Rajat Tandon*, Pithayuth Charnsethikul, Ishank Arora, Dhiraj Murthy, and Jelena Mirkovic*

I know what you did on Venmo: Discovering privacy leaks in mobile social payments

Abstract: Venmo is a US-based mobile social payments platform. Each Venmo transaction requires a "payment note", a brief memo. By default, these memos are visible to all other Venmo users. Using three data sets of Venmo transactions, which span 8 years and a total of 389 M transactions with over 22.5 M unique users, we quantify the extent of private data leaks from public transaction notes. To quantify the leaks, we develop a classification framework SENMO, that uses BERT and regular expressions to classify public transaction notes as sensitive or non-sensitive. We find that 41 M notes (10.5%) leak some sensitive information such as health condition, political orientation and drug/alcohol consumption involving 8.5 M (37.8%) users. We further find that users seek privacy by making their notes private, inconspicuous or cryptic. However, the large increase in Venmo's user base means that the number of users whose privacy is publicly exposed has grown substantially. Finally, the privacy of a user who transacts with a group on Venmo can be reduced or eliminated through the actions of other users. We find that this happens to around half of Alcoholics Anonymous, gambling and biker gang group members.

Our findings strongly suggest that public-by-default payment information puts many users at risk of unintended privacy leaks.

Keywords: Venmo, privacy, sensitive information

DOI 10.56553/popets-2022-0069

Received 2021-11-30; revised 2022-03-15; accepted 2022-03-16.

1 Introduction

Venmo [1] is a US mobile payment service, which allows a user to send or request payments from other registered users. Mobile payments are a convenient way to split a bill, pay one's share of rent, pay for some goods or services (e.g., a haircut), donate to a cause, etc. Venmo's popularity has steadily grown, exceeding 40 million users in 2019 [2]. Over the past two years, Venmo has doubled its revenue and tripled its annual payment volume [3].

Each Venmo transaction must be accompanied by a sender's note. By default these notes are public. Figure 1 shows anonymized snippets of a real user's public Venmo notes. Albeit short, these notes can be used to infer much private information. For example, the user has a child, James (fictitious name). The user and James went to Disneyland in May 2021. Anna Murphy (fictitious name) tutors James. James probably goes to Harvard Elementary school. The user probably has a sister Susan (fictitious name), because they split the cost of their parents' gifts. The user went to Flower Gardens Diving with Susan.

Public-by-default policies can create privacy risks to users. A study of five randomly chosen Venmo users [4] highlighted privacy leaks around drug use, relationship-related disputes, loan payment history, and eating habits, amongst other things. Like other public feeds on social platforms, Venmo's public feed likely helped to increase revenue, brand awareness, and helped support or even build community. Many users, however, want more privacy protections. In 2018, Mozilla-Ipsos polled [5] 1,009 Americans about their stance on public-by-default policy in payment apps – 77% were against it. Mozilla also delivered to Venmo a petition signed by more than 25,000 Americans urging them to change their public-by-default policy and the FTC investigated Venmo's privacy practices [6] and required changes in user privacy options. While Venmo has over the years provided more options for users to actively change their privacy settings, the default setting has not changed.

In this paper, we systematically quantify Venmo notes' privacy leaks with respect to sensitive information, such as drug and alcohol use, political orientation,

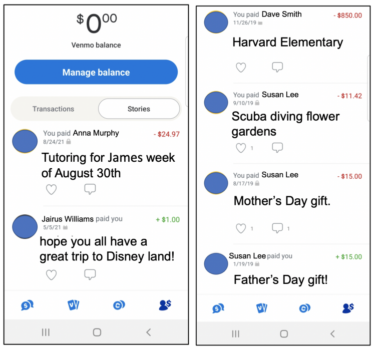
***Corresponding Author: Rajat Tandon:** University of Southern California Information Sciences Institute, Marina del Rey, CA, USA, E-mail: rajattan@usc.edu

Pithayuth Charnsethikul: University of Southern California Information Sciences Institute, Marina del Rey, CA, USA, E-mail: charnset@usc.edu

Ishank Arora: University of Texas, Austin, E-mail: ishankarora1100@utexas.edu

Dhiraj Murthy: University of Texas, Austin, E-mail: dhiraj.murthy@austin.utexas.edu

***Corresponding Author: Jelena Mirkovic:** University of Southern California Information Sciences Institute, Marina del Rey, CA, USA, E-mail: mirkovic@isi.edu



You paid Anna Murphy - $24.97
8/24/21
Tutoring for James week of August 30th

Jairus Williams paid you + $1.00
5/5/21
hope you all have a great trip to Disney land!

You paid Dave Smith - $850.00
11/26/19
Harvard Elementary

You paid Susan Lee - $11.42
9/10/19
Scuba diving flower gardens
1

You paid Susan Lee - $15.00
8/17/19
Mother's Day gift.
1 1

Susan Lee paid you + $15.00
1/19/19
Father's Day gift!

Fig. 1. An example of a user's Venmo notes.

adult activities, location, e-mail address, account numbers, etc. We also investigate users' privacy behaviors and discuss other risks of Venmo's public-by-default policy. To quantify the leaks, we develop a classification framework SENMO, that uses BERT [7] and regular expressions to classify public transaction notes as sensitive or non-sensitive. While we focus only on Venmo's practices, our findings are important for many other businesses that combine social network features with user transactions (e.g., fitness sites). While social features may help user engagement, they pose large risks to user privacy.

1.1 Contributions

Our main contribution is **quantification of privacy leaks on Venmo**. Towards this goal, we develop an accurate classifier for different types of sensitive user data in Venmo notes. We apply our classifier on Venmo data covering 8 years of transactions. We find that *41 M notes can be used to infer some private data, affecting millions of users*. Often, inferred data is about relationships (e.g., romantic, roommate, parental, service-oriented), buying or selling of drugs, gambling, adult activities and user location.

Our next contribution is **quantification of user actions to protect their privacy**. We find that *users increasingly adopt privacy-seeking behaviors*, either through making all their transactions non-public (25% of users had non-public profiles in 2013 versus 37% in 2018), or by making their notes cryptic (10% of users in 2013 vs 25% in 2018). On the other hand, the number of Venmo users grew 14-fold in the same time period. Therefore, in spite of users' efforts to protect their privacy, an increasing *number* of users are exposed to privacy risks.

We also **quantify privacy dangers of group payments on Venmo** through the lens of three select group types – Alcoholics Anonymous, gambling and biker gangs. We find thousands of groups in our select categories whose transactions are public. Slightly more than half of the users that issue payments to these groups post cryptic or inconspicuous content, seemingly seeking to protect their privacy. However, the notes of other users and sometimes the group's display name expose the sensitive nature of everyone's membership.

Finally, **we discuss other ways of leaking private data on Venmo, beyond notes, and provide recommendations for improvement.** In addition to public notes, Venmo's APIs and user profiles leak private information, and open users to scam and fraud attempts by others. We provide recommendations for users and Venmo, on improving their privacy practices. Venmo validated our findings about privacy leaks related to Venmo's APIs and user profiles, and has since fixed those issues.

2 Problem description

In this section, we discuss definitions of sensitive user data that should be protected and the negative effects of over-sharing and public-by-default policies.

2.1 Sensitive information

User privacy can be jeopardized if unauthorized entities can access the user's *sensitive* information. In many contexts the attacker must link some sensitive personal information (e.g, religious belief) about a user to the user's real identity (e.g., name, address, phone number, etc.) to do harm to the user. Laws and regulations around the world define *protected personally-identifiable information (PPI)* very broadly. Different tech companies provide their own definitions of what constitutes sensitive content, and encode them in their privacy policies or terms of service. Providers then create platform functionalities to protect sensitive data. Peddinti et al [8] found that data sensitivity is a complex measure, which should be viewed as continuous rather than binary.

Various laws and regulations around the world, and various tech companies define categories of information that are considered sensitive and that should be protected. We performed extensive review of this information, and for space reasons, we summarize it here. Table 14 and Table 15 in the Appendix summarize the types of information that are considered PPI by different laws and regulations (e.g., GDPR [9]), government agencies (e.g., DHS [10]) and tech companies (e.g., Google [11]). Sensitive categories include: race/ethnicity, adult content, sexual orientation, medical facts, religious/philosophical beliefs, trade union membership, political opinions, genetic/biometric information, personal financial information, personal idenifying information (e.g., e-mail, address, etc.), use of certain sensitive products and criminal history. Some definitions are very broad, while others are more specific. We start from these categories and from prior research on sensitive information [12–18]. Then, using the publicly available Venmo dataset D2 (described in Section 4), we manually narrow and refine these categories to those that frequently occur in Venmo notes. Our final list, shown in Table 1, has fourteen sensitive categories of data, which are shared often on Venmo. The categories capture sensitive information about a user's health, lifestyle, sexual orientation, race/ethnicity, political orientation, adult/criminal/violent behavior, alcohol and drugs consumption, location, relationship and family status, account numbers, e-mail addresses and phone numbers, physical addresses, and product/activity details.

To link sensitive information to user identity on Venmo, one can look up a user based on their name or phone number. Many users also include their photo in the profile. Thus, one can easily identify a user on Venmo [19].

2.2 Risks of oversharing

Oversharing one's personal information online can be dangerous. Sometimes the information alone can incriminate the user or put them in jeopardy. In other cases, the information that is shared can be combined with other, publicly or privately accessible information to harm a user. We provide some examples below.

Criminal investigation. In 2015, public Venmo posts of a Columbia University student were used as supporting evidence to arrest him under drug charges [20]. The Egyptian police created fake profiles on the dating app Grindr to bait and then arrest LGBTQ people [21]. U.S. Representative Matt Gaetz's Venmo transactions were used as evidence in an investigation of his adult relationship with three minors [22].

Theft. In 2019, robbers used social media to stake out houses in upscale Houston, Texas neighborhoods [23]. Robbers leveraged house owners' posts about vacation, work and parties to plan their theft for times when owners were away.

Health benefit loss. In 2016, a woman lost her disability pension for bragging about her active lifestyle on Facebook [24].

Job and opportunity loss. Businesses vet job [25] and colleges vet applicants using public social media profiles [26, 27]. 70% of employers screen candidates on social media during hiring, and 43% continue to monitor the social media activities of current employees [28].

Miss Florida USA 2017 was stripped of her title just six days after winning the beauty pageant [29], because her social media posts revealed that she violated the rules by using professional hair and makeup artists.

National security risks. U.S. President Joe Biden's Venmo account was discovered using the app's search tool [19]. While the President's transactions were private, Venmo at the time had no way for users to make their friends list private. Journalists were thus able to identify Biden's family members and senior White House officials [19]. As per the article [19], due to national security risks, Biden's Venmo account has been deleted.

Identity theft and financial scams. Active social media users are 30% more likely to be affected by **identity fraud** [30, 31]. In an incident from Nashville, Tennessee [32], family members and friends of a Venmo user, who was out-of-town, received Venmo payment requests. Attackers leveraged public Venmo note and friend information to set up a fake account with the user's picture and ask user's contacts for money. Scammers have also impersonated Venmo employees and sent phishing requests (see section 6.3).

Emotional effects. Effects of oversharing can cause social embarrassment, relationship problems, and regret [33–35]. For example, a student's inadvertent outing of sexual orientation on Facebook resulted in threats to cut off family ties [36].

2.3 Public-by-default and user recourse

Many social media set posts as public by default. Users are usually able to alter this setting to protect their privacy. A Venmo account is by default set to public, making transactions visible to everyone on the Internet. Users can either remain at the default setting or make their transactions visible only to friends (friends-only) or only to the other party in their transactions (private) [37]. A Venmo user can choose to change the visibility only of a specific note (at transaction issue time), or of all future notes, or of all past and future notes. While a user can choose stricter privacy settings than default on many social platforms, many users do not make these changes, because they may be unaware of the privacy options or confused by them.

Users want strong privacy. Hoofnagle et. al [38] show that users support the idea of strong privacy settings by default.

Users are unaware of privacy settings. Zhou et. al [39] show that users fail to use privacy features, because they are not aware of the features and how each feature would protect their privacy.

