Cold Hard E-Cash: Friends and Vendors in the Venmo Digital Payments System

Xinyi Zhang[†], Shiliang Tang[†], Yun Zhao[†], Gang Wang[†‡], Haitao Zheng[†] and Ben Y. Zhao[†]
[†]UC Santa Barbara, [‡]Virginia Tech
{xyzhang, shiliang_tang, yunzhao, htzheng, ravenben}@cs.ucsb.edu, gangwang@vt.edu

Abstract

For millions around the globe, digital payment apps such as Venmo are replacing cash as the preferred method of payment between friends and vendors. Apps like Venmo bring a unique blend of convenience and social interactions into financial transactions. In this paper, we study the role of social relationships in the adoption of the Venmo digital payment system. We collect records of all 91 million public transactions conducted on Venmo since its introduction, a social graph connecting most of its 10.5 million users, and analyze the interplay between social relationships and financial transactions. We find that Venmo communities are very densely connected compared to other interaction networks, and are often driven by specific niche applications. We are able to extract both user-to-user and user-to-vendor transaction communities, and show that they exhibit dramatically different structural properties.

Introduction

The mobile revolution has transformed how people handle financial payments, through a variety of mobile payment apps that are replacing cash and credit cards. These apps can be classified into several groups based on their target functionality. Digital wallets are essentially mobile wrappers around physical credit cards, including Apple Pay, Google Wallet, and Visa Checkout. Others, like PayPal, Stripe and Square focus on simplifying payments for vendors. Finally, apps like Venmo bring a unique blend of convenience and social interactions into payments, by supporting simple (and free) person-to-person payments (Sidel and Demos 2016).

The Venmo model for interpersonal payments has had tremendous success in the last few years. Venmo has 11 million users as of May 2016, and has seen transaction volume triple in 2015, reaching $1 billion USD in monthly transactions as of January 2016. Its success in the US has led to the very recent development of Zelle, a competing system created by major US banks including Chase, Citi and Bank of America (Rey 2016), as well as similar systems from Square (Perez 2016), Apple, and Facebook. In China, a similar person-to-person payment system exists in WeChat, which now includes more than 400 million users and $11 billion RMB ($1.65B USD) in transactions in 2014.

Beyond the convenience of a mobile app, what makes person-to-person payment apps like Venmo interesting is their tight integration with a *symbiotic* social network. On one hand, there is ample evidence that usage in social groups is a critical component of Venmo's fast adoption (Bird 2015). Friends who are users provide free advertising and awareness, and even peer pressure whenever payments are involved (*e.g.* sharing a meal). On the other hand, Venmo reduces friction between friends in financial matters, and its social features (comments on transactions) serve to reinforce social links with creativity and inside jokes.

But how has this symbiotic relationship affected users' social and financial behavior? This is the key question we seek to answer. In this paper, we report the results of a large-scale analysis of Venmo transactions, analyzing all public transactions in Venmo[1] totalling 91 million transactions over 6 years, all in the context of an underlying social network connecting 10.5 million users (all friend relationships are public in Venmo). From these traces, we can analyze both the Venmo social graph (composed of friendship links connecting Venmo users) and the Venmo transaction graph (composed of links representing transactions between users).

Our results include a number of surprising findings. *First*, we find that both normal users and businesses populate the Venmo transaction graph, and exhibit dramatically distinctive (and easy to identify) patterns in their transactions with others. *Second*, Venmo users form exceptionally dense communities in the transaction graph, with much higher than expected clustering coefficients. Using k-core decomposition, we find that Venmo transaction communities are similar to or denser than to all available datasets of user interactions (Twitter retweets, Facebook messages). *Third*, analysis of properties of communities show that many are "niche groups" that revolve entirely around a single type of transaction, *e.g.* rent, utilities, gambling or betting pools. Some of these groups are ephemeral and users turn dormant once the specific event (*e.g.* NFL Super Bowl) passes. This suggests Venmo is used by many as a specific application-driven utility rather than a social payment network.

To the best of our knowledge, this is the first large-scale analysis of financial transactions on person-to-person pay-

[1]By default, Venmo users have privacy settings set to share transactions (users, time, comments but no amounts) with all users.

ment systems like Venmo. In the remainder of our paper, we give background on mobile payments and the Venmo app, then describe our data collection and initial analysis. Next, we analyze the structure of Venmo's social and transaction graphs and show how they overlap. We use unsupervised learning to classify users by their transaction and social behavior; identify and study patterns found in communities in the transaction graph. Finally, we analyze Venmo transactions by both payment types and temporal dynamics.

Background & Related Work

Mobile Payments. Mobile payments fall into two general categories, one being mobile extension of credits cards, the other being mobile wallet services. For contact-less extensions of credit cards like Apple Pay, Samsung Pay and Android Pay, they only act as a wireless layer over credit cards. In comparison, mobile wallet services are heavier in functionality. Traditional mobile wallets such as PayPal and Alipay are spawn from large online shopping sites, with ecosystems built around merchants and customers.

Usage on mobile wallet services also attracts extensive research efforts. A lot of studies focus on the use case of different services in different countries, *e.g.* M-PESA in Kenya (Suri and others 2012), Bristol Pound in England (Ferreira and others 2015), bKash in Bangladesh (Hasan and Islam 2013), and mobile money offered by Network Operators in Uganda (Ndiwalana and others 2010). These studies typically deploy survey or interview to gather user data. Other study leverages on a theoretical framework called Technology Acceptance Model (Davis and others 1989), and looks at how different factors affect user adoption of these services (Park and Lee 2014).

In recent years, there emerges a new trend of social payments where a wallet builds a social network within itself. These services are eating into the market of traditional mobile wallets, examples being Venmo and WeChat Pay. Known for its convenient peer-to-peer transfer, Venmo quickly spread through word of mouth. In the year 2015, Venmo increased its transaction volume by 200% (PayPal 2016a), taking up 19% of the market share of mobile user-to-user payments in US (Sidel and Wakabayashi 2015).

