

Network-Based Decision Support for Cold-Start User Prioritization in P2P Platforms: The Value of Entry-Community Information

Pantelis Loupos

Moran Cerf

Graduate School of Management

Columbia Business School

University of California, Davis

Columbia University

Abstract

Predicting the behavior of newly acquired users is a central challenge for digital platforms because direct behavioral histories are sparse at entry. We examine whether relational information revealed by a user’s first transaction can improve decision support for cold-start prioritization in peer-to-peer (P2P) platforms. Using data on approximately 198,000 Venmo users, we compare models based on attributes of the focal user, the first counterparty, and the counterparty’s pre-existing community. We study two managerially relevant outcomes: near-term inactivity risk and future high-engagement-user identification. At entry, adding relational information increases top-decile lift from 1.11 to 1.20 for inactivity risk and from 1.22 to 1.53 for high-engagement identification. Community information also provides additional value beyond the first counterparty, particularly for identifying future high-engagement users. Decomposing the community signal shows that recent transactional activity is more informative than network structure alone. As an interpretable complement, a selection-adjusted analysis shows that entering through a more cohesive community is associated with a 5.0 to 6.5 percentage-point higher probability of becoming a high-engagement user. These findings show how information embedded in the community revealed by a user’s first transaction can support early prioritization when individual behavioral information is limited.

Keywords: Cold-Start Prediction; Decision Support; Peer-to-Peer Platforms; Social Networks; Customer Engagement; Predictive Analytics

1 Introduction

A central challenge for digital platforms is how to make early targeting and prioritization decisions for newly acquired users before they have accumulated meaningful observed behavior. Most conventional customer analytics tools rely heavily on users’ past behavior and transaction history to predict future behavior (Ascarza et al., 2018). This creates a cold-start prioritization problem at user entry: the platform must differentiate among newcomers before meaningful individual histories have accumulated.

This prediction problem matters because platforms face two distinct allocation constraints. On the one hand, basic onboarding can be delivered broadly, but resource-intensive interventions—such as promotional credits and higher-touch support—consume budget or operational capacity and must therefore be allocated selectively. On the other hand, low-cost digital actions—such as notifications, recommendations, and prominent placements—can be deployed at scale but still consume scarce user attention and may lose salience when overused, consistent with models of attention competition (Anderson and de Palma, 2012). Both constraints also shape monetization. On advertising-supported or freemium platforms, expected future engagement shapes audience value and sponsored-exposure opportunities, while advertising effectiveness depends on combining targeting with prominence rather than merely delivering another impression (Goldfarb and Tucker, 2011). Platforms must therefore decide which entrants to prioritize for incremental incentives, repeated outreach, high-salience placements, and monetization opportunities.

Prior research has approached the cold-start problem primarily using information tied to the entrant’s acquisition channel or initial transaction. Acquisition channels can capture systematic differences in the types of customers attracted to the platform (Verhoef and Donkers, 2005), while first-purchase characteristics provide an early behavioral signal of future customer heterogeneity (Padilla and Ascarza, 2021). These approaches are useful, but they largely characterize the entrant or the initial transaction in isolation. In networked services, however, entry can also

reveal relational information about the broader context surrounding the user. In peer-to-peer (P2P) platforms, a new entrant may transact with an existing user and, through that connection, reveal a pre-existing local community. This creates a network-informed cold-start setting: the platform observes little about the entrant directly, but the entry transaction can reveal the structure, activity, and composition of the community surrounding the first counterparty.

Our central question is whether the relational context revealed by a user’s first transaction can improve cold-start prediction and early user prioritization. The key idea is that the first transaction may reveal more than the identity and prior behavior of the initial counterparty. When that counterparty has prior platform history, the transaction also exposes a pre-existing community with observable network structure, recent transactional activity, and member characteristics. These community-level signals may contain information about the entrant’s future behavior that is not captured by the entrant’s sparse history or by the counterparty’s attributes alone. We therefore examine whether this broader entry-community information improves prediction beyond focal-user and first-counterparty information.

Our empirical analysis uses panel data on approximately 198,000 users from Venmo, a leading P2P mobile payment platform in the United States. We focus on two outcomes that matter for platform management: near-term inactivity risk and future high-engagement-user identification. At the moment of entry, adding relational information to focal-user information increases top-decile lift from 1.11 to 1.20 for inactivity risk and from 1.22 to 1.53 for high-engagement identification. The broader entry community also provides information beyond the first counterparty, with the additional value particularly pronounced for identifying future high-engagement users. Decomposing the community signal shows that recent transactional activity within the pre-existing community is more informative than its network structure alone. These gains have direct implications for constrained prioritization: when the platform can target only the highest-ranked 10% of new users, relational information allows it to identify substantially more future high-engagement users without increasing the number of users targeted.

We then examine one interpretable characteristic of the pre-existing community revealed by the entrant’s first transaction: its cohesiveness. We measure cohesion using the clustering coefficient of the first counterparty’s pre-entry neighborhood, which captures the extent to which

members of that community are connected to one another. Using propensity-score-based comparisons that adjust for observable differences in community size, age, and activity, and for the entrant’s own number of transaction partners, we find that entering through a more cohesive community is associated with a 5.0 to 6.5 percentage-point higher probability of becoming a high-engagement user within the acquisition cohort over the following year.

The paper makes three contributions to cold-start prediction and decision support. First, it shows that a user’s first transaction can reveal predictive information beyond both the entrant and the initial counterparty. When the counterparty has prior platform history, the transaction exposes a pre-existing community whose characteristics provide additional predictive information about the entrant’s future behavior, particularly for identifying future high-engagement users. This broadens the cold-start information set from attributes of the individual or initial dyad to the relational context revealed by entry. Second, we distinguish what aspects of that community are informative. Recent transactional activity within the community provides the strongest predictive signal, while network structure and community maturity contribute more modestly. This distinction shows that the value of relational context depends not only on how a community is connected, but also on how actively its members are using the platform before the entrant arrives. Third, we translate these predictive gains into a constrained prioritization problem. When the platform can target only a limited share of new users, incorporating relational information improves the identification of future high-engagement users without increasing the number of users targeted. This connects network-based cold-start prediction directly to the resource-allocation decisions faced by digital platforms.

2 Related Literature

This paper relates to three streams of research: network-based customer analytics, behavioral prediction under cold start, and decision support for customer prioritization and targeting.

First, our work contributes to research showing that social-network information can improve the prediction of customer behavior. Prior work shows that social-network data add predictive value beyond demographic and behavioral models (Goel and Goldstein, 2013), and that network

attributes can improve customer-retention prediction in financial-services and CRM settings (Benoit and Van den Poel, 2012; Ascarza et al., 2017). Our paper shifts this question to the moment of user entry, when the entrant’s own behavioral history is largely unavailable. Rather than evaluating network information as an additional predictor alongside conventional user-level information, we examine whether the relational context revealed by the first transaction can compensate for missing individual information. We further distinguish information contained in the first counterparty from that contained in the broader entry community.

Second, our paper contributes to the cold-start literature. Existing marketing research addresses cold start using information available before substantial customer histories accumulate, including acquisition channels, customer-base histories, and first-transaction characteristics (Verhoef and Donkers, 2005; Gopalakrishnan et al., 2016; Padilla and Ascarza, 2021). These approaches extract early signals from the entrant or the initial transaction, but do not consider the broader relational context that an entry event may reveal in a networked service. Our analysis complements this literature by showing that the surrounding entry community provides predictive information beyond focal-user and first-counterparty characteristics. Related information-systems research similarly uses contextual signals to support decisions under cold start; for example, Esmeli et al. (2024) incorporate session context into e-commerce recommendations. We apply a related contextual logic to user-level prioritization by using the relational context revealed at entry.

Third, the paper contributes to the literature on predictive decision support in customer analytics, which examines how data and models can inform managerial decisions (Phillips-Wren et al., 2021). Prior research shows how predictive and interpretable models can support decisions such as purchase targeting and customer prioritization (Chaudhuri et al., 2021; Coussement and Benoit, 2021). Much of this work, however, assumes that informative individual-level histories are already available. We study a setting in which users must be ranked before such histories exist and evaluate whether relational information improves prioritization under a fixed targeting capacity. This extends predictive decision support to settings in which the central challenge is information scarcity at the point of decision.

