2016 17th IEEE International Conference on Mobile Data Management

# Social Influence Analysis using Mobile Phone Dataset

Qiwei Han
*Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh PA, United States*
*Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal*
*Email: qiweih@cmu.edu*

***Abstract*—In this paper, we perform social influence analysis using a massive mobile phone dataset with two applications that are salient to the success of Mobile Network Operators (MNO). More specifically, we identify the role of social influence in subscriber churn and smartphone adoption, by applying robust identification strategies to separate social influence from confounding factors such as homophily. We also leverage network structures and position of each individual to account for heterogeneous social influence.**

***Keywords*-Social influence; Mobile data analytics; Mobile social networks; Churn; Smartphone adoption**

## I. Introduction

Mobile phones nowadays are ubiquitous and widely diffused. The rapid growth of mobile industry produces enormous amount of digital traces that record the human communications at an unprecedented level and resolution. In essence, these data reveal relational dynamics between millions of individuals with fine granularity [1]. Therefore, using mobile phone datasets to analyze the structure of mobile social networks (MSNs) has drawn great interest from diverse research areas ranging from statistical physics, computer science to public health and urban planning [2].

However, mobile phone datasets, as observational data by its nature, still have limitations when one needs to empirically estimate the degree to which social influence affects human behavior. In particular, the correlated behavior between connected individuals can be explained by both social influence and their inherent similarities - homophily, due to the endogenous formation of social ties [3]. Misattribution of homophily to social influence may lead to significant overestimation of the latter [4].

Moreover, as some individuals may exert disproportional influence to others and vice versa, individual heterogeneity should also be taken into account [5]. This requires detailed information about network structure as well as the deep understandings on how individuals' positions embedded within the network structure moderate the contagion process. Thus design and implementation of intervention strategies to either magnify or inhibit a social epidemic (*e.g.* viral marketing) still remain challenging.

MNOs typically archive transactional records about each subscriber's service usage for operational and billing purposes. These records include several categories of rich data that are useful for social influence analysis: i) tariff plans, the types of subscribed services, such as prepaid or postpaid plans, value-added services (*e.g.* mobile internet); ii) Call Detail Records (CDR), the call initiator and recipient, the timestamp and duration of the call, the device identifier and identifier for cell tower through which the call is connected (geolocation information); iii) demographic information: gender, age, etc. but it is mostly missing for prepaid subscribers. By aggregating individual's geolocation information over time, we can infer the subscribers' work and home locations as where they spend the most of days and nights, respectively. Further, we can infer socio-economic information (*e.g.* wage) by cross-referencing the census data.

## II. Construction of Mobile Social Network

We obtain a large-scale mobile phone dataset from a leading European MNO, which includes anonymized CDR for over 5 million subscribers between August 2008 and June 2009 in one European country (hereafter referred as EURMO for the sake of brevity). As previous studies show that phone contact represent social relationship (*e.g.* [6], the MSN is constructed as an undirected call graph, where nodes are subscribers and links are between two nodes who exchange calls. Specifically, two subscribers are denoted socially connected if they exchange at least one call in the same calendar month. The resulting network consists of 5.5 million nodes and 66 million edges. Figure 1 shows that empirical degree distribution of our MSN is highly skewed and heavy tailed. This implies that individuals have heterogenous degree of social connections and also reflects their dynamic calling behavior [7].

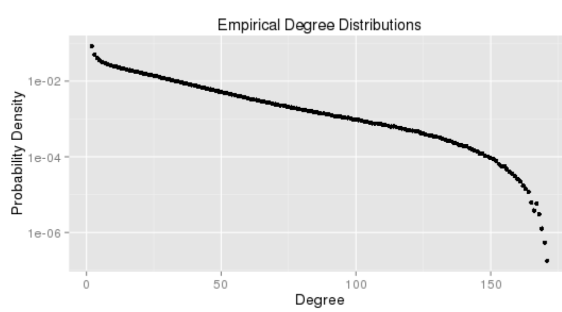



Figure 1. Empirical degree distribution for EURMO network



2375-0324/16 $31.00 © 2016 IEEE
DOI 10.1109/MDM.2016.82

## III. Social Influence in Subscriber Churn

### A. Background and Data Description

Churn measures the subscriber loss and is considered as the topmost challenge for MNOs [8]. In today's competitive wireless industry, subscribers can choose from many providers and easily transfer from one provider to another. Thus MNOs need to invest heavily in acquiring spectrums and upgrading their networks to provide quality communications and novel services to win customers from competitors. As a result, churn may cost the wireless industry billions of dollars every year and effective subscriber churn management becomes a priority for telecom managers as to ensure the sustainable growth of their companies.