The Venmo App. Venmo has two main functions: making transactions and socializing. First, Venmo lets user pay each other simply by specifying the receiver's Venmo ID, the amount, and a short descriptive text message associated to the payment. Transaction is made easy as users can quickly locate the receiver by searching among her Venmo friends. Second, Venmo users have the option to share their payments with their friends or with the public. Once shared, these transactions are streamed into a feed with the time, recipient, and message displayed to the audience. Fortunately, Venmo provides APIs to query public transactions and social connections, which makes it feasible to gather a dataset of all public activities on Venmo, and thus performing large-scale quantitative analysis on financial behaviors.

Interactions on Social Network. As a payment platform built on a social network, Venmo introduced a brand-new type of social interaction: making transactions. There have been extensive works studying different types of interactions in Online Social Networks. Interactions being studied including wall-post on Facebook (Wilson et al. 2009; 2012), retweets on Twitter (Kwak et al. 2010), reblog on Pinterest (Gilbert and others 2013) and Tumblr (Chang and others 2014), editing on Wikipedia (Crandall and others 2008), just to name a few. There are also works using detailed click-streams to study latent behavior that are not directly visible online, *e.g.*, profile browsing (Schneider and others 2009; Jiang et al. 2010; Metzger, Wilson, and Zhao 2017). Our work differs from theirs because our topic of study is a combination of financial activity and social activity, which introduces an interaction incentive that has long been present in the financial world, yet never seen in social networks.

Digital Transactions. Besides Venmo, Bitcoin is the only source of large-scale public records of transaction data. Most previous works utilizing this dataset is oriented towards the anonymity in Bitcoin (Ober and others 2013; Meiklejohn and others 2013). Ron *et al.* analyzed graph properties in Bitcoin (Ron and Shamir 2013). They provide basic distribution statistics for transactions on Bitcoin, and perform detailed studies on 364 transactions. These works struggled with the anonymous nature of Bitcoin, and did not perform behavior analysis beyond case studies. In contrast, transactions on Venmo are associated with real accounts and support in-depth analysis of user behavior. Finally, prior work has also analyzed the impact of social connections on the Overstock marketplace (Swamynathan et al. 2008).

Data & Initial Analysis

In this section, we start by describing our data collection methodology and datasets. Then we perform preliminary analysis to understand Venmo's user activities and growth trend. This provides context for studies in later sections.

Data Collection

We collect a complete set of public transaction records on Venmo over 6 years and its social network graph through public APIs (Venmo 2016). We received approval from our local IRB for our study, and carefully anonymized userIDs and user names in the collected dataset. We limited query rates to avoid disruption to Venmo's services. While the data we obtained is publicly accessible via Venmo's APIs, we are cognizant of deanonymization risks from releasing the entire dataset to the public. We are reaching out to Venmo to negotiate a possible release of a subset of the dataset.

Public Transaction Records. Venmo API allows us to query the *historical* public transaction stream of the entire network by specifying a time range. We use the API to sweep through the timestamps from Venmo's initial launch to May 5, 2016, and collect a complete set of public transaction records. In total, we obtain 91,355,414 transactions over 6 years from April 15, 2010 to May 5, 2016. Venmo went online in August 2009 as beta, posted its first public transactions in 2010, and was open to public users in March 2012. Each transaction record contains a transaction ID, sender,

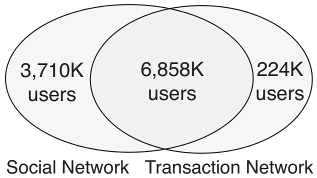
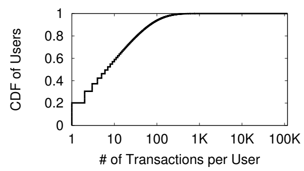


Figure 1: Venn diagram: social and transaction network.



(a) transaction count. (b) friend count.

Figure 2: Per user activity distribution.

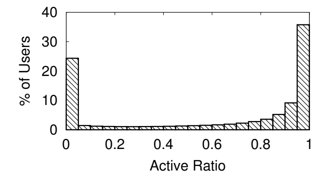


Figure 3: Distribution of active ratio of users.

receiver, transaction type, transaction time and related social activities (messages, likes, comments). For each user involved in the transaction, the record contains user profile information including userID, account creation time, and the first and last name of the user. In total, we extract 7,091,915 unique userIDs.

Note that all the transactions have a message to indicate the purpose of the transaction. Comments and likes, however, are less prevalent: only 2.7% transactions have comments and 11.3% of transactions have likes.

The Venmo Social Graph. Venmo's dual functionality as a payment network and a social network means that the two networks only overlap partially. Some users participated in transactions have no friends in Venmo, others have friends but have not participated in transactions. To build Venmo's social graph, we began by using public APIs to query each users' friend list. We observed that Venmo uses sequential numbers for userIDs (starting from 1). We validate this by creating a burst of 10 new accounts within 20 seconds, and confirming that the resulting userIDs are sequential integers. Thus, we can use a newly created userID to estimate the number of total registered users. As of May 5, 2016, the estimated total number of users is 10,586,252. We build a list of the entire Venmo user population by sequentially scanning the userID space, downloading each user's friend list for the complete social graph.

We crawl the social graph with focus on users registered before May 5, 2016, which gives us 10,568,274 users. Note that this number is slightly smaller than the total number of registered users (10,586,252) as of May 5, 2016. This is because 0.17% of userIDs are reported as "invalid" by the API, possibly due to account deletion. After excluding another 806,625 (7.6%) users who have no friends, the final social graph contains 9,761,649 users.

Coverage Estimation. Our social graph is complete, but our transaction data only cover public transactions. First, based on the sequential userID, we estimate there are 10,586,252 registered users as of May 5, 2016. Our public transaction dataset covers 67% of the user population. The rest of the users either did not make any transactions or only made private transactions.

Second, we estimate the number of private transactions. Just like the userIDs, we find the transaction IDs are also sequentially assigned. We validate this by creating 10 private transactions interleaving with 10 public transactions within 20 seconds. We find that the transaction IDs also increase monotonically. Based on the maximum transaction ID, we infer that there are 185,270,948 transactions up to our data collection time, and our dataset covers 49.3% of all transactions. The rest 50.7% of transactions are private. In this study, we seek to leverage the public transactions as a proxy to study the digital payment activities of Venmo users.

