Venmo: More Transactions, More Problems

Meghan McLaughlin
The University of Iowa
Iowa City, Iowa
meghan-l-mclaughlin@uiowa.edu

Matthew Haferbier
The University of Iowa
Iowa City, Iowa
matthew-haferbier@uiowa.edu

ABSTRACT

We present a detailed measurement of the transactions on Venmo, one of the world's most popular mobile payment services. Over the course of a week, from November 12, 2018 to November 18, 2018, we make 8 different types of measurements on transactions that occurred on Venmo. This measurement was made possible through the use of the Venmo Public API, which allowed us to collect data for over 397,000 transactions. Our findings include an analysis of the most popular words and emojis that are associated with messages on Venmo transactions, and a collection of other data relating to the general climate of Venmo. All of the analysis done throughout this research is intended to highlight the information, good or bad, that can be be found due to Venmo's transaction information being public to the masses.

1 INTRODUCTION

1.1 Background

Venmo is a mobile payment service that has been available since 2009. Currently, Venmo is only available in the United States, and has 22.9 million users (June 2018). In the first half of 2018, Venmo processed over $26 Billion in payments. Users can send, receive, and request money from users with their via their Venmo user-name, phone number, or QR code.

In the mobile app, users can view a social-network like interface, and are able to switch between three feeds: a public feed, friends feed, and private feed. The public feed shows transaction information (excluding transaction amount) for any user pair who has completed a transaction and both users' profiles are public. This feature is also available to any individual without creating a Venmo account.

In addition to being visible to any user logged in on the mobile application, any public transaction is accessible through the company's public API, which can be found at https://venmo.com/api/

v5/public . The API data contains more information than is available by viewing the corresponding transaction in the mobile app, such as each actor's user name (only their first and last name are visible in the app view) and account creation date. Until recently, it was possible for a single person to gather an unlimited amount of public user data. Recently, Venmo has implemented a rate limit for their API data, to approximately 1 API call every 30 seconds, per computer. However, due to the limit being imposed on a single machine, it would be possible for a single person to gather the data for every public transaction with enough machines. The API will be discussed further in Section 3.3 - Data Collection.

A Venmo transaction consists of a transaction initiator (the user who creates the transaction request), transaction recipient, amount, type, and message. The type may be a *payment*, where one user indicates the amount they would like to pay another user, or a *charge*, where one user indicates the amount they would like to be paid by another user. The message field is required to be filled in with plain-text or emojis. Users must create an account to perform transactions, and their information is usually publicly available, as the default privacy setting when an account is created is 'public',

1.2 Motivations & Contributions

Our initial goal for this research was to provide insight to the implications of allowing small amounts of data - such as a simple transaction descriptor - to be publicly available. Our initial methodology, roadblocks, and final methodology are described in Section 3.

The main purpose of this paper is to provide a general overview of the Venmo ecosystem strictly using data available via the public API. We gather data over a week-long period to analyze general trends in user activity, transaction category, and user privacy concerns based on user trends.

2 RELATED WORK

This study is inspired in part by *Public By Default* [1], a project by Hang Do Thi Duc. In Public By Default, the researcher gathered the transaction data from every public transaction in 2017 - 207,984,218 total transactions. By compiling and measuring trends in this data, the researcher was able to determine extremely personal information about users.

Through analysis of transactions over time, the researcher was able to track businesses, couples, and individuals as they traversed through their day to day actions.

For a food truck in California, the researcher was able to determine their most popular menu item, price, and most common customers. The customer who visited most often also had a trend of visiting every week at the same time. The idea of allowing any person on the internet to know where you may be at any given time is not a factor most users would consider when deciding their

social media privacy preferences. Another analysis in the study followed the transactions of a couple over the year. They would use Venmo to split grocery, utility, and other miscellaneous expenses. Through this information, the researcher was able to determine where the couple lives based on their utility company (mentioned by name in the transaction message). Along with this information, the couple could be providing advertisers with information about frequently visited stores and frequently purchased products.

3 DATA

3.1 Initial Goal

The initial goal of this research was to attempt to correlate Venmo activity with overall web privacy. We planned to collect all transactions over a 1-week period. From that data, we would find the 50 most active users and 50 least active users, and attempt to infer or collect further information about them based on the information collected from their public Venmo transactions. One of the ways that intended to do this is by collecting the user's Facebook ID. This information is included in the API image data for users who registered for a user account by connecting their existing Facebook account. The actor and target image field will contain a URL of the form https://graph.facebook.com/v2.10/<FB_USER_ID>/picture?type=large, where FB_USER_ID is the user's Facebook profile identifier. Using Facebook's Graph API, we intended to pull information about the user, including if their physical location, place of work, and school are publicly available on their profile.

