race/ethnicity, occupation, privacy setting) that significantly influence online self-disclosure, a stepwise regression was performed and suggested a removal of income and race/ethnicity variables. A stricter threshold for significance ($\alpha = 0.05/5 = 0.01$) was implemented as a result of Bonferroni correction. With respect to socio-demographic factors, the results were identical to the previous findings. The model further suggests that an increased disclosure was predicted among those who stated they activated the login verification setting, compared to those who answered "I don't know" (coefficient: 0.0764, $p < 0.01$).

# 10 Discussion

The current study investigates which of several socio-demographic factors (income, age, gender, race/ethnicity, education, occupation) are associated with the adoption of privacy and security settings and self-disclosing behaviors of 110 active Twitter users. We demonstrate interactions amongst these factors and disclosure as a function of disclosure content. We contribute to the existing literature by also examining multiple aspects of their stated/observed privacy behaviors from Twitter. The observed disparity between users' recollection and actual self-disclosure and users' lack of knowledge about the public or private status of their accounts raises concerns about personal data management on social networking sites.

## 10.1 Socio-Demographics and Privacy Behaviors

Twitter users' privacy behaviors were interpreted through the lens of privacy preferences and online self-disclosure. Firstly, statistical analyses on the login verification feature demonstrate that a user's socioeconomic status or demographic factors have no impact on their usage of the privacy protection feature, contradicting our expectations. Previous literature showed conflicting results that gender and age indicators are related to the use of privacy enhancing tools [24, 25, 92]. Additionally, previous work that analyzed attitudes toward privacy on Twitter [93, 94] solely focused on whether users set their accounts private or not, unlike our study which specifically tested for the login verification feature. Still, overall results for RQ1 did not confirm variations of privacy preferences in different SES or demographic populations observed by prior studies [24, 25, 92, 93].

For self-disclosure, our findings suggest that participants' gender and education level statistically affect their information sharing behaviors on Twitter. In line with previous findings, female participants demonstrated a higher level of self-disclosure than male participants. Also, participants holding a college diploma or higher shared less than those with less than a high school degree or equivalent. An analysis with the stepwise regression model showed that income and race/ethnicity are the least significant predictors of self-disclosure. The removal of income and race/ethnicity indicators produced two additional findings: (1) participants in their thirties were more likely to disclose personal information online than those of ages 18-20; and (2) participants engaged in management, business, and finance-related operations disclosed more about themselves than unemployed individuals. As a result, our study identified varying degrees of self-disclosure by some groups of populations, particularly in gender, education, age, and occupation.

To interpret our empirical findings on self-disclosure in the context of digital inequality, we note Madden's finding [1] that underrepresented populations report suffering from not finding adequate privacy-enhancing strategies required to better secure their data in online environments. Accordingly, we assumed that low-SES users would be more likely to share their personal information than high-SES users. Our assumption was supported in a sense of education level: participants with lower levels of education (less than a high school degree or equivalent) disclosed personal information more severely than those who attained a college degree or higher. Yet, in contrast to our expectation, unemployed individuals disclosed less than those reported to work in management, professional, and related occupations, which are typically the highest paying of the major occupational categories.

We observed no relation between privacy preferences and patterns of self-disclosure when examined in isolation using a one-way ANOVA. However, a linear regression model with other demographic variables suggested increased disclosure among those who claimed to be using the login verification setting, compared to those who answered "I don't know". This result is consistent with Liang, Shen, and Fu's report that individuals with advanced privacy settings were found to disclose large amounts of personal information online [93]. Additional insight is provided when considering participants' responses about their usage of the tweet protection feature. The results show that 10 users (9.09%) were not familiar with this feature by answering "Yes" or "I don't know" when no participants used it.