Users are confused with privacy settings. Users tend to either keep all their Facebook posts public or all private, since that is easier than post-level settings [40]. A user with all public settings could therefore inadvertently share something with a public audience. Similarly, users can be confused by Facebook's privacy settings and do not effectively make use of them [41]. Google also acknowledged that its users faced similar challenges [42]. Sleeper et al. [43] found that users may refuse to share because they cannot explicitly select their target audience on the social platform. Stenros et al. [44] found that users view privacy and sharing as a continuum, while platforms offer discrete choices.

A Venmo user that desires greater privacy, but is unaware of, or confused by Venmo's privacy settings, can post a note that is *inconspicuous*, such as a quote from a book or a note misstating the transaction's nature. A note may also be *cryptic*, such as consisting only of emojis, random letters or words, or interjections. We call the users that post inconspicuous or cryptic notes *privacy-seeking* users.

A Venmo user may also want to hide other details of their profile, such as their full name, their friend list or even that they are a registered user. The user can change their name at any time, but they cannot hide the fact that they are registered user on Venmo. Since June 2021, user can also change visibility of their friend list, but this change did not take effect on all Venmo UIs. We provide more details in Section 6.3.

Venmo users may also suffer privacy loss from their group transactions. A leader of a group (e.g., a club) can use their Venmo account as a convenient way to collect donations. Membership in some groups may be sensitive for a user (e.g., Alcoholics Anonymous, biker gangs, gambling pools). When a group's transactions are public on Venmo, this jeopardizes privacy of all users that interact with the group via public transactions. In some cases the group's display name reveals its purpose, and in other cases, transaction notes of other users can reveal the group's purpose. Once a group's purpose is revealed, all users with public transactions to that group are revealed as members.

3 Research questions and challenges

We study the following research questions: (RQ1) how many users risk their privacy by posting sensitive notes on Venmo, (RQ2) what do users do to protect their privacy on Venmo, (RQ3) how trends in user behavior change over time, (RQ4) do certain user behaviors open them more to privacy risks than others. Answering all these questions requires us to first define what we mean by "sensitive" information in Venmo's notes.

We consider the following types of *sensitive* information that can be embarrassing or harmful to users if shared publicly: (ADU) *adult*, such as talking about various types of sex or using sex-related jargon, (LGB) *LGBTQ*, such as referencing sexual orientations lesbian, gay, bisexual, and transgender, (HEA) *health*, such as referencing a particular disease, doctor or test, (DAG) *drugs/alcohol/gambling*, such as referencing a specific drug or game of chance (e.g., poker), (POL) *political opinions*, especially around US politics, (RET) *race/ethnicity*, such as posting content that may either reveal the user's membership in or feelings toward a given racial or ethnic group, (VCR) *violence/crime*, such as posting information about illegal activities or violence, (REL) *relationships*, such as sharing a house with another user, relationship status, or family information, (LOC) *location*, such as the businesses, cities or countries that the users visit, (ACC) *account details*, such as account ID, username or password for a non-Venmo account, (EMA) *e-mail address*, (PHO) *phone number*, (ADD) *address* and (PAD)*product/activity details, such as tracking number or invoice number, which may directly or indirectly leak personal information (e.g: finance, location, activities, products purchased etc.).* Notes that do not fall into any of our sensitive categories are classified as (NON) *non-sensitive*. Our categories and anonymized examples of notes are given in Table 1.

3.1 Challenges

Establishing if a Venmo note has a given type of sensitive information about the user or not is challenging. First, many notes are short. Around 93% of notes contain up to 5 words, 99% contain up to 10 words, and 99.9% contain up to 30 words. Classifying short messages is challenging, since they may lack context. For example "bar" may mean a bar where drinks are served or passing the bar exam. We address this challenge by

Category ID	Category name	Example
	Sensitive	
ADU	Adult	"sexual pleasures"
LGB	LGBTQ	"gay rights activist"
HEA	Health	"For aids treatment. Get well soon"
DAG	Drugs/alcohol/ gambling	"for the weed that we gon smoke"
POL	Political opinions	"Bush did 9/11"
RET	Race/ethnicity	"Acting like a black man!!"
VCR	Violence/crime	"Aggravated assault in an uber"
REL	Relations	"Your half of the divorce"
LOC	Location	"Train Rome to Salerno"
ACC	Account details	"[Name] man thank you 4 everything. The password to my Bank account is [Password], take what u want"
EMA	E-mail address	"Send it to my PayPal [Email]@gmail.com"
PHO	Phone number	"Call me [Phone number]"
ADD	Address	"August 2018 rent for [Unit, Street], Omaha NE 68142."
PAD	Product/activity details	"tracking: ups [Full tracking number]"
	Default	
NON	Non-sensitive	"hair styling for photo shoot"

Table 1. Categories of data we consider.

field	meaning
story ID	unique post ID, alphanumeric
payment ID	unique transaction ID, numeric
tr. note	user-specified, may leak personal/sensitive info
time	time when transaction occured
likes comments mentions	posted by other users for the transaction may leak personal/sensitive info not investigated in this paper
sender info receiver info	first and last name **display name**, date account created link to profile pic, **user ID**, joining date, external ID
last update	time of last app update
status	settled, pending or canceled
type	pay or charge

Table 2. Fields in the transaction record, those that we use are shown in boldface.

Name	Time period	Users	Transactions	Type
D1	03/2012 – 04/2018	22.5 M	338 M	complete
D2	07/2018, 08/2018, 10/2018	7.1 M	6.9 M	sampled
D3	01/2020 – 12/2020	13.5 M	43.7 M	sampled
D4	05/2021	681	11.5 K	limited-complete

Table 3. Our datasets.

being conservative with our sensitive label. If a note can be interpreted as non-sensitive (e.g., "Chinese" can relate to ethnicity or to food) we label it as NON.

Second, many notes contain colloquial expressions. We address this challenge by using BERT [7] for our sensitive note classification and we fine-tune it on a manually labeled fraction of our dataset to achieve good accuracy.

Third, long notes are often copied from unrelated public text (i.e., inconspicuous notes introduced in Section 2). Since, 99.9% of the Venmo notes across all our datasets contain up to 30 words, we label notes longer than 30 words as non-sensitive (NON).

Fourth, around 30% of Venmo transaction notes contain emojis, and around 25% consist only of emojis. While some emojis could be used to infer a user's intent (e.g., a heart may mean that the sender loves the receiver), this inference is highly speculative. We ignore emoticons in our note classification.

Fifth, a transaction note may relate to multiple categories, such as "queer asian", which may belong to both LGB and RET. We address this challenge by training our BERT classifier to perform multi-label classification. Overall, our conservative approach to classification means that our results likely undercount sensitive notes on Venmo, as well as affected users.

Finally, our datasets are limited and only capture public notes for a given time period. Some datasets are also *sampled* – they capture a random sample of public notes. We address this challenge by limiting our longitudinal study to those datasets that are not sampled, and inferring private notes and users with private profiles from global transaction and user identifiers on Venmo.

4 Datasets

We work with four datasets, containing public Venmo transactions, summarized in Table 3. Datasets D1–D3 were compiled using Venmo's public API (https://venmo.com/api/v5/public) to collect data over long time periods, since the collection speed is limited by Venmo. We collected D1 and D3 datasets via use of automated scripts. These datasets are stored at our institution. Dataset D2 was collected and shared publicly by Dan Salmon [45]. Dataset D4 was scraped by us using Selenium [46].

Dataset D1 covers the period of six years, ending in 2018. This dataset is *complete* – it has all public transactions in the given time period.

Dataset D2 is publicly available dataset obtained from Github [45], which was scraped by Dan Salmon [47]. This dataset covers three months in 2018, and is *sampled* – it has some, but not all public transactions from the given time period.

Dataset D3 covers the entire year 2020[1]. This dataset is *sampled* – it has some, but not all public transactions from the given time period.

[1] November and December of 2020 have fewer transactions than the other months, because Venmo limited the rate at which users were allowed to access its public APIs.

Dataset D4 covers only public transactions from selected sensitive groups: Alcoholics Anonymous (AA), gambling (G) and biker gang (BG) groups. We crawled Venmo using a Selenium-based [46] script and keywords related to AA, G and BG shown in Tables 9, 10 and 12, respectively. Our script used Venmo's search functionality to gather transactions of users whose user ID or display name contain these keywords of interest. This dataset is *limited-complete* – it contains all public transactions, but only for selected groups. We use this dataset only for calibration of our group identification heuristic (Section 5.4).

These datasets are all disjoint, and together cover the period of 8 years of Venmo transactions. Dataset D1 is the largest, and complete, thus it helps us measure longitudinal trends. However, it may be considered outdated since it ends in 2018. We complement it with dataset D3, collected in 2020, to verify that the trends we observed still hold. We include dataset D2, because it was publicly released, and it is disjoint from D1 and D3 (although sampled and from 2018). We use it simply as further confirmation of trends we observe.

Dataset D4 is different than others, because it focuses only on group transactions and let us answer research questions about user privacy when they pay to a group via Venmo.

4.1 Protecting user privacy

Each transaction in our datasets includes the fields shown in Table 2. Even though all data we analyze is public on Venmo, we take additional measures in our analysis to protect user privacy. We perform our analysis on a reduced set of columns, including only (a) user IDs of the transacting parties, (b) transaction notes, and (c) transaction date. These fields are shown in boldface in Table 2. We protect data privacy in the following ways: (1) we store the data at our institution's servers, which can only be accessed by authorized personnel, (2) we process a small fraction of D2 and D4 manually (0.2%), which is necessary to label the data and fine-tune our classifier, (3) the rest of the data (99.8%) is processed via automated scripts, and only aggregated, anonymous results are viewed by study personnel.

This study was fully reviewed by our IRB, and then approved under the exempt category. With regard to the ethics of using publicly available data, we see our study akin to the many published studies which use publicly available Twitter data. Moreover, we minimize risk to users by analyzing aggregated data [48, 49]. Without studies of publicly available data from social platforms, scholars are unable to uncover and highlight data privacy issues. As Fiesler and Proferes [50] show that many social media users do not want researchers to use their public social posts without consent, we did not target any individual users in our analysis. Where we needed to access individual notes (e.g., during labeling for training and testing of our classifier), we limited ourselves to a small sample of notes, and worked with note text only, omitting user identity. We have further attempted to contact Venmo through multiple channels, to obtain their permission for the study. We contacted Venmo directly via their support team and received an acknowledgement that our message was received, but there was no further response. We also filed a bug report for Venmo via HackerOne [51] vulnerability coordination and bug bounty platform, but it was closed as out of scope.

5 Methodology

In this paper, we aim to systematically quantify Venmo notes' privacy leaks, with respect to sensitive categories listed in Table 1. Our analysis of sensitive information leakage (research questions RQ1 and RQ3) focuses on English-language notes (including notes that contain only named entities).

We develop a classification framework, called SENMO (SENsitive content on venMO), which classifies a transaction note as one or more of the sensitive categories from Table 1. A note could also be classified as NON (non-sensitive), if it does not contain any sensitive information.

SENMO comprises the pre-processing module and the semantic engine. The pre-processing module cleans up and normalizes the notes, and the semantic engine classifies them into one or more categories using regular expressions and a machine-learning classifier. Figure 2 shows SENMO's architecture, and note classification process. All notes pass through the regular expression part of the semantic engine, and in parallel all notes are fed to the pre-processing module. The pre-processing module cleans up the notes, filters out some that are considered non-sensitive and feeds the rest to the ML classifier. The classifier outputs one or multiple classes for each note. The outputs of regular expression and classifier modules are combined, so that a note is only classified as non-sensitive if both modules output that class. Otherwise, the note's class is the union of all sensitive categories identified by the modules.

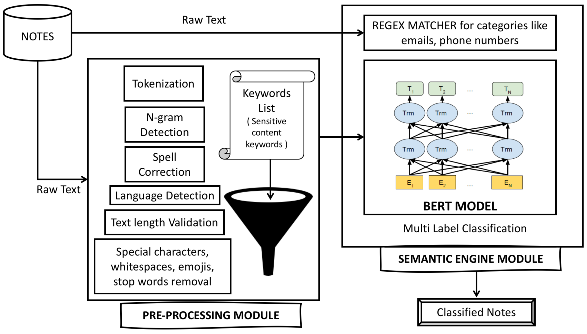


Fig. 2. Architecture diagram of SENMO.

5.1 Pre-processing module

SENMO's pre-processing module first removes any characters that are not in the English alphabet and are not digits. Next, we use the NLTK tokenizer [52] to produce words, bi-grams and tri-grams.