3.2 Limitations

Our initial goal and methodology was limited due to restrictions to both the Venmo and Facebook APIs. Following the release of Public By Default and subsequent attention to Venmo's privacy settings, Venmo imposed a rate limit for calls to their Public API. Through trial and error, we found that this limit is approximately 1 API call every 30 seconds per internet access point. If requests are made more frequently, the HTTP request will be returned with an error message, "Too Many Requests."

Additionally, the data contains fields "previous" and "next", which contain links of the form https://venmo.com/api/v5/public?until=<UNIX_TIME_STAMP>, implying that the next/previous set of data can be retrieved from the link, do not work as expected. Using the links alone to traverse and collect data resulted in transactions from various time ranges. Thus, we were not able to use this collection method to gather public data from a specific time interval. Because we are only able to gather a specific amount of transaction data over a specific time interval, we are not able to provide any analysis of variation in frequency of transactions (such as, do more transactions occur at a specific time or on a specific day).

```
{
  - paging: {
       next: "https://venmo.com/api/v5/public?until=1544155981",
       previous: "https://venmo.com/api/v5/public?since=1544155981"
    },
```

Figure 1: API Data - Next and Previous Paging

The effect of the rate limiting imposed by Venmo could be lessened or avoided completely given enough machines across networks. If enough data is collected and combined - then removing

duplicates based on transaction number - it would be possible to gather every single transaction.

The Facebook API also caused our data collection to be limited. Initially, we believed that we could use the Facebook Graph API to create an automated script to collect information about the user found on their Facebook profile. We believed the API, in conjunction with a user's Facebook ID found from their Venmo transaction data, would allow us to collect specific public fields from the user's Facebook profile. Unfortunately, if this functionality ever existed, it has been discontinued or blocked by Facebook. To gather data about the Facebook users from their Venmo data, we would be required to manually search using the user's first name and last name and attempt to find a profile image that matches the one stored by Venmo. We would then have to manually record the number and type of public fields available on their Facebook profile. For the scale of our project, this was no longer a feasible option and our data collection and research goals were adjusted accordingly.

3.3 Data Collection

Our data set was gathered by accessing the Venmo public API at https://venmo.com/api/v5/public. We retrieved the information from this API every 30 seconds over a 7-day period from November 12 - 18, 2018, with each API call returning the data for 20 transactions. We used a python script to download each set of transactions as a JSON file, each with a unique file name based on the time the file was downloaded. Over the course of our data collection we were able to gather information for 397,820 transactions, consisting of over 20,000 individual files. The data retrieved from the API contains information that is not visible from the mobile app, such as transaction ID (a unique identifier for each transaction), time-stamps, and information about whether either party involved in the transaction is a business. Additionally, specific information about the user is available. For example, rather than rendering the user's profile picture, the data retrieved from the API simply has a link to the image. The format of this link indicates how the user uploaded the image (Facebook authorization or direct image upload to Venmo). In the case of an account which uses Facebook connection, the image URL includes the user's profile ID. The comparison between data visible in the mobile application and the public API can be seen in Figure 2.

4 MEASUREMENT

4.1 Data Compilation

Following the data collection period, we were left with approximately 20,000 individual JSON files containing our data. In order to perform analysis, we combined all transaction pages into a single JSON file. We then ran a Python script that would remove all invalid or duplicate entries from our data set, which resulted in a our total of 397,820 individual transactions.

4.2 Collected Attributes

For our analysis, we chose to measure attributes in the following categories:

- Most Popular Actors and Targets
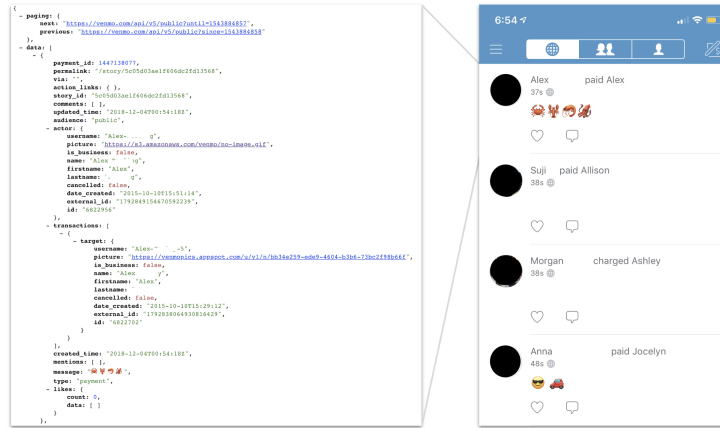- Profile Image Host