3 Conceptual Framework

A newly acquired user creates a ranking problem for a P2P platform: the platform must differentiate among entrants before meaningful individual histories have accumulated. The user’s first observed transaction identifies a counterparty and, when that counterparty has prior platform history, reveals a pre-existing community surrounding that relationship. We organize this relational information into three classes—first-counterparty characteristics, community structure, and community activity—and examine whether these signals improve cold-start prediction beyond the entrant’s own limited information.

3.1 Cold-Start Prioritization and the Decision Context

At entry, conventional behavioral predictors such as recency, frequency, and accumulated engagement provide little differentiation because users have not yet generated meaningful activity histories (Ascarza et al., 2018; Gopalakrishnan et al., 2016). Platforms may nevertheless need to rank entrants for interventions that consume monetary resources, operational capacity, or user attention. Cold-start prioritization is therefore a ranking problem under sparse individual information.

Unlike a conventional first purchase, a P2P transaction is relational. Beyond recording the entrant’s action, it identifies a counterparty and can reveal a broader relational context. When the first counterparty has prior platform history, the entry transaction exposes a pre-existing community consisting of that counterparty’s prior transaction partners, the connections among them, and their recent activity. When the first counterparty is also new, no such pre-existing community is observable at the moment of entry.

The relational context revealed at entry may carry predictive information for several reasons. Users may transact with similar others, shared payment occasions may create recurring opportunities for platform use, and connections to active peers may reflect exposure to established usage patterns (McPherson et al., 2001). Together, these channels suggest that the entry community can contain information about latent engagement propensity and future usage opportunities that is not yet visible in the entrant’s own behavioral history.

We examine this possibility through two prediction tasks. For retention-oriented decisions, entrants can be ranked by predicted near-term inactivity risk; for segmentation and growth decisions, they can be ranked by predicted future engagement. We evaluate whether relational information improves out-of-sample ranking for these tasks and, in particular, whether the predictive value revealed by the first transaction is exhausted by the first counterparty or extends to the structure and activity of the surrounding community.

3.2 Three Classes of Entry-Network Signals

We organize the primary relational signals revealed by the first transaction into three classes: information about the first counterparty, the structure of the counterparty’s pre-existing community, and the recent activity of that community. These categories distinguish where the information resides at entry—in the initial partner, in how the surrounding community is connected, or in how actively that community uses the platform.

First-counterparty signals: prior behavior and connectedness. The most immediate relational information comes from the user’s first transaction partner. Because individuals tend to connect to similar others (McPherson et al., 2001), an active or well-connected counterparty may signal that the entrant also has a greater latent propensity for platform use. This perspective complements cold-start approaches based on attributes of the entrant or the initial transaction by considering characteristics of the initial partner (Padilla and Ascarza, 2021). We capture first-counterparty information using the counterparty’s prior transaction frequency, recency, tenure, degree, transaction characteristics, and network connectedness. Model 4 augments these characteristics with a small set of counterparty-mediated measures describing the recent development and exclusivity of the counterparty’s local network.

Community-structure signals. The first counterparty’s prior relationships reveal how the surrounding community is organized before the entrant arrives. A closed community, in which members are connected to one another, may reflect repeated interaction, trust, or recurring shared payment occasions (Coleman, 1988; Granovetter, 1985). More open communities may instead provide greater reach into otherwise distinct parts of the network (Granovetter, 1973). Consistent with evidence that neighborhood structure can predict social-product adoption (Ugander et al.,

2012), we characterize the entry community using measures of closure, connectivity, redundancy, and outward reach.

Community-activity signals. Network structure alone does not indicate whether the surrounding community is active. Two similarly connected communities may differ substantially in how frequently their members transact, how recently they have been active, and how broadly activity is distributed across members. Recent community activity may therefore reveal ongoing opportunities for repeated platform use that are not captured by topology alone. We measure this dimension using pre-entry transaction activity, transactions per member, active-member share, transaction concentration, and member recency. Together with measures of community maturity and member composition, these signals form the entry-community information used in Model 5.

3.3 Empirical Expectations

The framework motivates two empirical questions. First, does relational information improve prediction when little is known about the entrant directly? If the first transaction reveals information about future behavior that is not contained in the entrant’s sparse observed history, then incorporating relational signals should improve cold-start prediction beyond focal-user information alone (*E1*).

Second, where does the predictive information revealed by the first transaction reside? If the initial counterparty and closely related counterparty-mediated signals capture most of the relevant information, then the counterparty-side benchmark should account for most of the relational value. If the transaction also reveals a broader pre-existing relational environment, then characteristics of the surrounding entry community should improve prediction beyond this benchmark (*E2*).

We further examine which dimensions of the entry community carry the most predictive information by separating its network structure, recent transactional activity, and maturity.

4 Overview of Data and Methods

4.1 Research Setting: Venmo

Our empirical setting is Venmo, a P2P mobile payment platform in the United States. Venmo combines P2P payments with social-feed functionality: public transactions display the sender, receiver, and accompanying text, and users can interact with transactions appearing in their feeds. The transaction graph therefore records payment interactions and the relational structure generated by those interactions.

Venmo is well suited for studying cold-start prioritization in a networked service. Transactions reveal relational ties, allowing us to observe the network context through which users enter the platform. At the same time, repeated usage is common: fewer than 2% of users in the sample transact only once during the one-year observation window. The setting therefore provides meaningful variation in both the relational context visible at entry and users’ subsequent levels of engagement.

The payment context gives these relational ties a direct behavioral interpretation. P2P transactions often arise from shared expenses—such as rent, restaurant bills, trips, and utilities—that can generate recurring occasions for platform use. The network surrounding an entrant may therefore reflect not only individual transactions, but also established patterns of shared payment activity. Moreover, when the first counterparty has prior platform history, the entry transaction makes aspects of that broader relational context visible to the platform.

Entrants differ in the amount of relational context observable at entry. For 72.5% of entrants, the first counterparty has prior platform history, making pre-existing network information available. The remaining 27.5% have a new first counterparty, whose network is empty at the moment of entry, so no pre-existing relational context is observable. This distinction determines whether community-level information is available to support cold-start prediction at the moment of entry.

4.2 Data

Our dataset contains transactions among Venmo’s public users from January 2014 through June 2015. For the main analysis, we focus on users who joined the platform between January 2014

and June 2014. This cohort definition allows us to observe each entrant for a full 12-month post-entry period. The final sample contains 197,970 users.

The transaction-level data include sender and receiver identifiers, timestamps, and accompanying messages. The user-level data include each user’s join date and whether the user signed up through Facebook. Table 1 summarizes the final dataset.

Dataset Overview	
Sample size	197,970
Observation window	January 2014 – June 2015
Acquisition cohorts	January 2014 – June 2014
Lifetime analysis window	0 – 12 months
Training assignment probability	70%
Testing assignment probability	30%

Table 1: Overview of the final dataset.

4.3 Variables

We organize predictors into three information groups: focal-user, counterparty-side, and relational-network information, with month-of-entry fixed effects included as common controls. The relational information set is stage-specific because the network observable around a user changes after entry. At lifetime 0, every predictor is measured at the entrant’s first transaction or strictly before it. At subsequent lifetimes, focal-user and relational-network predictors use only information available through the corresponding lifetime cutoff. Appendix B provides the complete variable definitions and timing conventions.

Focal-user variables characterize the entrant directly, including signup characteristics and observed transaction activity, timing, content, directionality, and number of transaction partners. These variables evolve as the entrant accumulates behavioral history. Counterparty-side variables characterize the entrant’s initial transaction partner and a small set of counterparty-mediated signals, all measured using information available strictly before the entrant’s first observed transaction and fixed at entry.

The construction of relational-network variables differs naturally between the point of entry and later lifetimes. At the first transaction, the entrant has essentially one observed direct

transaction partner, so the entrant’s own first-degree network contains little structural variation. The broader relational context is instead revealed through the first counterparty. We define the *entry community* as the set of users who had transacted with the first counterparty strictly before the entrant’s first transaction, excluding the entrant. Once the entry transaction occurs, these users constitute the entrant’s immediate second-degree neighborhood through the first counterparty. We characterize this pre-existing community using measures of its structure, recent transactional activity, maturity, and member composition. These measures capture information about the relational environment surrounding the first counterparty while remaining available at the moment of entry.