Our pre-processing removes all notes longer than 30 words, which is 0.1% of notes, and labels them as non-sensitive (NON). SENMO then uses Python's *enchant* library's dictionary to detect if a note is in English. We also check if the note contains a named entity (e.g., "Chicago"), using the NLTK Named Entity Recognizer (NER) [52]. Notes that pass either check are kept for further processing and we refer to them as "English notes" for short. We spell check those notes that do not match either check, and attempt to match them again (to English words or named entities). Our spell checking is simple, but accurate and efficient. We replace three or more consecutive characters with either two consecutive characters or a single character, e.g., "Phoenixxx" becomes "Phoenixx" and "Phoenix". If either version of the corrected word exists in the English dictionary or is identified as a named entity, we keep that correction and proceed further with the pre-processing module. Otherwise, we keep the original version of the word. Notes that do not match any English word or named entity, both before and after spell correction, are discarded.

Finally, the pre-processing module removes stopwords, converts all characters to lowercase and uses lists of keywords (described in Section 5.1) to identify notes that are likely sensitive. Such notes are passed to the semantic engine, and the rest of the notes are labeled as non-sensitive. We use the keywords to pre-filter likely sensitive notes, to improve classification accuracy for NON category. Without this step, due to the brevity of notes and a small size of our training set, our ML classifier can learn to associate categories with non-specific words that often occur in the given category. For example, "appointment" may become associated with HEA since it occurs in notes like "appointment for chemotherapy" , even though it can appear in other categories such as "haircut appointment". Similarly, "cream" may become associated with HEA because it occurs in notes like "hemorrhoid cream", while it can also occur in notes that are related to beauty products such as "face moisturizer cream", "hand cream" or "face cream".

Keyword selection: We build the list of keywords associated with sensitive content for different categories using various popular sources and prior published works, shown in Table 11 in the Appendix. We release this list of keywords as open source [53]. For example, for keywords related to ADU and LGB categories, we use reputable sources (e.g., [54–56]). As another example, we utilize the American Medical Association's glossary to build our keywords list for HEA. Similarly, we identify drug- and alcohol-related keywords using the vocabularies from popular sources such as National Institute on Drug Abuse [57] and National Institute on Alcohol Abuse and Alcoholism [58], respectively. For POL category, we restrict our keywords list to US politics, since Venmo is a US-based social payments platform. Details on our sources are shown in the Appendix. We further enrich our lexicons for different categories using the keyword lists from Wang et al. [18], which were used for tweet classification as sensitive or non-sensitive.

5.2 Semantic engine

SENMO's semantic engine module applies different classifiers to classify each note into one or more categories. For account, E-mail, phone number, address and product/activity details, the engine utilizes regular-expressions *on raw notes, without any pre-processing*, shown in Figure 7 in the Appendix. These content categories often include well-structured text that can be identified via regular expressions.

For the rest of the sensitive categories, and for the NON category, we train our classifier on our balanced training set (described below). We validate the classifier's accuracy on a smaller,balanced evaluation set and tune its hyperparameters. We then apply the trained classifier to the rest of our datasets. We use the Bidirectional Encoder Representations from Transformers (BERT) [7] for our classifier as it is designed to pretrain deep bidirectional representations from unlabeled text. Learning a model using BERT framework involves two steps: pre-training and fine-tuning. During the pre-training step, the model is trained on large, unlabeled data over different pre-training tasks. We use the English-language BERT$_{base}$ uncased model pre-trained on extremely large corpora for years (BooksCorpus (800 M words) [59] and English Wikipedia (2,500 M words)). We manually create two balanced, labeled datasets to *fine-tune* BERT to our custom domain (VENMO note classification) and to *evaluate* the model's classification accuracy.

Fine-tuning set: We create this set from the dataset D2, after it has undergone the pre-processing step. Three co-authors annotated this set. One annotator manually identified about 1,000 notes per category LGB, ADU, HEA, DAG, POL, RET, VCR, REL, LOC, NON, and labeled them. The manual identification process involved randomly selecting notes from the dataset D2 and deciding if they fall into any of the above categories or not. Though a labor-intensive process, our intention was to avoid biases that may have occurred if we used our keywords list to find such notes. The second annotator then independently assigned labels to the same set of notes. Notes with discrepancies were resolved first between annotators. In case this failed, the third annotator was asked to adjudicate. The pre-adjudicated Cohen's Kappa coefficient for the first two annotators was 0.95, indicating strong interannotator agreement. The final fine-tuning set has about 10.1 K notes. Out of these, 135 notes (1.3%), have more than one label.

Evaluation set: We create this set by removing the fine-tuning set from the dataset D2, and then selecting notes at random and labeling them, until we have about 100 notes per category LGB, ADU, HEA, DAG, POL, RET, VCR, REL, LOC and NON. Again three annotators participate in labeling, using the same process used for the fine-tuning set. 99% of the notes from our evaluation set with sensitive labels, matched at least one keyword from our keyword list (Section 5.1) corresponding to the label's category. This shows that our keyword lists are comprehensive. The pre-adjudicated Cohen's Kappa value for the first two annotators was 0.92. 64 notes (6%) have more than one label. Please note that, although we ensure no duplicate entries between the fine-tuning set and the evaluation set, some duplicate entries exist within these individual datasets. Since Venmo notes are usually short, multiple users post the same notes with high probability. D2 dataset has 62% duplicate notes. Examples of such notes are “Lung cancer”, “For bailing me out of jail” and “George bush doesn't care about black people”. Our classifier only applies to Venmo, since the fine-tuning set and the evaluation set reflect the composition and presence of duplicates in the original Venmo notes. We make no claims about the usefulness of this classifier to classify other sensitive content on other platforms.

During fine-tuning, each note's labels are encoded. We use only eight out of the nine categories: the notes in NON category are encoded as zeros in all categories. This approach works better than encoding NON as a separate category, because its notes are very diverse and BERT struggles to learn how to classify them. We evaluate our BERT model on the evaluation dataset, and compare its accuracy with the accuracy of three other approaches: (BoW-kw) bag-of-words using our keyword sets, (BOW-NB) Naive Bayes on bag-of-words, (TF-IDF-NB) Naive Bayes on TF-IDF. These approaches were used to accurately classify short posts on Twitter by Wang et al. [18]. We evaluate three flavors of our BERT model: (SENMO-npre) BERT without keyword pre-filtering, (SENMO-NONE) BERT with NON as a separate category in fine-tuning step and (SENMO) our chosen model – BERT with keyword prefiltering, and without NON as a separate category. Our results are shown in Table 13 in the Appendix and summarized here. SENMO classifies every note in the evaluation dataset with an aggregate per-note accuracy of 90%, which is higher than competing approaches: 73% for BoW-kw, 83% for BoW-NB and 76% for TF-IDF-NB. Further, SENMO without explicit NON category outperforms SENMO-NONE, which includes the NON category, by 4% in overall per-note accuracy. SENMO and SENMO-npre have comparable performance (90% in aggregate per-note accuracy), but SENMO is better in classifying notes in NON category (95% vs 87%).

Model sensitivity: To evaluate model sensitivity, we validate the three training hyper-parameters: batch size, learning rate and number of epochs. Results for these experiments are summarized, due to space reasons. There are many settings that achieve high accuracy, e.g., learning rate of $3 + 10^{-5}$ or lower, with more than 2 epochs, and both 16 and 32 batch sizes. In evaluation we use batch size of 32, $2 + 10^{-5}$ learning rate and 6 epochs, because this setting yields high accuracy for every category.

Spell checking: We measured how our spell-checker outputs, fed to our BERT model, compare with four other spell checkers: TextBlob [60], Pyspellchecker [61], Autocorrect [62] and JamSpell (free version) [63]. For space, we summarize these results. Our approach has best per-note accuracy (tied with Pyspellchecker) of 90%. This is because misspelling in Venmo notes is intentional, and includes character repetitions, and our spell-checker is tailored towards correcting this specific type of misspelling. Moreover, our spell-checker is much faster than other approaches (2–4 orders of magnitude), which enables us to process our large datasets.

5.3 Measuring user recourse

A user with the default Venmo setting (public transaction notes) may become more privacy aware over time (research question RQ2). Such a user may take two types of action. First, they may change their privacy settings to make their notes non-public (private or friends-only). We refer to a user that makes all their notes non-public as "non-public user". Second, users who may not know how to change their privacy settings, may instead create inconspicuous or cryptic notes. Since non-public transactions and non-public users are not visible to us, we have to infer their count from a dataset that has all public transactions from a given time period (in our case, only the D1 dataset is complete). To infer non-public transactions and users we use the approach proposed by Zhang et al. [64], which confirms that Venmo uses sequential numbers for user ID and payment ID fields. We verified that Venmo still uses this practice. Thus, based on user ID progression over time we can infer the number of all users/transactions, while a complete dataset contains all public users/transactions.

A user that does not know how to make their transactions non-public, may post inconspicuous or cryptic notes. Since we do not know the actual purpose of each transaction, we cannot identify inconspicuous notes. We define **cryptic notes** as notes that have one of the following patterns: (1) contain only emojis , (2) contain only random numbers that do not match our regex patterns, (3) contain only English interjections and greetings (e.g. "Hi", "Hey", "Aww"), (4) Contain only English stop words (e.g. "a", "the", "too"), (5) use English letters, but do not contain a vowel and (6) longer than 30 words. We measure cryptic notes before our preprocessing stage.

5.4 Measuring risks from group transactions

Venmo is not only popular for user-to-user transactions, but also as a means of splitting costs or collecting dues in a group, such as a fraternity/sorority, club, school, etc. Membership in some groups may be considered *sensitive* – it may pose privacy risk to a user, if it is publicly known. One such example are Alcoholics Anonymous (AA) groups, which go to great lengths to keep their membership private. To answer research question RQ4 we analyze payments to groups on Venmo. We study privacy protections of users that make payments to a group on Venmo, i.e., in our context the user is the sender in a transaction, and the group is the recipient. If a user's payments to the group (e.g., "AALongBeach") are public, but the notes contain just a generic word, like "hi", the observer can infer that the user is a member because the name of the group contains "AA" or because other people that make payments to the group, post notes using AA-specific phrases, like "7th tradition." We focus on three types of groups in this paper, whose memberships may be sensitive for many users: Alcoholics Anonymous, gambling and biker gang groups. We look for groups that may have many members and whose memberships may be sensitive. We select groups with sufficient samples to analyze and that comprise group-specific vocabularies. We apply a keyword/activity heuristic to identify sensitive groups consisting of three steps: (1) identification of candidate sensitive groups, (2) pruning of low-activity groups and (3) pruning of unrelated groups.

Identifying candidate sensitive groups. We identify candidate sensitive groups as: (1) users whose user ID or display name are sensitive, or (2) users whose *recipient notes* (notes in transactions where the given user is the recipient) are sensitive. We say that a string (a user ID, a display name or a note) is sensitive if it contains any of the *sensitive* keywords. We manually build lists of topic-specific sensitive keywords ("sensitive keywords" for short). Alcoholics Anonymous keywords are shown in Table 9 in the Appendix, and are mined from the names of AA meetings, which are publicly available via a Google Search. Gambling keywords are shown in Table 10 in the Appendix, and are the words directly related to the gambling activity. Biker gang keywords are shown in Table 12 in the Appendix, and include the names of different biker gangs [65, 66]. In addition to those sensitive keywords, we have manually identified from our D4 dataset some common patterns in recipient notes for AA (e.g., "donation", "thanks"), gambling groups (e.g., "money", "refund") and biker gang groups (e.g., "bike", "gear"). These common patterns are also topic-specific (see Tables 9, 10 and 12.).

The presence of sensitive keywords is necessary, but not sufficient to identify sensitive groups. That is because these keywords could sporadically appear in notes unrelated to AA, gambling or biker gangs. For example "7th" could appear in "Happy 7th birthday" or "I placed 7th in math competition". To reduce false positives from keyword-based classification, we apply the next two filters to prune low-activity and unrelated groups.

Pruning of low-activity groups. We focus only on groups that are more active than majority of Venmo's users, i.e., they are in 75th percentile of Venmo users by the number of transactions they receive and the number of users that send these transactions.