Figure 2: Public API Data and Corresponding In-App View

- Transaction Type (Charge or Payment)
- Most Popular Words
- Most Popular Emojis
- Account Creation Date

And for each day of the week,

- Most Popular Words
- Most Popular Emojis

In order to collect the data needed for these attributes and do analysis, we used Python. For each one of the results included in this study, we had a similar process. First, we would load the JSON file into Python. Next we would parse through this transaction information using loops and Python indexing. All of the transaction objects consisted of a hierarchy of dictionaries or lists, as can be seen in Figure 2, so indexing was fairly straightforward and the best option. With each transaction we parsed through, we would collect the piece of information that was valuable, add this information to a list, and then used a counting method to count the instances of each piece of information in the list.

For more clarity here is an example of our account creation dates were collected: First, we parsed through all of the distinct transactions and collected the 'date_created' field from both the 'actor' and the 'target' of the transaction. Then, we add these two dates retrieved to a Python list, and use the Counter method from Collections, which essentially counts up how much each item appears in the list. So if the date 2017-11 was in the list 9 times, Counter would return '2017-11': 9, along with the number of every other date in the list. We used this same approach for every single attribute that is highlighted in this paper. Also note that the parsing of URLs for Profile Image Hosts required the use of urlparse from urllib.parse, and we needed to use the emoji library to help collect unicodes and emojis.

Some issues also came up when we implemented the code for all of the different measurements we performed. Initially, we tried running all of these scripts simultaneously on a Virtual Machine, but kept running into memory errors. Following this, we moved these scripts onto our local machines and they ran fine, but it became hard to maintain and change over 20 different files when logic needed

to be changed or improved. As a result of this, we put all of the scripts into one file which will run one after the other, and this is the final file which we used. It is called 'analyze_script.py' and can be found on our github https://github.com/meghanmc/cs4980_project. Each of the measurements and their values from this file will be explained in further detail in the following sections.

4.2.1 Most Popular Actors and Targets. For each transaction, there is an Actor and a Target specified. The **Actor** is the user who has initiated the transaction. The **Target** is the user who is being paid or charged by the Actor. While transaction cannot exist without both of these values, it is possible for the Target to be specified without an existing Venmo account. For this to occur, the actor must simply specify a phone number, iMessage account, or email that is not yet associated with a Venmo account.

4.2.2 Profile Image Host. The value for Profile Image Host is gathered by analyzing the transaction participant's "picture" field. This field contains the URL of the image that is displayed as the user's profile field in the Venmo mobile app or desktop site. Our analysis values for this field include **Uploaded to Venmo**, **Connected to Facebook**, **Default Image**, and **Other**.

Uploaded to Venmo indicates that the user's picture is hosted by Venmo. This is determined by a "picture" field containing a value with "venmopics.appspot.com".

Connected to Facebook indicates that when creating their Venmo account the user selected the "login with facebook" option. This is indicated by the "picture" field containing a value with "graph.facebook.com". This value will contain a URL of the form https://graph.facebook.com/v2.10/<FB_USER_ID>/picture?type=large, where FB_USER_ID is the use's unique Facebook profile identifier, which can be used to further gather data about the user from Facebook's API.

Default Image indicates that the user has not provided a profile image for their Venmo account, or has deleted an existing profile image. This is indicated by the filed containing the value https://s3.amazonaws.com/venmo/no-image.gif which directs to Venmo's default user graphic.

Other indicates that the user has not yet created an account on Venmo. This is indicated by the field containing a phone number, email, or iMessage account value. Because users must create an account to initiate a transaction, this value can only be found for transaction recipients.

4.2.3 Transaction Type. The value for transaction type is found in the transaction details field called "type". There are exactly two possible values for this field - **Payment** and **Charge**. As of December 2018, the transaction type in the mobile app does not have a default values, which are displayed as "Pay" and "Request", respectively. When a user initiates a transaction, they must select between the two values before they can continue.

Payment indicates that the user initiating the transaction would like to make a payment to the user they have specified. **Charge** indicates that the user initiating the transaction would like to charge the specified user.