Preliminary Analysis

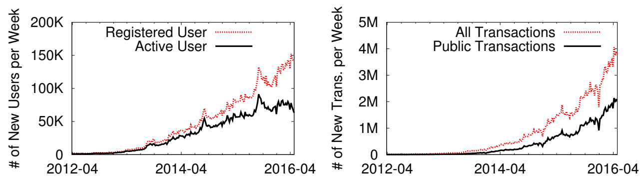
Social and Financial Activities. We first examine user participation in social and financial activities. Figure 1 uses a Venn diagram to show the overlap between users in the social graph and users in the transaction dataset. Most users (6.86M) participate in both financial transactions and social friending. This is only a lower bound — the 3,710K users who have no public transactions may still have private transactions. Only 224K users (2.1%) use Venmo for financial transactions but do not have any friends.

Figure 2(a) shows the number of transactions per user, which follows a long tail distribution. Most users (57%) have made less than 10 transactions, while certain users have made more than 10,000 transactions. A closer examination shows that these super active users are charity organizations and business owners. Compared to making transactions, Figure 2(b) shows users are more active in adding friends. Half of the users have at least 40 friends, and 30% of users have more than 100 friends.

Long-term vs. Short-lived Users. To examine the level of user engagement, we measure a user's *lifetime* which is the time difference between a user's first and last transaction. Only users with at least one transaction are considered. We find that 22.5% users used Venmo for less than a day. These are "try-and-quit" users who installed the app to make a transaction and then quickly abandoned it. In contrast, 30% of users have actively used Venmo for over a year.

To better depict the long-term and short-lived users, we calculate *active ratio*, which is the ratio of a user's active lifetime over her longest possible lifetime (time difference between the first transaction and the last day of our data collection). Figure 3 shows a clear bimodal distribution where most users are distributed to the two extremes. This indicates users would either like Venmo thus stay on the network for a long time, or quickly give it up after the initial try.

Venmo's Growth. Finally, we examine the growth trend of Venmo. In Figure 4, we can observe a *super linear* growth for both Venmo user population and the transaction count. The total number of registered users over time



(a) Number of new users. (b) Number of new transactions.

Figure 4: Weekly growth trend of Venmo.

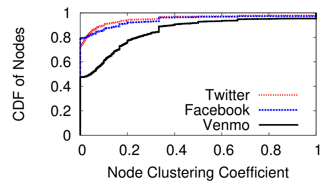


Figure 5: Distribution of clustering coefficient per user.

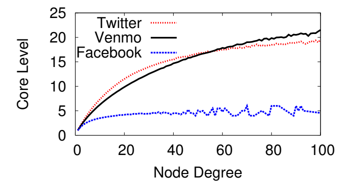


Figure 6: Average core level as a function of degree.

Dataset	# of Nodes	# of Edges	Avg. Degree	Cluster. Coeff.	Path Length	Assort. Coeff.
Venmo	9.8M	544M	111	0.130	4.3	0.38
FB	10.7M	408M	76.3	0.164	4.8	0.17
Renren	10.6M	200M	37.8	0.142	4.8	0.0045

Table 2: Comparing Venmo, Facebook (Wilson et al. 2009), Renren (Zhao et al. 2012) social graphs.

follows a power series model ($P(x) \propto ax^b$) with b=3.05 (R^2 = 0.9999). The total number of transactions follows a power series model with steeper increase, b=4.30 (R^2 = 0.9999). The growth of transaction count is highly consistent ($R^2 = 0.997$) with the reported growth in transaction volume (2013–2016) (PayPal 2016b), showing that our dataset is a faithful reflection of Venmo activities.

When measuring per user activity, we find users' average transaction frequency almost tripled in the four years since Venmo came exited beta in April 2012. Venmo is showing healthy growth in both overall scale and user engagement.

Transaction & Social Graphs

We now analyze our data to study the interplay between social relationships and financial transactions on Venmo. We seek to understand the role of social relationships in the adoption and usage of Venmo. In the rest of the paper, we focus on three sets of related questions. *First*, how much has Venmo's social component affected its functionality and design? What are the key differences between Venmo and other online social networks? How do social relationships shape the way users make financial transactions? We address these questions in this section. *Second*, how much do users' social friends and transaction patterns reveal about their identity as vendors or normal users? What drive users to form distinct social and financial communities? We answer these questions in the next section, Users & Communities. *Third*, what do users use Venmo to pay for? How does such spending pattern change over time? These questions are discussed in the section after, Payment Types & Dynamics.

In this section, we focus on the first set of questions, to examine whether and how Venmo differs from traditional online social networks. We build both a *social graph* and a financial *transaction graph* from Venmo, and compare the graph properties with those of existing online social networks. To further explore how social relationships impact financial transactions, we use social connections to divide Venmo's transaction graph into a friend-only transaction graph, and a stranger-based transaction graph. We examine the key differences between the two and their implications.

Transaction Graph

We start by constructing a financial transaction graph for Venmo, where each node is a user and each edge (directed) represents a payment relationship between two users. The weight of the edge represents the total number of (directed) financial transactions between the two users. While building the graph, we find 0.35% of the transactions reported their target as "a phone number" or "an email address," thus cannot be directed to any single entity. We omit these transactions from the graph.

We compare key graph properties of the Venmo transaction graph with interaction graphs of Facebook wall-posts (Wilson et al. 2009) and Twitter retweets (Xu et al. 2011) in the top half of Table 1. We find that Venmo's transaction graph shows strong "small-world" properties (Watts and Strogatz 1998) with high degree and clustering coefficient, small average path length and densely connected core user groups. These properties are commonly observed in social networks where a group of friends closely interact with each other. More importantly, compared to pure social interactions, Venmo displays a much higher local clustering, reflecting the structure of stronger friendships often required by financial relationships. Next, we briefly explain and compare key graph properties.

Clustering Coefficient. Clustering coefficient is the number of edges between a user's immediate neighbors divided by all possible connections that could exist among them. It measures the level of local connectivity between users. Venmo's clustering coefficient (0.147) is much higher than Facebook (0.059) and Twitter (0.048) (Figure 5).

However, even in Venmo, a significant number of users have clustering coefficients of 0. One major reason is that transactions in Venmo follow a long-tail distribution, with many Venmo users (29.34%) partaking in only one transaction, resulting in a clustering coefficient of 0. Despite that, we find 39.09% of Venmo users have clustering coefficients more than 0.1, whereas the numbers are only 12.84% and 9.42% for Facebook and Twitter, indicating that Venmo users are more likely to make financial transactions within tightly connected groups or communities.

