"The Company We Keep:" Endogenous Network Formation and Peer Effects in Churn

Yulia Nevskaya
Washington University in St. Louis
Yijun (Sally) Chen
Washington University in St. Louis

Cite as:
Nevskaya Yulia, Chen Yijun (Sally) (2020), "The Company We Keep:" Endogenous Network Formation and Peer Effects in Churn. *Proceedings of the European Marketing Academy*, 49th, (59621)

Paper from the 49th Annual EMAC Conference, Budapest, May 26-29, 2020.

"The Company We Keep:"
Endogenous Network Formation and Peer Effects in Churn

Abstract:

The paper addresses the endogeneity in social ties in empirical measurement of causal peer effects in customer churn from observational data. Peer effects in churn are of paramount importance for products consumed in a socially connected manner. Equipped with the knowledge of causality, managers can craft effective retention campaigns. To tackle the endogeneity, we directly model network formation. After accounting for the choice of peers, we model the interdependence of agents' churn decisions and estimate the causal peer effect in churn. Modeling network formation first allows to recover the latent individual-specific parameters that might affect both the selection of peers and churn. These parameters correct for tie endogeneity in the peer effects model. We use data from a popular online game and find a strong causal peer effect in churn decisions. Our post-estimation analysis informs a firm's policy of inducing formation of peer groups less prone to peer effects in churn.

Keywords: customer churn, peer effects, endogeneity

Track: Methods, Modelling & Marketing Analytics

1. Introduction

The social connections that can be recovered from typical observational data on consumption by peers are fundamentally endogenous. In this case, the social interactions emerge or get measured while being intermediated by peers' interactions with products and, therefore, are not independent from the outcomes an analyst might be interested in, such as product adoption, consumption intensity, or customer churn.

This paper contributes by addressing the issue of social network endogeneity in identification of peer effects in customer churn. Equipped with in-depth understanding of peer effects in churn behavior, managers can craft effective retention campaigns (Ascarza, Ebbes, Netzer, and Danielson, 2017). To leverage the full potential of social connections in retention campaigns, marketing practitioners have to move beyond the predictive power of social connections and toward isolating causality in churn behavior, where more work is needed (Ascarza, Neslin, and Netzer, 2018). Marketing academic researchers have been focusing mostly on peer effects in product adoption rather than on peer effects in product use and churn (e.g., Risselada, Verhoef, and Bijmolt, 2014, Hartmann, 2010; Nair, Manchanda, and Bhatia, 2010; Yang, Narayan, and Assael, 2006; Van den Bulte and Stremersch 2004; and others), likely due to relative abundance of data on purchases compared to, until recently, scarce data on after-purchase behavior and social connections. The few notable exceptions are Nitzan and Libai (2010) and Haenlein (2013).

Following the literature on identification of peer effects (Manski, 1993; Moffitt, 2001; and others), we recognize the challenges in identification of peer effects. First, the network of peers and the strength of connections between the peers depend on those individuals' inherent characteristics and preferences. The inherent individual characteristics that affect tie formation are often unobserved by the analyst but known to peers. Those unobserved characteristics might factor not only into selection of peers but also into consumer's decision to churn. As a result, the estimates of the effect of peer churn on consumer's decision will be biased if her selection of peers is not accounted for. Similarly, peer's choice is not exogenous as consumers are simultaneously affecting each other, which gives rise to the "reflection problem," as originally termed in Manski (1993). Again, it leads to the biased estimate of the peer effect if that simultaneity is ignored. Finally, there is an issue of correlated unobservables, i.e. the peers can be affected by some common churn drivers that lead consumers to abandon the product at the same time. These unobserved (to the analyst) common drivers can manifest as peer effects if not controlled for.

The peer connection endogeneity and the reflection issue have not been addressed by previous empirical studies of social effects in consumer churn, possibly for reasons related to challenges that this domain presents. The distinguishing nature of consumer churn is that it is observed at most once for each consumer; therefore a researcher does not have the opportunity to exploit the repeated observations of churn for a given customer across time. The panel data would have allowed to pin down (correlated) consumers' preferences and estimate the unobserved consumer heterogeneity which gives rise to the endogeneity concerns as peer connections might be formed based on preferences for the product (Hartmann, 2010; Nair et al., 2010).