From lifetime 1 onward, the entrant has accumulated an observable transaction network of their own. Relational predictors are therefore constructed directly from the entrant’s evolving ego network rather than from the first counterparty’s pre-entry community. These measures include local cohesion, first-degree neighborhood characteristics, second-degree or friends-of-friends characteristics, PageRank, and giant-component membership. First- and second-degree neighborhood measures summarize the activity, degree, and message characteristics of the entrant’s transaction partners and their surrounding networks. All such measures are recomputed using information available at or before the relevant lifetime cutoff.

Several lifetime-0 measures warrant additional explanation. *Cohesion* is the local clustering coefficient of the first counterparty’s pre-entry neighborhood,

$$C_p = \frac{2T_p}{k_p(k_p - 1)}, \quad (1)$$

where k_p is the number of members of the entry community and T_p is the number of edges among them; C_p is set to zero when $k_p < 2$ (Watts and Strogatz, 1998). Cohesion therefore captures the extent to which members of the community revealed by the entrant’s first transaction are connected to one another. From lifetime 1 onward, the analogous measure is computed on the focal user’s own accumulated neighborhood. Community-activity variables at lifetime 0 are constructed from pre-entry transactions of community members, excluding transactions involving the first counterparty.

Global network position follows the same stage-specific logic. At lifetime 0, the entrant does

not yet have a pre-existing global network position. We therefore use the first counterparty’s pre-entry position to characterize the broader network context exposed by the entrant’s initial connection. *PageRank* captures how centrally positioned the first counterparty is within the existing transaction network and thus provides information about the broader network context of the entrant’s initial connection. We compute PageRank on cumulative directed transaction-graph snapshots using GraphFrames with reset probability $\alpha = 0.15$ and use the most recent snapshot strictly preceding the entrant’s first transaction (Brin and Page, 1998). Giant-component membership similarly indicates whether the entrant’s first transaction connects the user to the platform’s largest connected component. From lifetime 1 onward, once the entrant has developed an observable network position of their own, PageRank and giant-component membership refer directly to the focal user, using the most recent network snapshot available at or before the corresponding lifetime cutoff.

4.4 Empirical Design

The unit of analysis is the user. We organize observations by user lifetime, defined as the number of months since joining Venmo. Lifetime 0 denotes the moment of entry, and users are followed through lifetime 12. At each lifetime t , predictors use only information observed through that point to predict subsequent outcomes. This structure allows us to examine how the predictive value of focal-user and relational-network information evolves as entrants accumulate direct behavioral histories.

We study two prediction tasks. The first is *near-term inactivity risk*: whether the user records no transaction in the following month. This outcome maps to early onboarding and retention decisions. The second is *future high-engagement status*: whether the user belongs to the top 20% of the acquisition cohort based on transaction frequency over the first year. This outcome maps to early segmentation, growth, and monetization decisions.

The second outcome measures engagement rather than realized monetary value because the data do not reveal transaction amounts or fee-generating product usage. Transaction frequency nevertheless captures meaningful usage intensity and represents an observable behavioral component of customer engagement (Kumar et al., 2010). We therefore interpret the top quintile as

the cohort’s most engaged future users, not its highest-revenue users. Appendix A.2 shows that the conclusions are unchanged when high engagement is defined using the top 10% or top 25%.

4.5 Modeling Framework

At each lifetime, we compare models using focal-user information, relational-network information, and their combination. Let $x_{i,t}$ denote focal-user predictors, $z_{i,t}$ the relational information available at lifetime t , and $y_{i,t}^{(q)}$ the outcome for task q . As described above, the composition of $z_{i,t}$ is stage-specific: at lifetime 0 it characterizes the pre-existing relational context revealed through the first counterparty, whereas from lifetime 1 onward it characterizes the focal user’s own evolving transaction network. We estimate

$$\text{Model 1: } y_{i,t}^{(q)} = F_q(x_{i,t}), \quad (2)$$

$$\text{Model 2: } y_{i,t}^{(q)} = F_q(z_{i,t}), \quad (3)$$

$$\text{Model 3: } y_{i,t}^{(q)} = F_q(x_{i,t}, z_{i,t}). \quad (4)$$

Model 1 uses focal-user information, Model 2 uses relational-network information, and Model 3 combines the two. Comparing Models 3 and 1 indicates whether relational information improves prediction beyond what is observed about the focal user at each stage of the lifecycle. We evaluate this comparison using out-of-sample ranking performance across lifetimes and, at lifetime 0, under fixed targeting capacity.

At lifetime 0, we additionally compare information associated with the first counterparty with information about the broader entry community. Let f_i denote first-counterparty and counterparty-mediated information, and let c_i denote characteristics of the counterparty’s broader pre-existing community. The latter includes community structure, transaction activity, maturity, and member composition, while excluding PageRank and giant-component membership.

We estimate

$$\text{Model 4: } y_{i,0}^{(q)} = F_q(f_i), \quad (5)$$

$$\text{Model 5: } y_{i,0}^{(q)} = F_q(c_i), \quad (6)$$

$$\text{Model 6: } y_{i,0}^{(q)} = F_q(f_i, c_i). \quad (7)$$

Model 4 provides the closest benchmark to prior cold-start approaches based on first-transaction information (Padilla and Ascarza, 2021). Comparing Models 4 and 6 tests whether the broader entry community provides predictive value beyond first-counterparty and counterparty-mediated information, while Model 5 shows how informative the community is on its own.

We estimate logistic regression, lasso, and ridge specifications. Logistic regression provides an unregularized baseline. The lasso is the primary specification because its ℓ_1 penalty performs embedded variable selection and produces sparse models suited to interpretable decision support (Coussement and Benoit, 2021). Ridge regression provides a complementary benchmark when predictors are correlated, as is common among relational-network measures. Appendix A.1 shows that the near-term inactivity-risk results are stable across these specifications.

Models 1–3 are estimated separately at each lifetime, allowing predictive relationships and the available information set to evolve over the user lifecycle. For lasso and ridge, the penalty parameter is selected through ten-fold cross-validation in the training sample by maximizing AUC, with predictors standardized internally. All models include month-of-entry fixed effects to account for calendar-time differences across acquisition cohorts; these common controls are omitted from the notation above for simplicity.

4.6 Model Evaluation

Users are independently assigned, using a fixed seed, to the training sample with probability 0.70 and to the holdout test sample with probability 0.30. Each user’s complete lifetime panel remains in the same partition, preventing information from a test user from entering estimation or tuning.

We evaluate out-of-sample performance using two complementary metrics. Our primary

decision-oriented metric is top-decile lift, which measures how strongly positive outcomes are concentrated among the highest-ranked 10% of users relative to random targeting. This corresponds to a setting in which the platform has limited capacity and must prioritize only a subset of entrants. We also report AUC as an overall ranking metric. For the inactivity-risk task, the positive class is next-period inactivity; for the high-engagement task, the positive class is future high-engagement status.

To make the decision value of the rankings more concrete, we also evaluate fixed-capacity targeting at alternative selection rates. At each capacity, users are ranked by their out-of-sample predicted scores, and we calculate the number and share of eventual positive outcomes captured among the selected users. This allows us to compare models under the same targeting constraint and quantify how many additional relevant users are identified by incorporating relational information.

We construct 95% confidence intervals using a nonparametric bootstrap of test-sample users with $B = 1,000$ replications. Fitted models are held fixed, and each metric and pairwise model difference is recomputed on the same resampled users, preserving the paired structure of the comparisons.

5 Empirical Results

5.1 Predictive Value Across Tasks and Lifetimes

Figure 1 compares the predictive value of focal-user, relational-network, and combined information across the user lifecycle. At lifetime 0, relational information improves ranking performance for both outcomes. For near-term inactivity risk, top-decile lift increases from 1.105 under the focal-user model (Model 1) to 1.195 under the combined model (Model 3), a gain of 0.090 (95% paired-bootstrap CI [0.053, 0.129]). For high-engagement identification, the increase is larger, from 1.217 to 1.529, a gain of 0.312 ([0.244, 0.383]). AUC likewise increases from 0.528 to 0.548 for inactivity risk and from 0.548 to 0.585 for high-engagement identification. These results support *E1*: relational information improves prediction at the cold-start decision point.