K-core Decomposition. K-core decomposition examines network connectivity by recursively stripping off peripheral nodes from the network. K-cores exist when users at level k have made transactions with at least k peers who are also at

Dataset	# of Nodes	# of Edges	Avg. Degree	Tie Strength	Clustering Coefficient	Avg. Path Length	Assortativity Coefficient	Avg. Reciprocity	Largest SCC	Largest WCC
Venmo Transactions	7.08M	35.0M	9.89	3.22	0.147	6.98	-0.0022	0.147	56.10%	95.50%
Facebook Wall Post	707K	1.26M	3.57	1.77	0.059	10.13	0.116	0.126	21.20%	84.80%
Twitter Retweet	4.32M	17.0M	7.86	2.07	0.048	5.52	-0.025	0.025	14.20%	97.20%
Venmo Friends Transactions	6.36M	24.9M	7.83	3.82	0.140	8.64	0.389	0.174	54.49%	93.69%
Venmo Strangers Transactions	4.26M	10.1M	4.75	1.92	0.036	7.84	-0.00552	0.087	34.43%	87.27%

Table 1: Comparison between Venmo transaction graph and the interaction graphs in Facebook (Wilson et al. 2009) and Twitter (Xu et al. 2011). Venmo transaction graph is further divided into friend- and stranger-transaction graphs based on whether a transaction is made between friends.

level k. Figure 6 compares K-core connectivity of Venmo, Twitter retweets and Facebook wall posts. Venmo and Twitter both show very dense local interaction groups of highly active users, while the local clusters for Facebook are much weaker. Prior analysis has shown that Twitter retweets form densely connected groups that capture real-world social relationships (Bild and others 2015). The similarly dense local clusters in Venmo's transaction graph suggest a strong correlation between real offline friendships and transactions.

Average Path Length. Average path length is the average shortest path length between all node pairs in the largest connected component. To estimate average path length, we randomly sample 1000 nodes and compute their shortest path to all the nodes in the graph. Venmo's average path length (6.98) is higher than Twitter's 5.52, indicating more focus on local connectivity; yet it is lower than Facebook's 10.13, possibly a result of Facebook's lower average degree.

Average Reciprocity. Reciprocity measures how likely interactions occur on both directions for a user pair. Venmo's reciprocity (0.147) is similar to Facebook (0.126) and higher than Twitter (0.025). This shows Venmo users are more likely to engage in bidirectional interactions — both sending (and receiving) money to (from) the other user, suggesting that person-to-person transactions are more prevalent than customer-vendor payments.

Assortativity. Assortativity measures the probability for nodes to connect to other nodes of similar degrees. A more positive assortativity indicates users tend to interact with other users of similar degrees. Venmo's assortativity is nearly zero (-0.0022) and lower than friend-only graphs like Facebook (0.116). This because interactions exist between both similar-degree nodes (*e.g.*, friends) and dissimilar-degree nodes (*e.g.* Vendors).

Additional Validation. Since the above analysis uses the interaction graphs from Facebook and Twitter covering only a three-month period, we further validate the conclusions by constructing a smaller Venmo graph using Venmo transactions in the most recent three months of our dataset. Both average degree and tie strength for Venmo show a notable dip (due to the reduced data volume); for all other features, we obtain the same conclusions as above.

Social Graph

Next, we compare Venmo's social graph with those datasets of existing online social networks, Facebook and Renren (Chinese Facebook), provided by (Wilson et al. 2009; Jiang et al. 2010). We also compare it to a more recent and complete Facebook social graph (Ugander and others 2011), leading to similar results (omitted for brevity). We did not include Twitter since its asymmetric follow relationships do not reflect offline friendships. Table 2 lists the key graph properties. We see that Venmo's social graph is very similar to traditional online social networks in clustering coefficient and average path lengths, with a slightly higher average degree. This can be partially attributed to the fact that Venmo allows user to import their Facebook friends to bootstrap their social network.

Notably, Venmo's social graph has an extremely high assortativity (0.38) compared to Facebook (0.17) and Renren (0.0045). This indicates that Venmo users have strong inclination to befriend users of similar degree. This high level of local homophily is also a key property of offline social relationships (McPherson and others 2001). In addition, its assortativity is much higher than that of the financial transaction graph (-0.0022). This is likely because users might have transactions with high-degree nodes like merchants and vendors, but typically do not add them as friends.

Transactions Between Friends

Venmo's social and transaction graphs resemble existing online social networks in some metrics, but differ in other key metrics. Next, we seek to better understand such differences by further exploring the impact of social relationships on financial transactions, examining key differences in transactions made between friends and strangers.

We first compare and contrast social and transaction graphs to see what portion of transactions take place between friends. Among all edges in the social graph, only 3.55% overlap with transaction graph. Even accounting for the possible private transactions, this still indicates users only make transactions with a small portion of their friends. In fact, most users (70%) only transfer money to or from less than 10% of their friends. On the other hand, among all edges in the transaction graph, 80% of them overlap with the social graph. This indicates transactions among friends are more common than those among strangers. This also explains why Venmo transactions exhibit similar properties to social interactions as observed in transaction graph analysis.

To understand the different transaction patterns among friends and strangers, we divide the transaction graph into two subgraphs: a *friend transaction graph* that captures transactions between social friends, and a *stranger transaction graph* that captures transactions between non-friends (strangers). Key graph properties are shown at bottom of Table 1. The friend transaction graph shows clear person-to-person transaction patterns with strong network effects, while the stranger graph captures a customer-vendor model.

Degree and Tie Strength. Tie strength measures the average number of transactions for all edges. The friend transaction graph has a much higher degree (7.83) and tie strength (3.82) than the stranger graph (4.75 and 1.92). This indicates sustainable financial relationships among friends. In contrast, transactions between strangers are more likely one-time payments between customers and vendors.

Assortativity. The friend transaction graph has an extremely high assortativity (0.389). This suggests a network effect on Venmo where users' financial transactions are heavily influenced by their friends, leading to strong local homophily. And the stranger graph's assortativity is close to zero (-0.00552), indicating no significant influence from strangers. It is worth noting that, while most Venmo transactions take place between friends, the high assortativity of friend transaction is hidden when inspecting the overall transaction graph (-0.0022). This demonstrates the benefit of integrating social information into transaction analysis.