Pruning unrelated groups. Out of high-activity groups, we prune those that have low incidence of sensitive posts or common patterns for the given sensitive topic. Our manual examination shows that such groups often interleave group payments with personal user payments, and this can pollute our analysis. We use a threshold on the percentage of sensitive/common-pattern recipient notes, and tune this threshold to achieve low false positives on D4 dataset, which we label manually with ground truth. Manual labeling of D4 dataset is performed by the same three annotators that have annotated the fine-tuning set and the evaluation set (Section 5.2). Overall they have labeled 41 groups from D4 dataset with ground truth (”AA”, ”biker gang”, ”gambling”). Threshold values of 20% and above result in 100% true positives and zero false positives (see Appendix, Table 8). We then apply this calibrated group identification process to D1–D3 datasets and report the results. While our keywords are specific to each topic of interest (AA, gambling and biker gangs), our group identification heuristic is general, and could be applied to identify other topics of interest on Venmo.

5.5 Limitations

Limitations stem from the type of data available (short transaction notes) and from the fact that data is passively collected (i.e., without interaction with users). First, we use keyword lists to pre-filter notes of interest. While we tried to be comprehensive in list creation, it is possible that we have missed some keywords, which would lead to false negatives. Another limitation is that our regular expressions for ADD and PHO category are geared towards US addresses and phone numbers.

Another limitation is that we only analyze English-language notes for research questions RQ1 and RQ3, and may miss some types and trends of data-sharing in other languages. English notes make up 62%-69% of all notes in our datasets. Our findings still apply to most, but not all Venmo notes. Further, since our datasets are passively collected, we have no way to ascertain if a user's note represents truth or is a joke or fake information. However, when the notes contain potentially sensitive information, even in jest, they can bring harm to the sender and/or the recipient. For example, if the note implies that the sender is paying for drugs, then regardless of veracity of the note, the sender's public reputation may suffer, e.g., during a job interview. Lastly, some notes that we classify as non-sensitive (e.g., food) may be sensitive to some users (e.g., a user trying to lose weight). Further research, involving direct user surveys, would be needed to estimate how often user notes are true, and how often our inferred sensitive categories are actually sensitive to a given user.

6 Results

In this section, we present our evaluation of privacy leakage from user notes and user recourse (Section 6.1), privacy leakage from group activities (Section 5.4) and other privacy and security concerns about Venmo's user data policy (Section 6.3).

Figure 3(a) illustrates the total number of notes per month in our datasets. Between 62% and 69% of all notes were in English (not shown in the Figure). Our analysis therefore covers the majority of Venmo notes.

6.1 Privacy leakage from user notes

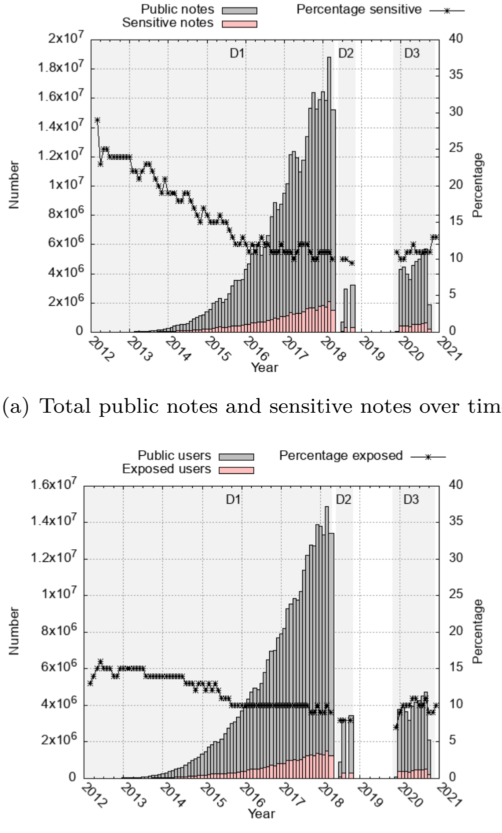
A sensitive note could affect only its sender (e.g., “paying for my hangover”) or both sender and receiver (e.g., “thanks for babysitting”). Since we cannot tell which party will be affected, we conservatively assume that only the sender is affected. Our results can thus underestimate the number and percentage of affected users.

Privacy leakage affects many transaction notes and users. From our dataset D1, we find that 35.8 M out of 338 M notes are sensitive (10.6%). Despite a low percentage of sensitive notes, many users are still affected: 37.8% of the D1 users (8.5 M out of 22.5 M). Similar findings apply to the D2 and D3 datasets, and are summarized in Table 4. The percentage of sensitive transactions is stable across these datasets, while the percentage of affected users increases from the smallest (D2) to the largest dataset (D1) (see Figure 3(b)). This is expected as observations across longer time periods are more likely to observe a user post a sensitive note.

Privacy leakage increases over time, in spite of user measures to contain it. We analyze two indicators to investigate whether users become more concerned about their privacy and modify their privacy settings over time. First, we analyze the number and percentage of users whose profiles were public, but

Metric	Total	Sensitive
D1: Notes	338 M	35.8 M (10.6%)
D1: Users	22.5 M	8.5 M (37.8%)
D2: Notes	6.9 M	708 K (10.3%)
D2: Users	7.1 M	640 K (9%)
D3: Notes	43.7 M	4.6 M (10.5%)
D3: Users	13.5 M	2.4 M (17.8%)

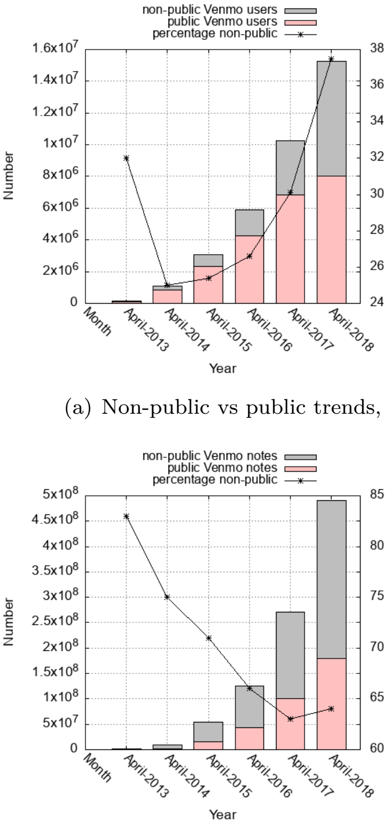
Table 4. Sensitive notes and affected users in our three datasets



(a) Total public notes and sensitive notes over time

(b) Exposed users over time

Fig. 3. Sensitive notes and users that post them, trends over time



(a) Non-public vs public trends, users

(b) Non-public vs public trends, notes

Fig. 4. Public and non-public user profiles and total notes from April 2013 to April 2018.

Dataset	Alcoholics Anonymous (AA)	Gambling (G)	Biker Gangs (BG)
D1	9.3 K	5.8 K	14
D2	26	1.7 K	2
D3	5.1 K	8.7 K	17

Table 5. Number of Alcoholics Anonymous (AA) and Gambling (G) groups identified.

are currently inaccessible. This may imply that the user changed their privacy settings to "private" or "friends-only", or that they closed their account. An increase in the number of privacy concerned users would lead to this indicator increasing. Second, we analyze the number and percentage of notes that are classified as sensitive over time. Here, an increase in the number of privacy concerned users would lead to percentage of sensitive notes decreasing.

Since non-public transactions are not visible to us, we apply the inference mechanism described in Section 5.3, on our only complete dataset, D1, to estimate non-public transactions and non-public users. Figure 4(a) shows the number and percentage of users that had public vs non-public transactions in years 2013–2018, and Figure 4(b) shows the same metrics for transactions.

In the first year of Venmo's operations, the percentage of users with only non-public transactions (shown as line in Figure 4(a)) decreased from 32% to 25% due to an influx of new users. Over the next four years, though, it has steadily climbed from 25% in 2014 to 37.5% in 2018. This indicates that users are increasingly concerned about their privacy and take action to make their transactions non-public. But a competing trend is a large rise in Venmo's user population. Thus, even though the percentage of users with public profiles declines, the *sheer number* of users with privacy risks rises over time (gray area in Figure 4(a)).

Note privacy shows a different trend than user privacy. This is probably due to influx of new users, and increase of user activity over time. Figure 4(b) shows the number (bars) and percentage (line) of non-public notes from 2015 to 2018. Over time the percentage of non-public notes decreased from 83% in 2013 to 64% in 2018, while the annual number of transactions increased more than 400 times.

Comparing the D1 and D3 datasets, we also find that 11.8 M users out of 22.5 M users (52%) that posted public notes in 2018, still post public notes in 2020. Posting public notes over longer time periods increases a user's privacy risk.

A user could post public transaction notes, but not post any sensitive information. Figures 3(a) and 3(b) show how notes and affected users changed over time. Drawing from longitudinal trends on the complete D1 dataset, the percentage of sensitive notes declined from around 25% in 2013 to around 10% in 2018. Similarly, the percentage of users exposed due to sharing sensitive content declined from 15% in 2013 to 10% in 2018. Users are thus increasingly trying to protect their privacy. However, increases in number transactions and users lead to a greater number of users affected by privacy leaks.

Another way to quantify user efforts to protect privacy is to measure prevalence of *cryptic* notes. We identify these notes using methodology described in Section 5.3. Figure 6(b) shows the count and percentage of cryptic notes over time, and Figure 6(a) shows the count and percentage of public users that post at least one cryptic transaction. Using the D1 dataset, cryptic notes increase from around 15% in 2013 to 45% in 2018. Similarly, users that posted at least one cryptic note increased from around 10% in 2013 to almost 30% in 2018. These trends indicate that users are increasingly trying to protect their privacy.

In summary, our analysis shows that users care about their privacy. Increasing fraction of users choose to make their accounts non-public (25% in 2014 to 37.5% in 2018). Out of the remaining users, whose accounts remain public (75% in 2014 down to 62.5% in 2018), half seem to manage their privacy by occasionally posting cryptic notes and a slightly smaller number do not post a single sensitive note, as per our classification approach. Thus, users seem to increasingly try to manage their privacy, but leaks still occur, and they affect a decreasing *portion*, but an increasing *number* of users. Using SENMO, we find that at least 8.5 M users across our datasets are affected.

Users who post more notes have a higher chance of posting sensitive content. We divide the users into activity categories based on the number of notes they post in our datasets – 1-5 public transactions, 6-10, 11-20, 21-50, 51-100, 101-500 and >500. For space reasons we summarize these results. Most users post between 1 and 5 notes across all the three datasets, and run only low risk of exposure (from 12% in D2 to 23% in D1). However, those users that post many notes (>500) run a much higher risk of exposure – up to 98% in the D1 dataset. Looking at the count and the percentage of sensitive transactions posted by users in different activity categories, the small datasets (D2 and D3), are dominated by users that post 1–5 transactions. The large dataset D1, is dominated by transactions from users that post more than 20 transactions. Even though such users represent only 23% of all users, they are responsible for about 80% of transactions. These users also run a higher privacy risk – 84%–98% expose some sensitive information in their notes. Even though all user groups post similar percentage of sensitive notes (9–13%), the sheer amount of activity increases privacy leakage. If we focus only on the more active users (75th percentile of Venmo users by the number of notes they post), they mostly post the notes in relationship (REL), location (LOC) and drugs/alcohol/gambling (DAG) categories. Moreover, the longer one's posts remain public, the greater the chance of potential misuse of that information to harm the user.