The additional value of relational information declines rapidly once users accumulate their

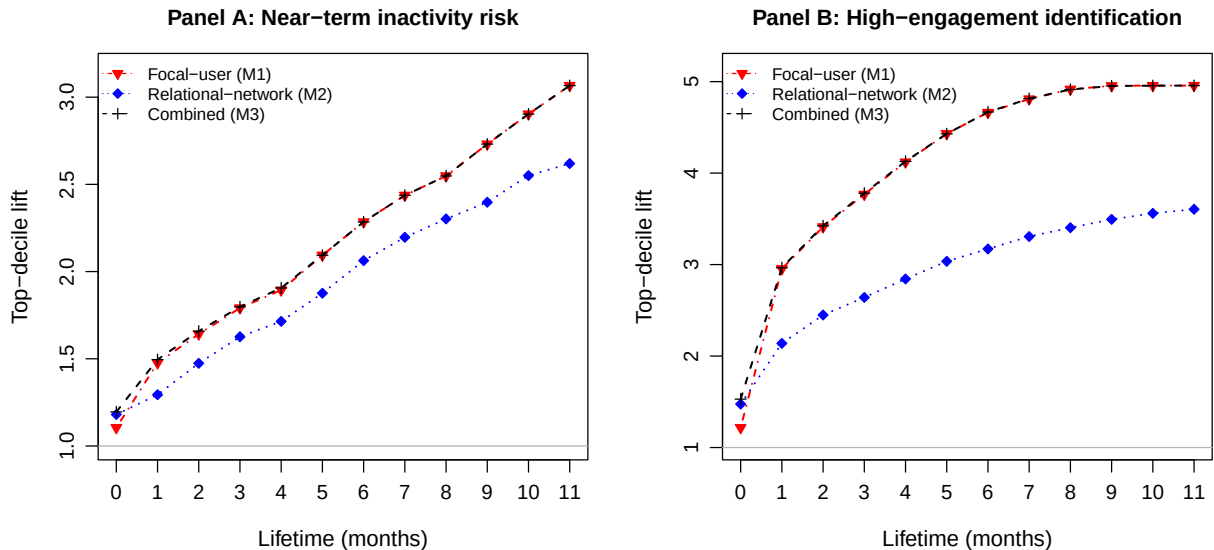


Figure 1: Top-decile lift by lifetime (lasso). Panel A reports near-term inactivity risk and Panel B high-engagement identification.

own behavioral histories. By lifetime 1, top-decile lift is 1.476 under the focal-user model and 1.495 under the combined model for inactivity risk, and 2.954 versus 2.966 for high-engagement identification. The corresponding gains fall to 0.019 and 0.012, and the 95% paired-bootstrap interval for Model 3 minus Model 1 includes zero at every subsequent lifetime for both tasks.

The standalone relational model nevertheless becomes increasingly predictive as the entrant’s own network develops. From lifetime 1 to lifetime 11, its top-decile lift rises from 1.294 to 2.619 for inactivity risk and from 2.139 to 3.605 for high-engagement identification. Yet it remains below the focal-user model throughout and provides little additional predictive value once focal-user information is included. Thus, after behavioral history begins to accumulate, much of the information contained in the entrant’s network overlaps with information already captured by the focal-user block.

The cold-start gain is especially pronounced for high-engagement identification. At entry, relational information produces a substantially larger improvement for this task than for near-term inactivity, whereas after lifetime 0 neither task shows a statistically distinguishable gain from adding relational information. The inactivity results are stable across alternative functional forms, and the high-engagement result is robust to alternative definitions of high engagement

(Appendix A.1 and Appendix A.2).

5.2 Decision Value Under Fixed Targeting Capacity

We next translate the ranking improvements into a fixed-capacity prioritization problem. Suppose the platform can act on only a limited share of newly acquired users. Entrants are ranked by their predicted scores, and the platform selects the highest-ranked users under the same capacity constraint for each model. Figure 2 reports the additional positives identified per 100,000 entrants as targeting capacity varies. The solid line measures the value of adding relational information to focal-user information (Model 3 versus Model 1), while the dashed line shows the value of entry-community information beyond the counterparty benchmark (Model 6 versus Model 4).

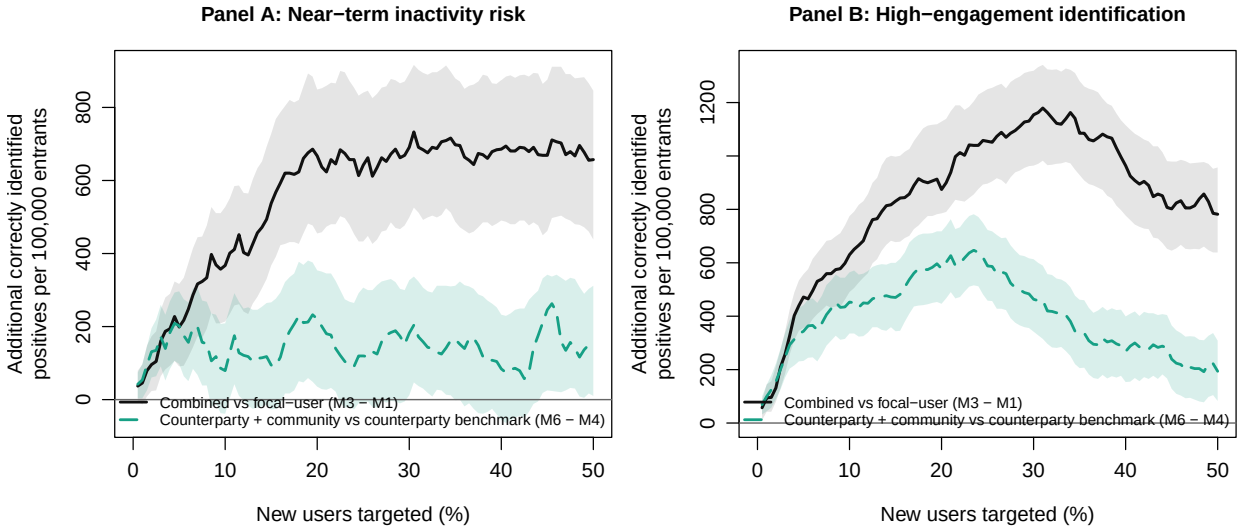


Figure 2: Additional correctly identified positives per 100,000 entrants at lifetime 0 across targeting capacities. Solid lines compare Model 3 with Model 1; dashed lines compare Model 6 with Model 4. Shaded areas show 95% paired-bootstrap intervals.

The decision value is strongest for high-engagement identification. At a 10% targeting capacity, the focal-user model achieves a top-decile lift of 1.217, whereas the combined model reaches 1.529. This yields 630 additional future high-engagement users identified per 100,000 entrants (95% CI [490, 775]) without increasing the number of users targeted. For inactivity risk, lift increases from 1.105 to 1.195, yielding 367 additional subsequently inactive users per 100,000

entrants (95% CI [214, 527]). Figure 2 shows that these gains persist across a range of targeting capacities.

The broader entry community also contributes decision value beyond the counterparty benchmark, particularly for high-engagement identification. At a 10% targeting capacity, adding entry-community information to Model 4 increases top-decile lift from 1.281 to 1.506 and identifies 453 additional future high-engagement users per 100,000 entrants (95% CI [335, 564]). For inactivity risk, the gain is smaller: 79 additional users per 100,000 entrants, with a confidence interval that includes zero. Appendix C reports the complete results at targeting capacities of 5%, 10%, and 20%.

5.3 Does Entry-Community Information Add Value Beyond the Counterparty Benchmark?

E2 asks whether the predictive information revealed by the first transaction extends beyond the initial counterparty to the surrounding entry community. We examine this question using the counterparty benchmark (Model 4), the community-only model (Model 5), and their combination (Model 6). As described in Section 4.5, Model 4 includes first-counterparty characteristics together with counterparty-mediated information.