As MNOs aim at controlling churn through proactive retention campaigns. they identify subscribers with high propensity to churn, evaluate the underlying reasons for churn and devise strategies to prevent it. The perplexing nature of churn, however, makes it very difficult to explain and address churn in an efficient and comprehensive manner. Subscribers may churn for many different reasons. [9] generally categorized these reasons into three streams. First, they may opt out due to the unsatisfication with the service quality. Second, they may get induced by competing carriers that provide more attractive service offerings or decide to acquire a new handset or service that is either not compatible with or not provided by their carrier. Third, changes in subscribers' personal communication needs may lead their valuation of existing service to become not attractive anymore. For example, they may be persuaded by close friends to switch to another carrier, simply because they need to maintain the communications with the friends, while also ensure their current arrangement meet their needs. Thus, wireless carriers can hardly provide one single solution to prevent all potential churners from leaving.

Churn management using mobile data analytics has long been studied (*e.g.* through data mining techniques to predict potential churners [10]). In particular, advances in studying the effect of social influence on subscriber churn have received considerable attention recently. [11] showed that propensity of a subscriber to churn depends on the number of friends that have already churned. [12] demonstrated that by integrating social factors such as influence from churners into machine learning models can greatly enhance the prediction performance. However, disentangling the role of social influence in subscriber churn from other reasons is still under-explored. As such, without fully understanding the true determinants of subscriber churn would preclude MNOs to design sound retention strategies.

We randomly sampled 10 thousand subscribers with prepaid plans for our analysis, because prepaid users have less contractual obligations and thus are more likely to churn without notice. Following industry standards, the subscriber is assumed to churn if she places no calls for 3 consecutive months. For each subscriber and in each calendar month, we extract usage patterns such as number of calls, number of airtime, expense and structural properties such as number of friends. More importantly, we account for social influence as number of churned friends. Over the period of analysis, the subscribers in our sample placed 3.75 million calls and 1,191 of them churned, which amounts to an average monthly churn rate of 2.04%.

### B. Matching Model

Propensity score matching (PSM) is widely used to control for selection bias when estimating treatment effects in observational data [13]. In our setting, social influence is associated with the presence of churned friends in subscriber's local social network (the treatment). With matching models, we can compare subscribers that are similar in relevant characteristics and differ only in the treatment levels. For example, [4] use dynamic matching to rule out homophily from social influence in an online social network setting. They dichotomize the treatment level (number of adopter friends) to explore the heterogeneity in treatment effects, *e.g.* they compared the users having more than three adopter friends with those otherwise. However, the results obtained at one treatment level may "absorb" the effect at the next treatment level, and thus complicates the evaluation of cumulative and marginal peer influence.

As in our case that friends' churn is not binary treatment but rather an integer, egos can be subject to cumulative amounts of treatment as they see more friends churn. Different treatment intensities can have different effects on the ego's churn (the outcome). Thus we apply the Generalized propensity score (GPS) to allow for matching on continuous levels of treatment [14]. Essentially, GPS provides a dose-response function (DRF) that measures the relationship between the outcome of interest and the intensity of treatment.