Relationship (REL) and drugs/alcohol/gambling (DAG) categories are the most frequent. Figure 5(a) shows the distributions of the number (log axis) and percentage of sensitive notes (out of all public notes) for each category over time, in our three datasets. The percentages in various categories remain roughly consistent over time. Relationship (REL) and drugs/alcohol/gambling (DAG) are the most popular categories, followed by location (LOC), adult content (ADU) , health (HEA), race/ethnicity (RET), violence/crime (VCR) and political opinions (POL). Though percentages of notes per category are small, the counts of sensitive notes number in the tens to hundreds of thousands per month. Some sensitive notes can leak information that can be used to impersonate a user, such as account number (ACC), address (ADD), phone number (PHO), E-mail (EMA) and product/activity details (PAD). Numbers and percentages of notes in these categories are small, as shown in Figure 5(b), but significant. E-mail address (EMA), phone number (PHO) and address (ADD) are leaked much more often than account details (ACC) and product/activity details (PAD). Figures 5(c) and 5(d) show the same measures for users affected by privacy leaks. While percentages of affected users (out of all public users) per category are modest (< 10%), the counts are increasing due to Venmo's popularity and reach. Our findings that many use Venmo for drug, alcohol, or gambling related transactions may look surprising, but these trends are supported by public polls [67]. LendEDU found that 32.6% of young adults use Venmo for drug purchases and 21% use it for gambling. Our data indicates a much smaller percentage (3.5%). This is due to the fact that our population includes all public users, does not include non-public

Dataset	D1			D2			D3		
Group category	AA	G	BG	AA	G	BG	AA	G	BG
Total senders	290 K	129 K	490	134	7 K	8	10 K	67 K	207
Sens/com senders	138 K	51 K	241	80	3 K	3	5 K	30 K	90
Total notes.	930 K	350 K	1.3K	148	8 K	9	27 K	215 K	456
Sens/com notes.	310 K	95 K	509	88	3 K	4	13 K	74 K	143

Table 6. Number of unique senders and transactions for the Alcoholics Anonymous (AA), Gambling (G) and Biker Gangs (BG) groups. Sens/com senders are the number of users who posted sensitive or common-pattern notes. Sens/com notes. are the sensitive or common-pattern notes.

users or non-public notes, and that some of users in our datasets post cryptic or inconspicuous notes.

6.2 Privacy leakage from groups

As described in Section 5.4, we apply a keyword/activity based heuristic to programmatically identify AA, gambling and biker gang groups in our datasets. Table 5 shows the number of groups we identified, and Table 6 shows the number of users and transactions in those groups. Around 40%–50% of users post at least one sensitive or common-pattern note, while the rest attempt to hide their membership by posting unrelated notes. However, all the users that send public payments to these groups (around 503 K of them in our datasets) have privacy leakage, because sensitive notes and group display names/user ID reveal the group's true nature.

6.3 Other security and privacy concerns

In this Section we discuss other security and privacy concerns around Venmo's publicly visible user data, beyond transaction notes.

User profile and friend list. Venmo has three ways to view a user's profile: (1) the current UI (https://account.venmo.com/u/<userID>, which is accessible through Venmo's search functionality, (2) the old UI, (https://venmo.com/<userID>), which was superseded by the current UI but was still active during our data collection, and (3) the app UI, accessible through Venmo app on a mobile phone. A user's public transactions can also appear in *global feed* (shown to random users as they log in) and in *public feed* (listed on a user's profile page for everyone to see). In June 2021 Venmo has allowed users to make their friend lists private (they were public before) and in July 2021 Venmo removed its global feed (public transactions of random users in a user's news feed). We tested the three Venmo's UIs with regard to visibility of a user's friend list and a user's transactions and feeds, from the viewpoint of logged in and non-logged in users, and show results in Table 7.

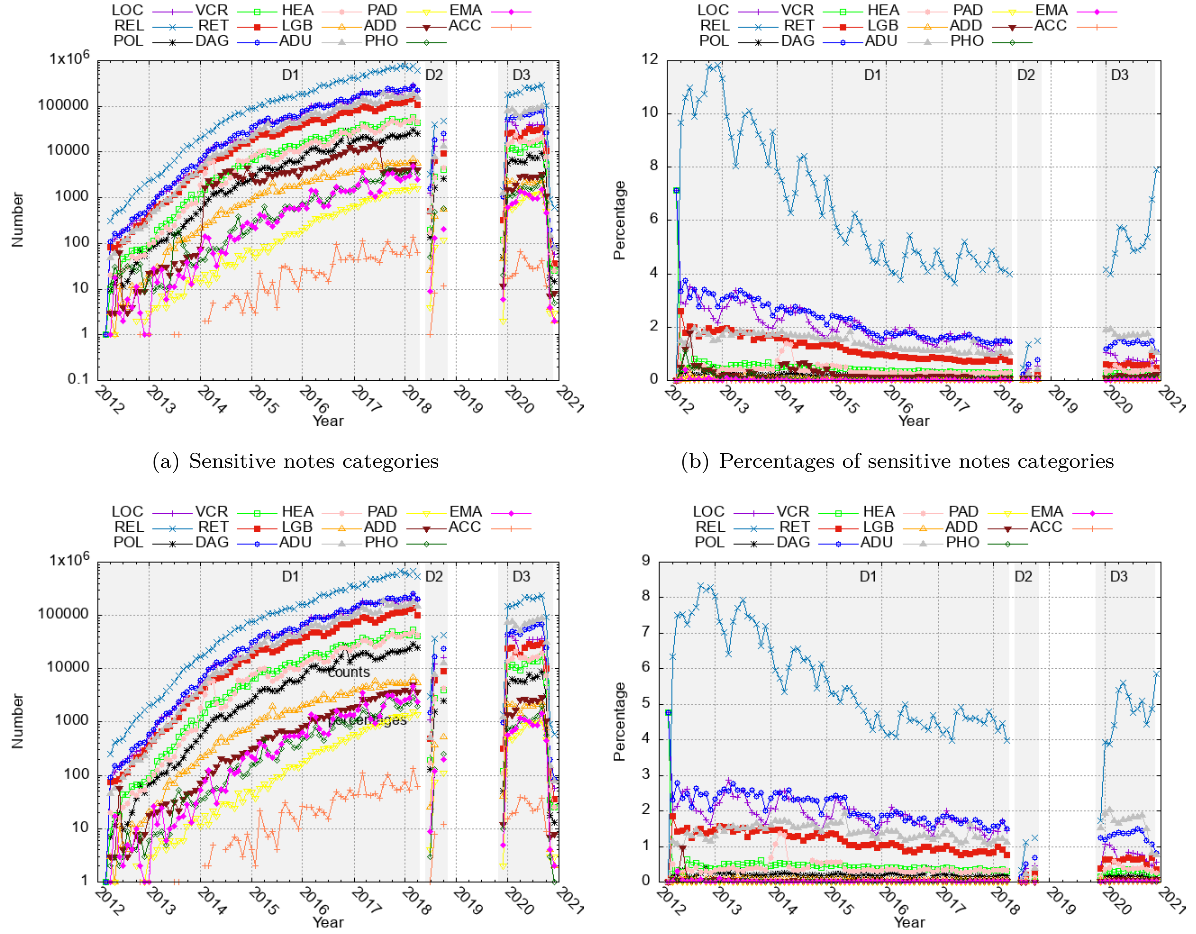
	VISIBILITY							
	Logged In				Not Logged In			
	old UI	current UI	mobile app: Aug21	mobile app: Jun21	old UI	current UI	mobile app: Aug21	mobile app: Jun21
Private Friends List	✓	✗	✗	✓	✗	✗	✗	✗
Public Friends List	✓	✗	✓	✓	✗	✗	✗	✗
Global Feed	✗	✗	✗	✓	✗	✗	✗	✗
Public Feed	✓	✓	✓	✓	5 recent transactions	✗	✗	✗
Private Transactions	✗	✗	✗	✗	✗	✗	✗	✗

Table 7. Visibility as of August 31, 2021: Different Scenarios.

For mobile app, we tested the version from June 2021 and then from August 2021. We highlight the fields that mismatch Venmo's privacy policy in yellow. There are two disturbing findings. First, the tested UIs offer *different privacy protections* to users. This is a problem because users expect the same protections from a given online service, regardless of the access mode. For example a logged in user can see public friends of another user on mobile app but not using the current Web UI. Second, in some cases these privacy protections *differ from Venmo's privacy policy.* This is a potential liability. For example a user's five recent public transactions are no longer visible to non-users via the current UI but they are still visible via old UI. Similarly, a user can make their friend list private and thus hide it from view via updated mobile app (Aug21) and the current UI, but it still remains visible via non-updated mobile app (Jun21) and the old UI. We believe these differences in privacy protections stem from different code bases, which should be synchronized. User privacy protections should also be enforced by Venmo's servers regardless of the client (version of the mobile app). We disclosed our findings to Venmo both via e-mail, and via PayPal's Bug Bounty Program. Our findings were validated and the issues were fixed by Venmo in November 2021, allowing us to claim the bug bounty.

Contact import. When Venmo app is installed on a user's phone, if the user consents, it may download users' complete contact list from the phone [68]. Venmo then automatically adds these contacts as friends. This creates large implications for user privacy when friend list is publicly visible (which is a default setting). Anyone logged in can crawl Venmo to build a list of phone contacts for any registered user.

Links to personal documents. While manually analyzing D2 dataset to devise regular expressions for personal note classification, we found out that sometimes a note may contain a URL. A very small percentage of notes in D2 dataset (0.01%) contains URLs. While most URLs seem to promote content, pointing to sites like YouTube, PornHub, etc., a few contain links to online documents with very personal data, such as



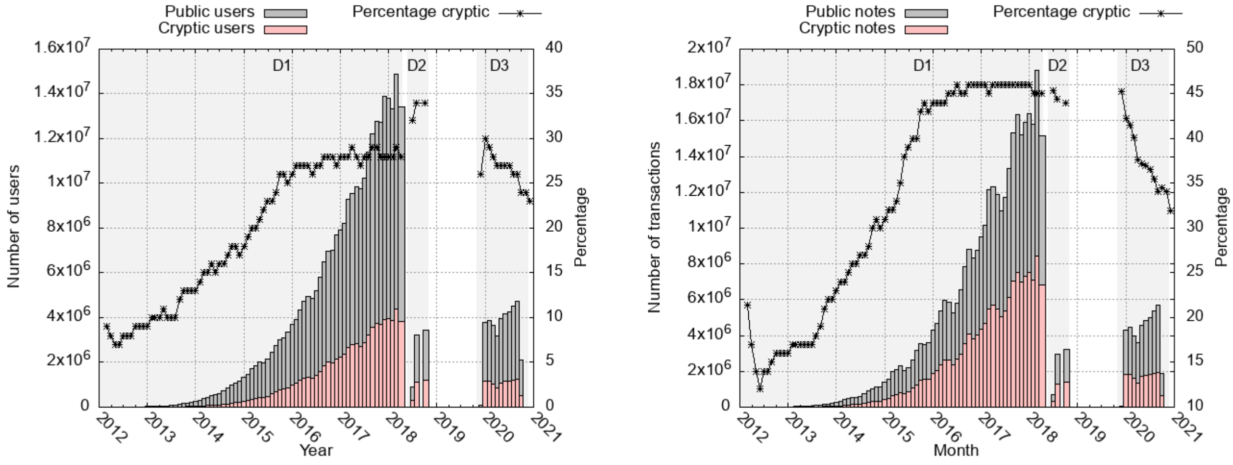
(a) Sensitive notes categories

(b) Percentages of sensitive notes categories

(c) Users who posted sensitive notes for different categories

(d) Percentages of users who posted sensitive notes for different categories

Fig. 5. Sensitive transaction and user counts and percentages, per category: smooth lines show counts (log axis), lines with points show percentages.



(a) Cryptic vs non-cryptic trends, users

(b) Cryptic vs non-cryptic trends, notes

Fig. 6. Cryptic and non-cryptic users and transactions – longitudinal trends.

credit card information, signatures, full Uber/Lyft receipts, split of shares of people who went to certain trips, receipts of all the restaurants and bars that they visited, etc. These online documents were also publicly visible. The user has likely relied on security through obscurity, since auto-generated URLs of online documents are difficult to guess. However, if one posts them in a public Venmo transaction note, it becomes easy for anyone to discover these URLs and access the online documents.

Scams. We came across a few scam posts on Venmo by chance, while we were building the fine-tuning set out of randomly drawn D2 notes. In these posts, the user is asked to help the alleged Venmo team by paying a small amount to help them make the app more secure.

Data breaches. While drawing notes at random to build our fine-tuning dataset, we found a few posts that indicate a payment for a settlement over data breach. Leaving such payments public can further increase the risk of identity theft for the user, since it advertises the fact that their data was part of the given data breach.

Vicemo [69] and D2 dataset: Vicemo pulls data from Venmo and publishes sensitive users posts related to adult content, drugs and alcohol, along with the usernames. Similarly, D2 dataset was originally collected to illustrate the risks of Venmo's public-by-default policy. While we agree with the intent of the data creators, posting such data publicly jeopardizes user privacy.