To tackle the peer network endogeneity, we model separately the process of social network formation and the choice to churn. Such an approach is in line with the developing stream of literature in economics that pairs empirical models of peer effects with modeling the formation of peer networks (Goldsmith-Pinkham and Imbens, 2013; Graham, 2015; Hsieh and Lee, 2016; Griffith, 2019; and others). A separate structural model for the selection of peers allows us to recover the unobserved consumer-specific parameters that drive the formation of peer relationships and possibly factor into a consumer's decision to churn. The latent parameters serve as a correction for peer endogeneity in the second-stage peer effects model of churn. We include those parameters into the consumer's utility function as otherwise omitted drivers of the churn decision. We specify the empirical churn model as a static simultaneous-move game of incomplete information, where consumer utility from continuing product use depends on the consumer's expectations about the continued use by her peers. As argued in Hartmann (2010), the interdependence of peer churn decisions via the structure of the game alleviates the reflection problem arising from the simultaneity in peer choices. Finally, we control for the correlated unobservables by a rich set of time and peer group fixed effects.

We apply the specified dual model of peer network formation and churn to data from the popular video game *World of Warcraft*, which has an important social component as its core value proposition (Rapp, 2018; Martončik & Lokša, 2016). Apart from video game publishing fast becoming one of the largest entertainment industries, many online games have defining characteristics of products typically featured in studies of customer churn: the product is consumed over a long period of time and consumers use subscriptions to be renewed periodically.

For the product we study, we find significant evidence of peer connections being explained by the unobserved heterogeneity, in addition to observables (in-game experience level, etc.). Estimation of the peer effects model demonstrates that the churn of others has a strong casual effect on gamer's product use. Based on the estimated utilities of network formation and product use, we run simulation studies to investigate how churn dynamic is affected by composition of a peer group (discussed in the full version of the paper).

2. Empirical Settings

2.1. The product

For this study we use individual-level data collected from a server of the *World of Warcraft* (*WoW*), a massively multiplayer online role-playing game (Lee , Chen, Cheng, and Lei, 2011). To play the game, a consumer creates an avatar to represent her in the virtual environment and to complete in-game tasks. Gamers progress through experience levels from 1 to 60 by playing the game. The fantasy world consists of gaming zones the avatars can travel to. In the gaming environment, the gamer can see, interact and cooperate with other avatars. Avatars can join guilds, which are self-organized communities of gamers who get together to complete in-game tasks and socialize within the game (Nardi, 2010).

Once a gamer logs onto a *WoW* server, we observe her experience level, the gaming zone she appears in, and the guild she belongs to with 4-5 minute intervals, until she logs off. There are 96 unique guilds and the gamers are observed in 87 gaming zones. In our analysis we focus on experienced gamers at level 30 or higher. We restrict our attention to fully functioning guilds that have at least 50 members. As such, we have 2,444 unique avatars and 23 unique guilds observed for 30 weeks which results in 307 guild-week observations. For each gamer-week as well as guild-week observation, we compute the average experience level, total time spent gaming, and other similar metrics.

2.2. Peer relationships

We define peers as gamers belonging to the same guild. A gamer can be a member of only one guild at a time. Since the game has many guilds, we observe multiple peer groups.

In our study, a gamer is socially connected to all other gamers in her guild, but with varying strength. We measure the strength of relationships using a *continuous* scale. A gamer reveals the strength of her relationship to different peers through her choice of gaming zones and gaming time relative to her peers. In other words, gamer *i* must have a strong relationship with gamer *j* if gamer *i* is observed spending a lot of her time with gamer *j*. At the same time, gamer *j* might not have a strong relationship with gamer *i* if the former is observed spending more of her gaming time with other gamers. As such, the social ties are *directed*.