*1) Description of Matching Panels:* The panel structure of our data poses several extra challenges when applying matching techniques that have been developed for cross-sectional data. First, although our data is a random sample, standard matching routines on panel data typically ignore the time dimension in the panel and pair observations of the same unit in different time-periods (i.e. subscriber-month). Thus the systematic within-panel dependence may violate the independence assumption between matched observations (a.k.a stable unit treatment value assumption (SUTVA) in PSM setting). Second, standard matching routines would discard unmatched observations from the middle of some panels that may cause missing data problems. Third, compressing the data to one observation per panel (*e.g.* averaging covariates across time) may help alleviate the concerns of matching the same unit at different periods, however, important information regarding the dynamic subscriber behavior is inevitably eliminated. In our case, a downward trend in usage may be a signal of eventual churn. Mismatch of subscribers with similar average usage but divergent trend may lead to the biased conclusion. Therefore, analogous to [15], we choose to estimate the panel-level GPS as the unit of analysis that accounts for both the systematic dependence between observations of single subscriber and dynamic behavioral trend across time periods. Specifically, we include the lagged values of covariates prior to treatment in the GPS model, such that treatment assignment is applied to the panel that include all observations having the same subscriber identifier, rather than the individual observation.

We split the period of analysis into three intervals: i) the Pre-Treatment Period (PTP), during which we observe the important subscriber characteristics; ii) the Treatment Exposure Period (TEP), during which egos observe, and count, their friends churning, represented by $frd\_churn$; iii) the Churn Observation Period (COP), during which we observe whether the ego churns. This definition of intervals can ensure the best possible match between treated and control panels because the matched pairs correspond to the same duration of intervals rather than the same calendar month. For example, a treated panel spanning the first three months may have a control panel spanning the last three months as its best match, such that the selection bias on observables is reduced to the most extent.

*2) Description of GPS Analysis:* Because we have 10 months of data and we need to observe subscribers for 3 months to determine whether they churn, we are limited to 7 months of data only. Thus we include all available yet well-balanced time periods as: $\Big\{\text{PTP:}\{1,2,3\}, \text{TEP:}\{4,5\}, \text{COP:}\{6,7\}\Big\}$. Then we compute GPS by controlling the same characteristics specified above for each subscriber and for each of the 3 levels of $n{-}call frd\_churn$, i.e. those who exchange at least 1 call, 3 calls, and 5 calls in the same calendar months. By testing whether the conditional means of subscriber's characteristics given propensity score are different across treatment intensities, we find that most characteristics become statistically similar after adjustment on propensity score, indicating that GPS significantly reduce the bias.

*3) Effects of Friends Churn:* Figure 2 shows the results obtained for $n = 1, 3, 5$. We observe that having more friends churn increases the likelihood of churn for any $n$ considered in our analysis. Also, we see that when considering the churn from the marginal effect of treatment, that is the effect of having one more of these friends churn, remains nearly positive with the number of friends that churn. This provides evidence of social influence in churn in wireless networks. Furthermore, the churn likelihood for 5-call increases well beyond the 1-call and 3-call, which provides some evidence that churn from stronger friends might be more important. This is a sensible result showing that enough strong friends churning makes a significant difference on the ego's probability to churn when enough friends churn.

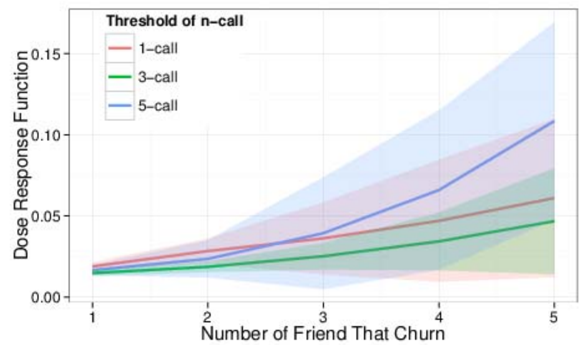



Figure 2. Estimated dose response function with PTP:{1,2, 3}, TEP:{4,5}, COP:{6,7} relative to having no friends churning. Ribbons represent the 95% confidence intervals. Standard errors are obtained via bootstrapping (100 repetitions)

## IV. Social Influence in Smartphone Adoption

### A. Background

MNOs nowadays are challenged by the continued decline in voice and SMS usage and heavy investment in network resources to handle capacity issues, due to the explosion of mobile data traffic. To ensure the profitability and sustainable growth, MNOs have to devise effective strategies to keep inducing subscribers to upgrade to newer generation of smartphones, in that smartphone users have higher willingness to choose more expensive tariff plans with mobile data service.