4.2.4 Most Popular Words and Emojis. The most popular words and emojis are gathered from the "message" field of the transaction. Venmo requires a message for each transaction, and encourages the use of emojis. In addition to the Unicode emoji set, Venmo has their created their own set of custom emojis, which are are visible in the mobile app but are displayed as ":<EMOJI_ID>:" in our data. An example of this type of emoji is ":uber:" which will display the Uber logo in the mobile app.

4.2.5 Account Creation Date. The value for account creation date is found for each Actor and Target in each transaction in the field "date created". This field contains a timestamp of the date and time that the user created their Venmo account, or there will be no date provided if the user has not yet created an account. We use only the month and year of this value for our analysis.

We will analyze the data gathered through these measurements in the following section.

5 ANALYSIS

5.1 Most Popular Actors and Targets

The images below show a collection of the most popular actors, or transaction initiators, and transaction targets.
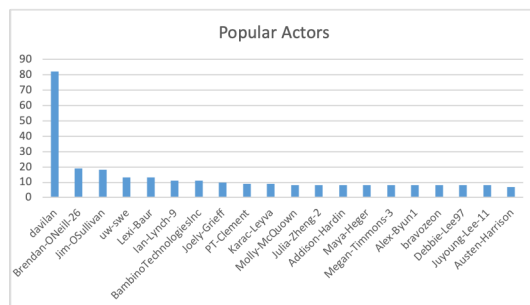

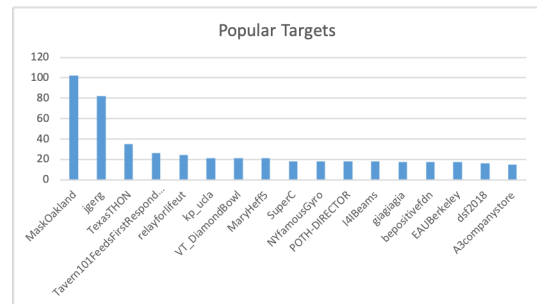**Figure 3: Most Active Actors**

Figure 4: Most Active Targets

Throughout the study, most observed users participate in a single transaction, but there are some users that are much more active, as indicated in the image. Overall, Actors are more active than Targets. This indicates that users who initiate more transactions are more likely to be more active overall. Additionally, users who are more likely to initiate transactions interact with more users overall than those who are transaction targets. Note that because our study estimates user activity based on completed transactions, users who did not complete any transactions are not considered in these totals.

In the image below, we break down targets that have an existing Venmo account compared to those that do not. As described in Section 4 above, a Venmo transaction request can be sent to a person who has not created an account by using the persons email or phone number. For the previous section, Actors, this analysis cannot be done because a user must have an existing Venmo account to initiate a transaction. From this data, we can see that only *.7%* of transactions are sent to a user without an existing account.//**TODO add note somewhere that all conclusions are extrapolations based on 1 week of data being representative of all data**
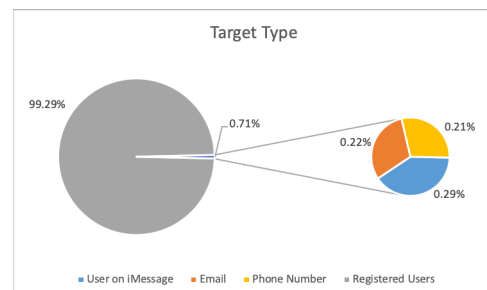

Figure 5: Target Type

5.2 Profile Image Host

Based on our collected data, we can see that the most common (*74%*) profile image upload preference is uploading an image directly to Venmo. Even though Venmo does not require users to upload an image, most active users opt to upload an image of their choice.

The next most popular user choice (*19%*) is not uploading an image at all, which will display a grey image with user initials in the mobile app, and the default shadow image on the web client.

A surprisingly unpopular (>7%) image upload choice is using the "Connect to Facebook" feature, which will automatically populate the user's Venmo profile image with their Facebook profile picture.