Let g_{ij} denote the directed tie from gamer i to gamer j. Based on the above principles, we define g_{ij} as follows:

$$g_{ij} = \sum_{d \in D, z \in Z} \frac{N_{zd}^{ij}}{\sum\limits_{k \in K_{zt}:\, k \neq i} N_{zd}^{ik}}, \tag{1}$$

where D is the set of time periods (days) in a week; Z is the set of gaming zones; N_{zd}^{ij} is the duration of time when both gamer i and gamer j are observed together in zone z in period d, $d \in D$, and captures joint gaming; K_{zd} is the set of gamers observed in zone z in period d. By weighing the frequency of gaming N_{zd}^{ij} by interactions with other gamers at the same time and place, described by the denominator in Equation 1, we capture the intensity of interactions between gamers i and j. For each pair of gamers i and j, the intensities are aggregated over all zones and time periods to produce a measure of tie strength g_{ij}.

2.3. Churn behavior

More than 21% of gamers churn at some point during the 6-month observation period in our data. We look at churned gamers and separate their peers into two groups: the peers who also churned and the peers who did not. We find that churned gamers have stronger ties with their churned peers than with their non-churned ones. Specifically, the average tie strength with the churned peers is 0.58 versus 0.22 ($p < 0.001$) with the non-churning peers.This positive association between tie strength and joint churning could indicate the existence of peer effects, but can also be explained, fully or in part, by the endogenous peer connections and common shocks.

3. Model and Estimation

To measure the peer effects in churn and tackle the issues associated with endogeneity of peer effects, we proceed in two stages. In the first stage, we estimate an empirical game of peer ties formation. In this game, players jointly decide on the strength of their connections with peers in their guild. This stage helps us recover the unobserved gamer characteristics that affect both formation of the ties between peers and might affect their churn decisions. In the second stage, conditional on the observed peer connections and recovered unobservables, we estimate a peer effects model of churn.

3.1. Network formation model

The network formation model is based on developments in Griffith (2019). In our model, gamers simultaneously decide on the tie strength with each peer in their guild. Gamer receives utility from her relationships with peers. That utility is determined by gamer's own characteristics, the characteristics of her peers, and the overall profile of ties with peers. Gamers incur cost in each tie since establishing connections requires effort. Each gamer has a fixed budget constraint on the total cost of establishing the ties. Under that budget constraint, the gamer trades off between forming ties with different peers and maximizes her utility from her connections.

Let S be the number of observed guild-weeks. For each guild-week s ($s \in S$), matrix X_s contains gamers' characteristics observed by the researchers in that guild-week. The dimensions of matrix X_s are $I_s \times K$, where I_s is the number of gamers and K is the number of observed characteristics. We assume that ties and observed gamer characteristics are independent across observed guild-weeks. Let vector $a_s = \{a_{is}\}$ of length I_s defined in a compact space contain individual unobserved (latent) characteristics of all gamers in s. Note, that while matrices X_s are data, vectors a_s are to be estimated. Unobserved characteristics of gamer i are assumed to be independent of the observed characteristics of all gamers in s, $\mathbb{E}[a_{is}|\mathbf{X_{ks}}] = 0\,\forall\, k, s$, as well as uncorrelated with the unobserved characteristics of other gamers, $\mathbb{E}\left[a_{is}|a_{js}\right] = 0\,\forall\, j \neq i\,\forall\, s$. For identification purposes, their conditional variance is set to $\sigma_a^2 = 1$. Let matrices $G_s = \{g_{ijs}\}$ with dimensions $I_s \times I_s$ collect information on directed tie strength between each pair of gamers i and j in s, as defined in Section . Since the ties are directed, matrices G_s are not symmetric, i.e. $g_{ijs} \neq g_{jis}$.

For gamer i, the utility of establishing a tie with gamer j is defined as:

$$u_{ijs} = g_{ijs}^{\alpha} g_{jis}^{\beta} e^{f(X_{is}, X_{js}, a_{is}, a_{js})}, \tag{2}$$

where $f(X_{is}, X_{js}, a_{is}, a_{js})$ is some function of observed and unobserved characteristics of gamers i and j to be defined later; α and β are parameters to be estimated, with the specification $0 < \beta < (1 - \alpha) < 1$[1]. This Cobb-Douglas utility function is capable of capturing the complementarity of ties: the utility gamer i receives is determined in part by the linking decision of gamer j.