## 10.2 Types of Personal Information of Self-Disclosure

Our study is novel in that it explores the contents of personal information in self-disclosure along socio-demographic lines. In prior work, only gender and age variables have been primarily explored [40, 41, 44]. As expected, the ANOVA suggested that self-disclosure varied across some contents of personal information. A descriptive analysis demonstrates that emotions and opinions are dominant topics of the conversation regardless of socio-demographic dimensions. Corresponding to Humphreys et al.'s finding [81], tweets tend not to include personally identifiable information: only one participant in our sample mentioned an email address. While tweets by people with different levels of income, age and ethnic groups did not vary in terms of categories of personal information, variances concerning the remaining variables (gender, education, and occupation) were found to be statistically significant. Our results reinforce prior observations of higher self-disclosure by women of more intimate and subjective topics [40, 41]. While topical patterns of disclosure along socio-demographic factors have not been thoroughly examined in previous literature, we did not pursue specific predictions or hypotheses in terms of differences within race/ethnicity, education, and occupation variables.

## 10.3 Implications for Future Studies

The present study has provided insights into how online self-disclosure varies across socio-demographic populations. We found that participants who are female or have lower levels of education tended to share their personal information more actively than the others, which makes them more susceptible to privacy harms. In line with Madden's finding [1], these findings highlight potential signs of digital inequality, particularly across education levels. Yet, findings associated with the occupation variable are inconsistent with our previous finding in the education variable: unemployed individuals disclosed less than those reported to work in management, professional, and related occupations. Regardless, overall results suggest that privacy has become part of the digital divide, separating those who can understand and control their privacy from those who cannot. That is, some populations are more adversely affected than others. Prior literature [26, 72–75] has demonstrated undesirable repercussions of intended or unintended disclosure. Future work should explore differential impacts of disclosure across SES and demographic subgroups.

Not only the intensity of self-disclosure but also the contents of shared information matter. We discovered a strong variation in topics of disclosure amongst different socio-demographic groups. Although participants generally did not reveal sensitive information such as phone number or email address, it does not guarantee the protection of their privacy. For example, if pieces of information, including "Birthday/Age" and "Location", retrieved from tweets are combined with other publicly available data, the particular user may be identified. That being said, we reckon it is critical to investigate technological interventions that could assist all individuals with privacy and security decision making.

Our additional findings open up new lines of research related to privacy nudges. We found that there exist significant discrepancies between participants' recollections of sharing and their actual behaviors and that some participants were unaware if their accounts were public or not. Hence, developing an automated mechanism that regularly informs users about shared categories and privacy settings can be useful to improve their privacy awareness.

### 10.4 Limitations

Despite our efforts to enroll diverse participants, our sample of Twitter users is not fully representative of the United States adult population. Our sample was demographically skewed in terms of race/ethnicity: none of our participants identified as Asian and a majority of participants (n = 87, 79.09%) identified as Caucasian. According to the Pew Research Center, the racial and ethnic breakdown of Twitter users is similar to the U.S adult population [89]. According to the 2019 U.S. Census Bureau Estimates,[9] 60.1% of the total U.S population is Caucasian, 13.4% is African American, and 18.5% is Hispanic. Asian Americans make up of 5.9% of the population, and multiracial Americans the remaining 2.8%. It is critical to note that the sample size of our study is small. That is, the uneven racial distribution of our sample and the small sample size might have limited our opportunities to uncover significant relationships with respect to our variables of interest. We are encouraged, however, by the significant effects we do find and suggest that they should be considered meaningful given the exploratory nature of the study.

We did not explore distinctions between professional and personal accounts. We acknowledge that these accounts may have differential patterns of self-disclosure and that this variable may mediate observed effects. In addition, we used a sequential recruitment approach, and included *follower look-alikes* as an ad targeting strategy later in the study. We are not sure the impact this strategy may have had on our findings, however, we have verified that participants recruited using this approach do not significantly differ from other study participants with respect to demographic variables.

This study is focused on capturing explicit forms of self-disclosure, signalled by the presence of first-person pronouns. We acknowledge that this approach does not exhaustively capture every instance of self-disclosure. This may have affected discrepancies between individuals' recollection of sharing and their actual disclosure, particularly for some subjective categories.