Clustering Coefficient. The clustering coefficient of the stranger graph (0.036) is much lower than that of the friend graph (0.140). This is likely the result of customer-vendor relationships in the stranger graph. Intuitively, a vendor's customer is unlikely to have financial transactions with other customers. Similarly, different vendors of the same customer are unlikely to transact with each other.

Average Reciprocity. The low reciprocity of the stranger graph (0.087) is only half of the friend graph (0.164), indicating a vendor-customer relationship: financial transactions between a customer and a vendor are highly directional. This also suggests the possibility of identifying distinct roles (*e.g.*, users vs. vendors) in the Venmo network.

Dynamics. Venmo also shows an increasing trend in stranger transactions, growing from 5.5% from the start of 2014 to 24.4% by May 2016, highlighting the importance of studying the different natures of interaction.

Users & Communities

Our graph analysis showed distinctive patterns in how social relationships affected user behavior. In this section, we explore whether and how much users' social relationships and transaction patterns *reveal who they are*.

In the following, we use various techniques to profile (or classify) users into semantically meaningful user groups. By analyzing these groups, we seek to understand different user types and communities on Venmo. More specifically, we experiment with three different ways to group users. First, we group users based on their behavioral features. By clustering users with similar behavioral patterns, we identify distinct user types in Venmo. Second, we search for communities in the *financial transaction graph* that capture frequent transactions within a group. We explore key factors that drive users to form such communities. Third, we use similar methods to identify communities in the *social graph*, and examine differences between social and financial communities.

Clustering Users based on Behavior

To identify prevalent user types in Venmo, we cluster users based on their behavior. Then we analyze identified clusters to infer and understand different user types.

Behavior Clustering via Similarity Graph. We cluster distinct user behaviors by constructing and partitioning a behavioral similarity graph (Wang et al. 2016). Each node is a Venmo user and each edge captures the similarity in behavioral traces of its two endpoints. We can identify groups of users with similar behavior by partitioning this similarity graph, with no need of pre-defined labels.

To build the similarity graph, we need to measure the behavioral *similarity* between any two users, capturing key aspects of user behavior. Based on results in the previous section, we select three key features:

- **Activity Level:** The number of transactions the user had.
- **Local Connectivity:** Clustering coefficient of the user in the transaction graph.
- **Transactions w/ friends:** Portion of the user's transactions that involved friends.

We compute a feature vector for each user (min-max normalized) and measure the similarity between any two users based on the Euclidean distance of their feature vectors and construct the similarity graph.

We detect clusters in the similarity graph by partitioning it using the Divisive Hierarchical Clustering algorithm (Kaufman and Rousseeuw 2009). This algorithm divides the similarity graph into small subgraphs by minimizing edge weight cut. We stop the graph partitioning process when the overall clustering quality, measured by modularity, plateaus.

We apply this clustering methodology to all Venmo users except those "try-and-quit" users (active ratio <0.23 in Figure 3), leaving $5{,}046{,}348$ users. Directly clustering all 5 million users is computationally challenging. Instead, we apply incremental clustering. We first randomly sample 100K users, and perform clustering to generate the initial clusters. Then we incrementally assign the remaining users to existing clusters based on their nearest neighbors in the sampled set. To validate our results, we calculate probability distributions of all features before and after incremental clustering. Results from the Kolmogorov–Smirnov test (Chakravarti and Laha 1967) show that the difference of distributions is insignificant ($p > 0.18$ for all features).

Understanding Behavior Clusters. Our clustering algorithm produces five clusters. We manually label each cluster leveraging two information sources: 1) feature distribution of each cluster and 2) keyword analysis for messages in users' transaction records. The distribution of the three features for each cluster is shown in Figure 7. For keyword analysis, we rank keywords for each cluster based on Chi-square statistics (Yang and Pedersen 1997), after stemming (Porter 1980) and stop words filtering[2]. Chi-square statistics mea-

[2] `http://www.textfixer.com/resources/common-english-words.txt`

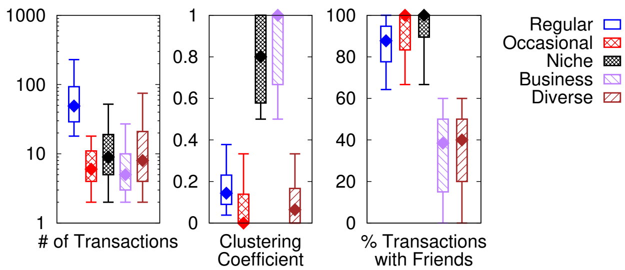


Figure 7: Comparing different clusters on 3 key features. We depict each distribution with box plot quantiles (5%, 25%, 50%, 75%, 95%).

Users	Top 15 Keywords
Regular Users	📺, cabl, comcast, 💡, electr, ⚡, cab, 🚖, water, ga, internet, beer, 🍻, za, uber
Occasional Users	phone, fantasi, 📱, loan, car, footbal, love, payment, test, ticket, cell, birthday, insur, mom, 🏠 💸
Niche Users	phone, 📱, cell, ga, electr, ⚡, ⛽, 💡, groceri, water, insur, internet, 💦, pizza, 🍕
Business Owners	splitwis, wte, ibotta, cashout, lunch, earn, id, brand, app, favorit, bachelorett, 🍴, way, formal, check
Diverse Users	cashout, ibotta, earn, app, brand, id, favorit, check, way, cash, fun, withdrew, signupbonu, thank, bonu

Table 3: Top keywords (after stemming) for user clusters.

sures how strongly (or exclusively) a keyword is associated to a particular cluster. Table 3 shows the top 15 keywords for each cluster. Combining Figure 7 and Table 3, we label the five major user types as:

- **Regular Users** (35.64**%).** Common user type on Venmo, with a large number of transactions, mostly with their friends. Keywords show they use Venmo to pay for utility, drink, transportation and other daily expenses.
- **Occasional Users** (35.66**%).** They use Venmo infrequently, but almost always transact with friends. Payments are often limited to occasional events like birthdays, tickets or sports betting (*e.g.*, fantasy football).
- **Niche Users** (7.97**%).** These users almost exclusively make transactions with their friends in tight-knit communities. Payments focus on utility bills and groceries, indicating they are groups of close friends or even roommates.
- **Business Owners** (1.78**%).** These users are likely to transact with strangers. However, people they interact with are also making transactions with each other (high clustering coefficients). We suspect these are "small business owners" dealing within a group of customers.
- **Diverse Users** (18.95**%).** These users mostly make transactions with strangers, and the people they interact with don't interact with each other (low clustering coefficient). We suspect they are mostly vendors.