Table 2: Counterparty Benchmark and Entry-Community Information at Lifetime 0

Model	Inactivity Risk		High Engagement	
	Lift	AUC	Lift	AUC
Counterparty Benchmark (M4)	1.174	0.540	1.281	0.561
Community-Only (M5)	1.178	0.542	1.476	0.575
Counterparty + Community (M6)	1.193	0.548	1.506	0.578

The value of entry-community information differs across prediction tasks. For inactivity risk, the counterparty benchmark and community-only model perform similarly, with top-decile lifts of 1.174 and 1.178. Adding community information to the counterparty benchmark increases lift to 1.193, although the difference is not statistically distinguishable from zero at the 10% targeting threshold. In overall ranking performance, Model 6 improves AUC over Model 4 by 0.007 (95% CI [0.004, 0.010]).

For high-engagement identification, the community provides a stronger predictive signal. The community-only model achieves a top-decile lift of 1.476 compared with 1.281 for the counterparty benchmark, and combining the two raises lift to 1.506. Relative to Model 4, Model 6 increases top-decile lift by 0.22 (95% CI [0.17, 0.28]) and AUC by 0.017 (95% CI [0.013, 0.021]).

These results support *E2*: the broader entry community provides additional predictive information beyond a benchmark that already incorporates first-counterparty and counterparty-mediated information, with the advantage particularly pronounced for identifying future high-engagement users.

5.4 Which Entry-Community Signals Matter?

We next decompose the entry-community information set into network structure, recent transactional activity, and maturity and composition. This allows us to examine which characteristics of the community account for its predictive value at entry.

Table 3: Predictive Contribution of Entry-Community Signal Classes at Lifetime 0

Signal Added	Inactivity Risk Δ Lift [95% CI]	High Engagement Δ Lift [95% CI]
Structure added to focal-user model	0.038 [0.005, 0.073]	0.133 [0.068, 0.200]
Activity added to focal-user model	0.071 [0.033, 0.110]	0.222 [0.157, 0.288]
Maturity/composition added to focal-user model	0.000 [−0.016, 0.020]	0.047 [−0.003, 0.102]
Full community block added to focal-user model	0.090 [0.053, 0.129]	0.310 [0.241, 0.373]
Structure added to activity	0.022 [0.003, 0.042]	0.003 [−0.027, 0.028]
Activity added to structure	0.056 [0.026, 0.086]	0.092 [0.044, 0.132]
Maturity/composition added to structure + activity	−0.003 [−0.021, 0.018]	0.085 [0.039, 0.134]

Note. Entries report changes in top-decile lift relative to the indicated baseline, with paired-bootstrap 95% confidence intervals.

Recent community activity provides the strongest predictive contribution for both tasks. Added to focal-user information, the activity block increases top-decile lift by 0.071 for inactivity risk and 0.222 for high-engagement identification, compared with gains of 0.038 and 0.133 from community structure. Activity also continues to improve ranking after structural information is included, increasing lift by 0.056 and 0.092, respectively.

The contribution of structure and maturity differs across tasks. Once recent activity is observed, additional structural information produces a small improvement for inactivity risk

and essentially no further improvement for high-engagement identification. Community maturity and composition contribute little to inactivity prediction but provide additional value for high-engagement identification when added to structure and activity, increasing lift by 0.085. Overall, the predictive value of the entry community reflects primarily its recent transactional activity, with structure and maturity providing more task-specific information.

5.5 Structural Embeddedness and Future Engagement

We next examine one interpretable structural characteristic of the entry community: its cohesion, measured by the local clustering coefficient of the first counterparty’s pre-existing neighborhood. The analysis asks whether users entering through more cohesive communities are more likely to become highly engaged after adjusting for observed differences in community characteristics, including size, age, and activity, and for the entrant’s own number of transaction partners. We use propensity-score adjustment to compare users entering through communities above a given cohesion threshold with observably similar users entering through communities below that threshold. We report three estimators—nearest-neighbor matching, inverse probability weighting (IPW), and augmented inverse probability weighting (AIPW).

Table 4 reports the adjusted differences across alternative clustering-coefficient thresholds. The estimates are positive and stable across thresholds and adjustment methods. Across thresholds of 0.3, 0.4, 0.5, and 0.6, entry through a more cohesive community is associated with a 5.0 to 6.5 percentage-point higher probability of becoming a future high-engagement user. The consistency of the estimates indicates that the association is not driven by a particular cohesion threshold or adjustment method.

Table 4: Selection-Adjusted Differences by Community-Cohesion Threshold

Clustering-Coefficient Threshold	NN	IPW	AIPW
0.3	0.053	0.053	0.061
0.4	0.050	0.057	0.058
0.5	0.056	0.060	0.061
0.6	0.059	0.065	0.064

Note. Adjusted differences in the probability of future high engagement. Models adjust for entry-community size, age, and activity and for entrant partner count. Restricting the sample to entrants with at least two partners yields estimates of 4.4–6.7 percentage points across all thresholds and estimators.

The results identify community cohesion as an interpretable entry-time marker of future high engagement after adjustment for observed community characteristics. Unobserved factors may nevertheless shape both entry-community cohesion and subsequent engagement, so the estimates should be interpreted as selection-adjusted associations rather than causal effects.

6 Discussion and Conclusion

6.1 Summary of Findings

This work examines whether relational information can improve decision support for cold-start user prioritization in a P2P platform. Four findings stand out. First, relational information improves prediction at the moment of entry, when little is known about the entrant directly, with the gain especially pronounced for identifying future high-engagement users. This advantage declines quickly once users begin generating their own behavioral histories. Second, the first counterparty does not exhaust the information revealed by the first transaction. The surrounding pre-existing community provides additional predictive value, particularly for future engagement. Third, recent transactional activity within that community provides the strongest predictive signal, while network structure and community maturity provide smaller and more task-specific contributions. Fourth, these predictive gains translate into meaningful prioritization improvements under fixed targeting capacity. At a 10% targeting rate, the combined model identifies 630 additional future high-engagement users per 100,000 entrants relative to focal-user information alone. As a complementary result, the selection-adjusted analysis identifies community cohesion as an interpretable entry-time marker of future engagement: entering through a more cohesive community is associated with a 5.0 to 6.5 percentage-point higher probability of becoming a future high-engagement user.

These findings yield three theoretical implications for cold-start prediction and decision support.

First, cold-start information in networked platforms can be distributed beyond the focal user and the initial dyad. Existing approaches typically extract early predictive signals from the entrant, the acquisition source, or the first observed transaction. In a P2P setting, however,

the first transaction can also reveal a pre-existing relational environment surrounding the initial counterparty. The entry event is therefore not only an early behavioral observation but also an information-revelation event that exposes community-level context. This broadens the relevant unit of information under cold start from the individual and dyad to the surrounding community.

Second, community context is multidimensional. Network topology alone does not characterize the information contained in the entrant’s surrounding environment. Communities that look similar structurally may differ substantially in how frequently their members transact, how recently they have been active, and how established their membership is. Our decomposition shows that recent transactional activity provides the strongest predictive signal, while structure and maturity contribute more selectively. This distinction suggests that the predictive value of relational context depends not only on how a community is connected, but also on whether its members are actively transacting before the entrant arrives.

Third, the relative value of relational and individual information changes over the user life-cycle. At entry, relational context helps compensate for the scarcity of direct behavioral information. Once users begin generating their own histories, focal-user information rapidly accounts for most of the combined model’s ranking performance. Cold-start decision support is therefore dynamic: the most informative inputs depend on when the decision is made. Relational context is most useful at the point of entry, whereas direct behavioral information becomes increasingly informative as the user accumulates platform history.

6.2 Managerial Implications

The results have two direct implications for platform decision support. First, platforms can use relational information as an early screening layer at the moment of a user’s first transaction, before meaningful behavioral history has accumulated. When the first counterparty has prior platform history, the platform can augment focal-user information with characteristics of that counterparty’s pre-existing community, particularly its recent transactional activity. When the first counterparty is also new, this community information is not yet available. As users accumulate their own behavioral histories, platforms can update the information set and rely increasingly on direct user behavior.