**9** https://www.census.gov/quickfacts/fact/table/US#

## 11 Conclusion

We presented a study of how 110 Twitter users perceived and managed their individual privacy. We examined effects of income, gender, age, race/ethnicity, education level and occupation on their privacy preferences and patterns of self-disclosure. Although differences in the usage of privacy and security feature were minimal, gender and education level were identified as important factors in predicting the levels of self-disclosure. Additionally, we described differences in the types of information shared by socio-demographic groups.

Our additional findings identify notable discrepancies in participants' knowledge of their privacy: 1) participants' stated behaviors do not match their actual self-disclosure, and 2) some participants were unaware if their accounts were public or not. Future work can investigate potential remedies to offer an equitable environment for all individuals' privacy understanding and control.

## 12 Acknowledgments

This research was supported in part by an Accelerator Award from the Center for Social Data Analytics at the Pennsylvania State University. We also acknowledge and thank the reviewers for their time in helping us revise this manuscript.

## References

[1] Mary Madden. Privacy, security, and digital inequality. *Data & Society*, 2017.

[2] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill. "privacy is not for me, it's for those rich women": Performative privacy practices on mobile phones by women in south asia. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pages 127–142, 2018.

[3] Tamy Guberek, Allison McDonald, Sylvia Simioni, Abraham H Mhaidli, Kentaro Toyama, and Florian Schaub. Keeping a low profile? technology, risk and privacy among

undocumented immigrants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2018.

[4] Alice Marwick, Claire Fontaine, and Danah Boyd. "nobody sees it, nobody gets mad": Social media, privacy, and personal responsibility among low-ses youth. *Social Media+ Society*, 3(2):2056305117710455, 2017.

[5] Paul DiMaggio, Eszter Hargittai, et al. From the 'digital divide'to 'digital inequality': Studying internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University*, 4(1):4–2, 2001.

[6] Jayati Dev, Pablo Moriano, and L Jean Camp. Lessons learnt from comparing whatsapp privacy concerns across saudi and indian populations. In *Sixteenth Symposium on Usable Privacy and Security ({SOUPS} 2020)*, pages 81–97, 2020.

[7] Anita L Allen and Erin Mack. How privacy got its gender. *N. Ill. UL Rev.*, 10:441, 1989.

[8] Mariea Grubbs Hoy and George Milne. Gender differences in privacy-related measures for young adult facebook users. *Journal of Interactive Advertising*, 10(2):28–45, 2010.

[9] Bas Hofstra, Rense Corten, and Frank van Tubergen. Understanding the privacy behavior of adolescents on facebook: The role of peers, popularity and trust. *Computers in Human Behavior*, 60:611–621, 2016.

[10] Yang Wang, Gregory Norice, and Lorrie Faith Cranor. Who is concerned about what? a study of american, chinese and indian users' privacy concerns on social network sites. In *International conference on trust and trustworthy computing*, pages 146–153. Springer, 2011.

[11] Nathan S Consedine, Shulamit Sabag-Cohen, and Yulia S Krivoshekova. Ethnic, gender, and socioeconomic differences in young adults' self-disclosure: Who discloses what and to whom? *Cultural Diversity and Ethnic Minority Psychology*, 13(3):254, 2007.

[12] Valerian J Derlaga and John H Berg. *Self-disclosure: Theory, research, and therapy*. Springer Science & Business Media, 1987.

[13] Prasanna Umar, Anna Squicciarini, and Sarah Rajtmajer. Detection and analysis of self-disclosure in online news commentaries. In *The World Wide Web Conference*, pages 3272–3278, 2019.

[14] Taylor Blose, Prasanna Umar, Anna Squicciarini, and Sarah Rajtmajer. Privacy in crisis: A study of self-disclosure during the coronavirus pandemic. *arXiv preprint arXiv:2004.09717*, 2020.

[15] JinYeong Bak, Suin Kim, and Alice Oh. Self-disclosure and relationship strength in twitter conversations. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 60–64, 2012.

[16] Azy Barak and Orit Gluck-Ofri. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior*, 10(3):407–417, 2007.