Case Studies. Although Venmo is designed for person-to-person payments between friends, we find distinct clusters that may represent vendors and business owners. For more insights on these users, we take a closer look at related behavior clusters (Business Owners and Diverse Users).

The *Business Owners* cluster contains small business owners. For example, one user we examined had 88 transactions, a clustering coefficient of 0.81 and 51% transactions conducted with friends. Since September 2015, this user started to charge fees from 7 other users (possible tenants) on a monthly basis. Many transactions are related to utility bills: 16 for the Internet, 21 for electricity, 10 for water, and 15 for TV services. The rest are related to personal expenses. We notice that a small number of non-business owners are grouped into this cluster because they did not bother to add social friends on Venmo.

Second, the *Diverse Users* cluster contains a mixture of large business owners, vendors and normal users who use Venmo for diverse purposes. For example, one user has 65 transactions, a clustering coefficient of 0.08, and 55% transaction with friends. She not only splits fees with her friends on food and groceries, but also use Venmo to make business payments, e.g. Airbnb.

There are also large vendors in *Diverse Users*. For example, Ibotta[3] is a business that uses Venmo to send cash rebates to customers. It has 119,123 transactions, with a clustering coefficient of 3×10^{-5}, and 0.85% of its transactions are with friends. All transactions have the same message ("Cashed out from Ibotta."). Possibly because these large vendors are relatively rare, the clustering algorithm did not make them a separated cluster.

Communities on Transaction Graph

During behavior clustering, we are able to identify distinct groups of users, separating merchants and customers, according to their behavior. We next analyze user groups (communities) based on their interconnectivity in the transaction graph, leveraging community detection algorithms. Our goal is to identify distinct types of financial communities and understand the roles they play in Venmo's ecosystem.

Identifying Communities. A community on the transaction graph represents a group of people who constantly make financial transactions among each other but barely interact with the rest of the world. We apply Louvain (Blondel and others 2008), a popular modularity-based community detection algorithm, on the transaction graph. We tested alternative algorithms such as Infomap (Rosvall and Bergstrom 2008), and the overall results are consistent, thus omitted for brevity.

Our community detection produces 815 communities with modularity 0.836. In practice, modularity > 0.3 already indicates meaningful community structures (Kwak and others 2009). Venmo's transaction graph has an extremely high modularity, with 85.7% of the transactions taking place within communities. The sizes of communities are skewed, shown as the dotted red line in Figure 8.

[3] https://ibotta.com/

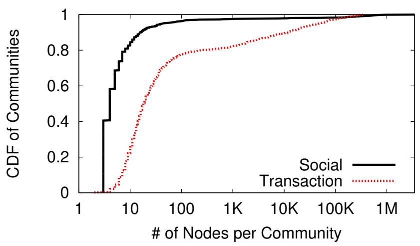


Figure 8: Size distribution of communities.

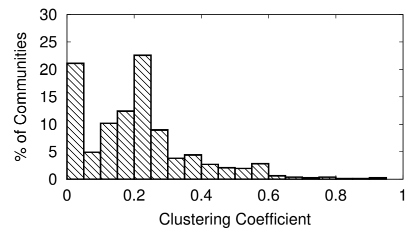


Figure 9: Clustering coefficient of financial communities.

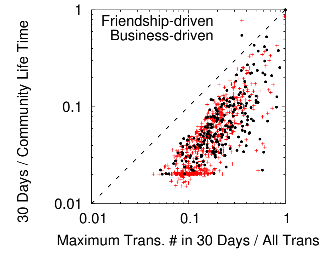


Figure 10: Transaction burstiness in financial communities.

Categorizing Communities. We characterize different communities based on various graph metrics in Table 1, and find there are two major community types (*business-driven* and *friend-driven*), most effectively identified using clustering coefficients. As shown in Figure 9, these two types are located near clustering coefficients of 0 (business) and 0.2 (friend). Simple parameter testing shows a threshold of 0.11 can identify the type of community with 90% accuracy on manually labeled sample set of 100 communities.

We apply this threshold to all the communities and identify 592 friend-driven communities and 223 business-driven communities. We observe that most business-driven communities have a "star" structure with the business owner in the center. This is reflected in business-driven communities' *assortativity* (-0.52, SD=0.24) being much lower compared to that of friend-driven communities (-0.16, SD=0.32). In addition, we also observe a significant difference in *user active ratio*, defined in the Initial Analysis section to measure how long users actively use Venmo to make payments. Business-driven communities have a much lower median user active ratio (0.43, SD=0.32) compared to friendship-driven communities (0.69, SD=0.26). This indicates friendship-driven communities have a higher level of stickiness, better able to keep users in the system, consistent with previous finding that social ties play an important role in retaining and engaging users (Katona and others 2011). We find most friendship-driven communities are mesh-like, built for friends to split fees on food and rent.

Case Studies. Next, we discuss examples for specific communities and examine what drives their formation. Many *business-driven* communities are formed because of major events. For example, one community (186 users) revolves around the "New Year Event" in Chicago and Rochester, where two sellers formed two connected star-shapes (low clustering coefficient 0.079). These two sellers are grouped into one community because some users purchase tickets from both. The community is ephemeral, gets active around the new year in both 2015 and 2016. Here, 104 out of 186 users only used Venmo once to buy the tickets, and almost half of all transactions take place within one-month before the new year.

We take a closer look at the ephemeral nature of communities. We locate for each community i the 30-day period that has the maximum transaction count T_i, and study two metrics: the ratio of T_i and all the transactions of the community, and the ratio of 30 days and the lifespan of the community. Figure 10) plots these two metrics across all the communities. Here a community with a constant rate of transactions will produce a point on the diagonal line. The further away a community is from the diagonal line, the more bursty the community is. When we look closer, these outlier communities are generally formed because of specific events, *e.g.* graduation, sports betting, group trips.