7 Recommendations

This section discusses the actions that can be taken by users and Venmo to mitigate the current privacy risks, and what design decisions should be considered by Venmo and others as they develop their products.

Private by default. Making any user data public by default on an online platform creates potential privacy leaks as the platform grows. From a user safety point of view, private-by-default is the best policy and it would completely address the problems we noted in the paper. Users that want to share some details can make an active decision to do so by changing their settings. While private-by-default solution is best for user privacy, it may conflict with Venmo's business preferences, and thus is unlikely to be deployed. We therefore explore other possible solutions.

User actions. Venmo users should immediately make all their past and future transactions private, as well as their friends list, by changing the settings in their profile. This is especially important for users that provide services to others or collect money for a group (e.g., AA). By making their data private, these users also protect their clients or group members.

Evaluate social features. Many platforms add social features and may resist private-by-default policies as they believe such changes could decrease user engagement. In Venmo's case, social features can be valued by users and the platform can be considered a type of social media [70]. However, we recommend that Venmo explores these questions, potentially by testing alternative UI design that retain social features within trusted groups, but reduces larger privacy risks through the removal of global note visibility. Friend lists and contact mining could be kept to allow users to find their friends by phone number on the platform, but such data should remain private to the given user. If social features are important, they could also be kept in a reduced form (e.g., show number of transactions/senders, but not individual details or notes).

Full control of user profiles. Venmo's users today have no way to make their profile private. A user's profile is considered private if no information about that user is visible to public (e.g., name, friends list, transaction history) and only name and mutual transactions are visible to friends. This feature should be added to enable users to fully control their privacy settings.

Clarify data use agreements. When we started this research, we studied Venmo's data use agreements, but did not find clauses that regulate research on public data. Given that Venmo data is being academically studied, it would be prudent to specify acceptable use policy for research. We also recommend designating a contact at Venmo, who can help facilitate internal patching of researcher-discovered issues before they become known to general public. We further recommend that Venmo proactively seeks to remove publicly visible Venmo data/datasets on third-party storage services or websites that reveal real user identities and transaction notes, as they remain a large privacy risk to users in those datasets. Researchers should keep their Venmo datasets on secure cyberinfrastructure rather than repositories such as GitHub. Finally, since Venmo's API was used to harvest much of the data used in our work, retiring that API could help protect user privacy. Venmo has stopped admitting new users to its API and has limited API request rate per IP address. However, these steps cannot fully protect user privacy. User profile and transaction data still remains accessible via browser and scraping can be automated via Selenium. User data can only be fully protected through policy and platform changes by Venmo and cooperation with researchers can help in this process.

8 Related work

Previous works have discovered various security issues with Venmo. Kraft et al. [71] find evidence of information leaks, which users might not want to publicize and vulnerabilities which could allow adversaries to steal users' money. Siddiqui et al. [72] analyze user activity and identify patterns and messaging that may expose personal activities and budget habits. Using a dataset of 2.12 M users and 20.23 M Venmo notes, Yao et al. [73] examine security issues in more detail, by inferring user locations with high levels of accuracy. They look at users who had Firstname+Lastname in their Venmo profiles and search Facebook to get a user's home location. Like our work, these studies make clear that there are leaks of real names, locations, and potentially sensitive activities on Venmo. Our analysis looks into a broader set of sensitive activities and at longitudinal trends in user privacy on Venmo, and thus complements these existing works. An application called "Public By Default" (https://publicbydefault.fyi/) [74] examines all transactions in 2017, collected from the Venmo public API at that time. It presents a statistical overview of common types of transactions, including by time (e.g., rent transactions naturally peak on the first of month) and singles out some specific, redacted, case studies (e.g., a cannabis retailer and the messages used to pay them). Our findings complement Do Thi Duc's [74] observations in that sensitive information about drugs, relationships, etc. are being publicly revealed on Venmo. Our approach examines data leaks across many categories and time periods.

Acker et al. [75] illustrate how transaction feeds of mobile payments support social practices, communication, and commerce with mobile devices and wireless networks. Huang et al. [76] observe that the social aspects of novel peer to peer (P2P) payment systems, such as Venmo, can even play a role in how consumers judge a business. Through interviews with 14 Venmo users and surveys of 164 peer-to-peer payment app users and 80 Venmo users, Caraway et al. [49] find uses consistent with other social awareness streams (SASs), and uncover novel uses that reflect the unusual inclusion of an SAS within a social payments app. People write purely functional transaction descriptions (i.e., notes) with strangers, while in transactions with friends, they sometimes craft playful descriptions that enhance their relationships. These studies emphasize that the social features are important not only to Venmo's popularity, but lead to a certain intimacy, which could be lead to inadvertent public sharing of sensitive information.

Zhang et al. [64] find that Venmo communities are densely connected compared to other interaction networks, and are often driven by specific niche applications. Using a newer and larger dataset, Unger et al. [77] sought to replicate Zhang et al.'s [64] methods and found that most network properties like density and clustering have been stable over time on Venmo. They notice an increase in users who quit after making a single Venmo transaction, and more communities with a smaller member pool.These studies have found that Venmo has had some consistent structural and network properties, which we leverage in our longitudinal investigation. Wang et al. [18] study how to classify private tweets into different, potentially sensitive topics. We leverage their keywords in part in our work, but revise them to achieve a higher confidence on the sensitivity of data being shared. Our classification approach also outperforms Wang et al., as shown in Table 13 in the Appendix and as discussed in Section 5. Wang et al. [78] quantify the level of sensitivity in tweets. They use multiple sets of keywords to come up with the candidate list of sensitive tweets and then score each tweet in terms of how much it exposes user privacy. This is direction we plan to explore in our future work. Deb et al. [79] designed a Twitter bot that leverages machine learning to identify users posting pro-tobacco tweets. They achieve a binary classification accuracy of 74% using Char-CNN on tobacco and drug-related tweets.

9 Conclusions

Our analysis of multiple, large Venmo datasets and user data privacy practices highlights serious risks from a public-by-default policy for mobile social payments. Mandatory notes and human challenges in navigating Venmo's user interface have led to significant sensitive data leakage. Though social features can have real benefits to users on social platforms (including Venmo), companies are accountable for providing a safe default settings to users. This is important given that most users do not have the technical literacy to change default privacy settings or the awareness of how their information leaks add up over time. The public-by-default model poses real privacy risks for users. It is critical that users opt into a public setting, and are kept appraised of the data they are sharing publicly (e.g., through a periodic summary compiled by the platform). We encourage Venmo and other social platforms to pro-actively work with researchers to collaboratively develop better privacy practices.

Acknowledgement

We would like to thank Mr. Shrivu Shankar for helping us with the data collection. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Venmo. What is Venmo?, 2021. https://help.venmo.com/hc/en-us/articles/221011388-What-is-Venmo-.

[2] David Curry. Venmo revenue and usage statistics. *Business of Apps*, 2021. https://www.businessofapps.com/data/venmo-statistics/.

[3] Kate Rooney. Venmo has 40 million users. *CNBC*, 2019. https://www.cnbc.com/2019/04/24/venmo-has-40-million-users-paypal-reveals-for-first-time.html.

[4] Hang Do Thi Duc. Venmo Stories of 2017, 2017. https://publicbydefault.fyi.

[5] Ashley Boyd. 25,000 Americans urge Venmo to update its privacy settings. *Mozilla*, 2018. https://tinyurl.com/samvatzr.

[6] Federal Trade Commission. PayPal settles FTC charges, 2018. https://tinyurl.com/9tr3cah2.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, 2018.

[8] Sai Teja Peddinti, Aleksandra Korolova, Elie Bursztein, and Geetanjali Sampemane. Cloak and swagger: Understanding data sensitivity through the lens of user anonymity. In *2014 IEEE Symp. on Security and Privacy*, pages 493–508, 2014.

[9] European Commission. What personal data is considered sensitive?, 2021. https://tinyurl.com/6cr7p23u.

[10] U.S. Department of Homeland Security. Handbook for Safeguarding Sensitive PII, 2017. https://tinyurl.com/4dxs4yvp.

[11] Google. Google Policies and Principles, 2021. http://www.google.com/policies/privacy/key-terms/.

[12] Kostas Drakonakis, Panagiotis Ilia, Sotiris Ioannidis, and Jason Polakis. Please forget where I was last summer. *arXiv:1901.00897*, 2019.

[13] Spyros Boukoros, Mathias Humbert, Stefan Katzenbeisser, and Carmela Troncoso. On (the lack of) location privacy in crowdsourcing applications. In *28th USENIX Security Symposium*, pages 1859–1876, 2019.

[14] Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proc. of the 2013 ACM Conference on Computer & Communications Security*, pages 901–914, 2013.

[15] Abhinav Palia and Rajat Tandon. Optimizing noise level for perturbing geo-location data. In *Future of Information and Communication Conference*, pages 63–73, 2018.

[16] Changchang Liu and Prateek Mittal. LinkMirage: Enabling Privacy-preserving Analytics on Social Relationships. In *NDSS*, 2016.

[17] Roxana Geambasu, Tadayoshi Kohno, Amit A Levy, and Henry M Levy. Vanish: Increasing Data Privacy with Self-Destructing Data. In *Proc. of USENIX Security Symposium*, volume 316, 2009.

[18] Qiaozhi Wang, Jaisneet Bhandal, Shu Huang, and Bo Luo. Classification of private tweets using tweet content. In *2017 IEEE 11th Int. Conf. on Semantic Computing (ICSC)*, pages 65–68, 2017.

[19] Martin Pengelly. Joe Biden's Venmo account discovered in 'less than 10 minutes' – report. *The Guardian*, 2021. https://tinyurl.com/t5fzkdnz.

[20] Conor Skelding. Students fret over Venmo payments to alleged drug dealer. *Politico*, 2015. https://tinyurl.com/2fbhumbr.

[21] Donald Padgett. Egyptian police use grindr to arrest, assault gays. *Out*, 2020. https://tinyurl.com/hrt2a4n4.

[22] Jose Pagliery and Roger Sollenberger. Gaetz Paid Accused Sex Trafficker, Who Then Venmo'd Teen. *The Daily Beast*, 2021. https://tinyurl.com/279crv8e.

[23] Steve Almasy. Houston burglary gang used social media to find houses with high-end art to steal, police say. *CNN*, 2021. https://tinyurl.com/3t4t4s32.

[24] Lucie Rychla. Danish woman loses disability benefits over facebook photos. *CPH Post*, 2016. https://cphpost.dk/?p=69466.

[25] PRNewswire. 79% of Businesses Have Rejected a Job Candidate Based on Social Media Content; Job Seekers Should Post Online Carefully. https://tinyurl.com/cr2a356a, 2020.

[26] IvyWise. Yes, College Admissions Officers Are Looking At Social Media, 2020. https://tinyurl.com/7bp9tybx.

[27] Natasha Singer. They Loved Your G.P.A. Then They Saw Your Tweets. *The New York Times*, 2013. https://tinyurl.com/zra62d2y.

[28] Business Daily. Keep it clean: Social media screenings gain in popularity.

[29] Aja Romano. One of these pageant queens lost her crown over her social media use. The other did not. Why? *Vox*, 2016. https://tinyurl.com/rj58bxbs.

[30] Identity Fraud Hits Record High with 15.4 Million U.S. Victims in 2016, Up 16 Percent According to New Javelin Strategy & Research Study. *Javelin*, 2017. https://tinyurl.com/vkxxbxsv.

[31] Sam Cook. Identity theft facts & statistics: 2019-2021. *Comparitech*, 2021. https://tinyurl.com/52tdxnh7.

[32] Jeremy Finley. Trained hacker: This is how criminals steal money on Venmo. *News4 Nashville*, 2021. https://tinyurl.com/b7w7989z.

[33] Nicolás Emilio Díaz Ferreyra, Rene Meis, and Maritta Heisel. Learning from online regrets: from deleted posts to risk awareness in social network sites. In *27th Conference on User Modeling, Adaptation and Personalization*, pages 117–125, 2019.

[34] Amandeep Dhir, Puneet Kaur, Sufen Chen, and Kirsti Lonka. Understanding online regret experience in Facebook use–Effects of brand participation, accessibility & problematic use. *Computers in Human Behavior*, 59:420–430, 2016.