The total utility for gamer i is the sum of the utilities she derives from all her ties with peers: $U_{is}(G_s, X_s) = \sum_{j \in s, j \neq i} u_{ijs}$. Simultaneously with other gamers, gamer i maximizes her total utility by deciding on her peer connections profile, $g_{is} = \{g_{ijs}\}$, under the budget constraint on her total effort in establishing ties with peers:

$$\begin{gathered} \max U_{is}(G_s, X_s) \quad s.t. \\ \sum_{j=1:I_s:\, j \neq i} c_{ijs} g_{ijs} \leq M_{is}, \end{gathered} \tag{3}$$

where c_{ijs} denotes the costly effort for gamer i of forming a tie with gamer j. The cost c_{ijs} is assumed to vary between connections and be mean-independent of individual observed and unobserved characteristics, $\mathbb{E}\left[\log c_{ijs} | \mathbf{X_{ks}}, a_{ks}\right] = 0 \forall k, s$. The term M_{is} denotes the budget of effort for gamer i.

The budget constraint, which is binding at the optimum, ensures that agents in the model make trade-offs when deciding on their connection profiles. In this simultaneous-move setting, the connection strategy of gamer i depends on the connection strategy of gamer j. The connection strategy of gamer i is affected by connection strategies of *all* other gamers in the guild. The game has a unique strictly positive Nash equilibrium (see Griffith, 2019), that is $g_{ijs} > 0$ for all i and j.

Next, we define function $f(X_{is}, X_{js}, a_{is}, a_{js})$ that enters the utility specification in Equation 2:

$$f(\cdot) = \gamma_1 X_{is} + \gamma_2 a_{is} + \delta_1 X_{is} X_{js} + \delta_2 X_{is} a_{js} + \delta_3 X_{js} a_{is} + \delta_4 a_{is} a_{js} + \gamma_3 X_{js} + \gamma_4 a_{js}. \tag{4}$$

The specification allows for the individual observed gamers' characteristics $X_{is} = \{X_{is}^k\}$ and $X_{js} = \{X_{js}^k\}$ $(k = 1, \ldots, K)$, unobserved latent gamers' characteristics a_{is} and a_{js}, as well as for a number of interactions between them to affect the utility of a tie between gamers i and j. Vector $\gamma = \{\gamma_n; n = 1, \ldots, 4\}$ contains parameters that describe the main effects of the two gamers' characteristics on the utility of the tie. Parameters $\delta = \{\delta_n; n = 1, \ldots, 4\}$ measure the interaction effects. In our empirical setting, gamers' observed characteristics include level,

[1] The restrictions ensure the concavity of the utility function to establish the equilibrium existence and uniqueness. See Griffith (2019) for a detailed discussion.

total gaming time, and the number of days spent at the highest available level of 60.

The model is estimated using generalized method of moments (see Griffith (2019) for the discussion of the instrumentation strategy and the analytical proof of identification).

3.2. Peer effects model of churn

We specify a peer effects model in which a gamer's decision to churn is a function of her own characteristics, other gamers' characteristics, her connections to peers, and her expectation over peers' probability to churn. These recovered unobservables will be used as explanatory variables to control for endogeneity in peer connections.

We define the utility of continuing the use of the game for gamer i in guild-week $s = (t, m)$, where $t \in T$ denotes a week and $m \in M$ denotes a guild, as:

$$U_{is}(\alpha, \Theta, \kappa | X_s, a_s G_s) = \boldsymbol{\theta}_1[X_{is}, a_{is}] + \boldsymbol{\theta}_2[\bar{X}_{is}, \bar{a}_{is}] + \kappa_T + \kappa_M + \alpha \bar{P}_{j \neq i, s} + \epsilon_{is}, \tag{5}$$

where the recovered unobservable characteristics a control for endogeneity in peer connections; vector of coefficients $\boldsymbol{\theta}_1$ captures the effect of gamer i's own observed and unobserved characteristics on her gaming utility; vector $\boldsymbol{\theta}_2$ measures the effect of peers' characteristics; κ_t and κ_m are vectors of week and guild fixed effects, respectively; coefficient α captures the effect of peers' churn on gamer i's utility of product use; variable $\bar{P}_{j \neq i, s}$ is the arithmetic mean over the probability of other gamers to churn in guild-week s:

$$\bar{P}_{j \neq i, s} = \sum_{j, j \neq i} P_{js} g_{ijs} \tag{6}$$

In Equation 6weight g_{ijs} is the tie strength between gamers i and j, which is the outcome of the network formation game. The stronger the tie between gamers i and j, the stronger is the influence j will have on i. We define the weighted variables $\bar{X}_{is}$ and $\bar{a}_{is}$ representing peer characteristics in Equation 5 similarly. Finally, ϵ_{is} is the idiosyncratic shock to the utility that follows the Gumbel distribution. We denote the non-random part of the utility as V_{is} and normalize the utility of churning to 0. Then the probability of churning for gamer i in guild-week s is:

$$P_{is} = \frac{1}{1 + e^{V_{is}}} \tag{7}$$

We recognize that a peer's choice is not exogenous as the peer herself is being affected by the choices of others ("reflection problem"). To address this endogeneity we follow Hartmann (2010) and Brock and Durlauf (2006, 2001) in setting up an equilibrium model of peer choices. As such, we formulate a simultaneous static discrete game of incomplete information. The incomplete information implies that gamers react to their rational beliefs about their peers' churn. As such, we set a system of simultaneous equations describing the peers' churn choices:

Table 1: Estimated parameters of the network formation model

Variable	Parameter	Estimate
Connection offered by gamer j, $\dot{g}^i_{jis}$	$\widetilde{\beta}$	0.970 (0.015)
Observable gamer j's characteristics ($\dot{X}^i_{js}$):		
experience level	$\widetilde{\gamma}^1_3$	–0.026 (0.023)
gaming time	$\widetilde{\gamma}^2_3$	0.140 (0.014)
time spent at level 60	$\widetilde{\gamma}^3_3$	0.020 (0.016)
Unobservable characteristics:		
gamer j's unobservable, $\dot{a}^i_{js}$	$\widetilde{\gamma}_4$	0.890 (0.021)
interaction $a_{is}\dot{a}^i_{js}$	$\tilde{\delta}_4$	0.006 (0.002)
Interactions of observable characteristics of gamers i and j ($X_{is}\dot{X}^i_{js}$):		
experience levels	$\widetilde{\delta}^1_1$	0.080 (0.005)
gaming times	$\widetilde{\delta}^2_1$	0.012 (0.008)
times spent at level 60	$\tilde{\delta}^3_1$	0.030 (0.006)
Interaction of observables of gamer i with unobservables of gamer j ($X_{is}\dot{a}^i_{js}$):		
experience level of $i \times a_j$	$\tilde{\delta}^1_2$	–0.790 (0.050)
gaming time of $i \times a_j$	$\tilde{\delta}^2_2$	–0.830 (0.014)
time spent at level 60 of $i \times a_j$	$\tilde{\delta}^3_2$	–1.010 (0.016)
Interaction of unobservables of gamer i with observables of gamer j ($a_{is}\dot{X}^i_{js}$):		
a_i × experience level of j	$\tilde{\delta}^1_3$	–0.014 (0.021)
a_i × gaming time of j	$\tilde{\delta}^2_3$	–0.011 (0.010)
a_i × time spent at level 60 of j	$\tilde{\delta}^3_3$	0.003 (0.012)

$$C_s = \mathcal{F}(C_s|X_s, a_s, G_s, \Theta) \tag{8}$$

where C_s denotes the churn profile in guild-week s, $\{X_s, a_s\}$ are matrices of observed and unobserved characteristics, and $\Theta = \{\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \kappa_T, \kappa_M\}$ are parameters to be estimated. We use the mathematical programming with equilibrium constraints approach (Su and Judd, 2012) to estimate parameters Θ by maximizing the log-likelihood of observed choices subject to the constraint described by Equation 8. This approach assumes that the equilibrium played in the data is the one that maximizes the likelihood of observed choices (Ellickson & Misra, 2011).

4. Results and Analysis

4.1. Estimation results for the network formation model

The ultimate goal for the estimation of the network formation model is to recover unobserved by the analyst individual-level parameters $A = \{a_s : s = 1, ..., S\}$ which might define peer ties and possibly affect churn decisions. These parameters will control for the endogeneity in peer ties in the peer effects model of churn. The estimates for the network formation model are in Table 1.