Although traditional technology diffusion and acceptance studies acknowledge that social factors are strongly related to consumer's decision to adopt high-technology products, only until recently [16] and [17] evaluate the role of social influence in smartphone adoption using mobile phone datasets. However, both of the studies fail to account for individual heterogeneity in the tendency to influence (or be influenced by) peers, as this requires extra knowledge of network structure. [18] broadly described a taxonomy of ”macro”-level and ”micro”-level characteristics emerged from social network that can affect product diffusion. On one hand, the macro-level patterns relate to the global properties of the network, such as network densities, degree distributions, path lengths, and so on. These summary statistics capture the essential graph topologies of the network. However, such overall characterizations of network overlook the richness of dyadic social interaction information between pairs of individuals. On the other hand, the micro-level patterns refer to the local network structure among individual's connections. Moreover, when it comes to learning about how individual's position matters, we may need information that extends beyond individual's own local network structure. Therefore, [18] pointed out the “blur” between macro and micro measures of network structure and individual position that needs further exploration.

We choose Apple's iPhone 3G as the exemplary model in our analysis because its release time coincides with the timespan of CDR data and EURMO is the sole partner with Apple with exclusive arrangement, such that we are able to capture the full cycle of the adoption. In our period of analysis, there are 20,570 iPhone adopters with complete profiles.

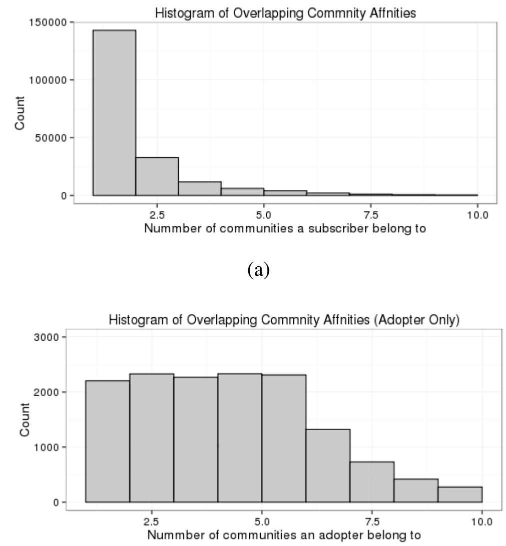
### *B. Discovering Social Circles in Adopter Network*

Recent empirical analyses on statistical properties of real world MSN provide evidence of community structure embedded within the network, as individuals tend to form closely connected social circles because of their homophlious characteristics [19]. Moreover, individuals may belong to multiple overlapping social circles [20]. As [3] suggest that uncovering social circles may help control for group-level unobserved homophily, we resort to state-of-the-art method CESNA to discover overlapping community from the MSN [21] as follows: i) for each iPhone adopter, we construct the ego-network that contains adopter and their direct neighbors; ii) for each subscriber in the ego-network, we extract a list of binary valued covariates to represent pluralistic homophily, including gender and wage (socio-demographic homophily); tariff plan, phone technology and mobile broadband (contextual homophily); and home location (spatial homophily); iii) we then apply CESNA on each ego-network using both node and edge information with the optimal number of communities identified through cross-validation; iv) we remove duplicated and nested communities and only retain those having iPhone adopters. As a result, we obtain 11,454 communities with 202,743 subscribers, 14,685 (71%) of which are iPhone adopters.

### *C. Core-Periphery Structure*

As Figure 3a shows, over 70% of subscribers belong to only one community and nearly 90% of those belong to two, while only about 5% of subscribers belong to more than 5 communities. However, we find clearly different patterns of community memberships for iPhone adopters alone (Figure 3b), that they are more likely to belong to multiple communities. This provides us with extra implications that iPhone adopters tend to link with others with the shared properties through overlapping social circles. Moreover, we believe that the intersection of overlapping communities may reveal another type of "meso"-level organizing principle of network: core-periphery structure [22]. In general, *core nodes* refer to set of central nodes that are connected to other core nodes as well as peripheral nodes, while *periphery nodes*, by contrast, are only loosely connected to the core nodes but not to each other.

