# Do Referral Programs Drive Loyalty? *

Xintong Han[†], Shaojia Wang[‡], and Tong Wang[§]

**Abstract**

Using unique data from a leading Chinese content platform with more than 300,000 users, we propose a structural approach to evaluate the effect of the structure of a referral network on users' renewal decisions. Referral networks provide essential identification sources, which enable us to embed the expectation of network peers' behavior into the utility function as an important component to capture the decision variations. We find that these networks play an essential role in users' renewal decisions, which are significantly and positively associated with the renewal decisions of both referrers and referrals. Our counterfactual analysis has important implications on the referral policies of digital platforms. First, we find that the referral-targeted discount discrimination policy is more effective than the uniform discount policy. More optimistic expectations for referrals' decisions due to the price discount generate a snowball effect on referral networks, which in turn increases renewal rates. Compared to a uniform discount policy, a more referral-targeted discount policy would significantly increase renewal rates while reducing overall revenue loss. Second, our results highlight the importance of the structure of a referral network. With the same beta index, a high-centrality network implies a reduction in the chain hierarchy, which is detrimental to customer retention. We suggest that an efficient referral network should be highly connected with a lower degree of closeness-based centrality.

**Keywords:** referral network, renewal decision, structural estimation, word-of-mouth marketing

*Acknowledgements: We thank Jan Victor Dee, Thierry Magnac, Lei Xu, and our colleagues at Concordia and Edinburgh as well as seminar and conference participants at Canada IO conference for helpful comments. We gratefully thank ShenZhen UNNOO Information Technology Co.,Ltd (the owner of the platform) for data provision, and Pu Zhao for his research assistance. The conclusions drawn from the data are those of the researchers and do not reflect the views of the platform. All remaining errors are our own.

†Concordia University and CIREQ, Department of Economics, 1455 Boulevard de Maisonneuve Ouest, Concordia University, Montreal, H3G 1M8, Canada. Email: xintong.han@concordia.ca

‡Concordia University, Department of Economics, 1455 Boulevard de Maisonneuve Ouest, Concordia University, Montreal, H3G 1M8, Canada. Email: shaojia.wang@concordia.ca

§University of Edinburgh, Business School, 29 Buccleuch Pl, Edinburgh EH8 9JS. United Kingdom. Email: tong.wang@ed.ac.uk

# 1 Introduction

*Price does not rule the web; trust does.*

— *Reichheld and Schefter, 2000, Harvard Business Review.*

Many products in the digital economy rely on users' loyalty (e.g., Dropbox, Fizz and Netflix), where the platform faces both the problem of customer acquisition and retention. Among many platform policies, the referral program is a common word-of-mouth (WOM) marketing tool designed to motivate existing users to refer their family, friends, and contacts to become new customers (e.g., Heskett et al. (1994); Reichheld and Schefter (2000)). The referral program allows both referrers and referrals to receive discounts or rewards when joining the platform, and successful referral programs often generate considerable business value. For example, Schmitt et al. (2011b) suggested companies would earn more than 16% profit from referred customers. Despite the abundant existing literature on referral program studies (e.g., Bryan et al. (2015); Belo and Li (2018); Van den Bulte et al. (2018); Han et al. (2019)), the link between referral programs and customer loyalty remains largely unexplored. In the 2013 Nilson Report, the authors stated that consumers are four times more likely to buy a product referred by a trusted friend than without a referral. In 2014, Forbes reported that more than 90% of consumers think that products recommended by friends are better than any form of advertising. By this logic, users' renewal decisions may also be affected by the decision of the referrer or referrals. This paper addresses the following questions through an empirical study: How does the referral relationship influence renewal decisions, and what are the most efficient referral policies given a specific type of referral network?

Theories have suggested two channels that might positively link the referral network and customer retention. First, referred customers are more likely to match the product more effectively. Since the referrers have a social connection with the referrals, the referral decision could thus be based on underlying characteristics of the referrals that are only observable to the referrers. Many studies have pointed out that referred customers, in general, fit the product better than non-referrals (e.g., Montgomery (1991); Kornish and Li (2010); Pallais and Sands (2016)). Second, the trust between referrers and referrals can be regarded as social capital (e.g., Karlan (2005)). The social capital theory contends that such social relationships are resources (Fernandez et al. (2000); Schmitt et al. (2011a)). Therefore, remaining in the same social network is likely to generate positive values for both referrers and referrals. Nevertheless, matching and social capital theory generate slightly different predictions on the influential power direction in a referral-referrer network. Matching theory implies a unilateral influence from the referrers to the referrals; that is, the extent to which referrers are familiar with the referrals will determine the effectiveness of matching. However, the social capital theory implies a bilateral influence between these two parties, since both referrers and referrals receive some positive value by purchasing the same product, the referrals' purchase decision should therefore also positively affect the referrers.

Both matching and social capital theories are extensively studied in current literature (e.g., Burks et al. (2015)). Distinguishing between these studies is challenging, since we will also need to separate the impact of the referrer's purchase decisions on referrals, and the impact of referral's decisions on referrers. In this paper, we use data provided directly by the KnowledgePlanet platform to study the effect of referral relationships on renewal decisions.[1] KnowledgePlanet is a well-known content provision platform in China, where users subscribe to the communities by paying annual fees. Each community has an owner who has access to specific pricing policies, while the owner is also responsible for managing the community and providing the content. A community's annual subscription fee varies from 50 RMB to 5,000 RMB.[2] A key and interesting feature of our data is that the platform implements a specific referral program, such that users who recommend others to join the community will share the subscription fee with the community owner. Furthermore, after completing their payment, referrals will receive a specific discounted annual fee. Since the referral program in our data is mutually beneficial for both referrals and referrers, all the referral behaviors are captured and recorded by the platform. Of all the referral links sent by users, 60% are sent via private chat software (QQ and WeChat[3]), so the referral relationship in our data also reflects the user's social network. The final data includes rich and abundant information about the payments, community characteristics, and referral relationships among more than 300,000 users from 204 randomly selected communities on the platform.[4] We construct the referral network for each community, and Figure 1 illustrates the referral network for a community of 5,332 members in the data.

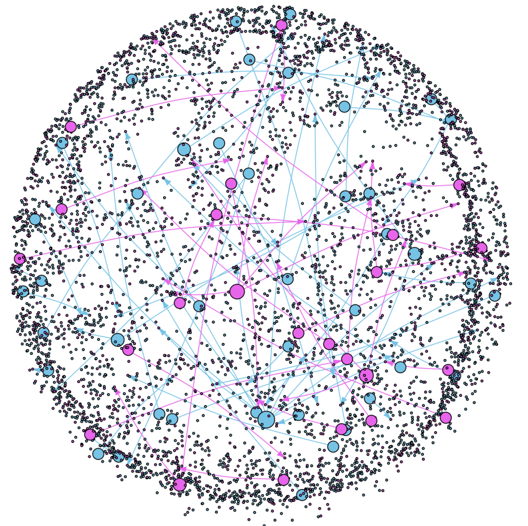
We first study the influence of price and referral networks on renewal decisions by using regression models. To deal with potential endogeneity problems, we adopt the instrumental variable (IV) method. Our instrumental variable is constructed based on the rate at which the community pays apple taxes on gross revenue. Android users and iPhone users pay the same price for their renewal. However, communities are required to pay additional

[1]Source: https://zsxq.com/

[2]One RMB is equal to approximately US$0.15. In 2020, Chinese residents' median per capita disposable income was approximately ¥27,540 (around US$4,262).

[3]QQ and WeChat are very similar to WhatsApp. They are two leading Chinese chat software applications, with approximately 80 percent (1.1 billion) of the Chinese population communicating through these two applications.

[4]Observing the entire network also helps to identify how shocks propagate through the network, since incomplete networks lead to biased estimates due to missing observations (e.g., Demir et al., 2018).



Notes: Each vertex in the above network diagram represents a user, with pink color for users who have joined in the last year and blue for users who have made at least one renewal decision. The vertex size represents the rewards that users receive through referrals, and the arrows represent the referral relationships between users. Among the 5,332 users, 3,569 are old users who made at least one renewal decision and 1,735 are new users who joined the community in the past year.

Figure 1: An illustration of a referral network in data

"Apple taxes" on the money they receive from the Apple platform, which further increases the renewal price set by the community owner if the ratio of iPhone users is high. The IV regression results show solid evidence that both the decision of the upstream network peer (referrer) and the number of downstream network peers (referrals) significantly and positively affect the focal user's renewal decisions. The bilateral nature of our results provides solid empirical evidence of the social capital theory. In terms of price, the estimated elasticity by IV estimation is 0.143. Our reduce-form results indicate that if the upstream referrer decides to renew, the impact on the probability that his or her referrals will also renew is equivalent to a 1% price discount.

In order to simulate the effect of referral policies on renewal rates, we further build a structural model to describe how users are making their renewal decisions and interacting with their network peers. The model explicitly imposes that when a user is faced with the renewal decision, she/he not only considers the renewal price and quality of the community but also provides feedback to her/his referrers' and anticipates the referrals' decisions. Under such a setting, the equilibrium status guarantees that each user in the community makes a renewal decision in line with the expectations of their network peers. Compared to traditional discrete choice models, the complexity of the network structure brings computational burdens to the estimation process. We propose a computational algorithm for estimating model parameters with given network structures, based on Su (2014) and Han and Xu (2018). Our estimation results show that the fitness of our structural model incorporating rational expectations through referral networks is significantly better than the traditional discrete selection models. Both upstream and downstream decisions in the referral network significantly and positively affect the renewal decisions of the focal user.

The structural model estimation results indicate that the effect of price changes will be transmitted through the referral network, which causes a "snowball" effect (e.g., Hada et al. (2014)). In the counterfactual analysis, we consider two scenarios of price changes: first, unified discount, where all users enjoy the same level of renewal price discount; and second, discount discrimination, where only referrals are entitled to a price discount when they decide to renew. The simulation results quantify the effect of price changes in the network and show that the referral network has contributed an additional 9% to 11% of the renewal rate. However, renewals come at the cost of less revenue: we find that a 1% increase in renewals may cost 4% of revenue.

Interestingly, we find that the network spillover benefit has a concave shape. We attribute this to a tug-of-war between price and network effects. When prices start to drop, only a small number of users decide to renew, and their decisions to renew affect their referrer and referrals. The network effect then begins to kick in and increases as prices fall further and the number of users affected increases. When the price drops sharply, the price becomes the dominant lever, and the network effect becomes insignificant. Even though most users are willing to renew their subscription at this time and continue to influence each other through the network, the main reason for renewing is the sharp drop in the price rather than the decision changes of others in the network. In order to maximize the network benefits while mitigating the loss of revenue, we examine the discount discrimination policy and find referral-targeted discount discrimination is more effective than the unified discount. Assigning special renewal discounts for referrals takes advantage of the snowball effect of the referral network, which dramatically reduces the revenue loss caused by the price decrease while increasing the renewal rate.

Last, we put all the results together to answer the question we posed earlier: What kind of referral structure is efficient for customer retention? We reconstruct the new referral network in Figure 1 based on the order in which users join the community. Four special types of network structures are considered: first, Serpentine, in which everyone has a referrer and a referral; second, Pyramid of second degree, where each user is referred by someone else and each of them also recommend two other users to join the community; third, Pyramid of third degree, each user is referred by someone else and each of them also recommend three other users to join the community; and fourth, Dictator, where all users are referred to as one user (i.e., a super influencer). These four network structures all have the same level of beta index (i.e., they have the same beta index so that the community gets exactly the same revenue when users first join), but their closeness-based centrality increases with type. Our counterfactual gets a counter-intuitive result. We find that high centralization in fact damages the propagation of network effects. In the case of the same degree of beta index, the highly centralized network is not conducive to the diffusion of effects. Although both the referrer's and the referral's decisions mutually influence each other, centralization also leads to a reduction in the chain hierarchy; this then, reduces the number of times the positive effect travels through the network, thus reducing the snowball effect even further.

Our paper contributes to the literature in the following aspects.