[35] 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 *Proc. of the Seventh*

Symposium on Usable Privacy and Security, SOUPS '11, New York, NY, USA, 2011. Association for Computing Machinery.

[36] Geoffrey A. Fowler. When the Most Personal Secrets Get Outed on Facebook. *The Wall Street Journal*, 2012. https://tinyurl.com/4xdwbddy.

[37] Venmo. Payment Activity & Privacy, 2021. https://help.venmo.com/hc/en-us/articles/210413717-Payment-Activity-Privacy.

[38] C. Hoofnagle and Jennifer M. Urban. Alan Westin's Privacy Homo Economicus. *Information Privacy Law eJournal*, 2014.

[39] Yun Zhou, Lianyong Qi, Alexander Raake, Tao Xu, Marta Piekarska, and Xuyun Zhang. User attitudes and behaviors toward personalized control of privacy settings on smartphones. *Concurrency and Computation: Practice and Experience*, 31(22), 2019.

[40] Casey Fiesler, Michaelanne Dye, Jessica L. Feuston, Chaya Hiruncharoenvate, C.J. Hutto, Shannon Morrison, Parisa Khanipour Roshan, Umashanthi Pavalanathan, Amy S. Bruckman, Munmun De Choudhury, and Eric Gilbert. What (or Who) Is Public? Privacy Settings and Social Media Content Sharing. In *Proc. of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, page 567–580, 2017.

[41] Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding Privacy Settings in Facebook with an Audience View. *UPSEC*, 8:1–8, 2008.

[42] Tyler Sonnemaker. Google's own engineers said the company 'confuses users' on privacy settings that are now the subject of a lawsuit. *Insider*, 2020. https://tinyurl.com/jjm2y2ys.

[43] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. The Post That Wasn't: Exploring Self-Censorship on Facebook. In *Proc. of the Conference on Computer Supported Cooperative Work*, CSCW '13, page 793–802, 2013.

[44] Jaakko Stenros, Janne Paavilainen, and Jani Kinnunen. Giving Good 'Face': Playful Performances of Self in Facebook. In *Proc. of the International Academic MindTrek Conference: Envisioning Future Media Environments*, page 153–160, 2011.

[45] Dan Salmon. Venmo transaction dataset, 2018. https://github.com/sa7mon/venmo-data.

[46] Selenium. Selenium webdriver, 2012. https://tinyurl.com/y6a4czhe.

[47] Zack Whittaker. Millions of Venmo transactions scraped in warning over privacy settings. *TechCrunch*, 2021. https://tinyurl.com/dneuxmzz.

[48] David R Bild, Yue Liu, Robert P Dick, Z Morley Mao, and Dan S Wallach. Aggregate characterization of user behavior in twitter and analysis of the retweet graph. *ACM Transactions on Internet Technology (TOIT)*, 15(1):1–24, 2015.

[49] Monica Caraway, Daniel A Epstein, and Sean A Munson. Friends don't need receipts: The curious case of social awareness streams in the mobile payment app venmo. *Proc. of the ACM on Human-Computer Interaction*, 1(CSCW):1–17, 2017.

[50] Casey Fiesler and Nicholas Proferes. "Participant" perceptions of Twitter research ethics. *Social Media+ Society*, 4(1), 2018.

[51] Hackerone. Paypal, 2020. https://hackerone.com/paypal.

[52] NLTK. Extracting Information from Text, 2020. https://www.nltk.org/book/ch07.html.

[53] Sensitive Keywords List, 2022. https://github.com/STEELISI/SENMO/tree/main/data/Lexicon.

[54] Luis von Ahn. Offensive/Profane Word List, 2021. https://www.cs.cmu.edu/~biglou/resources/bad-words.txt.

[55] Mohammadreza Rezvan, Saeedeh Shekarpour, Lakshika Balasuriya, Krishnaprasad Thirunarayan, Valerie L Shalin, and Amit Sheth. A quality type-aware annotated corpus and lexicon for harassment research. In *Proceedings of the 10th ACM Conference on Web Science*, pages 33–36, 2018.

[56] Mohammadreza Rezvan, 2018. https://github.com/Mrezvan94/Harassment-Corpus.

[57] National Institute on Drug Abuse. Commonly Used Drugs Charts, 2020. https://www.drugabuse.gov/drug-topics/commonly-used-drugs-charts.

[58] National Institute on Alcohol Abuse and Alcoholism. Drug-Alcohol List, 2021. https://www.niaaa.nih.gov/sites/default/files/Drug_alcohol_list_final.pdf.

[59] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the International Conference on Computer Vision*, pages 19–27, 2015.

[60] TextBlob, 2021. https://github.com/sloria/TextBlob.

[61] Pyspellchecker, 2021. https://github.com/barrust/pyspellchecker.

[62] Autocorrect, 2021. https://github.com/fsondej/autocorrect.

[63] JamSpell, 2021. https://github.com/bakwc/JamSpell.

[64] Xinyi Zhang, Shiliang Tang, Yun Zhao, Gang Wang, Haitao Zheng, and Ben Zhao. Cold hard E-cash: Friends and vendors in the Venmo digital payments system. In *Proc. of the International AAAI Conference on Web and Social Media*, 2017.

[65] Wikipedia. List of gangs, 2021. https://en.wikipedia.org/wiki/List_of_gangs_in_the_United_States.

[66] Automole, 2021. https://automole.net/the-most-dangerous-biker-gangs-in-america/.

[67] Mike Brown. Nearly a Third of Millennials Have Used Venmo to Pay for Drugs, 2017. https://tinyurl.com/582me8dc.

[68] Venmo. Phone contacts, 2021. https://help.venmo.com/hc/en-us/articles/217532217-Adding-Removing-Friends-#.

[69] Vicemo. See who's buying drugs, booze, and sex on Venmo, 2020. https://www.vicemo.com/.

[70] Amelia Acker and Dhiraj Murthy. What is venmo? a descriptive analysis of social features in the mobile payment platform. *Telematics and Informatics*, 52, 2020.

[71] Ben Kraft, Eric Mannes, and Jordan Moldow. Security research of a social payment app, 2014. https://www.benkraft.org/files/venmo.pdf.

[72] Husna Siddiqui, Aspen Olmsted, and Brendan Keane. Venmo: Exposing a user's lifestyle. In *Proc. of the International Conference for Internet Technology and Secured*

Transactions (ICITST), 2017.

[73] Xin Yao, Yimin Chen, Rui Zhang, Yanchao Zhang, and Yaping Lin. Beware of What You Share: Inferring User Locations in Venmo. *IEEE Internet of Things Journal*, 5(6):5109–5118, 2018.

[74] Hang Do Thi Duc. Public By Default, 2018. https://22-8miles.com/public-by-default.

[75] Amelia Acker and Dhiraj Murthy. Venmo: Understanding mobile payments as social media. In *Proc. of the 9th international conference on social media and society*, pages 5–12, 2018.

[76] Liang Huang, Anastasiya Pocheptsova Ghosh, Ruoou Li, and Elise Chandon Ince. Pay Me with Venmo: Effect of Service Providers' Decisions to Adopt P2P Payment Methods on Consumer Evaluations. *Journal of the Association for Consumer Research*, 5(3):271–281, 2020.

[77] Clive Unger, Dhiraj Murthy, Amelia Acker, Ishank Arora, and Andy Chang. Examining the evolution of mobile social payments in Venmo. In *Proc. of International Conference on Social Media and Society*, pages 101–110, 2020.

[78] Qiaozhi Wang, Hao Xue, Fengjun Li, Dongwon Lee, and Bo Luo. #DontTweetThis: Scoring Private Information in Social Networks. *Proceedings on Privacy Enhancing Technologies*, 2019:72–92, 10 2019.

[79] Ashok Deb, Anuja Majmundar, Sungyong Seo, Akira Matsui, Rajat Tandon, Shen Yan, Jon-Patrick Allem, and Emilio Ferrara. Social bots for online public health interventions. In *Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1–4, 2018.

[80] Bacem A Essam. Compiling a lexicon of pornography using web, WordNet and FrameNet to develop an individual pornographic index. *Sexuality & Culture*, 21(2):534–548, 2017.

[81] American Medical Association. AMA glossary of medical terms, 2021. https://tinyurl.com/282rcpsp.

[82] Ildado. Casino gambling glossary, 2021. https://www.ildado.com/casino_glossary.html.

[83] Wikipedia. List of presidents of the United States, 2021. https://en.wikipedia.org/wiki/List_of_presidents_of_the_United_States.

[84] Wikipedia. List of vice presidents of the United States, 2021. https://en.wikipedia.org/wiki/List_of_vice_presidents_of_the_United_States.

[85] National Park Service. First Ladies of the United States, 2021. https://www.nps.gov/fila/learn/historyculture/first-ladies.htm.

[86] Wikipedia. List of ethnic slurs, 2021. https://en.wikipedia.org/wiki/List_of_ethnic_slurs.

[87] Racial Equity Tools Glossary, 2020. https://www.racialequitytools.org/glossary.

[88] R Bhopal. Glossary of terms relating to ethnicity and race: for reflection and debate. *BMJ Journals*, 2004. https://jech.bmj.com/content/58/6/441.

[89] The Racial Slur Database, 2021. http://www.rsdb.org/races.

[90] Vocabulary: Crime & Criminals, 2021. https://tinyurl.com/rhrtwncj.

[91] FindLaw. List of Criminal Charges, 2020. https://www.findlaw.com/criminal/criminal-charges/view-all-criminal-charges.html.

[92] Wikipedia. Crimes, 2020. https://en.wikipedia.org/wiki/Category:Crimes.

[93] Clarifacts. Federal crimes list, 2021. https://tinyurl.com/rm6wh3b6.

[94] Members of the Family, 2021. https://tinyurl.com/ewpvxw83.

[95] Wikipedia. Family, 2021. https://en.wikipedia.org/wiki/Family.

[96] English Vocabulary for Dating and Relationships, 2021. https://tinyurl.com/4kds546b.

[97] Family Relationship Names in English Pdf, 2020. https://tinyurl.com/kevwn62d.

[98] Glossary of Family Law Terms, 2021. https://tinyurl.com/3stesmt6.

[99] World Cities Database, 2020. https://simplemaps.com/data/world-cities.

[100] Countries and Regions of the World, 2021. https://tinyurl.com/3ctywvp4.

[101] Wikipedia, 2021. https://en.wikipedia.org/wiki/Motorcycle_components.

[102] Ultimate Motorcycling, 2021. https://tinyurl.com/2y7wbx9a.

[103] International Telecommunication Union. DRAFT Southern African Development Community Model Law on Data Protection, 2011. https://tinyurl.com/4un66969.

[104] Repro APAC. Personal Data Protection Act in the South East Asia Region. https://tinyurl.com/b8s6h26b.

[105] Privacy laws in Southeast Asia, 2021. https://tinyurl.com/325jybnc.

[106] Aruba Marketing. Data privacy laws in APAC: What You Need to Know, 2019. https://tinyurl.com/y56v9v65.

[107] Office of the Australian Information Commissioner. What is personal information?, 2021. https://tinyurl.com/m4p86rcr.

[108] U.S. Department of Health & Human Services. HHS Privacy Policy Notice, 2021. https://www.hhs.gov/web/policies-and-standards/hhs-web-policies/privacy/index.html.

[109] Office of the Privacy Commissioner of Canada. Form of Consent, 2015. https://tinyurl.com/2vzfmc2w.

[110] Office of Privacy and Open Government, U.S. Department of Commerce. Safeguarding information, 2020. https://www.osec.doc.gov/opog/privacy/PII_BII.html.

[111] International Association of Privacy Professionals. Brazilian General Data Protection Law, 2019. https://tinyurl.com/2vw4mjtj.

[112] Apple. App privacy details on the app store, 2021. https://developer.apple.com/app-store/app-privacy-details.

[113] Microsoft. Data loss prevention reference, 2021. https://tinyurl.com/4peem2ad.

[114] Amazon. Your profile and sensitive products, 2021. https://tinyurl.com/yp9k7u3d.

[115] Sony. Playstation network terms of service and user agreement, 2020. https://www.playstation.com/en-us/legal/psn-terms-of-service/.

[116] WeChat. WeChat Privacy Protection Summary, 2020. https://www.wechat.com/en/privacy_policy.html.

[117] Facebook. Advertising Policies, 2021. https://www.facebook.com/policies/ads.