Core-periphery structure captures individuals' network positions that current centrality measures do not account for.




(a)

(b)

Figure 3. Histogram of overlapping community affinities for all subscribers (upper) and iPhone adopters (lower).

For example, if individuals with high degrees (hubs) exist at the periphery of a network distant from densely connected core, they will still have insignificant impact in the spreading process as adoptions are likely to be confined to their affiliated communities, whereas less connected individuals who are strategically positioned at the overlaps of communities and thus become more central, then adoptions may percolate across communities and to other central individuals who are also placed at the overlaps of communities and so on. Therefore, we argue that individuals who are placed at the core of the network are more likely to be influential compared to those at rather isolated peripheral parts, such that peer influence between core and periphery members of the network may appear to be asymmetric.

### *D. Instrumental Variable Probit Model*

We organize the subpopulation data into a panel where each individual is a subscriber and each period is a calendar month and observations after the first adoption are removed from the sample. In this way, we can estimate the discrete-time hazard model using standard binary choice specification. The dependent variable is an indicator for when a subscriber first starts to use iPhone. Our model specification includes subscriber-specific characteristics such as social-demographic indicators, wireless technological aptitude, service usage and cumulative adoptions from friends that are deemed as either core or periphery.

We stratify subscribers based on core/peripheral network positions and separately measure how subscribers respond to adoption by their core and peripheral friends in the previous month, given their own network positions. More specifically, we denote a subscriber as core if she belongs to at least 5 communities and as periphery if otherwise. Among 9,194 core nodes, 5,548 (60%) are iPhone adopters, whereas only 5% of peripheral nodes are adopters. With panel data, we are also concerned about controlling for the unobserved heterogeneity as much as we can, so we introduce dummy variables to control for fixed effects across home region, time and community memberships. These variables can help reduce systematic differences across locations where subscribers may have different experience of smartphones for using mobile broadband because urban areas normally support better network coverage, across time of periods due to the seasonal effects (e.g., campaigns during Christmas), and across communities for common traits at group level that we explained in the preceding section.

*1) Identification Strategy:* Still, unobserved heterogeneity such as latent homophily might bias our estimation. We introduce instrumental variable (IV) approach derived from network structure to alleviate endogeneity concerns [23]: for a pair of connected subscribers, one has a third friend who is not a friend with the other, then this third friend's decision to adopt iPhone is correlated to the pair only through the one she is connected with. In particular, we use "cumulative adoptions by friend of friend not friend of the focal subscriber" as the IV and estimate the model using 2-stage residual inclusion (2SRI), together with the pooled Probit model as the comparison.

*2) Empirical Results:* We present the estimated results in Table I about the effects of social influence using both pooled Probit (column [1] and [3]) and 2SRI models (column [2] and [4]), respectively. We find that the likelihood of subscriber to adopt iPhone is positively associated with the cumulative friends' adoption. More importantly, we clearly see *asymmetric* peer influence between the focal subscribers and their core/periphery friends. Specifically, all subscribers are more likely to get influenced by their core friends, regardless of their own network positions. More interestingly, periphery subscribers are more likely to get influenced by core friends than vice versa. This suggests that subscribers who occupy the central positions are likely to be more influential, while those who are located at the peripheral parts of the network are more susceptible to influence from the core friends.

## V. Conclusion

This paper explores the role of social influence in subscriber's churn and smartphone adoption behavior by leveraging a large scale mobile phone dataset. Our results show that social influence may have strong impact on subscriber behavior even given the presence of confounding factors. This suggest that MNOs could benefit from more intelligent business analytics efforts to improve customer experience and carve out new revenue streams from their valuable dataset.

## Acknowledgment

This work is partially supported by Portuguese Foundation for Science and Technology through Carnegie Mellon Portugal Program under Grant SFRH/BD/51153/2010.

## References

[1] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *PNAS*, vol. 106, no. 36, pp. 15 274–15 278, 2009.

[2] V. Blondel, A. Decuyper, and G. Krings, "A survey of results on mobile phone datasets analysis," *EPJ Data Science*, vol. 4, no. 10, 2015.