First, our paper complements the literature on digital economics and digital marketing. Recent studies show that network effects play a defining role in the digital world (e.g., Han and Xu (2018); Goldfarb and Tucker (2019); Boudreau et al. (2019)) and has important implications for business strategies. However, most of the studies are only limited to the effects from direct peers in the network (e.g., Chu and Manchanda (2016); Bailey et al. (2019); Shi et al. (2021)). How users' strategic behaviors interact and their influence propagates in the network is often neglected due to three reasons; first, it is difficult to observe the complete network in the data; second, the agent's decision-making process in the network is often endogenous; and third, depending on the size of the network, computational burdens are often considerable. In this paper, we provide a feasible method to measure the influence of network structure on renewal choice. Methodologically, our structure model is similar to the endogenous adoption models under the network (e.g., Gowrisankaran and Stavins (2004) ; Han and Xu (2018)). An important feature is that the referral network in our data not only indicates the recommendation relationship between users, but also implies the order in which users join and their decision-making time, which enables us to improve the estimation process to a large extent. Compared with the network literature in the field of marketing (Max Wei (2020) ; Hu et al. (2019)), our structural approach not only quantifies the effects of different pricing policies on the network, but also highlights the important role of social interaction in user relationship management (e.g., Ben Rhouma and Zaccour (2018)). The structural model based counterfactual analysis allows us to evaluate how the network higher-order statistics (e.g., closeness-based centrality) affect digital business revenue.

Second, we contribute to the literature on customer retention. Firms invest large amounts of resources into inducing customers to repurchase. An entire thread of marketing literature further studies the role of loyalty programs and different types of price promotions to encourage the retention of established customers (e.g., Bolton et al. (2000) ; Leenheer et al. (2007)). More recently, Hristakeva and Mortimer (2019) study the effects of legacy discounts in the national television advertising market. They find that firms that have longer-term relationships with broadcasters face lower prices in these networks. Despite these pioneering studies in the customer retention literature, empirical works on repeat-purchase are scarce, primarily due to the lack of data. Our paper links the customer acquisition policy to the customer retention problem. As an efficient tool for customer acquisition (e.g., Belo and Li (2018) ; Jung et al. (2020)), and for increasing product sales (Blanchard et al. (2018)), we find that the referral program also significantly and essentially affects the renewal decision of users since it creates a trust relationship between the referrer and referrals. The findings of our paper provide two important policy guidelines: first, we highlight that the long-term price discrimination policy for the referral program can effectively reduce the loss of platform revenue while increasing the renewal rate; and second, we find that the platform should guide and pursue the network structure with a low degree of centralization while improving the beta index. This makes the transmission of price policy smoother in the network and maximizes the renewal rates.

Third, we provide empirical evidence on the social capital theory. Other theories, such as the matching theory, may explain why referrals are more likely to be loyal customers. However, if social ties are resources, as stated by social capital theory, the (expected) referrals' purchase decisions should have an influence on that of the referrers too. The current literature contains only limited studies on the impact of this reverse link, mainly due to data limitation, and the literature on WOM generally focuses on the characteristics of referrals (e.g., Buttle (1998) and Trusov et al. (2009)). Our structural estimation finds that social ties increase the renewal rates of both referrers and referrals, which implies that customers do have social capitals concerns when making a purchasing decision.

# 2 Institutional Background

Our focal platform is KnowledgePlanet.[5] Since its online launch in 2015, KnowledgePlanet has grown to become one of the most widely-used Chinese text content provision platforms on knowledge and expertise sharing. The number of registered users is near 50 million, with

[5] Han et al. (2019) also uses this platform as their empirical environment for economic analysis. In their paper, the platform remains using its former name: "Little Secret Communities." The paper mainly focuses on the effect of the referral program on the community owners' pricing strategies.

500,000 daily active users. On the platform, there exist entities called 'communities', that include an owner, users, and (primarily text) content. The platform provides a marketplace where community owners can monetize their knowledge or expertise. The owners are often successful content providers on other platforms where they have accumulated a significant number of followers. By announcing the existence of a community to their followers, the owners obtain the first group of users and start to make profits.

Communities provide the content on the platform. Each community has an owner with absolute management and pricing power, who can set prices, discounts, and other relevant policies. In this paper, we focus on paid communities in which users pay an annual subscription fee for full access to content. Furthermore, we focus on those communities within which at least one user has renewed their subscription. In other words, these community owners have experience in both customer acquisition and retention. In the following subsections, we discuss the platform-designed tools utilized by the owners to attract new users and then to retain them when their subscription is about to expire. It is worth mentioning that user joining must be completed by binding personally identifiable information, so that there is no opportunity for a user to set up multiple accounts and to join the same community as a new user each time.

## 2.1 Referral Program as a Tool of Customer Acquisition

To help community owners to attract more users, the platform develops the referral program. Owners set up a sharing percentage between 0%~50%, which establishes the referral reward to the referrers and the referrals. Han et al. (2019) evaluate the effect of the referral program on pricing strategies. In summary, the referral program is beneficial for users because referrers receive additional financial returns for their referrals, and referrals receive a specific discount when they join the community for the first time. However, the referral program is not always good for community owners, because while it incentivizes customer acquisition it also causes users (who would otherwise pay the full price to join) to opt-in through referrals, indirectly reducing the income of the community owners.

Table 1 illustrates the case of how the referral program and the renewal discount work on the platform. In this case, the annual fee stays at US$100 over two years, and the referral reward is 50%, while the renewal discount is 20%. Suppose that an existing user (referrer) in a community posts a sharing link for the referral. A person (referral) who joins the community has to pay $100 for annual access to content through this link. The community owner takes 50% × ($100) = $50. The remaining $50 referral reward is further split according to a fixed ratio preset by the platform, whereby the referrer obtains 70% (i.e., $35) and the referral

| | owner | referrer | referral | new user without referral |
|---|---|---|---|---|
| Year 1 (with referral) | \$50 | \$35 | -\$85 | |
| Year 1 (without referral) | \$100 | | | -\$100 |
| Year 2 | \$80 | | -\$80 | -\$80 |

Notes: In practice, community owners and users will have to pay an additional 30% commission fee to the platform for all transactions incurred, which we will not consider here.

Table 1: Pricing under \$100 initial price, 50% referral reward, and "20% off" renewal discount

obtains 30% (i.e., \$15).

## 2.2 Renewal Discount as a Tool of Customer Retention

The platform allows all owners to set up a renewal discount for customer retention purposes. The community owner can set up the renewal discount rate and the discounted renewal price is then immediately observed on the same page. A user receives a renewal reminder when the yearly subscription is about to expire. Once the user allows the subscription to expire without renewal, she will not be able to review any new content posted afterwards. The renewal discount is notably visible under the renewal notification where the number of days since expiration is counted. Figure 2 illustrates the renewal process. In Table 1, when a "new" user becomes "old" (one year after joining the community), she will face a renewal decision. In this case, she needs to pay \$80 to the owner under the 80% renewal discount (i.e., "20% off"). It is worth mentioning that the referral program is only beneficial to both referrers and referrals in the first year. From the second year on, whether or not a user is referred will have no effect on the renewal price.

# 3 Data Description and Reduced Form Evidence

Our data are directly provided by the KnowledgePlanet platform. All communities in our data have at least one user renewed for the subscription. As of 2020, the platform has nearly 2,000 active communities and has amassed more than 2 million paying users. Communities in the platform are tagged according to their type, which represents eight fields of expertise: Art, Economics, Education, Entertainment, Fashion, Health, Life, and Science. Among all the communities, the number of communities with economics- and science-oriented topics is relatively large, and users in these communities are also more rational and cautious when they make payments. Therefore, we focus on the economics-oriented and science-oriented communities that satisfy the following criterion: first, the communities are more than 1.5


Notes: The community owner can determine the level of renewal price. The figure on the top illustrates that when users' subscriptions are about to expire, the platform will alert the user (i.e., "Service About to Expire"), and users have two options: close the window or renew their membership. If the users decide to renew, they will be directed to the renewal page. The figure on the bottom illustrates how they are notified about the renewal price (i.e., "¥199/year"), and the transaction will end when users click on "Pay now" to complete the payment.

Figure 2: Illustration of the renewal process

years old so that at a certain number of users have made a renewal decision at least once. Second, the communities have more than 50 users, and the owner has updated the content at least once in the past week. This is to ensure that the community owner is still actively offering content.

Overall, there are more than 600 communities that meet the above conditions. We randomly select one-third of the communities and require the platform to provide complete data of the users from these selected communities. The final data includes more than 300,000 users from 204 randomly selected communities.

## 3.1 Price Matching and Descriptive Statistics

It is important to note that in the data, the renewal price for the users who decide not to renew their subscriptions are unobserved. Without corresponding prices, we cannot identify the price elasticity associated with user renewal decisions. We take the necessary steps to match price data for these users. First, for all the users who decide not to renew, we look for the users who renew successfully in the same month. Then, we take the renewal price of these users as the price that the user who did not renew should have paid. Second, in rare cases, there are no renewals during the month. We then search the renewal prices paid by the users who agree to renew in the previous month and the next month. We average these prices as the renewal prices users should pay in the current month. According to the above algorithm, the prices of more than 96% of missing prices are matched successfully.

| Statistics | Mean | St.Dev. | Min | Median | Max |
|---|---|---|---|---|---|
| All Communities (N=204) | | | | | |
| Number of Users | 1,661.21 | 4,994.84 | 54.00 | 409.50 | 59,185.00 |
| Number of Articles | 4,443.25 | 8,289.31 | 1,673.00 | 96.00 | 74,692.00 |
| Number of Answers | 1,315.79 | 3,714.49 | 0.00 | 109.00 | 26,399.00 |
| Community Type | 0.10 | 0.30 | 0.00 | 0.00 | 1.00 |
| Age | 3.02 | 0.79 | 1.67 | 3.00 | 5.17 |
| All users (N=338,708) | | | | | |
| Initial Price (RMB) | 342.06 | 285.53 | 0.00 | 284.00 | 5,888.00 |
| Renewal Price (RMB) | 402.18 | 640.67 | 25.00 | 209.30 | 5,888.00 |
| Renewal Decision | 0.38 | 0.49 | 0.00 | 0.00 | 1.00 |
| Renewal Time | 0.56 | 0.82 | 0.00 | 0.00 | 4.00 |
| Smartphone Model | 0.09 | 0.29 | 0.00 | 0.00 | 1.00 |
| Referred | 0.08 | 0.26 | 0.00 | 0.00 | 1.00 |
| Number of Referrals | 0.07 | 4.57 | 0.00 | 0.00 | 1321.00 |

Notes: For community-level statistics: 1) Community Type equals 1 for science-oriented communities and 0 for economic-oriented communities. 2) There are only three communities with zero questions and answers. The main content of these communities is presented in the form of articles provided by the community owners so that these community owners close the Q&A function. 3) We collected data in June 2021. The age is the time from the establishment of the community to the collection of data. 4) There are only five communities in which the number of pictures is less than 10. For user-level statistics: 1) One RMB is equal to about US$0.15 US; 2) The minimum renewal price is 50 RMB, but the community owner can discount up to 50%. Therefore, the minimum renewal price is 25 RMB; 3) Smartphone Model is 1 for iPhone users and 0 for non-iPhone users; and 4) The variable Referred equals 1 when a user is referred and 0 otherwise.

Table 2: Statistics of key variables for communities and users

Table 2 gives the summary statistics of key variables. A representative community in our data has 1,661 users and is approximately three years old, while 10% of all communities are science-oriented, and 90% are economics-oriented. On average, more than 1,000 articles appear in the community each year, which equates to roughly three articles posted each day. On the user-side, 28% of users have joined the platform in the last year. When users join the community, the average price is 342 RMB, however, the average renewal price is 402 RMB, which is higher than the initial price. The reason for this is that the community owners may generally increase the price over time with the development and prosperity of the community. Only 8% of our users are referrals, and there is one referral per 14 users. The average renewal rate for users is 38%, and the average number of renewals is only 0.56. iPhone users account for less than 10% of the total.

## 3.2 Reduced Form Evidence

We first reveal the relationship between price, referral network, and renewal decision through a linear regression model. For a user $i$ in the community $m$ who faces a renewal decision at year $t$, we consider the following linear probability model (LPM):

$$\begin{aligned} d_{i,t} =& \beta_p \times \ln price_{i,t} + \underbrace{\alpha_r \times R_i + \alpha_{r,d} \times RD_{i,t} + \alpha_p \times nR_{i,t}}_{Network\,Effect} + \\ & \gamma_x' x_{i,t} + \gamma_m' z_{m,t} + u_{i,t}, \end{aligned}$$

where $d_{i,t}$ is a binary outcome variable that indicates whether user $i$ decides to renew; $price_{i,t}$ is the renewal price that the users get at the time of making renewal decisions; $R_i$ is a binary variable that indicates if user $i$ is referred by others; $RD_{i,t}$ is another binary variable that shows whether user $i$'s referrer decides to renew at year $t$; $nR_{i,t}$ is a variable indicates the number of users that $i$ has referred at year $t$; $x_{i,t}$ is a vector of user specific characteristics including when the user first joined the community and the initial price she paid; $z_{m,t}$ is a vector of community specific characteristics including the type of community, the total number of articles, and the number of Q&As in the past year; $u_{i,t}$ is an unobserved error term. In the above equation, $\beta_p$ captures the effect of price changes; $\alpha_r$, $\alpha_{r,d}$, and $\alpha_p$ capture the linear correlation between the renewal decision of user $i$ and her peers (i.e., $i$'s referrer and referrals) and in the referral network. $\gamma_x$ and $\gamma_m$ represent the effect of personal characteristics and community quality on $i$'s renewal decision, respectively.