The additional category, "Other", is a target that has not yet created a profile, as indicated in the previous section.
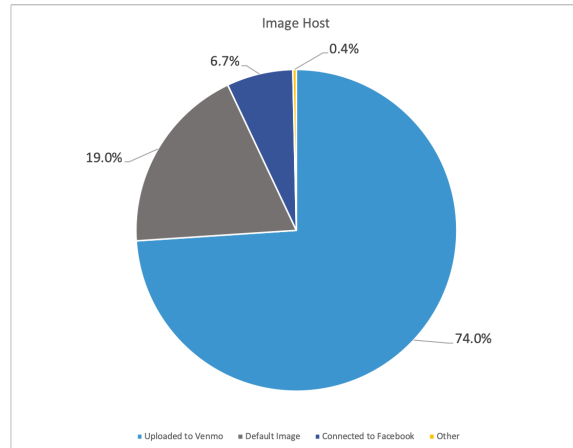

Figure 6: Profile Image Hosts

5.3 Transaction Type

Venmo offers two transaction types, with no default selection. Clearly, users much more prefer to use the Payment feature than the Charge (i.e. "Request") feature. Over *85%* of transactions are initiated as a payment rather than a charge.
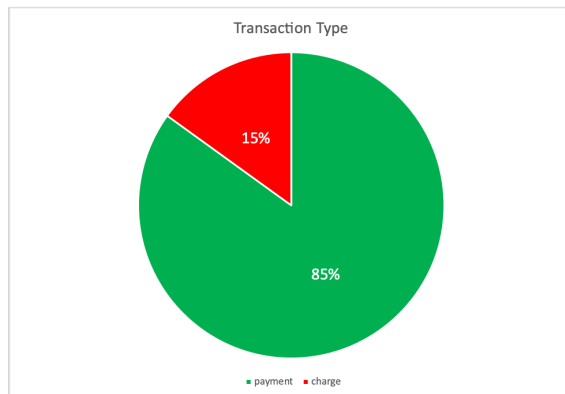

Figure 7: Transaction Type

5.4 Most Popular Words and Emojis

From the "message" field for all 397,820 transactions, we calculated the most frequently used, meaningful words - excluding words like 'for', 'the', 'my', 'your', etc. This distribution can be seen in Figure 8. From this data, we attempted to infer the type of transaction that was being made. For example if the transaction message included words like 'rent', 'bills', 'November', we concluded categorized these transactions together. We also categorized words relating to food -

'dinner', 'pizza' - and transportation - 'uber', 'gas'. Based on these classifiable words, we can approximate that *38%* of transactions on Venmo relate to food, *18%* relate to rent and bill-related expenses, *20%* are transportation related, and the remaining *24%* cannot easily be categorized.
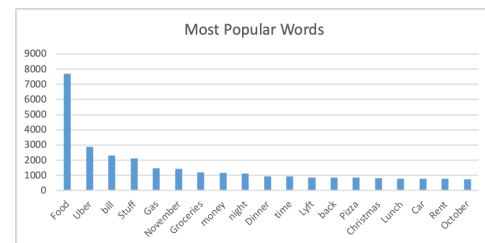

Figure 8: Popular Words

We also performed a separate analysis of frequently used emojis. Because these show up in our dataset as their Unicode value, we did not combine the analysis of popular words with emojis. For example, the pizza emoji does not contribute to the total for the word 'pizza'. However, the frequency of emojis did have a similar distribution when analyzed by category: *38%* of messages contained food-related emojis (pizza slice, taco, etc.), *12%* are for bill-related expenses (electric plug, house emoji) *9%* are transportation related, and the remaining *47%* could not easily be categorized.

Note that all categorizations and conclusions in this section are implied based on message content, as intent cannot be confirmed by the transaction participants.
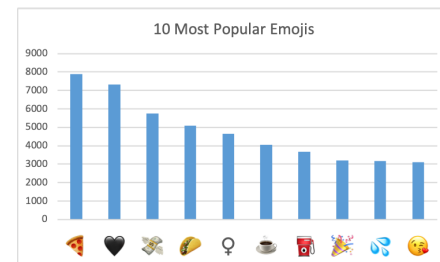

Figure 9: Popular Emojis

5.5 Account Creation Date

Figure 10 on the following page shows the general distribution of transactions by Venmo account creation date (monthly). Note that transaction targets who had not yet created a user account are not included in this diagram. This distribution shows a general trend of more Venmo accounts being created beginning in 2010 (the company was started in 2009), with almost all accounts being created after mid-2012, when the application was released publicly. Because this figure and analysis is based on transactions rather than individual users, it could also be concluded from this chart that users who have created accounts more recently are more active.