Communities on Social Graph

Applying Louvain on the social graph produces 925 communities, with a modularity of 0.56. The sizes of communities exhibit a highly-skewed distribution, as shown in Figure 8. To further understand the nature of these communities, we manually examine all 62 communities with more than 30 users, labeling 26 of them as business-driven and 25 as friendship-driven, while the remaining 11 are not identified due to lack of information. We discovered that these business-driven communities tend to be smaller in size, with only one business-driven community having more than a thousand users whereas only 3 friendship-driven communities fall below this size. This is likely because merchants rarely add their customers as friends. In general, business-driven communities tend to have very low assortativity (-0.727, SD=0.271), forming a star shape around businesses. Whereas friendship-driven communities display strong social influence, as indicated by their high assortativity (0.265, SD=0.214).

Payment Types & Dynamics

In this section, we analyze the types of payments to understand what people use Venmo for. We first classify payment types based on text/emoji in each transaction message, and then examine the dynamics of these different types.

Classifying Payment Scenarios

Inferring payment types from messages is challenging due to the extremely short message length. As shown in Figure 11, 99% of messages have less than 10 words. Meanwhile, emojis are very helpful for classifying transactions: 34% messages contain at least one emoji. Here, we analyze both keywords and emoji to classify payment types.

We first identify key categories of payments by manually examining top keywords and all the emojis. Six most used categories are thus identified:

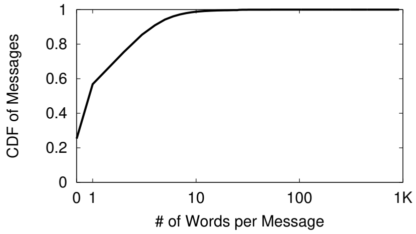


Figure 11: Number of words per message.

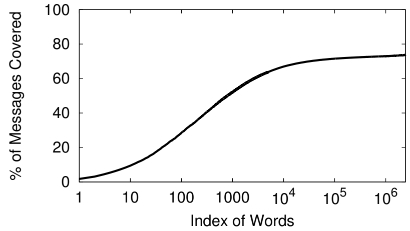


Figure 12: Cumulated # of messages hit by top keywords.

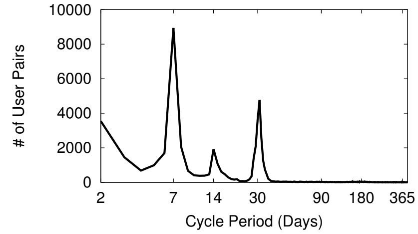


Figure 13: Recurrence period distribution among user pairs.

Category	Un-identified(%)	Food & Drink (%)	Transport (%)	Utilities (%)	Entertainment (%)	Life (%)	Home (%)
Word (68,292K)	47,592K (69.69%)	9,552K (13.99%)	3,745K (5.48%)	3,187K (4.67%)	2,009K (2.94%)	1,707K (2.50%)	1,386K (2.03%)
Emoji (31,475K)	13,233K (42.04%)	10,291K (32.70%)	2,716K (8.63%)	1,910K (6.07%)	2,674K (8.50%)	1,269K (4.03%)	441K (1.40%)
Word+Emoji (81,278K)	433,64K (53.35%)	19,358K (23.82%)	6,325K (7.78%)	4,974K (6.12%)	4,755K (5.85%)	2,928K (3.60%)	1,807K (2.22%)

Table 4: Transaction categorization using keywords and emoji. Messages without any word or emoji are not included. If a transaction belongs to multiple categories, we count it multiple times. Thus the sum of each row may be greater than 100%.

- **Food & drink:** dining, groceries, liquor, etc.
- **Transportation:** gas, parking, airfare, etc.
- **Utilities:** cleaning, electricity, phone, etc.
- **Entertainment:** game, sports, movie, music, etc.
- **Life:** gifts, clothing, insurance, medical, etc.
- **Home:** electronics, furniture, rent, etc.

We use a list of keywords and emojis for each category to classify transactions. We build the keyword list by manually assigning the most frequently used 500 English words in Venmo into the above categories. For example, we have keyword “food” under *Food & Drink* category, and “uber” under *Transportation*. 165 words out of 500 were classified, and the remaining words were generic terms such as “thanks” and “great.” For emojis, we manually assign categories to 247 emojis and emoji combinations. For example, “🏠💸 ” is interpreted as “rent” and thus belongs to *Home*.

Using keywords and emojis, we are able to classify 47% of all transaction messages (results in Table 4). Note that a single message can be classified into multiple categories, *e.g.* “gas + rent” belongs to both *Transportation* and *Home*. Remaining unidentified messages either have no English words/emojis or refer to inside jokes or acronyms. Even labeling a thousand more keywords still would not significantly increase classification coverage (Figure 12).

The most popular category is *Food & Drink*, with 19 million transactions, more than half of all identified transactions. This often corresponds to splitting bills after a group gathering or dinner out; *Transportation* is also very popular and often related to carpools or Uber rides. Least popular is *Home*, which involves infrequent payments such as rent.

Transaction Dynamics

We then study the temporal dynamics of different types of transactions. We start by studying periodic patterns of global trends, then turn our focus to the user-level to analyze the dynamics of recurring payments between users.

Global Periodic Trends. We find many types of transactions have clear *periodic* patterns, by looking at monthly transaction count for each category (figure omitted for space constraints). Clear annual patterns are visible for transactions under *Life*, *Entertainment* and *Food & Drink*. *Life* related payments have significant increase in the end of each year, likely matching gift exchanges for the holidays. *Entertainment* has significant increases every March and August, related to social betting on sports. To validate this, we examine the daily transactions of major sports in 2015, and find that events like “Super Bowl” and “March Madness” create huge spikes in February and March, while “Fantasy Football” is responsible for most betting activities in August. Finally, transaction numbers in *Food & Drink* are at their lowest during summers and winter breaks each year, when students are traveling or otherwise away from friends.

User-Level Transaction Dynamics. Next, we focus on the user-level, and examine how likely transactions between two users exhibit periodicity. We detect periodicity using an off-the-shelf algorithm (Vlachos and others 2005) for all user pairs with at least ten transactions (1.54 million pairs in total). Among them, 48,188 (3.1%) exhibit periodicity. The recurring frequencies are shown in Figure 13, with the most common recurrence frequencies being weekly and monthly. Monthly transactions are highly tied to rent (26.1%) and utility bills (20.3%). Weekly recurring transactions are more diverse, involving activities like baby-sitting, pet walking or dining out, and no single category stands out.