The estimation results are reported in Table 3. Since the referrals join the community because of the referrer's marketing, if the quality of the community is not as good as expected, their willingness to renew would be low. Both the referrer's renewal decision and the number of referrals positively and significantly affect the user's renewal decision. Compared with science-oriented communities, users in economics-oriented communities are more willing to renew their payments because economics-oriented communities often provide content that rewards users financially (e.g., stocks, investment information).

Furthermore, we find a significant positive correlation between the initial price paid by users when they join the community and their willingness to renew. This shows that the initial price is a good measure for controlling the type of user: since users join the community at different times, they pay different prices when they join, even in that community. Users who are willing to pay a higher initial price to join the community tend to be those who have a stronger renewal preference. In column (1) to column (4), the correlation between price and users' willingness to renew is significantly negative. Although this result is intuitive, some possible endogeneity issues still need to be addressed. For example, some factors related to community quality that cannot be captured in the data may affect both users' decisions and community owners' pricing strategies.

### 3.3 Instrumental Variables: "Apple Tax"

To deal with the endogeneity problem related to price, we propose "Apple Tax" as an instrumental variable. Using different tax rates as an exogenous source to identify the effect is common in social science. For example, Verboven (2002) chose automotive taxes as instrumental variables to measure the price elasticity on the demand of diesel and gasoline vehicles. In our data, we adopt a similar logic. If a user logs in to the platform with an iPhone and completes the payment, Apple will charge a commission of 30% on the purchases made in the app. Therefore, while all users pay the same amount of annual subscription fee, the existence of Apple Tax results in the community owner receiving less from Apple users. In the example in Table 1, in the absence of Apple tax, the renewal price for an Android user is \$80, and the community owner receives \$80. With the Apple tax, the iPhone user still pays \$80 but the community owner has to pay 30% (i.e., \$24) of her revenue to the Apple platform, so she only receives \$56.[6]

Suppose there is a high proportion of iPhone payments in the community. In that case, the owner will have more willingness to increase the price to reduce the loss caused by tax, making our instrumental variable meet the *relevance condition*. At the same time, since iPhone users pay exactly the same price as non-iPhone users, the percentage of iPhone users does not directly affect a user's renewal decision. Therefore, our instrumental variable also satisfies the *exclusion restrictions*. In the data, iPhone users account for about 10% of the total number of users.

Columns (4) and (5) in Table 3 show the IV estimation results. After correcting the endogeneity issues, the IV estimation shows that the OLS method underestimates the price effect. In fact, users are much more flexible with the IV estimate, showing that the effect of price reductions is slightly larger than the OLS estimate: a one percentage point increase in price leads to a decrease in the probability of renewal by 0.143%. At the same time, the coefficients before the other variables remain unchanged when applying instrumental variables, which shows that our instrumental variable is stable. The first stage IV regression results in column (5) are also in line with our expectations: as the number of iPhone users

[6] The community owner actually receives less in the real world because she has to pay an additional commission fee to the platform.

increases, community owners will have a higher incentive to raise prices to share the revenue loss caused by the "Apple Tax" with users.

| *Dependent variable:* | *Renewal Decision* ($d$) | | | | $\ln(Price)$ |
|---|---|---|---|---|---|
| | OLS | Probit | Logit | IV | IV First Stage |
| | (1) | (2) | (3) | (4) | (5) |
| $\ln(Price)$ | $-0.111^{***}$ | $-0.338^{***}$ | $-0.573^{***}$ | $-0.143^{***}$ | |
| | (0.001) | (0.004) | (0.007) | (0.003) | |
| $RD$ | $0.125^{***}$ | $0.307^{***}$ | $0.481^{***}$ | $0.121^{***}$ | $-0.036^{***}$ |
| | (0.007) | (0.019) | (0.031) | (0.007) | (0.007) |
| $Apple\,Tax$ | | | | | $2.702^{***}$ |
| | | | | | (0.007) |
| $\ln(Price_0 + 1)$ | $0.051^{***}$ | $0.175^{***}$ | $0.293^{***}$ | $0.058^{***}$ | $0.561^{***}$ |
| | (0.001) | (0.002) | (0.004) | (0.001) | (0.001) |
| $Community\,Type$ | $-0.239^{***}$ | $-0.753^{***}$ | $-1.371^{***}$ | $-0.240^{***}$ | $-0.063^{***}$ |
| | (0.004) | (0.014) | (0.025) | (0.004) | (0.004) |
| $\ln(N\_Answers + 1)$ | $0.007^{***}$ | $0.022^{***}$ | $0.038^{***}$ | $0.007^{***}$ | $0.032^{***}$ |
| | (0.001) | (0.002) | (0.003) | (0.001) | (0.001) |
| $\ln(N\_Article + 1)$ | $0.061^{***}$ | $0.162^{***}$ | $0.269^{***}$ | $0.057^{***}$ | $-0.073^{***}$ |
| | (0.001) | (0.003) | (0.005) | (0.001) | (0.001) |
| Network Statistics | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES | YES |
| Observations | 302,197 | 302,197 | 302,197 | 302,197 | 302,197 |
| $R^2$ | 0.103 | | | 0.101 | 0.458 |
| Adjusted $R^2$ | 0.103 | | | 0.101 | 0.458 |
| Log Likelihood | | $-191,483.600$ | $-191,220.900$ | | |

Notes: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01. Standard errors are reported in brackets. We use $\log(x+1)$ for the logarithmic transformation of variable $x$ to avoid the situation where $x = 0$, and $\log(x)$ does not exist. $Price_0$ is the initial price that the user paid when join the community. If some users get a special offer they may get the price free for the first year, but there is no such offer at the time of renewal. Community Type equals 1 for the science-oriented communities and 0 for economics-oriented communities. Year FE and Month FE are created based on when users first join the community. We do not include users who join the community within the last year and were not able to renew for the first time. In column (4), we construct an instrumental variable showing the percentage of iPhone users in the same community and same year for each user ($IV$). Network Statistics include whether the user is referred to join the community and the number of referrals. Year FE and Month FE are created based on when users first join the community

Table 3: Reduced-form regression results

| Statistic | *Mean* | *St.Dev* | *Min* | *Median* | *Max* | *Percentage* |
|---|---|---|---|---|---|---|
| Chat Software | 727.2 | 3,604.7 | 0.0 | 37.0 | 54,802.0 | 59.2% |
| Social Platform | 12.4 | 82.5 | 0.0 | 0.0 | 1,622.0 | 1.0% |
| Hyperlink | 488.7 | 2,160.2 | 0.0 | 26.0 | 32,454.0 | 39.8% |

Notes: There is total of 966,717 shared links from the community from 2016 to 2021 and 59.2% were sent via chat software. Summary statistics are based on the community annual data. Chat software includes WeChat and QQ; Social Platform includes Weibo; Hyperlink includes Invite-URL (directly share invite links) and Copy-URL (copy invite links and share to others).

Table 4: Descriptive statistics

## 3.4 Nature of Referral Links

The results in Table 3 indicate that the referral program generates a positive significant effect on renewal decisions. Thus, a natural question is: do users preferentially recommend the community to people they trust more? In the data, information on whether users who join by a referral link know the person is not directly provided. Referrals can be either followers of the referrer on other platforms or family or friends of the referrer. However, the platform provides us with additional information on the channels where the referral links are shared through the platform.

A referral link can be shared through the following three channels: first, chat software (i.e., WeChat and QQ, which are the main social networking software commonly used by Chinese citizens); second, social platforms (e.g., Weibo); and third, hyperlinks to share information/websites. Additional data provided by the platform clearly indicates the type of channels through which the referral links were sent. We consider that the links sent through chat software channels (WeChat and QQ) are links that are sent based on trust relationships because the chat software contains most friends and relatives that users know in real life. This is also important for the underlying mechanisms: if the link is shared via social platforms, the referrers are less likely to personally know the referrals, thus showing that information matching and social ties tend to be weaker.

Out of 966,717 links shared by users, 572,345 (59.2%) were sent via chat software. Such findings suggest that most of referral relationships often come from real-life social networks (e.g., family members or friends), reflecting the trust relationship between users. In the next section, we provide a structural model to describe how users on the same referral network interact in making their renewal decisions. The following section will quantify this using a structural model approach.

# 4 Structural Model of Decision Making

In this section, we use a simple structural model to describe the service renewal decision for each user whose term of service is coming to an end. There is a total number of $M$ communities. We define $\mathcal{I}_{m,t}$ as a set of users in the community $m$ who have to make their renewal decision at year $t \in \{1, ..., T\}$. The total number of users in the community $m$ at year $t$ is calculated by by $n_{m,t} = \#\mathcal{I}_{m,t}$, for $m \in \{1, ..., M\}$, with $\#$ an operator that measures the cardinality of set. A user $i \in \mathcal{I}_{m,t}$ has to decide whether to renew her community service for one more year.[7] The decision variable is a binary choice variable: $d_{i,t} \in \{0, 1\}$, such that

$$d_{i,t} = \begin{cases} 1 & i \; decides \; to \; renew \; at \; year \; t, \\ 0 & i \; leaves \; the \; community \; at \; year \; t. \end{cases}$$

User $i$ will choose $d_{i,t} = 1$ if the utility of renewal is higher than leaving the community.

**Referral Network** A user in the community can be either a referrer or a referral. The referral relationship between users in community $m$ can be represented by a squared matrix $\mathcal{R}_{m,t} \in \mathcal{M}_{n_{m,t} \times n_{m,t}}$, where each element $r_{i,j} \in \{0, 1\}$ is a binary indicator such that

$$r_{i,j} = \begin{cases} 1 & if \; i \; is \; j's \; referrer \; at \; year \; t, \\ 0 & otherwise \end{cases}.$$

Each user can only be recommended by one person, but can be the referrer of many users. The total number of referrals of referrer $i$ at time $t$ is $\sum_{k \in \mathcal{I}_{m,t}} r_{i,k}$.

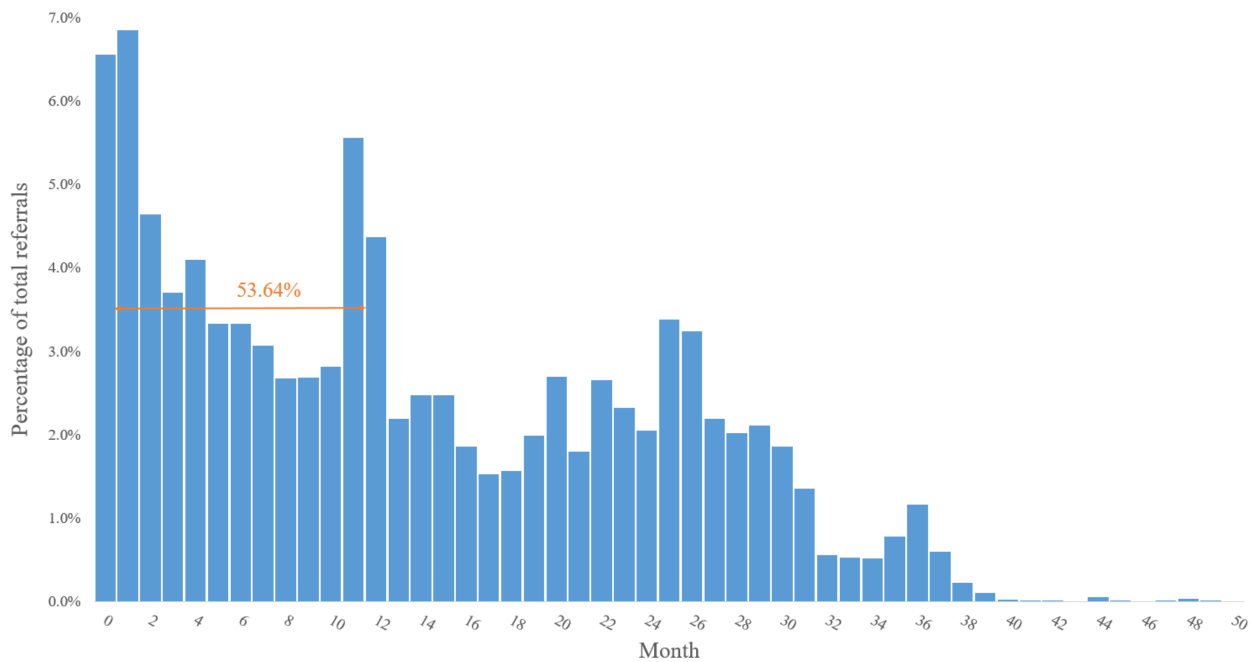
## 4.1 Utility Function

In this paper, we assume the renewal decisions of users who renew multiple times to be independent decisions.