From this data, we also found that there are increases in account creation for August and September of each year for 2014 to 2018.
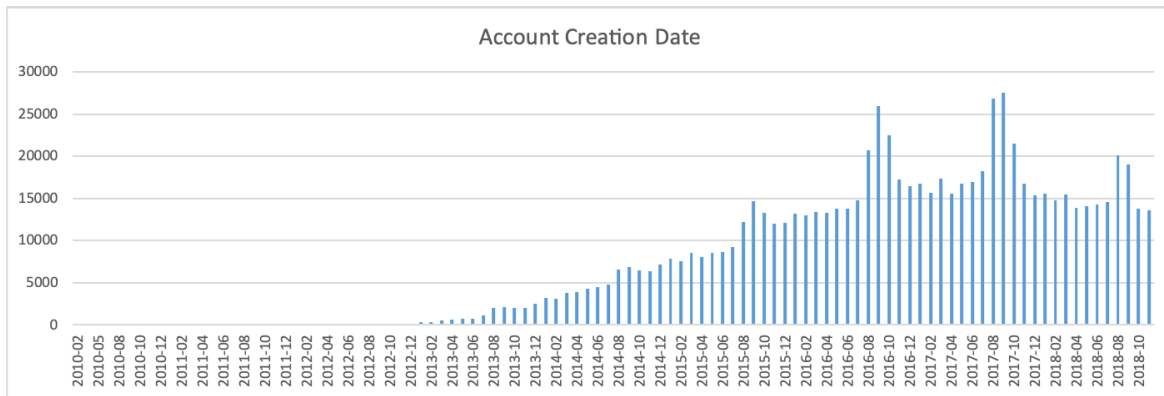
Figure 10: Account Creation Date

At this time we are unable to conclude the reason or meaning of these sharp increases, by we hypothesize that it may be due to college students (one of Venmo's major demographics) starting the school year in apartments in which they have to pay rent to new roommates.

6 CONCLUSION

6.1 Why This Matters

The initial goal of our research was to focus on the correlation between web activity - specifically on the Venmo app - and overall web privacy. The more time that an individual spends on the internet, the more data they are likely to have available to share. Although we were not able to complete our initial research goal, we did provide meaningful data indicating the general state of the Venmo ecosystem. We found that a surprising number of active accounts do not connect their account to Facebook, limiting the amount of information that could be uncovered through that connection.

Bringing attention to the implications of web privacy is important for internet users and companies. Following the release of Public By Default, attention was brought to Venmo's default privacy settings.

Venmo has made several improvements such as adding interactive user applications to help users better understand privacy settings, and making it simple for users to retroactively apply new privacy settings to past transactions (note, if the data were stored by another party by using the public API, there is no way for the user to reclaim that data). [2] However, there is still room for improvements. They continue to defend their choice to have transactions default to "Public" with no review required, even though most social media at a minimum require the user to review their security preferences upon signing up. Venmo said in a statement that, "it is very clear in each payment what audience it is being shared with and we have made this even more prominent in recent years." [3] However, as of December 2018 the only indicator of the privacy of the transaction is a small icon and label in the lower right corner of the transaction dialogue as indicated in the figure below.
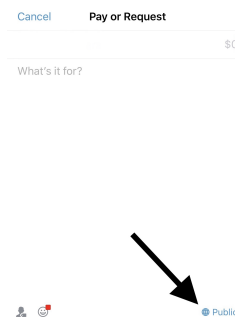

Figure 11: Transaction View - Privacy indicator marked by arrow

6.2 Future Work

Ideally, this analysis would have been performed on the complete set of transactions over a specific time interval. That would have allowed analysis of frequency of transactions. Because our data collection was limited to 20 transactions every 30 seconds, we are not able to be sure of the percentage of transactions we were able to gather and analyze out of the total transactions completed in that time frame.

Additionally, in order to complete our original goal of comparing Venmo activity to overall web privacy, we would need to discover some methodology for gathering Facebook profile information using a user's Facebook ID. At this time, this functionality does not exist.

7 ADDITIONAL MATERIAL

The repository contianing our data collection and analysis scripts can be found at https://github.com/meghanmc/cs4980_project/ . Our dataset of transactions and analysis raw results can be found in the following Google Drive folder https://goo.gl/z8Z6xk .

REFERENCES

[1] Hang Do Thi Duc. 2018. Public By Default. Retrieved December 2, 2018 from https://publicbydefault.fyi
[2] Venmo. 2018. Payment Activity Privacy. Retrieved December 2, 2018 from https://help.venmo.com/hc/en-us/articles/210413717-Payment-Activity-Privacy

[3] Zack Whittaker. 2018. Venmo has no good reason to make user transactions public by default. Retrieved December 2, 2018 from https://www.zdnet.com/article/venmo-refuses-to-say-why-transactions-are-public-by-default/