[17] Olga Abramova, Amina Wagner, Hanna Krasnova, and Peter Buxmann. Understanding self-disclosure on social networking sites-a literature review. 2017.

[18] Adam N Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B Paine Schofield. Privacy, trust, and self-disclosure online. *Human–Computer Interaction*, 25(1):1–24, 2010.

[19] Tobias Dienlin and Miriam J Metzger. 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, 2016.

[20] Kun Liu and Evimaria Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Trans. Knowl. Discov. Data*, 5(1):6:1–6:30, December 2010.

[21] Terence Chen, Abdelberi Chaabane, Pierre Ugo Tournoux, Mohamed-Ali Kaafar, and Roksana Boreli. How much is too much? leveraging ads audience estimation to evaluate public profile uniqueness. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 225–244. Springer, 2013.

[22] MM Rahman and Muhammad Abdullah Adnan. Two step verification system of highly secure social media: Possible to breach the security. In *2017 International Conference on Networking, Systems and Security (NSysS)*, pages 185–190. IEEE, 2017.

[23] Curt J. Dommeyer and B. Groß. What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing*, 17:34–51, 2003.

[24] Amanda Nosko, Eileen Wood, Lucia Zivcakova, Seija Molema, Domenica De Pasquale, Karin Archer, et al. Disclosure and use of privacy settings in facebooktm profiles: Evaluating the impact of media context and gender. *Social Networking*, 2(01):1, 2013.

[25] Isabelle Oomen and Ronald Leenes. Privacy risk perceptions and privacy protection strategies. In *Policies and research in identity management*, pages 121–138. Springer, 2008.

[26] Natalya N Bazarova and Yoon Hyung Choi. 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, 2014.

[27] Tracy Alloway, Rachel Runac, Mueez Quershi, and George Kemp. Is facebook linked to selfishness? investigating the relationships among social media use, empathy, and narcissism. *Social Networking*, 2014, 2014.

[28] Rebecca Sawyer and Guo-Ming Chen. The impact of social media on intercultural adaptation. 2012.

[29] Hanna Krasnova, Sarah Spiekermann, Ksenia Koroleva, and Thomas Hildebrand. Online social networks: Why we disclose. *Journal of information technology*, 25(2):109–125, 2010.

[30] Sonja Utz. The function of self-disclosure on social network sites: Not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. *Computers in Human Behavior*, 45:1–10, 2015.

[31] Natalya N Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. Social sharing of emotions on facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 154–164, 2015.

[32] Mesfin A Bekalu, Rachel F McCloud, and K Viswanath. Association of social media use with social well-being, positive mental health, and self-rated health: disentangling routine use from emotional connection to use. *Health Education & Behavior*, 46(2_suppl):69S–80S, 2019.