[118] Adobe. Adobe General Terms of Use, 2020. https://www.adobe.com/legal/terms.html.

[119] Mert Aktas. 8 Biggest Tech Companies of 2021, 2021. https://tinyurl.com/4spv68tm.

[120] World Top 1000 Companies List and World Ranks as on Jan 1st 2021 from Value.Today, 2021. https://www.value.today/.

[121] Largest Companies by Market Cap, 2021. https://companiesmarketcap.com.

A Classification criteria

Table 11 shows sources we used to extract keywords for our sensitive note categories. Figure 7 shows the regular expressions we use to identify sensitive note categories.

B Classification accuracy

Evaluation for our BERT model, compared with three other approaches: (BoW-kw) bag-of-words using our keyword sets, (BOW-NB) Naive Bayes on bag-of-words, (TF-IDF-NB) Naive Bayes on TF-IDF. We evaluate three flavors of our BERT model: (SENMO-npre) BERT without keyword pre-filtering, (SENMO-NONE) BERT with NON as a separate category in fine-tuning step and (SENMO) BERT with keyword prefiltering, and without NON category in fine-tuning. Our results are shown in Table 13.

Threshold	True Positives			False Positives			True Positives Percentage			False Positives Percentage		
	AA	G	BG	AA	G	BG	AA	G	BG	AA	G	BG
>10%	15	4	1	0	3	0	100%	100%	100%	0%	14%	0%
>20%	15	4	1	0	0	0	100%	100%	100%	0%	0%	0%
>30%	15	1	0	0	0	0	100%	25%	0%	0%	0%	0%
>40%	15	1	0	0	0	0	100%	25%	0%	0%	0%	0%
>50%	15	1	0	0	0	0	100%	25%	0%	0%	0%	0%

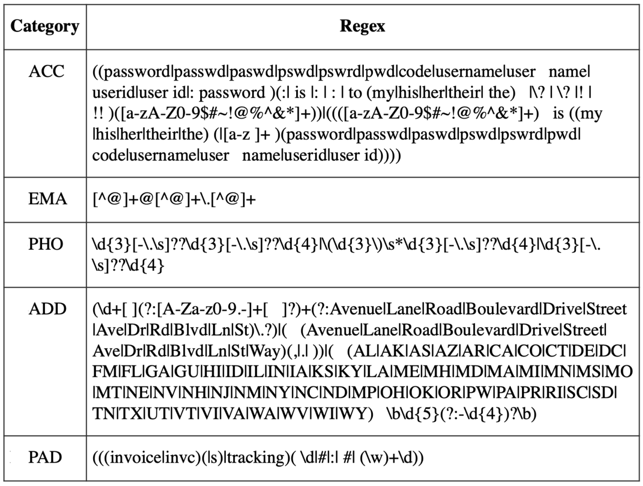
Table 8. Thresholds for percentages of transactions that match AA keywords for the AA groups, gambling keywords for the gambling groups and biker gang keywords for the biker gang groups.

Sensitive (AA)		Common Patterns	
alcoholics anonymous	sun up	only Emojis	paying
book study	grounded in the cloud	time (e.g. 5pm)	my share
7th tradition	a gathering of men	dates (e.g. 11/21)	greetings /interjections (examples are below)
7th trad	joy of living	lunch bunch	hello
7 th trad	our common welfare	donation	hi
7 th tradition	live and let live	donate	hey
seventh tradition	Early birds	donations	wassup
seventh trad	11th step	contribute	thank you
attitude adjustment	eleventh step	contribution	thank u
sunrise meeting	11 th step	contributing	thanks
daily reprieve	decade day	stuff	thnx
as bill sees it	keep it simple	meeting	thx
step workshop	reflections	meetings	ty
hole in the sky	higher power	meet	hurray
winners attitude	gratitude	due	hurrah
a new start	sobriety	dues	bye
eye opener	awakening	payment	yes
back of the book	AA	pay	sorry

Table 9. List of keywords for AA groups.

Sensitive (Gambling)		Common Pattern	
poker	play	only Emojis	emails
gamble	game	time (e.g. 5pm)	names of games (eg: football)
gambling	**darts**	dates (e.g. 11/21)	greetings /interjections (examples are below)
casino	baccarat	refund	hello
betting	jackpot	Vegas	hi
blackjack	keno	vip membership	hey
pokr	black jack	money	thank you
bet	pool	dollar	thank u
roulette	cards	ticket	thanks

Table 10. List of keywords for gambling groups.



<table>
<tr><th>Category</th><th>Regex</th></tr>
<tr><td>ACC</td><td>((password|passwd|paswd|pswd|pswrd|pwd|code|username|user name|userid|user id|: password)(:| is |: | : | to (my|his|her|their| the) |\? | \? |! | !! |!!)([a-zA-Z0-9$#~!@%^&*]+))|((([a-zA-Z0-9$#~!@%^&*]+) is ((my|his|her|their|the) (|[a-z]+)(password|passwd|paswd|pswd|pswrd|pwd|code|username|user name|userid|user id))))</td></tr>
<tr><td>EMA</td><td>[^@]+@[^@]+\.[^@]+</td></tr>
<tr><td>PHO</td><td>\d{3}[-\.\s]??\d{3}[-\.\s]??\d{4}|\(\d{3}\)\s*\d{3}[-\.\s]??\d{4}|\d{3}[-\.\s]??\d{4}</td></tr>
<tr><td>ADD</td><td>(\d+[ ](?:[A-Za-z0-9.-]+[ ]?)+(?:Avenue|Lane|Road|Boulevard|Drive|Street|Ave|Dr|Rd|Blvd|Ln|St)\.?)|((Avenue|Lane|Road|Boulevard|Drive|Street|Ave|Dr|Rd|Blvd|Ln|St|Way)(,|.|))|((AL|AK|AS|AZ|AR|CA|CO|CT|DE|DC|FM|FL|GA|GU|HI|ID|IL|IN|IA|KS|KY|LA|ME|MH|MD|MA|MI|MN|MS|MO|MT|NE|NV|NH|NJ|NM|NY|NC|ND|MP|OH|OK|OR|PW|PA|PR|RI|SC|SD|TN|TX|UT|VT|VI|VA|WA|WV|WI|WY) \b\d{5}(?:-\d{4})?\b)</td></tr>
<tr><td>PAD</td><td>(((invoice|invc)(|s)|tracking)(\d|#|:| #| (\w)+\d))</td></tr>
</table>

Fig. 7. Regexes for ACC, EMA, PHO, ADD and PAD.

Category ID	Sources used to identify keywords for the different categories
ADU and LGB	A quality type-aware annotated corpus and lexicon for harassment research [55, 56] Luis von Ahn's research group: offensive/profane word list [54] Pornography lexicons [80]
HEA	American Medical Association [81]
DAG	National Institute on Drug Abuse [57] National Institute on Alcohol Abuse and Alcoholism [58] Gambling Il Dado - casino gambling glossary [82]
POL	List of presidents of the United States [83] List of vice presidents of the United States [84] First ladies of the United States [85] A quality type-aware annotated corpus and lexicon for harassment research [55, 56]
RET	List of ethnic slurs [86] Racial equity tools glossary [87] Glossary of terms relating to ethnicity and race for reflection and debate [88] The racial slur database [89] A quality type-aware annotated corpus and lexicon for harassment research [55, 56]
VCR	Rebecca's vocabulary of crime & criminals [90] FindLaw's team of legal writers and editors [91] Crime list [92] Federal crimes list [93]
REL	Members of the family [94] Family [95] Really learn English: English vocabulary for dating and relationships [96] English grammar here: family relationship names [97] Divorce law info [98]
LOC	World cities database [99] Countries and regions of the world [100]

Table 11. Various popular sources that we use to identify keywords for the different categories.

Sensitive (BG)	Common Pattern
It comprises the list of biker gangs from the following sources: Outlaw motorcycle clubs [65] The most dangerous biker gangs in America [66]	It comprises the list of motorcycle components and motorcycle names from the following sources: Motorcycle components [101] Common motorcycle names [102] It also includes terms like: gas, gasoline, bike payment/rental/service, donation, funds.

Table 12. List of keywords for biker gang groups.

	Category ID	BoW-kw	BoW-NB	TF-IDF-NB	SENMO-npre (without Keyword Filter)	SENMO (with Keyword Filter)	SENMO-NONE (with explicit NON category)
Per Note Accuracy		0.68	0.77	0.65	**0.90**	**0.90**	0.86
True Positives	LGB	**1.00**	0.59	0.70	0.94	0.94	0.94
	ADU	**0.96**	0.80	0.66	0.95	0.94	0.94
	HEA	**0.99**	0.83	0.57	0.95	0.95	0.92
	DAG	**0.97**	0.70	0.40	0.93	0.93	0.83
	POL	0.95	0.80	0.67	**0.96**	0.95	0.90
	RET	**0.99**	0.70	0.76	0.95	0.95	0.91
	VCR	**0.96**	0.80	0.60	0.89	0.89	0.95
	REL	**0.96**	0.49	0.84	0.94	0.93	0.93
	LOC	**0.98**	0.81	0.89	0.87	0.87	0.91
	NON	0.88	0.89	**0.97**	0.87	0.95	0.68
False Positives	LGB	0.05	**0.00**	**0.00**	**0.00**	**0.00**	0.01
	ADU	0.03	0.02	**0.00**	0.01	0.01	0.01
	HEA	0.13	0.01	0.00	**0.00**	**0.00**	0.01
	DAG	0.01	0.01	**0.00**	0.01	0.01	0.03
	POL	0.02	0.01	**0.00**	**0.00**	**0.00**	0.01
	RET	0.03	0.01	**0.00**	**0.00**	**0.00**	0.01
	VCR	0.06	0.02	**0.00**	0.01	0.01	0.02
	REL	0.01	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	LOC	0.00	**0.00**	**0.00**	**0.00**	**0.00**	0.03
	NON	**0.01**	0.16	0.34	0.02	0.03	0.14

Table 13. Classification accuracy on the Ground Truth data.

Different Regions of the world consider the following as sensitive content:						
Region	Africa	Asia	Australia	Europe	North America	South America
Laws and regulations/ Govt. Agencies	HIPSSA [103]	PDPA [104], APPI [105, 106], PIPA [105, 106]	OAIC [107]	GDPR [9]	HIPAA [108], PIPEDA [109], DHS [10], OPOG [110]	LGPD [111]
Race/Ethnicity	✓	✓	✓	✓		✓
Adult content/Sexual orientation	✓	✓	✓	✓		✓
Medical Facts	✓	✓	✓	✓	✓	✓
Religious/Philosophical beliefs	✓	✓	✓	✓		✓
Trade union membership	✓		✓	✓		✓
Political opinions	✓	✓	✓	✓		✓
Genetic/Biometric information			✓	✓	✓	✓
Personal financial information					✓	
Government ID number					✓	
Data that may cause discrimination	✓				✓	
Data that may cause embarrassment					✓	
Data that may cause unfairness					✓	
Data that may cause inconvenience					✓	
Date of birth, Place of birth					✓	
Mother's maiden name					✓	
Employment Information					✓	
Criminal history			✓		✓	

Table 14. Interpretation of sensitive content by different laws, regulations and government agencies across the globe.

Different Online service providers consider the following as sensitive content:								
Online Service Providers	Apple Inc. (App Store [112])	Alphabet Inc. (Google [11])	Microsoft (Office 365 [113])	Amazon (Delivery Products [114])	Sony (Playstation [115])	Tencent (WeChat [116])	Facebook Inc. (Advertising policies [117])	Adobe [118]
Race/Ethnicity	✓	✓			✓	✓	✓	✓
Adult content/Sexual orientation	✓	✓		✓	✓		✓	✓
Medical Facts	✓	✓	✓		✓	✓	✓	✓
Religious/Philosophical Beliefs	✓	✓			✓	✓	✓	✓
Political opinions	✓	✓						
Genetic/Biometric information	✓							✓
Illegal content/activity					✓		✓	
Personal financial information			✓					✓
Violence/Crime					✓		✓	
Personal information (e.g. email, phone number, address)					✓			
Personal care products				✓				
Intimate clothing				✓				
Jewelry				✓				
Personal protection products				✓				
Government ID number			✓					

Table 15. Interpretation of sensitive content by some of the largest tech companies [119–121] across the globe.