**Assumption 1.** *Users are myopic so that they take the referral network at each year as given.*

The reason for making Assumption 1 is twofold: first, it allows us to largely avoid the computational burdens. In particular, it is almost impossible to embed the anticipation of the referral network's dynamic changes in the model; second, it does not cause us to lose the generality of the results. More than 85% of renewed users renew only once in our data,

[7] In the paper, we uniformly use "she" to refer to users in our data.



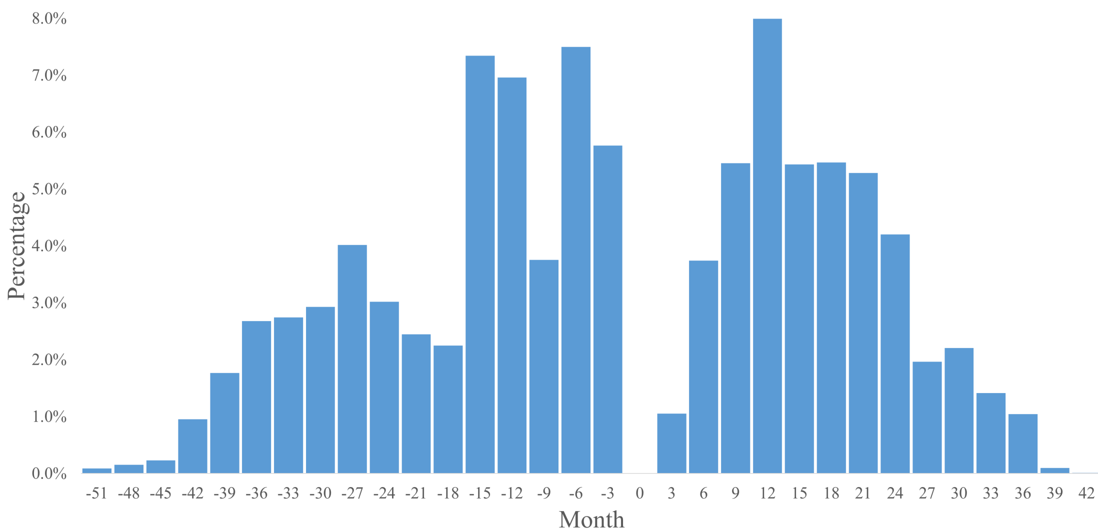

Notes: The x-axis provides the difference in join time between each referrer,, and the y-axis provides the frequencies. The join time gap is computed by subtracting the join time of the referral from that of the referrer. For example, the referrer joined the community in April 2020, and she invites a referral to join the community in April 2021: so, the join time gap is 12 months in this case.

Figure 3: Distribution of referrals' join time

and the average life of most communities is less than three years. Therefore, we believe that the choice of digital customers is relatively short-sighted: for the most part, users make renewal choices based on current circumstances rather than on long-term forecasts. Instead of using an infinite-horizon dynamic discrete choice model to add meaningless complexity, we use a classic random utility model to describe the scenario. Figure 3 further provides some empirical justifications supporting the validity of Assumption 1. It is shown that more than 50% of referrals join the community within a year of the referrer's joining. In particular, subscriptions of referrals most often take place within two months after the referrer completes their payment. If referrers are not short-sighted, they should optimally choose their time to join the community and recommend others: they would join when the community price is low and complete referral services when the community price is high to maximize their referral income, which is contrary to the statistical evidence in Figure 3.

**Assumption 2.** *Users do not coordinate with each other so that they do not make collective decisions.*

Assumption 2 is important to model users' decisions. It assumes that users do not collectively decide when to leave the community while joining the community through trust




Notes: The horizontal axis represents the difference of leaving time between two users in a referral relationship. The left part of the chart is that the referrer leaves the community earlier than the referrals (i.e., if the referral left the community in July 2019 and the referrer left in July 2020, then the month is 12). The right area on this graph shows the referrer who quits the community later than the referrals (i.e., if the referral left the community in July 2020 and the referrer left in July 2019, then the month is -12).

Figure 4: Distribution of the time gap in leaving the community between a referrer and her referral

and referral relationships. The reason to support this assumption is that the referrer and referrals do not benefit from negotiating the time to leave, as they will no longer be able to receive the community's content after they leave. At the same time, even if there is such an agreement between two users, after a user leaves, she has no way to bind the other to fulfill the promise. At the same time, a referrer usually recommends several referrals, which makes the cost of coordination higher. In the data, only 17% of couples in a referral relationship decide to leave in the same month. Figure 4 provides further evidence by showing the distribution of time gaps in leaving the community between a referrer and her referral. The data shows that half of the referrals leave before their referrer leaves, and the other half leave after. There is no significant evidence that referrers and referrals leave the community within an agreed period.

For user $i$ in community $m$, the utility of leaving the community at time $t$ is $u_m(i, d_{i,t} = 0) = \varepsilon^0_{i,t}$ with $\varepsilon^0_{i,t}$ represents $i$'s preference specific shock of choosing $d_{i,t} = 0$. The utility of renewing the membership is assumed to be the following:

$$u_m(i, d_{i,t} = 1) = \underbrace{\beta_p \times \ln price_{i,t}}_{price\ effect} + \underbrace{\alpha_r \times \left( \sum_{j \in \mathcal{I}_{m,t}} r_{j,i} d_{j,t} \right) + \alpha_p \times \mathbf{E} \left( \sum_{k \in \mathcal{I}_{m,t}} r_{i,k} d_{k,t} / \sum_{k \in \mathcal{I}_{m,t}} r_{i,k} | d_{i,t} \right)}_{network\ effect}$$

$$\underbrace{\gamma_x' x_{i,t} + \gamma_m' z_{m,t}}_{individual\ and\ community\ characteristics} + \varepsilon_{i,t}^1,$$

where $\beta_p$ measures the renewal price elasticity, $x_{i,t}$ is a vector of individual characteristics at time $t$ and $z_{m,t}$ is a vector of community characteristics that captures the quality of community at time $t$. $\varepsilon_{i,t}^1$ represents $i$'s preference specific shock of choosing $d_{i,t} = 1$. Preference specific shocks are assumed to follow Extreme Value Type 1 distribution, they are only observed by users but not by econometricians.

Interestingly, we add into the utility function two specific terms: first, the effect of the choice from the referrer ($\alpha_r$); and second, the effect of choices from the referrals ($\alpha_p$). $\alpha_r$ measures whether the referrer's renewal decision motivates referral $i$ to choose to stay in the community. $\mathbf{E}\left(\sum_{k \in \mathcal{I}_{m,t}} r_{i,k} d_{k,t} / \sum_{k \in \mathcal{I}_{m,t}} r_{i,k} | d_{i,t}\right) = \frac{\mathbf{E}\left(\sum_{k \in \mathcal{I}_{m,t}} r_{i,k} d_{k,t} | d_{i,t}\right)}{\sum_{k \in \mathcal{I}_{m,t}} r_{i,k}}$ is the expected proportion of $i$'s referrals who choose to renew after observing $i$'s decision at $t$ given the network structural $\mathcal{R}_{m,t}$. Referrers always need to make decisions earlier than their referrals on the platform since users first need to join the community to be eligible for the referral program. Therefore, $\alpha_p$ measures the influence of referrals' decisions on the referrer's choice and both $\alpha_r$ and $\alpha_p$ are considered as effects of peers in a referral network.

## 4.2 Decision Process

We now describe the users' decision-making process. To better understand this process, we start with a simplest illustrative scenario in which there are in total three users: $A$, $B$ and $C$ in the community $m$. $A$ is the referrer of $B$ such that $r_{A,B} = 1$, the referral network is given by the following matrix:

$$\mathcal{R} = \begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} .$$

The network structure is illustrated in Figure 5.

Let $v_i = \varepsilon_i^1 - \varepsilon_i^0$ being the difference of error terms, for $i \in \{A, B, C\}$. $v_i$ is logistically

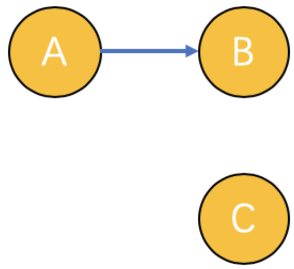



Figure 5: An illustration of the network structure of community $\{A, B, C\}$

distributed. For the user $C$, the user simply maximizes her utility function by choosing

$$d_c = \begin{cases} 1 & \beta_p \times \ln price_C + \gamma_x' x_C + \gamma_m' z_m + v_C \geq 0 \\ 0 & otherwise \end{cases}.$$

The decision of $A$ and $B$ is very similar to the Stackelberg model. For the user $B$, since $B$ is referred by $A$, after observing the decision of $A$ (i.e., $d_A$), $B$'s best response is

$$d_B\left(d_A\right) = \begin{cases} 1 & \beta_p \times \ln price_B + \alpha_r d_A + \gamma_x' x_B + \gamma_m' z_m + v_B \geq 0 \\ 0 & otherwise \end{cases}.$$

When $A$ has to make her decision, $A$ knows about $B$'s best response function but does not observe $v_B$. The value of $\mathbf{E}\left(\sum_{k\in\{A,B,C\}} r_{A,k} d_k | d_A = 1\right)$ is equivalent to

$$\mathbf{E}\left(\sum_{k\in\{A,B,C\}} r_{A,k} d_k / \sum_{k\in\{A,B,C\}} r_{i,k} | d_A = 1\right) = \mathbf{P}\left(d_B = 1 | d_A = 1\right) = \frac{e^{\beta_p \times \ln price_B + \alpha_r + \gamma_x' x_B + \gamma_m' z_m}}{1 + e^{\beta_p \times \ln price_B + \alpha_r + \gamma_x' x_B + \gamma_m' z_m}}.$$

More generally, in a referral network, a referrer $i$ will always make the decision first since she joins the community first. $i$ anticipates her referrals' decisions and chooses $d_i$ to maximizes her utility. After observing $d_i$, her referrals follow, they maximize the utility function by deciding whether to renew the memberships. Then the referrals of referrals will follow, and so on.

**Definition 1.** For a community $m$, we define the market equilibrium at time $t$ is a set of decisions $\mathcal{E}_{m,t} = \left\{d_{1,t}, ..., d_{i,t}, ..., d_{n_{m,t}}\right\}$ such that:

$$\forall i \in \mathcal{I}_{m,t}, d_{i,t} \in \arg\max_{d\in\{0,1\}} u_m(i, d|\mathcal{R}_{m,t}, \sum_{k\in\mathcal{I}_{m,t}} r_{k,i} d_k).$$

The computation of the market equilibrium becomes burdensome when the network expands, and referral relationships become complex. Moreover, when some users are both referrers and referrals, the decisions of their upstream referrers can have a particularly large snowball effect on the downstream.

## 4.3 Computation of Equilibrium

Here we propose an approach to compute the market equilibrium by adopting the idea of Han and Xu (2018). Han and Xu (2018) study the version adoption problem in a hierarchy network. The setting of their model is very similar to ours: they use a tree-like network, and the decisions of upstream users is observed by downstream users and affect downstream users. For a given community $m$, and its network $\mathcal{R}_{m,t}$ at year $t$, we start by classifying users in $\mathcal{I}_{m,t}$ according to their referral hierarchy. The lowest level is $\mathcal{L}_0^m = \left\{i \in \mathcal{I}_{m,t} | \sum_{j\in\mathcal{I}_{m,t}} r_{i,j} = 0\right\}$. $\mathcal{L}_0^m$ contains all the users who never recommend the community to anyone. $\mathcal{L}_1^m = \left\{i \in \mathcal{I}_{m,t} | \prod_{j\in\mathcal{L}_0^m} (1 - r_{i,j}) = 0\right\}$ contains all the users who recommend the community to users in $\mathcal{L}_0^m$. We define the highest level as $\bar{m}$, so $\mathcal{L}_{\bar{m}}^m$ is a set such that $\mathcal{L}_{\bar{m}}^m = \left\{i \in \mathcal{I}_{m,t} | \prod_{j\in\bar{m}-1} (1 - r_{i,j}) = 0 \; and \; \sum_{j\in\mathcal{I}_{m,t}} r_{j,i} = 0\right\}$. $\mathcal{L}_{\bar{m}}^m$ includes all the users who recommend the community to users in $\mathcal{L}_{\bar{m}-1}^m$, and these users have no referrers.