We see that unobserved characteristics $A = \{a_s : s = 1, ..., S\}$ are strong drivers of the peer connections, as the coefficients associated with them are large in magnitude and are statistically significant. As gamers tend to link stronger to individuals with higher a, those individuals display stronger relationship in the peer effects model (Equations 5 and 6). If those a play a role in churn behavior, the estimate of the peer effect in churn would be biased if we were to omit a from the peer effects model. The rest of the estimated coefficients of the utility for tie formation are logically consistent. A complete discussion of the estimation results for the network formation model is available in the full version of the paper.

4.2. Estimation results for the model of peer effects in churn

The estimation results for the model of peer effects in churn are in Table 2. The estimated parameters enter the utility of gaming and therefore a negative coefficient indicates a positive effect on probability of churning.

Table 2: Estimation results: Parameters of gaming utility

Variable	Estimate (std. error)
Peers' churn	-11.00 (0.59)
Observed characteristics X:	
experience level:	
gamer's own	-1.25 (0.04)
peers', weighted	-0.09 (0.10)
gaming time:	
gamer's own	1.18 (0.03)
peers', weighted	-0.01 (0.07)
time spent at level 60:	
gamer's own	0.02 (0.05)
peers', weighted	-0.02 (0.06)
Unobserved characteristics a:	
gamer's own	0.19 (0.09)
peers', weighted	-0.32 (0.26)
Guild and time fixed effects are included.	

The statistically significant estimate $\alpha = -11.00$ confirms that a higher expectation for peers' churn, denoted by $\bar{P}_{j\neq i,s}$ in Equation 5, has a negative impact on gamer's probability of continued gaming. Therefore, peer effects in churn are strong. Unobserved gamer's own characteristics a have a statistically significant negative effect on churn probability. Given that observables X (after the transformation performed for estimation) and unobservables a are on the same scale, the larger magnitude of parameters for observed characteristics indicates that they are of higher importance in explaining utility from gaming. The unobserved heterogeneity measured by a, which is found to strongly drive network formation, is less important in explaining churn, but it does have a statistically significant impact. We note that this finding is specific to the empirical context we study and the unobserved heterogeneity might be less or more important in explaining churn in other settings. Estimated coefficients for gamer's own observed characteristics have expected signs. Peers' characteristics are not important in explaining churn decisions.

6. Conclusion

We investigated causal peer effects in customer churn decisions in the empirical setting of an online video game. This product illustrates very well an increasingly common situation where consumers are exposed to consumption decisions of others. Knowledge and ability to measure causal peer effects becomes important for managers when they plan retention campaigns relying on peer influence.

Our focus was to measure causal peer effects in churn using observational data while controlling for endogeneity in peer ties, as well as addressing the "reflection problem" and accounting for common unobservables affecting all peers. We find that some characteristics of peers, which an analyst (the firm) typically can't observe, guide network formation. Their importance in our empirical application is significant. That unobserved heterogeneity also partly explains consumers' churn decisions. As such, we find evidence of limited endogeneity in peer connections in our empirical settings. To obtain the results, we separately modeled the network formation process before estimating the peer effects in churn. The model of peer network formation in the first stage of our analysis allowed us to recover the unobserved gamer characteristics. These unobservables affect the strength of ties between peers and and might drive churn decisions as well, thus giving rise to the issue of endogeneity in peer connections. This research also sheds light on the optimal network composition, based on our model estimation results.

References

1. Haenlein, M. (2013). Social interactions in customer churn decisions: The impact of relationship directionality. International Journal of Research in Marketing, 30, 3, 236-248.

2. Hartmann, W.R. (2010). Demand estimation with social interactions and the implications for targeted marketing. Marketing Science, 29, 4, 585-782.

3. Griffith, A. (2019). Random assignment with non-random peers: A structural approach to counterfactual treatment assessment. Working paper.

4. Nair, H.S., Manchanda, P. & Bhatia, T. (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. Journal of Marketing Research, 47, 5, 883-895.

5. Nitzan, I. & Libai, B. (2011). Social effects on customer retention. Journal of Marketing, 75, 6, 24-38.