Conclusion

Our analysis on Venmo provides several interesting observations that would benefit the research community and application developers. First, Venmo’s transaction graph provides a close representation of strong real-world friendship, demonstrated by the higher clustering effect and higher reciprocity compared to interaction graphs of traditional online social networks. Venmo’s friend transaction graph can be seen as a stronger and more meaningful dataset of social relationships. Second, Venmo transactions exist between both friends and strangers. The two types of transactions exhibit drastically different graph properties that are hidden when analyzing transactions as a whole. Finally, for application developers, the ability to identify different user types provide grounds for targeted advertisement. We demonstrate that user types, *e.g.* businesses and niche users, can be identified using clustering and community detection.

Acknowledgments. The authors wish to thank the anonymous reviewers for their comments. This project was supported by NSF grants CNS-1527939 and IIS-1321083. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies.

References

Bild, D. R., et al. 2015. Aggregate characterization of user behavior in twitter and analysis of the retweet graph. *ACM Trans. Internet Technol.* 15(1):4:1–4:24.

Bird, J. 2015. How venmo won in one of the most crowded spaces in tech. TechCrunch.

Blondel, V. D., et al. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008(10).

Chakravarti, I. M., and Laha, R. G. 1967. Handbook of methods of applied statistics. John Wiley & Sons.

Chang, Y., et al. 2014. What is tumblr: A statistical overview and comparison. *ACM SIGKDD Explorations Newsletter* 16(1):21–29.

Crandall, D., et al. 2008. Feedback effects between similarity and social influence in online communities. In *Proc. of KDD*.

Davis, F. D., et al. 1989. User acceptance of computer technology: A comparison of two theoretical models. *Management Science* 35(8):982–1003.

Ferreira, J., et al. 2015. Spending time with money: From shared values to social connectivity. In *Proc. of CSCW*.

Gilbert, E., et al. 2013. I need to try this?: a statistical overview of pinterest. In *Proc. of CHI*.

Hasan, M. M., and Islam, M. R. 2013. Assessing service quality of mobile money transfer in bangladesh: A case study on bkash. *International Journal of Innovative Research and Development* 2(7).

Jiang, J.; Wilson, C.; Wang, X.; Huang, P.; Sha, W.; Dai, Y.; and Zhao, B. Y. 2010. Understanding latent interactions in online social networks. In *Proc. of IMC*.

Katona, Z., et al. 2011. Network effects and personal influences: The diffusion of an online social network. *Journal of marketing research* 48(3):425–443.

Kaufman, L., and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Kwak, H., et al. 2009. Mining communities in networks: a solution for consistency and its evaluation. In *Proc. of IMC*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. B. 2010. What is twitter, a social network or a news media? In *Proc. of WWW*.

McPherson, M., et al. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1):415–444.

Meiklejohn, S., et al. 2013. A fistful of bitcoins: characterizing payments among men with no names. In *Proc. of IMC*.

Metzger, M.; Wilson, C.; and Zhao, B. Y. 2017. Passive social interaction in social networking sites. In *Proc. of International Communication Association Conference (ICA)*.

Ndiwalana, A., et al. 2010. Mobile money use in uganda: A preliminary study. *M4D* 121.

Ober, M., et al. 2013. Structure and anonymity of the bitcoin transaction graph. *Future internet* 5(2):237–250.

Park, A., and Lee, K. J. 2014. Critical success factor of noble payment system: Multiple case studies. *Journal of Intelligence and Information Systems* 20(4):59–87.

PayPal. 2016a. Paypal annual report 2015.

PayPal. 2016b. Paypal financial releases. `https://investor.paypal-corp.com/releases.cfm/`.

Perez, S. 2016. Square cash's fast payments app can now hold a balance. TechCrunch.

Porter, M. F. 1980. An algorithm for suffix stripping. *Program* 14(3):130–137.

Rey, J. D. 2016. America's biggest banks have announced their venmo competitor, zelle. Recode.

Ron, D., and Shamir, A. 2013. Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography*, 6–24.

Rosvall, M., and Bergstrom, C. T. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105(4):1118–1123.

Schneider, F., et al. 2009. Understanding online social network usage from a network perspective. In *Proc. of IMC*.

Sidel, R., and Demos, T. 2016. The battle heats up in person-to-person payments. Wall Street Journal.

Sidel, R., and Wakabayashi, D. 2015. Apple, banks in talks on mobile person-to-person payment service. Wall Street Journal.

Suri, T., et al. 2012. Documenting the birth of a financial economy. *PNAS* 109(26):10257–10262.

Swamynathan, G.; Wilson, C.; Boe, B.; Almeroth, K. C.; and Zhao, B. Y. 2008. Do social networks improve e-commerce: a study on social marketplaces. In *Proc. of WOSN*.

Ugander, J., et al. 2011. The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*.

Venmo. 2016. Venmo public API. `https://developer.venmo.com/`.

Vlachos, M., et al. 2005. On periodicity detection and structural periodic similarity. In *Proc. of SDM*.

Wang, G.; Zhang, X.; Tang, S.; Zheng, H.; and Zhao, B. Y. 2016. Unsupervised clickstream clustering for user behavior analysis. In *Proc. of CHI*.

Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature* 393(6684):409–10.

Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K. P.; and Zhao, B. Y. 2009. User interactions in social networks and their implications. In *Proc. of EuroSys*.

Wilson, C.; Sala, A.; Puttaswamy, K. P. N.; and Zhao, B. Y. 2012. Beyond social graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web* 6(4).

Xu, T.; Chen, Y.; Jiao, L.; Zhao, B. Y.; Hui, P.; and Fu, X. 2011. Scaling microblogging services with divergent traffic demands. In *Proc. of Middleware*.

Yang, Y., and Pedersen, J. O. 1997. A comparative study on feature selection in text categorization. In *Proc. of ICML*.

Zhao, X.; Sala, A.; Wilson, C.; Wang, X.; Gaito, S.; Zheng, H.; and Zhao, B. Y. 2012. Multi-scale dynamics in a massive online social network. In *Proc. of IMC*.