The equilibrium of market $m$ described as Definition 1 can be computed by using the following "backward anticipation and forward optimization" steps:

1. Start from $\mathcal{L}_0^m$, for each $i \in \mathcal{L}_0^m$, compute their best response function such that their probability of choosing $d_{i,t} = 1$ is a function of upstream decisions. We use $p_{i,t,d}^1$ and $p_{i,t,d}^0$ to denote the probabilities:

$$p_{i,t,d}^1 = \mathbf{P}\left(d_{i,t} = 1 | \left\{d_{j,t}\right\}_{j\in\mathcal{L}_1^m}\right), \quad p_{i,t,d}^0 = 1 - p_{i,t,d}^1.$$

   We notice that there are in fact two sets of probabilities for each $i$: one is the probability of renewing/leaving when $i$'s referrer decides to renew $p_{i,t,1}^1/p_{i,1,1}^0$: and the other is the probability of renewing/leaving when $i$'s referrer decides to leave $p_{i,t,0}^1/p_{i,t,0}^0$.

2. Given a set of predicted probabilities $\left\{p_{i,t,1}^1\right\}_{i\in\mathcal{L}_0^m}$, we move up one level. For all the $i \in \mathcal{L}_1^m$, we first calculate the value of $\mathbf{E}\left(\sum_{j\in\mathcal{L}_0^m} r_{i,j} d_{j,t} | d_{i,t} = 1\right)$ as a function of

$\left\{p^1_{i,t,1}\right\}_{i\in\mathcal{L}^m_0}$ and $d_{i,t}$, we use $E^d_{i,t}$ to denote it. Since $\mathbf{E}\left\{d_{j,t}|d_{i,t}\right\} = \mathbf{P}\left(d_{j,t}=1|d_{i,t}\right)$, we have

$$E^d_{i,t} = \sum_{j\in\mathcal{L}^m_0} r_{i,j}p^1_{j,t,1}.$$

3. We repeat Step 1 and 2, for each class $\mathcal{L}^m_k$, $k = 1, ..., \bar{m}-1$, we compute the choice probabilities for each user $i \in \mathcal{L}^m_k$ and get $\left\{p^1_{i,t,1}\right\}_{i\in\mathcal{L}^m_k}$. For a given set of $\left\{p^1_{i,t,1}\right\}_{i\in\mathcal{L}^m_k}$, we move to the upper level $\mathcal{L}^m_{k+1}$ and compute the value of $E^d_{i,t}$ for $i \in \mathcal{L}^m_{k+1}$.

4. For users in the highest class $\mathcal{L}^m_{\bar{m}}$, we use $\left\{E^d_{j,t}\right\}_{j\in\mathcal{L}^m_{\bar{m}-1}}$ to compute the level of utility and determine $\left\{d_{i,t}\right\}_{i\in\mathcal{L}^m_{\bar{m}}}$.

5. The last step is the forward optimization loop. For each class $\mathcal{L}^m_k$, $k = \bar{m}-1, ..., 0$, we use $\left\{d_{i,t}\right\}_{i\in\mathcal{L}^m_{k+1}}$ and $\left\{E^d_{i,t}\right\}_{i\in\mathcal{L}^m_k, d\in\{0,1\}}$ that we obtained from the previous steps to determine $\left\{d_{i,t}\right\}_{i\in\mathcal{L}^m_k}$ for each class.

6. The market equilibrium is given by $\mathcal{E}_{m,t} = \cup_{k\in\{0,...,\bar{m}\}}\left\{d_{i,t}\right\}_{i\in\mathcal{L}^m_k}$.

# 5 Model Identification and Estimation Strategy

We observe market equilibria $\left\{\mathcal{E}_{m,t}\right\}_{m=1,...,M,t=1,...,T}$ from the data. Each market equilibrium $\mathcal{E}_{m,t}$ is described by a set of decision variables $(d_{i,t})_{i\in\mathcal{I}_{m,t}}$ given the community specific network structure at time $t$ $\mathcal{R}_{m,t}$, a set of individual characteristics $(x_{i,t})_{i\in\mathcal{I}_{m,t}}$ and a vector of community features $z_{m,t}$. The structural model is embedded by a vector of parameters $\theta = (\beta_p, \alpha_r, \alpha_p, \gamma'_x, \gamma'_m)$ in which $\gamma_x$ and $\gamma_m$ are identified through dynamic changes in $x_{i,t}$ and $z_{m,t}$.

We are mainly interested in identifying $(\beta_p, \alpha_r, \alpha_p)$ where $\beta_p$ captures the price elasticity and $\alpha_r$ and $\alpha_p$ capture the effect of the decisions of referrer and referrals on a given user's renewal decision, respectively. The key identification problem is twofold: first, $\mathbf{E}\left(\sum_{k\in\mathcal{I}_m} r_{i,k}d_k/\sum_{k\in\mathcal{I}_m} r_{i,k}|d_i\right)$ cannot be directly observed, and its calculation relies on the value of $\theta$. Our data contains more than 300,000 users, and the network structure of each community is extremely complex, thus the direct estimation of model parameters is nearly impossible due to the computational burden; second, the renewal price is an endogenous factor. In this section, we discuss the solution to each issue one by one.

## 5.1 Step Estimation Algorithm

The full likelihood estimation relies on

$$\max_{\theta} \sum_{t=1,\ldots,T} \sum_{i\in\mathcal{I}_{m,t}} \sum_{m=1,\ldots,M} \ln \mathbf{P}\left(d_{i,t}|x_{i,t}, z_{m,t}, \mathcal{R}_{m,t}, \sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}, \mathbf{E}\left(\sum_{k\in\mathcal{I}_{m,t}} r_{i,k}d_{k,t}(\theta)\,|d_{i,t}=1\right);\theta\right)$$

where $\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}$ is the referrer's decision that is observed in the data and the method computing $\mathbf{E}\left(\sum_{k\in\mathcal{I}_{m,t}} r_{i,k}d_{k,t}(\theta)\,|d_{i,t}=1\right)$ is described in Section 4.3, which is somewhat burdensome. Inspired by the method provided in Su (2014), we use the following sequential estimation algorithm to tackle the computational issue.

1. We estimate a fake auxiliary model to build up our initial guess of $\theta$. For example, one can adopt:

$$\hat{\theta}_{old} \in \arg\max_{\theta} \sum_{t=1,\ldots,T} \sum_{i\in\mathcal{I}_{m,t}} \sum_{m=1,\ldots,M} \ln \frac{\exp\left\{\beta_p \times \ln price_{i,t} + \alpha_r \times \left(\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}\right) + \gamma_x' x_{i,t} + \gamma_m' z_{m,t}\right\}}{1+\exp\left\{\beta_p \times \ln price_{i,t} + \alpha_r \times \left(\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}\right) + \gamma_x' x_{i,t} + \gamma_m' z_{m,t}\right\}}.$$

   The above formula is the estimation of a classic logit model by ignoring the term of expectation.

2. Based on the previous estimate of $\theta$, we compute the predicted renewal probabilities of all user $i \in \mathcal{I}_{m,t}$ given that their referrer decides to renew (i.e., $\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t} = 1$), for all $m \in \{1, ..., M\}$ and $t \in \{1, ..., T\}$. We use $p^1_{i,t,1}\left(\hat{\theta}_{old}\right)$ to denote the probability.

3. We use the predicted probability to compute the expected value of $\sum_{k\in\mathcal{I}_{m,t}} r_{i,k}d_{k,t}(\theta)$ given $d_{i,t}=1$, we have:$\hat{E}^d_{i,t}\left(\hat{\theta}_{old}\right) = \sum_{k\in\mathcal{I}_{m,t}} r_{i,k}p^1_{k,t,1}\left(\hat{\theta}_{old}\right)$.

4. Based on the predicted values of $E^d_{i,t}$s, we update the estimation of $\theta$ by maximizing:

$$\hat{\theta}_{new} \in \arg\max_{\theta} \sum_{t=1,\ldots,T} \sum_{i\in\mathcal{I}_{m,t}} \sum_{m=1,\ldots,M} \ln \frac{\exp\left\{\beta_p \times \ln price_{i,t} + \alpha_r \times \left(\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}\right) + \alpha_p \hat{E}^d_{i,t}\left(\hat{\theta}_{old}\right) + \gamma_x' x_{i,t} + \gamma_{m,t}' z_{m,t}\right\}}{1+\exp\left\{\beta_p \times \ln price_{i,t} + \alpha_r \times \left(\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}\right) + \alpha_p \hat{E}^d_{i,t}\left(\hat{\theta}_{old}\right) + \gamma_x' x_{i,t} + \gamma_{m,t}' z_{m,t}\right\}}.$$

5. We update the expected probability measurer by using the new estimated results and obtain $p^1_{i,t,d}\left(\hat{\theta}_{new}\right)$ for all $i$, $m$ and $t$.

6. We update the expected value of $\sum_{k\in\mathcal{I}_{m,t}} r_{i,k}d_{k,t}(\theta)$ given $d_{i,t}=1$ by $\hat{E}^d_{i,t}\left(\hat{\theta}_{new}\right) = \sum_{k\in\mathcal{I}_{m,t}} r_{i,k}p^1_{k,t,1}\left(\hat{\theta}_{new}\right)$.

7. We take the distance between old expectations and updated expectations as our criterion, and measure:

$$\mathcal{Q}\left(\hat{\theta}_{old},\hat{\theta}_{new}\right) = \sum_{t=1,\dots,T}\sum_{k\in\mathcal{I}_{m,t}}\sum_{m=1,\dots,M}\left\{\left(\hat{E}^d_{i,t}\left(\hat{\theta}_{old}\right)-\hat{E}^d_{i,t}\left(\hat{\theta}_{new}\right)\right)^2\right\}.$$

8. We repeat Step 2~7 until $\mathcal{Q}\left(\hat{\theta}_{old},\hat{\theta}_{new}\right)<\varepsilon$ for a $\varepsilon$ small enough that is selected by the econometrician.

Our method is similar to that of Aguirregabiria and Mira (2007), where the convergence requires that, for all the users in the platform, the referrals choice probability should be consistent with the expected value in the referrer's utility. Without taking the iterations, the method is also equivalent to the approaches used in Dubé et al. (2010) and Ryan and Tucker (2012), where they estimate the best response function directly from the data without requiring the consistency.[8]

## 5.2 Endogeneity Issue

As we mentioned earlier, the price of community renewals is often associated with some community quality components that the econometrics model fails to capture, leading to the endogeneity problem. To deal with the endogeneity issue, we adopt the control function approach in the estimation of structure model (Petrin and Train (2010); Wooldridge (2015)). We consider the following auxiliary regression model:

$$\ln price_{i,t} = \delta_I \times instrument_{i,t} + \delta_E \hat{E}^d_{i,t} + \delta_C\text{'}controls_{i,t} + \eta_{i,t}, \tag{1}$$

where $instrument_{i,t}$ is the instrumental variable that we have defined in Section 3.3 and $controls_{i,t}$ is a set of control variables in the structural model that include $\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}$, $x_{i,t}$ and $z_{m,t}$. The above equation allows us to obtain the residual terms $\hat{\eta}_{i,t}$. The idea is that $\hat{\eta}_{i,t}$ controls the potential unobserved factors that may affect the $\ln price_{i,t}$ and lead to the bias estimation.

Compared with the above algorithm, with the presence of endogeneity issues, the estimation procedure contains a few extra steps. Since $\eta_{i,t}$ is a function of $\hat{E}^d_{i,t}$ that also depends

[8]The condition they impose is stronger than in Bajari et al. (2007).

on the estimation of $\theta$, we initially set $\hat{E}^d_{i,t} = 0$ for all observations. We first estimate the auxiliary regression model to get $\hat{\eta}_{i,t}\left(\hat{E^d_{i,t}} = 0\right)$. We then add $\hat{\eta}_{i,t}$ into Step 4 and maximizes the following new objective function:

$$\sum_{t,i,m} \ln \frac{\exp\left\{\beta_p \times \ln price_{i,t} + \alpha_r \times \left(\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}\right) + \alpha_p \hat{E}^d_{i,t}\left(\hat{\theta}_{old}\right) + \gamma_x' x_{i,t} + \gamma_m' z_{m,t} + \alpha_\eta \hat{\eta}_{i,t}\right\}}{1 + \exp\left\{\beta_p \times \ln price_{i,t} + \alpha_r \times \left(\sum_{j\in\mathcal{I}_{m,t}} r_{j,i}d_{j,t}\right) + \alpha_p \hat{E}^d_{i,t}\left(\hat{\theta}_{old}\right) + \gamma_x' x_{i,t} + \gamma_m' z_{m,t} + \alpha_\eta \hat{\eta}_{i,t}\right\}}.$$

Once we obtain $\hat{\theta}_{new}$ and compute $\hat{E}^d_{i,t}$ in Step 6 by using $\hat{\theta}_{new}$, we plug the updated $\hat{E}^d_{i,t}\left(\hat{\theta}_{new}\right)$ into Equation 1 and re-estimate the auxiliary regression to update $\hat{\eta}_{i,t}\left(\hat{E}^d_{i,t}\left(\hat{\theta}_{new}\right)\right)$. In each iteration, we update $\hat{E}^d_{i,t}$ and re-estimate $\theta$ by using Step 2~7. Compared with the original method, our estimation algorithm has higher convergence requirement after adding the control function. The final estimator ensures that both the values of $\hat{E}^d_{i,t}$ and the fitted values of $\eta_{i,t}$ are consistent between two iterations while the likelihood function is also maximized.