Second, the value of relational information is greatest when the platform has limited targeting capacity. The models need not predict individual outcomes perfectly to improve decisions; they only need to rank entrants better among the users the platform can act on. This is especially relevant for future high-engagement identification: at a 10% targeting capacity, adding relational information allows the platform to identify 630 additional future high-engagement users per 100,000 entrants without increasing the number of users targeted. The same logic applies to retention-oriented decisions, although the gains are smaller. These scores identify users who are more likely to experience the outcome, not those who would respond most strongly to a particular intervention. A natural decision-support architecture is therefore two-stage: predictive models first prioritize candidate users, after which experimental or credible quasi-experimental evidence determines which interventions generate the greatest incremental response.

6.3 Limitations and Future Research

The network-informed approach developed here may extend to other P2P and networked platforms in which entry events reveal relational information. An important next step is to examine its performance in settings where relational ties are weaker, only partially observed, or less central to subsequent usage, and where network information is available only in incomplete form or alongside richer behavioral histories. Such comparisons would clarify the boundary conditions and portability of network-based cold-start prioritization.

The present study also does not identify the causal mechanisms underlying the predictive value of entry-community information. The propensity-score comparisons adjust for observed community characteristics, but they do not establish that cohesion causally increases engagement. More generally, community structure, transactional activity, and maturity arise jointly in naturally evolving networks, making it difficult to isolate the causal effect of any one relational dimension. Future research could exploit experimental or quasi-experimental variation in network formation, community exposure, or surrounding-peer activity to estimate how specific relational conditions affect subsequent engagement.

Finally, the models predict inactivity risk and future engagement under observed conditions rather than responsiveness to intervention. Future work could estimate heterogeneous treat-

ment effects within the priority segments identified by the models and design targeting policies that combine predicted outcomes with estimated incremental effects. This would extend the analysis from identifying informative cold-start signals to estimating the incremental value of interventions targeted using those signals.

A Robustness Checks

A.1 Comparison of Alternative Functional Forms for Near-Term Inactivity-Risk Prediction

Table A1 compares logistic regression, lasso, and ridge specifications for the near-term inactivity-risk task, reporting top-decile lift (Panel A) and AUC (Panel B). The results are very similar across functional forms on both metrics, indicating that the main classification results are not driven by a particular regularization choice.

Panel A: Top-decile lift									
Lifetime	Model 1			Model 2			Model 3		
	Logistic	Lasso	Ridge	Logistic	Lasso	Ridge	Logistic	Lasso	Ridge
0	1.10	1.10	1.10	1.18	1.18	1.18	1.19	1.20	1.20
1	1.48	1.48	1.48	1.31	1.30	1.31	1.49	1.49	1.49
2	1.64	1.64	1.64	1.48	1.48	1.47	1.66	1.66	1.65
3	1.78	1.79	1.78	1.63	1.63	1.62	1.79	1.80	1.78
4	1.90	1.89	1.91	1.71	1.72	1.71	1.91	1.91	1.91
5	2.10	2.09	2.10	1.88	1.88	1.86	2.09	2.09	2.10
6	2.28	2.28	2.29	2.07	2.07	2.05	2.29	2.28	2.29
7	2.44	2.44	2.43	2.20	2.20	2.20	2.43	2.44	2.44
8	2.55	2.55	2.55	2.31	2.31	2.29	2.55	2.55	2.56
9	2.73	2.73	2.73	2.40	2.40	2.40	2.73	2.73	2.73
10	2.90	2.90	2.90	2.55	2.55	2.54	2.90	2.90	2.90
11	3.06	3.07	3.07	2.62	2.62	2.62	3.06	3.07	3.07

Panel B: AUC									
Lifetime	Model 1			Model 2			Model 3		
	Logistic	Lasso	Ridge	Logistic	Lasso	Ridge	Logistic	Lasso	Ridge
0	0.529	0.528	0.530	0.542	0.542	0.542	0.549	0.548	0.549
1	0.723	0.723	0.723	0.643	0.644	0.643	0.725	0.725	0.724
2	0.764	0.764	0.764	0.678	0.677	0.677	0.765	0.765	0.765
3	0.781	0.781	0.781	0.692	0.692	0.691	0.782	0.782	0.781
4	0.790	0.791	0.790	0.699	0.699	0.698	0.791	0.791	0.790
5	0.805	0.806	0.805	0.713	0.713	0.711	0.805	0.806	0.804
6	0.819	0.820	0.818	0.726	0.725	0.723	0.819	0.820	0.817
7	0.828	0.829	0.827	0.736	0.736	0.733	0.828	0.829	0.826
8	0.837	0.840	0.836	0.743	0.742	0.739	0.837	0.840	0.835
9	0.839	0.842	0.838	0.745	0.745	0.742	0.839	0.842	0.837
10	0.846	0.849	0.845	0.749	0.749	0.745	0.846	0.849	0.843
11	0.846	0.850	0.845	0.753	0.753	0.750	0.846	0.850	0.844

Table A1: Top-decile lift (Panel A) and AUC (Panel B) for alternative functional forms, near-term inactivity-risk task.

Note. Models use the stage-specific information sets defined in Section 4.3.

A.2 Robustness to the high-engagement cutoff

The main analysis defines high-engagement users as the top 20% of transaction frequency within the acquisition cohort. Table A2 re-estimates the lifetime-0 models with the positive class defined as the top 10%, top 20% and top 25% of the cohort, and reports top-decile lift for the focal-user and combined models.

Table A2: Cold-Start Lift Across Alternative High-Engagement Definitions (Lifetime 0, Lasso)

High-engagement definition	Focal-user (M1) lift	Combined (M3) lift	Δ Lift [95% CI]
Top 10%	1.193	1.678	0.485 [0.343, 0.573]
Top 20% (main)	1.217	1.529	0.312 [0.244, 0.383]
Top 25%	1.198	1.461	0.262 [0.202, 0.321]

Note. Lifetime-0 test users ($n = 59,350$); 6,019, 11,974 and 14,920 positives under the three definitions. The positive class is defined within the entrant’s month-of-entry cohort. Δ Lift is the combined model minus the focal-user model, with a paired bootstrap 95% interval from $B = 1,000$ resamples of test users with the fitted models held fixed.

The additional value of relational information at entry is present under every definition and is largest when the positive class is narrowest.

B Variable Description

Table B1 defines the predictors used at each lifecycle stage. Variable labels are descriptive rather than implementation names. Table B2 reports summary statistics at user entry (lifetime 0).

Table B1: Variable Description

Variable Type	Stage	Predictor	Description
Focal User	All lifetimes	Facebook sign-in	Indicator for signup through Facebook.
		Transaction recency	Days since the user’s most recent transaction; zero at lifetime 0.
		Night-time share	Share of transactions occurring between 6:00pm and 4:00am.
		Weekend share	Share of transactions occurring on weekends.
		Transaction frequency	Transactions per day of tenure; zero at lifetime 0.
		Charge count	Number of charge transactions.
		Payment count	Number of payment transactions.
		Partner count	Number of distinct transaction partners.
		Emoji share	Share of transaction descriptions containing an emoji.
		Users brought	Number of users whose first observed transaction was with the focal user.
		Outgoing share	Share of transactions initiated by the focal user; at lifetime 0 this is the direction of the first transaction.

continued on next page

Table B1 continued from previous page

Variable Type	Stage	Predictor	Description
First Counterparty and Counterparty-Mediated Information	Lifetime 0	Transaction frequency	First counterparty's transactions per day of tenure.
		Transaction count	Number of pre-entry transactions by the first counterparty.
		Recency	Days since the first counterparty's most recent pre-entry transaction.
		Tenure	Days since the first counterparty's first transaction.
		Degree	Number of distinct pre-entry transaction partners.
		Outgoing share	Share of prior transactions initiated by the first counterparty.
		Charge share	Share of prior transactions recorded as charges.
		Emoji share	Share of prior transaction descriptions containing an emoji.
		Transactions, prior 30 days	Number of transactions during the 30 days before entry.
		PageRank	First counterparty's PageRank at the most recent pre-entry network snapshot.
		New counterparty	Indicator that the first counterparty had no prior transactions.
		Community growth, 30 days	Share of community members first connected to the counterparty during the prior 30 days.
		Community growth, 90 days	Share of community members first connected to the counterparty during the prior 90 days.
		Counterparty-only member share	Share of community members whose only pre-entry transactions were with the first counterparty.
Entry Community: Structure	Lifetime 0	Cohesion	Share of community-member pairs connected by a transaction tie.
		Internal ties	Number of transaction ties among community members.
		Network constraint	Burt constraint of the first counterparty in the pre-entry network.
		Isolate share	Share of community members with no tie to another community member.
		Largest-component share	Share of community members belonging to its largest connected component.
		Two-hop reach	Number of users outside the community connected to at least one community member.
		Expansion ratio	Two-hop reach divided by community size.
Entry Community: Activity	Lifetime 0	Internal transactions, 30 days	Transactions among community members during the prior 30 days.
		External transactions, 30 days	Transactions between community members and users outside the community during the prior 30 days.