[33] Avner Levin, Mary Foster, Bettina West, Mary Jo Nicholson, Tony Hernandez, and Wendy Cukier. The next digital divide: Online social network privacy. *Privacy and Cyber Crime Institute, Ryerson University*, 2008.
[34] Joseph Henrich, Steven J Heine, and Ara Norenzayan. Most people are not weird. *Nature*, 466(7302):29–29, 2010.
[35] Eden Litt. Understanding social network site users' privacy tool use. *Computers in Human Behavior*, 29(4):1649–1656, 2013.
[36] Vasileios Lampos, Nikolaos Aletras, Jens K Geyti, Bin Zou, and Ingemar J Cox. Inferring the socioeconomic status of social media users based on behaviour and language. In *European Conference on Information Retrieval*, pages 689–695. Springer, 2016.
[37] Daniel Preoţiuc-Pietro, Vasileios Lampos, and Nikolaos Aletras. An analysis of the user occupational class through twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1754–1764, 2015.
[38] Márton Karsai Abitbol, Jacob Levy and Eric Fleury. Location, occupation, and semantics based socioeconomic status inference on twitter. In *In 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1192–1199, 2013.
[39] Bradley J Bond. He posted, she posted: Gender differences in self-disclosure on social network sites. *Rocky Mountain Communication Review*, 6(2), 2009.
[40] Xiao Ma, Jeffery T Hancock, Kenneth Lim Mingjie, and Mor Naaman. Self-disclosure and perceived trustworthiness of airbnb host profiles. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 2397–2409, 2017.
[41] Charles T Hill and Donald E Stull. Gender and self-disclosure. In *Self-Disclosure*, pages 81–100. Springer, 1987.
[42] James Caverlee and Steve Webb. A large-scale study of myspace: observations and implications for online social networks. In *ICWSM*, 2008.
[43] Yi-Chia Wang, Moira Burke, and Robert Kraut. Modeling self-disclosure in social networking sites. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pages 74–85, 2016.
[44] Erin E Hollenbaugh and Marcia K Everett. The effects of anonymity on self-disclosure in blogs: An application of the online disinhibition effect. *Journal of Computer-Mediated Communication*, 18(3):283–302, 2013.
[45] Anatoliy Gruzd and Ángel Hernández-García. Privacy concerns and self-disclosure in private and public uses of social media. *Cyberpsychology, Behavior, and Social Networking*, 21(7):418–428, 2018.
[46] Flavius Kehr, Daniel Wentzel, and Tobias Kowatsch. Privacy paradox revised: Pre-existing attitudes, psychological ownership, and actual disclosure. 2014.
[47] Mary Madden. Privacy management on social media sites. *Pew Internet Report*, pages 1–20, 2012.
[48] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. " i regretted the minute i pressed share" a qualitative study of regrets on facebook. In *Proceedings of the seventh symposium on usable privacy and security*, pages 1–16, 2011.
[49] Sameer Patil, Greg Norcie, Apu Kapadia, and Adam J Lee. Reasons, rewards, regrets: privacy considerations in location sharing as an interactive practice. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, pages 1–15, 2012.
[50] Alessandro Acquisti and Jens Grossklags. Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behavior. In *2nd Annual Workshop on Economics and Information Security-WEIS*, volume 3, pages 1–27, 2003.
[51] Andreas Krause and Eric Horvitz. A utility-theoretic approach to privacy and personalization. In *AAAI*, volume 8, pages 1181–1188, 2008.
[52] Alyson Leigh Young and Anabel Quan-Haase. Privacy protection strategies on facebook: The internet privacy paradox revisited. *Information, Communication & Society*, 16(4):479–500, 2013.
[53] Alessandro Acquisti and Ralph Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58. Springer, 2006.
[54] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.
[55] Alessandro Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce*, pages 21–29, 2004.
[56] Miriam J Metzger. Effects of site, vendor, and consumer characteristics on web site trust and disclosure. *Communication Research*, 33(3):155–179, 2006.
[57] Jonathan A Obar and Anne Oeldorf-Hirsch. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1):128–147, 2020.
[58] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Tech. LJ*, 30:39, 2015.
[59] Ramnath K Chellappa and Raymond G Sin. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information technology and management*, 6(2-3):181–202, 2005.
[60] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs*, 41(1):100–126, 2007.
[61] Jessica Vitak and Jinyoung Kim. " 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*, pages 461–474, 2014.
[62] Casey Fiesler, Michaelanne Dye, Jessica L Feuston, Chaya Hiruncharoenvate, Clayton J Hutto, Shannon Morrison, Parisa Khanipour Roshan, Umashanthi Pavalanathan, Amy S Bruckman, Munmun De Choudhury, et al. What (or who) is public? privacy settings and social media content sharing. In *Proceedings of the 2017 ACM Conference on*

*Computer Supported Cooperative Work and Social Computing*, pages 567–580, 2017.

[63] Fred Stutzman, Robert Capra, and Jamila Thompson. Factors mediating disclosure in social network sites. *Computers in Human Behavior*, 27(1):590–598, 2011.

[64] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. Where is the digital divide? a survey of security, privacy, and socioeconomics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 931–936, 2017.

[65] Sidney M Jourard and Paul Lasakow. Some factors in self-disclosure. *The Journal of Abnormal and Social Psychology*, 56(1):91, 1958.