## 5.3 Estimation Results

We report the estimation results in Table 5. Columns (1) and (2) show the estimated results of the structural model, and columns (3) and (4) show the estimation results of probit and logit models without adding $\hat{E}$ as control variable. In column (1), we do not iterate to get stable values of $\hat{E}$ so that the result can be approximately considered as the estimation of parameters under imperfect information. The results in column (2) are estimated by running iterations, and the parameters depict the model equilibrium with perfect information.

There is no significant difference between the estimation results with perfect and imperfect information. The estimation results show that a 1% decrease in price results in a 0.707% increase in willingness to renew, and the effects of other variables are similar to those obtained by the Table 3. In particular, we find that the variable $nR$ becomes insignificant after taking into account users' anticipation of their referral decisions, and the proportion of expected renewed referrals becomes an important explanatory variable. The difference between the first, second, and third (and fourth) columns is that the last two columns do not contain the prediction for the referrals' decisions. We test the model's goodness-of-fit through the likelihood ratio test. The test result shows that the model containing the prediction items (i.e., $E$). is statistically significantly better than the traditional logit model at the 1% level. In addition, the coefficient of $\hat{\eta}$ is statistically significant, which proves that our control function approach correctly solves the potential endogeneity issue.

| *Dependent variable:* | *Renewal Decision* ($d$) | | | |
|---|---|---|---|---|
| | Incomplete information | Complete information | Probit | Logit |
| | (1) | (2) | (3) | (4) |
| $\ln(Price)$ | $-0.707^{***}$ | $-0.707^{***}$ | $-0.445^{***}$ | $-1.766^{***}$ |
| | (0.015) | (0.014) | (0.008) | (0.014) |
| $E$ | $2.523^{***}$ | $2.621^{***}$ | | |
| | (0.048) | (0.051) | | |
| $RD$ | $0.425^{***}$ | $0.425^{***}$ | $0.294^{***}$ | $0.458^{***}$ |
| | (0.031) | (0.031) | (0.019) | (0.031) |
| $\ln(Price_0 + 1)$ | $0.323^{***}$ | $0.324^{***}$ | $0.198^{***}$ | $0.335^{***}$ |
| | (0.0048) | (0.005) | (0.003) | (0.005) |
| $Community\,Type$ | $-1.355^{***}$ | $-1.356^{***}$ | $-0.763^{***}$ | $-1.93^{***}$ |
| | (0.025) | (0.025) | (0.014) | (0.025) |
| $\ln(N\_Answers + 1)$ | $0.035^{***}$ | $-0.035^{***}$ | $-0.024^{***}$ | $-0.043^{***}$ |
| | (0.003) | (0.003) | (0.002) | (0.003) |
| $\ln(N\_Article + 1)$ | $0.240^{***}$ | $0.240^{***}$ | $0.148^{***}$ | $0.244^{***}$ |
| | (0.005) | (0.005) | (0.003) | (0.005) |
| $\hat{\eta}$ | $0.188^{***}$ | $0.188^{***}$ | $0.140^{***}$ | $0.249^{***}$ |
| | (0.016) | (0.016) | (0.010) | (0.016) |
| Network Statistics | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES |
| Observations | 302,056 | 302,056 | 302,056 | 302,056 |
| Log Likelihood | −189,370.700 | −189,385.300 | −191,255.200 | −191,000.900 |

Notes: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01. Standard errors are reported in brackets. We use $\log(x+1)$ for the logarithmic transformation of variable $x$ to avoid the situation where $x = 0$, and $\log(x)$ does not exist. $Price_0$ is the initial price that the user paid when join the community. If some users get a special offer they may get the price free for the first year, but there is no such offer at the time of renewal. Community Type equals 1 for the science-oriented communities and 0 for economics-oriented communities. We do not include users who join the community within the last year and are not able to renew for the first time. Network Statistics include whether the user is referred to join the community and the number of referrals. Year FE and Month FE are created based on when users first join the community.

Table 5: Structure estimation results

# 6 Counterfactual Analysis: Price Changes and Network Structure

The structural model estimation results obtained in the previous section indicate that the referral relationship has a direct effect on the renewal decisions of users. This also means that according to different referral relationships, the effect of price changes on the renewal rates varies across different network structures and are generated by referral relationships in each community. In particular, we ask whether it is better for the community to offer long-term discounts to users who are recommended to join? Moreover, what kind of referral network structure is best for the platform? In this section we illustrate these problems through three counterfactual analyses.

## 6.1 Effect of Price Changes

We start by analyzing the effect of price changes. We assume that the price of renewals has been reduced by $d \in [0, 1]$ for all the users across the platform and seek to evaluate the effect of such price changes on community income and renewals. The key issue is to develop an algorithm to compute the equilibrium change due to the price reduction since price changes affect both the upstream and downstream renewal probabilities, which then leads to a network propagation effect. Starting from the initial equilibrium that we observe in the market $m$: $\mathcal{E}_{m,t} = \left\{d_{1,t}, ..., d_{i,t}, ..., d_{n_{m,t}}\right\}$, we recompute the equilibrium of price changes by using the following steps:

1. Recompute the model outcomes by set $price^{CF}_{i,t} = price_{i,t} \times (1 - d)$, predict fitted renewal probabilities by the model.

2. Given a set of preference shocks, recalculate the market equilibrium $\mathcal{E}^{CF}_{m,t} = \left\{d_{1,t}, ..., d_{i,t}, ..., d_{n_{m,t}}\right\}$;

3. Update referrer's status based on $\mathcal{E}^{CF}_{m,t}$, update the value of $\hat{E}^{d}_{i,t}$s based on the fitted value of probabilities.

4. Recompute the equilibria and repeat Steps 2 and 3 until both $\mathcal{E}^{CF}_{m,t}$ and $\hat{E}^{d}_{i,t}$s converge to stable values.

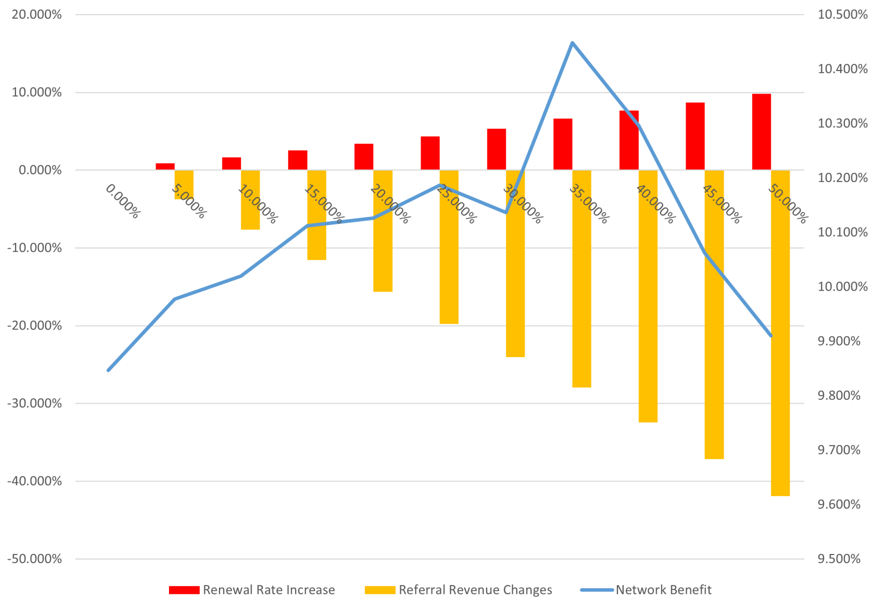
It is worth noting that our model assumes that users are myopic, so two consecutive renewals of a user will be treated as two renewals of different users. This is because user departures do not have a long-term effect on counterfactual analysis; for example, users in each period are treated as new users. Even if they leave in the previous year, they still choose to rejoin in the next year and recommend the community to others. Although repeated purchases are very rare in the data, our counterfactual analysis may underestimate the snowball effect of users' departures to a certain extent.

Figure 6 illustrates the case of price changes. The price drop will increase the user's willingness to renew the fee and affect the user's expectations. When the price is reduced by 50%, the overall renewal rate increases by 9.84% Compared to the renewal rate increase, the total revenue decreases by 40.48% when the discount rate is 50%. Such results provide a quantitative basis for pricing strategies, as community-based platforms often balance direct revenue from users with advertising revenue based on the number of users (e.g., Benzell and Collis (2020)). Our results suggest that a 1% increase in renewals may cost 4.26% of revenue.

In particular, we measure the renewal gain from network spillover effects. We compare the effect of network disconnection (i.e., the user ignores the referrals and the referrer's decisions when making a decision) on renewal rates. We find that the network spillover benefit has a concave shape. When prices drop by 35% to 40%, the network spillover benefit reaches the maximum by adding more than 10.45% of the renewal rate. The benefit gradually declines after the price decreases by more than 36%. We attribute this to a tug of war between price and network effects. When prices drop less, only a small number of users decide to renew, and their decisions to renew affect their referrer and referrals. Network effect begins to kick in and increases as prices fall and the number of users affected increases. When the price drops sharply, the price becomes the dominant lever, and the network effect becomes insignificant. Even though most users are willing to renew their subscription at this time and they still affect each other through the network, the main reason for renewing is the sharp drop in the price rather than the decision changes of others in the network.

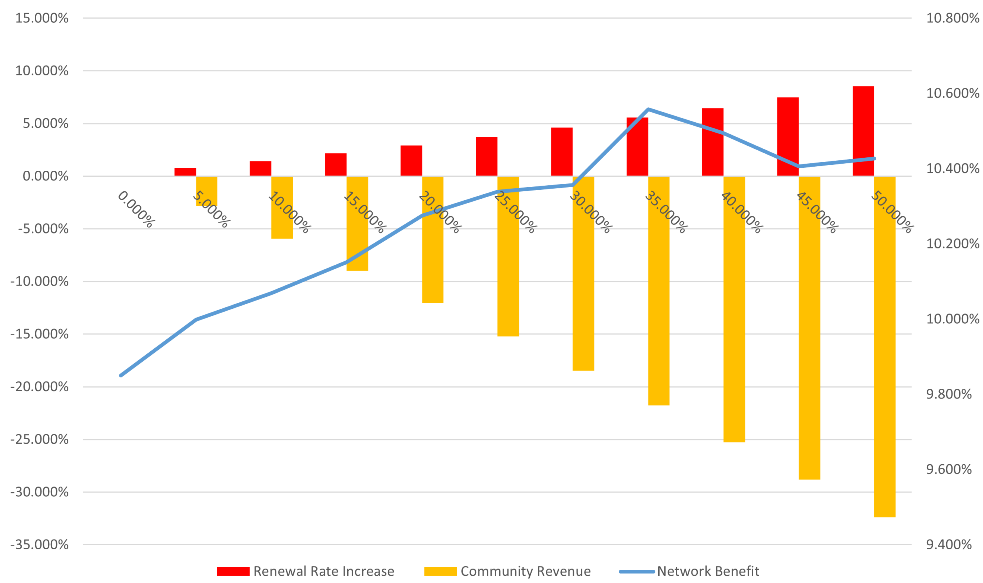
## 6.2 Discount Discrimination Policy

Although the uniform price discount policy can increase the renewal rate, it also negatively affects the revenue. Therefore, we seek a more effective pricing strategy in this section, hoping to reduce the revenue loss as much as possible while increasing the renewal rate. Referrals typically receive a discount on their first purchase in many business models, which is usually limited to the first purchase (for example, many referred bank credit card users are exempt from paying an annual fee for the first year). Inspired by Hinz et al. (2011), we examine the effect of price changes when the discount policy is available only to the referrals in an alternative counterfactual analysis. Figure 7 illustrates the effect of providing referral-targeted discount discrimination. Compared with Figure 6, a 50% price discount results in an increase of 8.55% in renewals under the new pricing policy. The rate increase under the new policy is only 1.29% lower than the 9.84% increase under the uniform discount policy. The referrals discount also makes the network effect more significant, which is maximized when the renewal discount reaches 35% to 40%. Surprisingly, we find that the loss of income under the new policy is much smaller than before. When the renewal discount reaches 50%, the uniform discount results in a referral revenue loss of 40.48%, while the referral-targeted discrimination discount results in only 32.40%. The discount discrimination policy reduces the loss of total revenue by 2.66% at the cost of a 1.29% reduction in the renewal rate when




Notes: Our counterfactual analysis is based on a group of users with network relationships (i.e., each user has at least one referrer or one referral), which accounts for about 10% of the total number of users. The horizontal axis represents the change of the overall price discount. The red bars represent the renewal rate increases as the price drops, and the yellow bars represent the changes in the overall revenue. The blue line measures the effect of the network by comparing the difference in renewal rates with and without the referral network.