continued on next page

Table B1 continued from previous page

Variable Type	Stage	Predictor	Description
		Internal transaction share	Share of community transactions occurring between community members.
		Transactions per member	Prior-30-day community transactions divided by community size.
		Active-member share	Share of community members active during the prior 30 days.
		Activity concentration	Herfindahl concentration of transactions across community members during the prior 30 days.
		Internal transactions, full history	Transactions among community members over their complete pre-entry history.
		Mean member recency	Mean days since community members' most recent qualifying transaction.
		No recent member activity	Indicator that no community member has a qualifying transaction for the recency measure.
All entry-community activity measures exclude transactions involving the first counterparty.			
Entry Community: Maturity and Composition	Lifetime 0	Mean community age	Mean platform tenure of community members.
		Tenure dispersion	Standard deviation of community-member tenure.
		New-member share, 30 days	Share of community members who joined during the prior 30 days.
		New-member share, 90 days	Share of community members who joined during the prior 90 days.
		Mean member degree	Mean number of distinct transaction partners among community members.
		Degree dispersion	Standard deviation of community-member degree.
		Mean member frequency	Mean transaction frequency among community members.
		Frequency dispersion	Standard deviation of community-member transaction frequency.
		Mean member emoji share	Mean emoji-use share among community members.
		Emoji-share dispersion	Standard deviation of community-member emoji-use share.
Pre-Entry Network Position	Lifetime 0	Giant-component connection	Indicator that the first counterparty belongs to the giant component before entry.
Entrant-Centered Relational Network	Lifetimes 1–11	Cohesion	Share of pairs of the entrant's transaction partners connected to one another; zero with fewer than two partners.
		PageRank	Entrant's PageRank at the most recent network snapshot available by the lifetime cutoff.
		Giant-component membership	Entrant's giant-component membership at the most recent available network snapshot.
		Partner frequency, mean	Mean transaction frequency across the entrant's transaction partners.
		Partner frequency, SD	Standard deviation of transaction frequency across transaction partners.

continued on next page

Table B1 continued from previous page

Variable Type	Stage	Predictor	Description
		Partner degree, mean	Mean number of transaction partners of the entrant's direct partners.
		Partner degree, SD	Standard deviation of degree across the entrant's direct partners.
		Partner emoji share, mean	Mean emoji-use share across the entrant's direct partners.
		Partner emoji share, SD	Standard deviation of emoji-use share across direct partners.
		Second-degree frequency, mean	Mean transaction frequency across second-degree contacts.
		Second-degree frequency, SD	Standard deviation of transaction frequency across second-degree contacts.
		Second-degree degree, mean	Mean degree across second-degree contacts.
		Second-degree degree, SD	Standard deviation of degree across second-degree contacts.
		Second-degree emoji share, mean	Mean emoji-use share across second-degree contacts.
		Second-degree emoji share, SD	Standard deviation of emoji-use share across second-degree contacts.
Common Control	All lifetimes	Entry month	Calendar month in which the user joined Venmo.

Table B2: Summary Statistics at User Entry (Lifetime 0)

Variable	Mean	SD	Median	P10	P90
<i>Focal user</i>					
Facebook sign-in (indicator)	0.131	0.338	0	0	1
Days since last transaction	0	0	0	0	0
Night-time transaction share	0.637	0.481	1	0	1
Weekend transaction share	0.252	0.434	0	0	1
Transaction frequency per day	0	0	0	0	0
Number of charge transactions	0.096	0.294	0	0	0
Number of pay transactions	0.904	0.295	1	1	1
Emoji-use share	0.117	0.321	0	0	1
Outgoing-transaction share	0.512	0.5	1	0	1
Number of transaction partners	1	0.008	1	1	1
Number of new users brought	0.275	0.446	0	0	1
<i>First counterparty and counterparty-mediated information</i>					
First-counterparty transaction frequency	0.295	1.22	0.078	0	0.545
First-counterparty transaction count	14.35	112.175	3	0	32
First-counterparty recency (days)	7.088	15.23	1	0	21
First-counterparty tenure (days)	80.038	116.566	28	0	237
First-counterparty degree	7.187	94.86	2	0	13
First-counterparty outgoing share	0.342	0.355	0.286	0	1
First-counterparty charge share	0.082	0.187	0	0	0.313
First-counterparty emoji share	0.091	0.203	0	0	0.333
First-counterparty transactions, prior 30 days	5.061	51.129	2	0	10
First-counterparty PageRank	1.206	18.294	0.368	0	2.416
New first counterparty (indicator)	0.275	0.446	0	0	1
Community growth share, prior 30 days	0.373	0.407	0.2	0	1
Community growth share, prior 90 days	0.574	0.431	0.667	0	1
Share of members tied only to the counterparty	0.143	0.264	0	0	0.5

continued on next page

Table B2 continued from previous page

Variable	Mean	SD	Median	P10	P90
<i>Entry community: structure</i>					
Entry-community clustering coefficient	0.097	0.198	0	0	0.333
Ties among community members	4.673	27.153	0	0	9
Burt constraint of the first counterparty	0.309	0.342	0.189	0	1
Share of isolated community members	0.472	0.429	0.375	0	1
Largest-component share	0.447	0.372	0.429	0	1
Two-hop reach	130.213	999.229	6	0	75
Expansion ratio	16.671	151.553	2	0	9.389
<i>Entry community: activity</i>					
Internal transactions, prior 30 days	3.337	17.382	0	0	8
External transactions, prior 30 days	51.115	539.748	3	0	46
Internal transaction share	0.07	0.154	0	0	0.25
Transactions per member, prior 30 days	6.55	58.105	1.222	0	6.889
Active-member share, prior 30 days	0.475	0.41	0.5	0	1
Activity concentration (HHI)	0.338	0.369	0.213	0	1
Internal transactions, full history	18.277	117.315	0	0	34
Mean member recency (days)	15.025	27.229	6.6	0	38.333
No qualifying member activity (indicator)	0.05	0.218	0	0	0
<i>Entry community: maturity and composition</i>					
Community age (days)	99.951	105.811	76	0	244
SD of member tenure	56.596	70.512	26.163	0	156.016
New-member share, prior 30 days	0.176	0.301	0	0	0.667
New-member share, prior 90 days	0.347	0.378	0.2	0	1
Mean member degree	18.42	151.841	3.857	0	13.2
SD of member degree	32.617	260.116	1.258	0	9.883
Mean member transaction frequency	0.243	0.938	0.107	0	0.431
SD of member transaction frequency	0.255	1.299	0.038	0	0.343
Mean member emoji share	0.086	0.148	0.005	0	0.277
SD of member emoji share	0.073	0.114	0	0	0.237
<i>Network position</i>					
Counterparty in giant component (indicator)	0.645	0.478	1	0	1

Note. $N = 197,970$. User recency and transaction frequency are zero at entry by construction. PageRank is reported on the native GraphFrames scale.

C Targeting Value Under Fixed Capacity

Table C1 reports lifetime-0 targeting performance at capacities of 5%, 10%, and 20% of the entrant cohort. Table C2 compares entry-community performance with the counterparty benchmark.