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

[67] Lisa Collins Tidwell and Joseph B Walther. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human communication research*, 28(3):317–348, 2002.

[68] Sarah Rajtmajer, Anna Squicciarini, Christopher Griffin, Sushama Karumanchi, and Alpana Tyagi. Constrained social-energy minimization for multi-party sharing in online social networks. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 680–688, 2016.

[69] Sarah Rajtmajer, Anna Squicciarini, Jose M Such, Justin Semonsen, and Andrew Belmonte. An ultimatum game model for the evolution of privacy in jointly managed content. In *International Conference on Decision and Game Theory for Security*, pages 112–130. Springer, 2017.

[70] Xiao Ma, Jeff Hancock, and Mor Naaman. Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 3857–3869, 2016.

[71] Tom Postmes, Russell Spears, Khaled Sakhel, and Daphne De Groot. Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, 27(10):1243–1254, 2001.

[72] Adam N Joinson, David J Houghton, Asimina Vasalou, and Ben L Marder. Digital crowding: Privacy, self-disclosure, and technology. In *Privacy online*, pages 33–45. Springer, 2011.

[73] Ann E Schlosser. Self-disclosure versus self-presentation on social media. *Current opinion in psychology*, 31:1–6, 2020.

[74] Jun-Ming Xu, Benjamin Burchfiel, Xiaojin Zhu, and Amy Bellmore. An examination of regret in bullying tweets. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 697–702, 2013.

[75] Wenjing Xie and Cheeyoun Kang. See you, see me: Teenagers' self-disclosure and regret of posting on social network site. *Computers in Human Behavior*, 52:398–407, 2015.

[76] Rebecca A DiVerniero and Angela M Hosek. Students' perceptions and communicative management of instructors' online self–disclosure. *Communication Quarterly*, 59(4):428–449, 2011.

[77] Meifen Wei, Daniel W Russell, and Robyn A Zakalik. Adult attachment, social self-efficacy, self-disclosure, loneliness, and subsequent depression for freshman college students: A longitudinal study. *Journal of counseling psychology*, 52(4):602, 2005.

[78] Baiyun Chen and Justin Marcus. Students' self-presentation on facebook: An examination of personality and self-construal factors. *Computers in Human Behavior*, 28(6):2091–2099, 2012.

[79] Yongjun Sung, Jung-Ah Lee, Eunice Kim, and Sejung Marina Choi. Why we post selfies: Understanding motivations for posting pictures of oneself. *Personality and Individual Differences*, 97:260–265, 2016.

[80] C Akiti, Sarah Rajtmajer, and Anna Squicciarini. Contextual representation of selfdisclosure and supportiveness in short text. In *Proceedings of the 3rd Workshop on Affective Content Analysis@ AAAI (AffCon2020), New York, New York (February 2020)*.

[81] Lee Humphreys, Phillipa Gill, and Balachander Krishnamurthy. How much is too much? privacy issues on twitter. In *Conference of International Communication Association, Singapore*. Citeseer, 2010.

[82] Balachander Krishnamurthy and Craig E Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 7–12, 2009.

[83] Lee Humphreys, Phillipa Gill, and Balachander Krishnamurthy. Twitter: a content analysis of personal information. *Information, Communication & Society*, 17(7):843–857, 2014.

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

[85] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282, 2012.

[86] Myriam Munezero, Calkin Suero Montero, Erkki Sutinen, and John Pajunen. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE transactions on affective computing*, 5(2):101–111, 2014.

[87] Teagen Nabity-Grover, Christy MK Cheung, and Jason Bennett Thatcher. Inside out and outside in: How the covid-19 pandemic affects self-disclosure on social media. *International Journal of Information Management*, 55:102188, 2020.

[88] Koustuv Saha, John Torous, Eric D Caine, and Munmun De Choudhury. Social media reveals psychosocial effects of the covid-19 pandemic. *medRxiv*, 2020.

[89] S Wojcik and A Hughes. Sizing up twitter users. pew research center, 2019.

[90] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing*, 19(1):27–41, 2000.