Figure 6: Price changes, renewal rate, and network benefit




Notes: Our counterfactual analysis is based on a group of users with network relationships (i.e., each user has at least one referrer or one referral), which accounts for about 10% of the total number of users. The horizontal axis represents the change of the price discount for referrals. The red bars represent the renewal rate increases as the price drops, and the yellow bars represent the changes in the overall revenue. The blue line measures the effect of the network by comparing the difference in renewal rates with and without the referral network.

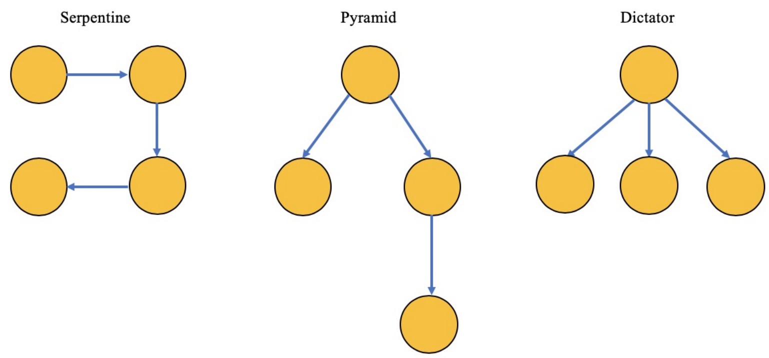
Figure 7: Referral-targeted discrimination, renewal rate, and network benefit

the discount rate is 50%.

## 6.3 Network Structure Changes

We now discuss the effect of network structure changes on renewal rates. Our structural model estimates show that a discount policy on a network with high beta index increases retention more significantly. Most of the literature on network effect studies is limited to peer or size effect based on the beta index measure (e.g., Gowrisankaran and Stavins (2004) ; Ryan and Tucker (2012)). However, the beta index is only one of the most basic statistics in the network structure. Especially when the decisions are endogenously made in the network, studies based on the beta index alone may not fully capture the overall network influence on the output variables. In Figure 8, we illustrate three different network shapes: Serpentine, Pyramid, and Dictator. These networks have the same beta index level, but the degree of closeness-based centralization increases from left to right.

We take the community in Figure 1 and reconstruct the network structure according to the order in which users joined the community. In a Serpentine network, we assume that

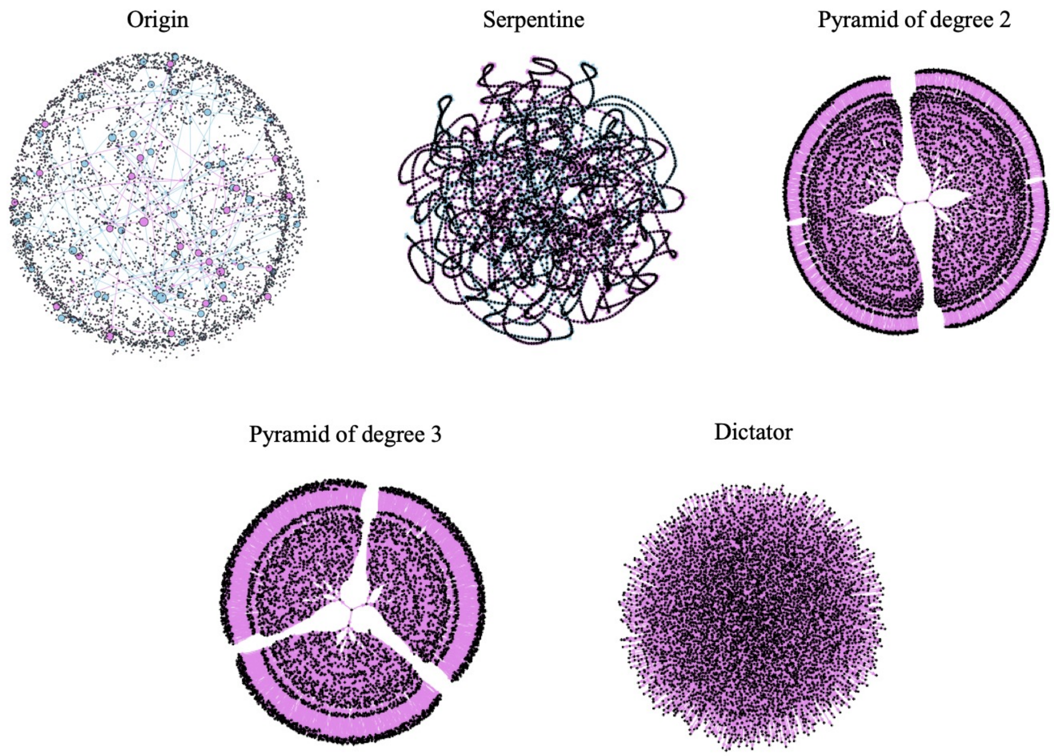



Notes: The three network structures in the above figure all have the same level of beta index and number of nodes, but the centralization increases once from left to right. We refer to their structures as Serpentine, Pyramid, and Dictator, respectively.

Figure 8: Illustration of different network structure with the same level of beta index

each new user will recommend the community to the following user. In a Pyramid network of degree 2, we assume that each new user will be the referrer of the next two users who join. In a Pyramid network of degree 3, we assume that each new user will be the referrer of the next three users who join. Finally, in a Dictator network, we assume that all users are recommended to join the platform by the first user. The remade networks under different structures are respectively represented in Figure 9. As we mentioned earlier, the remade networks all have the same level of beta index, with only the first original user under each network not being recommended and all the others joining through referral links. In an extremely centralized Dictator network, one can imagine the existence of a super influencer, such that all the other members join the community through her referral link.

Keeping other conditions unchanged, we re-estimate the equilibrium state under different network structures. Table 6 reports the results of the counterfactual analysis. These results suggest that the degree of network centralization is a key factor in determining the user renewal rate. Under the same level of beta index (Beta index), the renewal rate decreases with the increase in centralization. The results show that the Serpentine network structure is the best for maximizing network effects and achieving the highest retention. On the contrary, despite its high level of beta index, the Dictator network only achieves a 51.13% user retention rate, which is slightly higher than the rate of the original network. At the same time, we also find that the decline in the renewal rate decreases with the increasing degree of network centralization. Compared with the one-to-one referral network (Serpentine), the renewal rate under the one-to-two referral network (Pyramid of degree 2) decreases by 29.65% (96.67%-67.02%). However, compared with the one-to-two referral network, the overall rate




Notes: In the figure above, we show five different network structures based on the community in Figure 1: the original network structure on the upper left, the Serpentine network structure on the upper middle, the Pyramid network structure of degree 2 on the upper right, the Pyramid network structure of degree 3 on the bottom left, and the Dictator network structure on the bottom right.

Figure 9: Network structures with the same beta index and different centrality degrees

reduction under the one-to-three network (Pyramid of degree 3) and Dictator network is only 3.38% (67.02%-63.64%) and 15.89% (67.02%-51.13%), respectively.

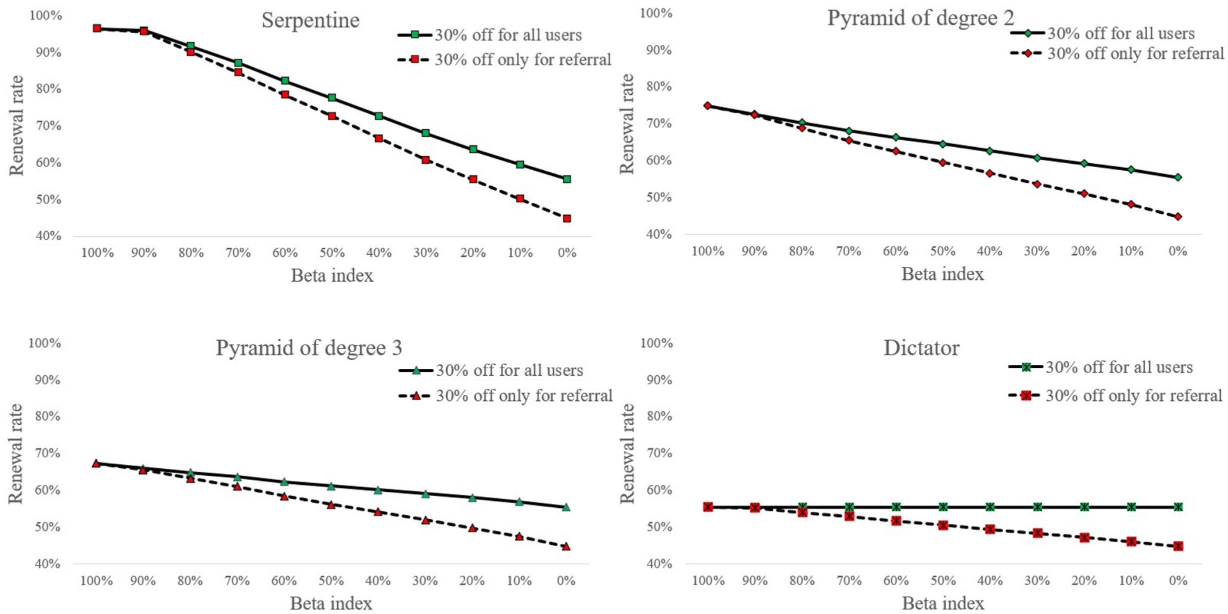
It is worth noting that although the last four network structures in Table 6 have very different effects on the renewal decision, they generate the same amount of revenue when users first join the community. For some time, business executives have focused on the effect of the referral program on user acquisition and too little on the program's long-term effect on user retention. Our results contradict the conventional business wisdom that platforms tend to attract super influencers. Rather than incentivizing super influencers, encouraging each user to exert their influence in a chain shape is more likely to maximize user retention. Although both the referrer's and the referrals' decisions mutually influence each other, centralization also leads to a reduction in the chain hierarchy, which reduces the number of times the positive effect travels through the network, thus further reducing the snowball effect.

| Network Structure | Original | Serpentine | Pyramid (degree 2) | Pyramid (degree 3) | Dictator |
|---|---|---|---|---|---|
| Renewal rate | 45.80% | 96.67% | 67.02% | 63.64% | 51.13% |
| Number of nodes | 6065 | 6065 | 6065 | 6065 | 6065 |
| Number of edges | 91 | 6064 | 6064 | 6064 | 6064 |
| Beta index | 0.015 | 0.999 | 0.999 | 0.999 | 0.999 |
| Closeness Centrality | $2.15 \times 10^{-7}$ | 0.00028 | 0.084 | 0.128 | 1 |

Notes: The Beta index is a classic measure of a network's level of beta index, it takes the ratio of the number of edges over the number of nodes. The Centrality degree ($C_C$) of the graph in terms of closeness is designed by Leavitt (1951): $C_C = \frac{\sum_{i=1}^{n}[C'_C(p^*) - C'_C(p_i)]}{(n^2-3n+2)/(2n-3)}$, where $n$ denotes number of points, $C'_C(p_i)$ denotes the point centrality of a point $p_i$ (Beauchamp (1965) suggested using $C'_C(p_i) = [\frac{\sum_{j=1}^{n} d(p_j,p_i)}{n-1}]^{-1}$, where $d(p_j, p_i)$ denotes the number of edges in the geodesic connecting $p_j$ and $p_i$), and $C'_C(p^*)$ denotes the largest value of $C'_C(p_i)$.