Table C1: Lifetime-0 Targeting Performance at Fixed Capacities

Task	Model	Capacity	Targeted	Positives identified	Precision	Cumulative gain	Lift	Additional vs M1 per 100,000 [95% CI]
Inactivity risk	Focal-user (M1)	5%	2,968	1,337	0.451	0.056	1.11	—
	Counterparty benchmark (M4)	5%	2,968	1,410	0.475	0.059	1.17	+123 [+8; +254]
	Community-only (M5)	5%	2,968	1,453	0.490	0.060	1.21	+195 [+76; +310]
	Counterparty + community (M6)	5%	2,968	1,530	0.515	0.064	1.27	+325 [+202; +457]
	Relational-network (M2)	5%	2,968	1,454	0.490	0.060	1.21	+197 [+81; +318]
	Combined (M3)	5%	2,968	1,455	0.490	0.060	1.21	+199 [+91; +317]
	Focal-user (M1)	10%	5,935	2,660	0.448	0.111	1.10	—
	Counterparty benchmark (M4)	10%	5,935	2,826	0.476	0.117	1.17	+280 [+108; +455]
	Community-only (M5)	10%	5,935	2,836	0.478	0.118	1.18	+297 [+121; +462]
	Counterparty + community (M6)	10%	5,935	2,873	0.484	0.119	1.19	+359 [+197; +544]
	Relational-network (M2)	10%	5,935	2,842	0.479	0.118	1.18	+307 [+131; +468]
	Combined (M3)	10%	5,935	2,878	0.485	0.119	1.20	+367 [+214; +527]
	Focal-user (M1)	20%	11,870	5,147	0.434	0.214	1.07	—
	Counterparty benchmark (M4)	20%	11,870	5,407	0.456	0.225	1.12	+438 [+216; +605]
	Community-only (M5)	20%	11,870	5,442	0.459	0.226	1.13	+497 [+273; +708]
	Counterparty + community (M6)	20%	11,870	5,538	0.467	0.230	1.15	+659 [+388; +827]
	Relational-network (M2)	20%	11,870	5,434	0.458	0.226	1.13	+484 [+256; +684]
	Combined (M3)	20%	11,870	5,543	0.467	0.230	1.15	+667 [+455; +849]
High-engagement	Focal-user (M1)	5%	2,968	723	0.244	0.060	1.21	—
	Counterparty benchmark (M4)	5%	2,968	773	0.260	0.065	1.29	+84 [−39; +174]
	Community-only (M5)	5%	2,968	981	0.331	0.082	1.64	+435 [+310; +527]
	Counterparty + community (M6)	5%	2,968	977	0.329	0.082	1.63	+428 [+307; +516]
	Relational-network (M2)	5%	2,968	979	0.330	0.082	1.63	+431 [+307; +524]
	Combined (M3)	5%	2,968	1,003	0.338	0.084	1.68	+472 [+356; +558]
	Focal-user (M1)	10%	5,935	1,457	0.245	0.122	1.22	—
	Counterparty benchmark (M4)	10%	5,935	1,534	0.259	0.128	1.28	+130 [−10; +270]
	Community-only (M5)	10%	5,935	1,767	0.298	0.148	1.48	+522 [+372; +661]
	Counterparty + community (M6)	10%	5,935	1,803	0.304	0.151	1.51	+583 [+431; +728]
	Relational-network (M2)	10%	5,935	1,766	0.298	0.147	1.47	+521 [+381; +672]
	Combined (M3)	10%	5,935	1,831	0.308	0.153	1.53	+630 [+490; +775]
	Focal-user (M1)	20%	11,870	2,824	0.238	0.236	1.18	—
	Counterparty benchmark (M4)	20%	11,870	2,994	0.252	0.250	1.25	+286 [+174; +532]
	Community-only (M5)	20%	11,870	3,268	0.275	0.273	1.36	+748 [+644; +1013]
	Counterparty + community (M6)	20%	11,870	3,337	0.281	0.279	1.39	+864 [+755; +1122]
	Relational-network (M2)	20%	11,870	3,291	0.277	0.275	1.37	+787 [+669; +1045]
	Combined (M3)	20%	11,870	3,343	0.282	0.279	1.40	+874 [+785; +1129]

Note. Lifetime-0 holdout sample ($n = 59,350$). Intervals are paired-bootstrap 95% confidence intervals. Additional counts are relative to Model 1 at the same targeting capacity and are not treatment effects.

Table C2: Entry-Community Information Relative to the Counterparty Benchmark

Task	Capacity	Δ Lift (M6 – M4)	Additional per 100,000 (M6 – M4) [95% CI]	Additional per 100,000 (M5 – M4) [95% CI]
Inactivity	5%	+0.10 [+0.05; +0.15]	+202 [+97; +296]	+72 [−57; +178]
	10%	+0.02 [−0.01; +0.05]	+79 [−38; +208]	+16 [−117; +171]
	20%	+0.03 [+0.00; +0.04]	+220 [+35; +353]	+59 [−96; +251]
High engagement	5%	+0.34 [+0.23; +0.44]	+343 [+232; +441]	+350 [+240; +458]
	10%	+0.22 [+0.17; +0.28]	+453 [+335; +564]	+392 [+264; +513]
	20%	+0.14 [+0.11; +0.18]	+577 [+446; +717]	+461 [+325; +616]

Note. Model 4 is the counterparty benchmark, comprising first-counterparty and counterparty-mediated information; Model 5 uses entry-community information, and Model 6 combines the two. Intervals are paired-bootstrap 95% confidence intervals.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with language editing and clarity of presentation. The authors reviewed and edited the output as needed and take full responsibility for the content of the publication.

References

- Anderson, S.P., de Palma, A., 2012. Competition for attention in the information (overload) age. *RAND Journal of Economics* 43, 1–25.
- Ascarza, E., Ebbes, P., Netzer, O., Danielson, M., 2017. Beyond the target customer: Social effects of customer relationship management campaigns. *J. Mark. Res.* 54, 347–363.
- Ascarza, E., Neslin, S.A., Netzer, O., Anderson, Z., Fader, P.S., Gupta, S., Hardie, B.G., Lemmens, A., Libai, B., Neal, D., et al., 2018. In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions* 5, 65–81.
- Benoit, D.F., Van den Poel, D., 2012. Improving customer retention in financial services using kinship network information. *Expert Syst. Appl.* 39, 11435–11442.
- Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 30, 107–117.
- Chaudhuri, N., Gupta, G., Vamsi, V., Bose, I., 2021. On the platform but will they buy? predicting customers’ purchase behavior using deep learning. *Decision Support Systems* 149, 113622.
- Coleman, J.S., 1988. Social capital in the creation of human capital. *American journal of sociology* 94, S95–S120.
- Coussement, K., Benoit, D.F., 2021. Interpretable data science for decision making. *Decision Support Systems* 150, 113664.

- Esmeli, R., Abdullahi, H., Bader-El-Den, M., Can, A.S., 2024. Session context data integration to address the cold start problem in e-commerce recommender systems. *Decision Support Systems* 187, 114339.
- Goel, S., Goldstein, D.G., 2013. Predicting individual behavior with social networks. *Marketing Science* 33, 82–93.
- Goldfarb, A., Tucker, C., 2011. Online display advertising: Targeting and obtrusiveness. *Marketing Science* 30, 389–404.
- Gopalakrishnan, A., Bradlow, E.T., Fader, P.S., 2016. A cross-cohort changepoint model for customer-base analysis. *Marketing Science* 36, 195–213.
- Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91, 481–510.
- Granovetter, M.S., 1973. The strength of weak ties. *American Journal of Sociology* 78, 1360–1380.
- Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., Tillmanns, S., 2010. Undervalued or overvalued customers: Capturing total customer engagement value. *Journal of Service Research* 13, 297–310.
- McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27, 415–444.
- Padilla, N., Ascarza, E., 2021. Overcoming the cold start problem of customer relationship management using a probabilistic machine learning approach. *Journal of Marketing Research* 58, 981–1006.
- Phillips-Wren, G., Daly, M., Burstein, F., 2021. Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. *Decision Support Systems* 146, 113560.
- Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J., 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109, 5962–5966.
- Verhoef, P.C., Donkers, B., 2005. The effect of acquisition channels on customer loyalty and

- cross-buying. *Journal of Interactive Marketing* 19, 31–43.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. *nature* 393, 440–442.