[91] Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding privacy settings in facebook with an audience view. *UPSEC*, 8:1–8, 2008.

[92] TR Soron and MA Tarafder. The relation between facebook use pattern and demographic factors. *Journal of Psychiatry*, 18(6):1–5, 2015.

[93] Hai Liang, Fei Shen, and King-wa Fu. Privacy protection and self-disclosure across societies: A study of global twitter users. *new media & society*, 19(9):1476–1497, 2017.

[94] Taraneh Khazaei, Lu Xiao, Robert E Mercer, and Atif Khan. Understanding privacy dichotomy in twitter. In *Proceedings of the 29th on Hypertext and Social Media*, pages 156–164. 2018.

# A Survey Questions

We present the questions asked to the participants in the surveys. Some answer options are omitted due to space constraints.

## A.1 Screening Survey Questions

1. What is your gender?
2. What category below includes your age?
3. Which race/ethnicity best describes you?
4. In what state or U.S territory do you live in?
5. What is the highest level of school you have completed?
6. What is your combined total household income? (eg: 50000)
7. How many adults are in your household (including yourself)?
8. How many children are in your household?
9. How often do you use Twitter?
10. How often do you visit the Twitter website or use a Twitter app?
11. Do you use 'Protect Your Tweets' setting on Twitter?

## A.2 Entry Survey Questions

1. Please provide your Twitter user name (eg: '@johnsnow')
2. Please provide your email address. We will use this to follow-up with you and send out the compensation.
3. Which of the following categories best describes your current occupation?
4. Please enter your job title.
5. Please describe your job in one or two sentences.

## A.3 Exit Survey Questions

1. During the study period, how many tweets did you delete per week on average?
   - (0, 1-5, 6-10, 11-15, more than 16)
2. Please think about the most recent time when you deleted a tweet. What was it about?
3. Why did you delete the tweet you described in question 2?
4. Do you use login verification setting on Twitter?
   - (Yes, No, I don't know)
5. Please describe the reason.

6. During the study period, do you recall sharing any of the following?

   - Age (Yes, No, I don't know)
   - Gender (Yes, No, I don't know)
   - Ethnicity (Yes, No, I don't know)
   - Marital Status (Yes, No, I don't know)
   - Education (Yes, No, I don't know)
   - Employment Status (Yes, No, I don't know)
   - Income (Yes, No, I don't know)
   - Hobbies (Yes, No, I don't know)
   - TV Programs (Yes, No, I don't know)
   - Leisure Activities (Yes, No, I don't know)
   - Shopping / Purchasing Habit (Yes, No, I don't know)

   - Phone number (Yes, No, I don't know)
   - Postal code (Yes, No, I don't know)
   - Street Address (Yes, No, I don't know)
   - City (Yes, No, I don't know)
   - State (Yes, No, I don't know)
   - Email address (Yes, No, I don't know)
   - Workplace location (Yes, No, I don't know)
   - SSN (Yes, No, I don't know)
7. Please select any online social networks you use other than Twitter.
8. How often do you use any online social networks?

# B Statistical Results for RQ3

We demonstrate a mean percentage of disclosing tweets in each personal information category across predictors (see Table 9 in the next page). A One-way ANOVA was used for significance tests. Significance codes ( '***' 0.001 '**' 0.01 '*' 0.05) are marked on all of the values in a significance set.