Table 6: Network structures and corresponding renewal rates

Figure 10 illustrates the effect of changes in the beta index on the renewal rates. A fully connected network has a beta index equaling 100%, the index decreases as the number of referrals decreases. In the figure, the corresponding renewal rates of Serpentine, Pyramid degree 2, Pyramid degree 3, Dictator under a 30% uniform-price discount policy are 96.43% and 74.85%, 67.48 and 55.42%, respectively. As the connectivity declines, the renewal rates for Serpentine, Pyramid degree 2 and Pyramid degree 3 also drop simultaneously. For every 10 % decrease in the beta index, the renewal rate for Serpentine decreased by 4.01 %, Pyramid degree 2 decreased by 1.95 %, and Pyramid degree 3 decreased by 1.19 %. In addition, the decreasing trend of the renewal rate for the three structures is almost linear, and we find that the slope of the decreasing renewal rate decreases as the closeness centrality decreases. When the beta index decreases to 0, the renewal rate converges to 55.42% for all four different network structures. When we implement the referral-targeted discount discrimination policy, the decreasing trends are similar to those under the uniform discount policy. However, the overall renewal rates are smaller since the number of users covered by the policy decreases. Such findings suggest that an efficient referral network that maximizes the effect of price discount policies on renewal rates should achieve a high degree of connectivity while keeping the closeness-based centrality as low as possible. In other words, the effect of price discount policies would be more efficient once the network becomes more hierarchical (i.e., high degree of connectivity and low degree of closeness-based centrality), so that the "snowball" effect is maximized.




Notes: Our counterfactual analysis is based on same group of users with different network structures (i.e.,the Serpentine network structure on the upper left, the Pyramid network structure of degree 2 on the upper right, the Pyramid network structure of degree 3 on the bottom left, and the Dictator network structure on the bottom right), which accounts for 5,332 users. The horizontal axis represents the change of beta index for the four network structures. The vertical axis represents the change of renewal rate for the four network structures. Beta-index varies in [0%, 10%, . . . , 100%]. For each beta index (e.g., when beta index = 0.9), the last 10% of the original network will be cut. We then recalculate the equilibrium incomplete market for networks in each level of beta index. When the beta index = 0, this is the case when there is no network effect. The solid line with green dots measures the effect of the renewal rate by comparing the difference in beta index when giving all users a 30% off discount on the renewal price. The dashed line with red dots measures the effect of the renewal rate by comparing the difference in beta index when only offering referrals a 30% off discount on the renewal price.

Figure 10: Beta index and renewal rate

# 7 Conclusion

Two decades ago, economists began to explore the influences of the network on economics and social welfare (Economides (1996)). Although the referral model is lucrative but costly in traditional business models, the emergence of digital platforms has expanded its profit and reduced its costs. The network effect generated by frequent referrals also drives the platform's revenue and enables us to express the relationship between users through the referral relationship network. We use more than 300,000 user data provided directly by a contentgeneration platform to describe user renewal decisions in a structural model. Observing complete and different referral networks under which users are making decisions helps us to identify the influence of network structure on users' strategical behaviors.

Both reduced-form evidence and structural estimation results demonstrate that referral relationships significantly affect the renewal rates of digital products bilaterally, since price changes would propagate through the network by influencing the user's expectations of the network peers' renewal decisions. Such findings suggest that companies/platforms should take into account the users' network effect when setting pricing or discount strategies, as the snowball effect will lead to significant differences across the network. We find that discount discrimination policies based on referral relationships is cost-efficient in customer retention, which is often ignored by managers in their real business models.

An important finding is that counterfactual analysis probes into the profound influence of network structure on the overall renewal rate. Counter-intuitively, we do not find that highly centralized networks improve retention: a less centralized but more hierarchical network is more likely to maximize externalities and thus improve retention. The results indicate that platforms should encourage more one-to-one referral relationships rather than relying on the contributions of a small number of super-influencers. Our findings can easily be implemented in digital marketing as "Refer A Friend" is becoming a popular digital marketing tool that online platforms have widely adopted. Unlike sending links via email or social media, digital platforms are increasingly focusing on motivating users to recommend products to their closest peers. For example, Lingoda is a large global online language learning platform. If a user refers a friend to Lingoda, the friend receives a discount of €50 off their first month with Lingoda, and the user receives five free group class credits on their active account. In the meantime, the platform also notices that it is against the terms to post the invitation link publicly on coupon or review websites. Referral networks based on one-to-one relationships are more likely to result from the recommendation of friends and family, as shown by Lingoda, which is more beneficial to the business development of those platforms that mainly rely on renewal revenue.[9]

# References

**Aguirregabiria, Victor and Pedro Mira**, "Sequential estimation of dynamic discrete games," *Econometrica*, 2007, *75* (1), 1–53.

[9]Source: https://lingoda-students.elevio.help/en/articles/250-can-i-recommend-lingoda-to-my-friends

**Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong**, "Peer effects in product adoption," Working Paper 25843, National Bureau of Economic Research 2019.

**Bajari, Patrick, C Lanier Benkard, and Jonathan Levin**, "Estimating dynamic models of imperfect competition," *Econometrica*, 2007, *75* (5), 1331–1370.

**Beauchamp, Murray A**, "An improved index of centrality," *Behavioral Science*, 1965, *10* (2), 161–163.

**Belo, Rodrigo and Ting Li**, "Referral programs for platform growth: Evidence from a randomized field experiment," *Available at SSRN 3224330*, 2018.

**Benzell, Seth and Avinash Collis**, "How to govern Facebook: A structural model for taxing and regulating big tech," *Available at SSRN 3619535*, 2020.

**Blanchard, Simon J, Mahima Hada, and Kurt A Carlson**, "Specialist competitor referrals: How salespeople can use competitor referrals for nonfocal products to increase focal product sales," *Journal of Marketing*, 2018, *82* (4), 127–145.

**Bolton, Ruth N, P KandBramlettMatthewD Kannan, and Matthew D Bramlett**, "Implications of loyalty program membership and service experiences for customer retention and value," *Journal of the Academy of Marketing Science*, 2000, *28* (1), 95–108.

**Boudreau, Kevin, Lars Bo Jeppesen, and Milan Miric**, "Competing on freemium: Digital competition with network effects," *Available at SSRN 2984546*, 2019.

**Bryan, Gharad, Dean Karlan, and Jonathan Zinman**, "Referrals: Peer screening and enforcement in a consumer credit field experiment," *American Economic Journal: Microeconomics*, 2015, *7* (3), 174–204.

**Burks, Stephen V, Bo Cowgill, Mitchell Hoffman, and Michael Housman**, "The value of hiring through employee referrals," *Quarterly Journal of Economics*, 2015, *130* (2), 805–839.

**Buttle, Francis A**, "Word of mouth: understanding and managing referral marketing," *Journal of Strategic Marketing*, 1998, *6* (3), 241–254.

**Chu, Junhong and Puneet Manchanda**, "Quantifying cross and direct network effects in online consumer-to-consumer platforms," *Marketing Science*, 2016, *35* (6), 870–893.

**Demir, Banu, Beata Javorcik, Tomasz K Michalski, Evren Ors et al.**, "Financial constraints and propagation of shocks in production networks," *Work. Pap., Univ. Oxford, UK*, 2018.

**den Bulte, Christophe Van, Emanuel Bayer, Bernd Skiera, and Philipp Schmitt**, "How customer referral programs turn social capital into economic capital," *Journal of Marketing Research*, 2018, *55* (1), 132–146.

**Dubé, Jean-Pierre H, Günter J Hitsch, and Pradeep K Chintagunta**, "Tipping and concentration in markets with indirect network effects," *Marketing Science*, 2010, *29* (2), 216–249.

**Economides, Nicholas**, "The economics of networks," *International Journal of Industrial Organization*, 1996, *14* (6), 673–699.

**Fernandez, Roberto M, Emilio J Castilla, and Paul Moore**, "Social capital at work: Networks and employment at a phone center," *American Journal of Sociology*, 2000, *105* (5), 1288–1356.

**Goldfarb, Avi and Catherine Tucker**, "Digital economics," *Journal of Economic Literature*, 2019, *57* (1), 3–43.

**Gowrisankaran, Gautam and Joanna Stavins**, "Network externalities and technology adoption: Lessons from electronic payments," *RAND Journal of Economics*, 2004, *35* (2), 260–276.

**Hada, Mahima, Rajdeep Grewal, and Gary L Lilien**, "Supplier-selected referrals," *Journal of Marketing*, 2014, *78* (2), 34–51.

**Han, Xintong and Lei Xu**, "Technology adoption in input-output networks," Working Paper, NET Institute 2018.

**_ , Pu Zhao, and Jun Hyun Joseph Ryoo**, "Pay for content or pay for referral? An empirical study on content pricing," Working Paper, NET Institute 2019.

**Heskett, James L, Thomas O Jones, Gary W Loveman, W Earl Sasser, Leonard A Schlesinger et al.**, "Putting the service-profit chain to work," *Harvard Business Review*, 1994, *72* (2), 164–174.

**Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan U Becker**, "Seeding strategies for viral marketing: An empirical comparison," *Journal of Marketing*, 2011, *75* (6), 55–71.

**Hristakeva, Sylvia and Julie Holland Mortimer**, "Impacts of legacy discounts in the market for national television advertising," 2019.

**Hu, Mandy Mantian, Sha Yang, and Daniel Yi Xu**, "Understanding the social learning effect in contagious switching behavior," *Management Science*, 2019, *65* (10), 4771–4794.

**Jung, Jaehwuen, Ravi Bapna, Joseph M Golden, and Tianshu Sun**, "Words matter! Toward a prosocial call-to-action for online referral: evidence from two field experiments," *Information Systems Research*, 2020, *31* (1), 16–36.

**Karlan, Dean S**, "Using experimental economics to measure social capital and predict financial decisions," *American Economic Review*, 2005, *95* (5), 1688–1699.

**Kornish, Laura J and Qiuping Li**, "Optimal referral bonuses with asymmetric information: Firm-offered and interpersonal incentives," *Marketing Science*, 2010, *29* (1), 108–121.

**Leavitt, Harold J**, "Some effects of certain communication patterns on group performance.," *Journal of Abnormal and Social Psychology*, 1951, *46* (1), 38.

**Leenheer, Jorna, Harald J Van Heerde, Tammo HA Bijmolt, and Ale Smidts**, "Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members," *International Journal of Research in Marketing*, 2007, *24* (1), 31–47.

**Montgomery, James D**, "Social networks and labor-market outcomes: Toward an economic analysis," *American Economic Review*, 1991, *81* (5), 1408–1418.

**Pallais, Amanda and Emily Glassberg Sands**, "Why the referential treatment? Evidence from field experiments on referrals," *Journal of Political Economy*, 2016, *124* (6), 1793–1828.

**Petrin, Amil and Kenneth Train**, "A control function approach to endogeneity in consumer choice models," *Journal of Marketing Research*, 2010, *47* (1), 3–13.

**Reichheld, Frederick F and Phil Schefter**, "E-loyalty: your secret weapon on the web," *Harvard Business Review*, 2000, *78* (4), 105–113.

**Rhouma, Tarek Ben and Georges Zaccour**, "Optimal marketing strategies for the acquisition and retention of service subscribers," *Management Science*, 2018, *64* (6), 2609–2627.

**Ryan, Stephen P and Catherine Tucker**, "Heterogeneity and the dynamics of technology adoption," *Quantitative Marketing and Economics*, 2012, *10* (1), 63–109.

**Schmitt, Philipp, Bernd Skiera, and Christophe Van den Bulte**, "Referral programs and customer value," *Journal of Marketing*, 2011, *75* (1), 46–59.

**_ , _ , and _** , "Why customer referrals can drive stunning profits," *Harvard Business Review*, 2011, *89* (6).

**Shi, Huanhuan, Rajdeep Grewal, and Hari Sridhar**, "Organizational herding in advertising spending disclosures: Evidence and mechanisms," *Journal of Marketing Research*, 2021, *58* (3), 515–538.

**Su, Che-Lin**, "Estimating discrete-choice games of incomplete information: Simple static examples," *Quantitative Marketing and Economics*, 2014, *12* (2), 167–207.

**Trusov, Michael, Randolph E Bucklin, and Koen Pauwels**, "Estimating the dynamic effects of online word-of-mouth on member growth of a social network site," *Journal of Marketing*, 2009, *73* (5), 90–102.

**Verboven, Frank**, "Quality-based price discrimination and tax incidence: evidence from gasoline and diesel cars," *RAND Journal of Economics*, 2002, pp. 275–297.

**Wei, Yanhao Max**, "The similarity network of motion pictures," *Management Science*, 2020, *66* (4), 1647–1671.

**Wooldridge, Jeffrey M**, "Control function methods in applied econometrics," *Journal of Human Resources*, 2015, *50* (2), 420–445.