**Table 9.** Mean percentage of disclosing tweets for RQ3

| Variable | | N | Birthday/Age | Race/Ethnicity | Gender | Marital Status | Education | Employment Status | Location | Emotion/Opinion | Interest |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Income | low (less than $45,000) | 30 | 0.234 | 0.049 | 0.207 | 0.464 | 0.456 | 0.301 | 0.504 | 26.899 | 1.516 |
| | middle ($45,000-149,999) | 60 | 0.235 | 0.048 | 0.245 | 0.188 | 0.315 | 1.585 | 0.181 | 22.877 | 1.493 |
| | high (more than $150,000) | 20 | 0.106 | 0 | 0.107 | 0.537 | 0.203 | 1.102 | 0.881 | 24.513 | 1.458 |
| Gender | Female | 57 | 0.201 | 0.052 | 0.262 | 0.42 | 0.367 | 0.715 | 0.425 | 27.937*** | 1.472 |
| | Male | 50 | 0.149 | 0.027 | 0.15 | 0.228 | 0.254 | 1.632 | 0.371 | 19.604*** | 1.429 |
| | Non-binary | 3 | 1.452 | 0 | 0.208 | 0.208 | 1.014 | 1.278 | 0.264 | 32.411*** | 2.969 |
| Age | 18-20 | 13 | 0.175 | 0.155 | 0.541 | 0.099 | 0.271 | 0.23 | 0.078 | 22.191 | 1.776 |
| | 21-29 | 24 | 0.271 | 0.057 | 0.213 | 0.154 | 0.434 | 0.33 | 0.033 | 24.484 | 1.583 |
| | 30-39 | 25 | 0.03 | 0.004 | 0.169 | 0.463 | 0.432 | 3.035 | 1.138 | 26.513 | 1.585 |
| | 40-49 | 15 | 0.181 | 0 | 0.091 | 0.471 | 0.332 | 1.5 | 0.167 | 25.621 | 1.419 |
| | 50-59 | 22 | 0.316 | 0.038 | 0.242 | 0.391 | 0.285 | 0.673 | 0.329 | 23.76 | 1.683 |
| | 60 or older | 11 | 0.369 | 0 | 0 | 0.338 | 0.059 | 0.192 | 0.325 | 20.352 | 0.473 |
| Education | Less than high school degree or equivalent | 7 | 0.48** | 0 | 0.397 | 0.626 | 0.357 | 0 | 0 | 30.72 | 1.865 |
| | Attending college | 22 | 0.489** | 0.154 | 0.237 | 0.208 | 0.361 | 0.404 | 0.112 | 24.778 | 1.574 |
| | Bachelor's degree / Associate degree | 43 | 0.084** | 0.008 | 0.26 | 0.304 | 0.164 | 1.548 | 0.111 | 22.523 | 1.337 |
| | Graduate degree | 38 | 0.145** | 0.015 | 0.102 | 0.366 | 0.504 | 1.335 | 0.956 | 24.768 | 1.555 |
| Race/Ethnicity | White / Caucasian | 87 | 0.203 | 0.021 | 0.202 | 0.354 | 0.358 | 1.244 | 0.47 | 24.65 | 1.55 |
| | Hispanic American | 6 | 0 | 0 | 0 | 0.158 | 0.559 | 1.011 | 0.136 | 23.427 | 1.658 |
| | Multiple Ethnicity | 14 | 0.389 | 0.162 | 0.144 | 0.29 | 0.091 | 0.821 | 0.134 | 23.361 | 1.306 |
| | Black or African American | 3 | 0.043 | 0.086 | 1.15 | 0.043 | 0.276 | 0.147 | 0 | 19.221 | 0.38 |
| Occupation | Unemployed | 16 | 0.208 | 0.218* | 0.623* | 0.465 | 0.515 | 0.087 | 0.12* | 24.687 | 1.168 |
| | Natural resources, construction, maintenance | 5 | 0.65 | 0* | 0.139* | 0.547 | 0.056 | 0.172 | 0* | 17.465 | 1.594 |
| | Sales and office | 7 | 0.307 | 0* | 0.067* | 0.034 | 0.147 | 0.175 | 2.53* | 20.239 | 1.264 |
| | Service | 11 | 0.159 | 0.023* | 0.072* | 0.195 | 0.226 | 0.331 | 0.163* | 26.3 | 1.723 |
| | Professional and related | 59 | 0.187 | 0.01* | 0.132* | 0.34 | 0.27 | 1.958 | 0.197* | 23.569 | 1.512 |
| | Management, business, financial operations | 12 | 0.145 | 0* | 0.277* | 0.278 | 0.724 | 0.294 | 0.876* | 30.496 | 1.711 |