[3] C. Shalizi and A. Thomas, "Homophily and contagion are generically confounded in observational social network studies," *Sociological Methods and Research*, vol. 40, no. 2, pp. 211–239, 2011.

[4] S. Aral, L. Muchnik, and A. Sundarajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *PNAS*, vol. 106, no. 51, pp. 21 544–21 549, 2009.

[5] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," *Science*, vol. 337, no. 6092, pp. 337–341, 2012.

[6] X. Bao, J. Yang, Z. Yan, L. Luo, Y. Jiang, E. M. Tapia, and E. Welbourne, "Commsense: Identify social relationship with phone contacts via mining communications," in *MDM*, 2015, pp. 227–234.

[7] Z.-Q. Jiang, W.-J. Xie, M.-X. Li, B. Podobnik, W.-X. Zhou, and H. E. Stanley, "Calling patterns in human communication dynamics," *PNAS*, vol. 110, no. 5, pp. 1600–1605, 2013.

[8] FCC, "13th Annual report and analysis of competitive market conditions with respect to mobile wireless, including commercial mobile services," Federal Communication Commission, WT Docket 10-133, 2009.

[9] T. J. Gerpott, N. Ahmadi, and D. Weimar, "Who is (not) convinced to withdraw a contract termination announcement? A discriminant analysis of mobile communications customers in Germany," *Telecommunication Policy*, vol. 39, no. 1, pp. 38–52, 2015.

[10] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2354–2364, 2011.

[11] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in *EDBT*, 2008, pp. 668–677.

[12] C. Phadke, H. Uzunalioglu, V. Mendiratta, D. Kushnir, and D. Doran, "Prediction of subscriber churn using social network analysis," *Bell Labs Technical Journal*, vol. 17, no. 4, pp. 63–75, 2013.

| Covariates | Probit [1] | 2SRI [2] | Probit [3] | 2SRI [4] |
|---|---|---|---|---|
| $Core\_friend\_adopt_{t-1}$ | 0.125*** | 0.318*** | 0.200*** | 0.706*** |
| | (0.007) | (0.04) | (0.013) | (0.032) |
| $Periphery\_friend\_adopt_{t-1}$ | 0.042** | 0.074*** | 0.047*** | 0.141*** |
| | (0.012) | (0.015) | (0.013) | (0.009) |
| $Core\_friend$ | 0.03*** | 0.036*** | 0.099*** | 0.109*** |
| | (0.002) | (0.002) | (0.003) | (0.004) |
| $Periphery\_friend$ | 0.005*** | 0.013*** | 0.016*** | 0.025** |
| | (0.0006) | (0.003) | (0.0004) | (0.001) |
| $Gender\_male$ | 0.116*** | 0.104*** | 0.153*** | 0.163*** |
| | (0.022) | (0.022) | (0.009) | (0.01) |
| $Gender\_female$ | -0.024 | -0.018 | -0.055*** | -0.053*** |
| | (0.03) | (0.031) | (0.013) | (0.013) |
| $Prepaid\_plan$ | -0.129*** | -0.035* | -0.054*** | -0.016* |
| | (0.022) | (0.02) | (0.009) | (0.01) |
| $Phone\_2.5G$ | 0.317*** | 0.316*** | 0.317*** | 0.331*** |
| | (0.046) | (0.046) | (0.018) | (0.018) |
| $Phone\_3G$ | 0.378*** | 0.338*** | 0.435*** | 0.452*** |
| | (0.046) | (0.047) | (0.018) | (0.018) |
| $Phone\_3.5G$ | 0.461*** | 0.439*** | 0.679*** | 0.718*** |
| | (0.050) | (0.051) | (0.022) | (0.022) |
| $Phone\_other$ | 0.415*** | 0.501*** | 0.504*** | 0.547*** |
| | (0.096) | (0.099) | (0.049) | (0.049) |
| $Mobile\_internet$ | 0.132*** | 0.125*** | 0.125*** | 0.142*** |
| | (0.023) | (0.023) | (0.017) | (0.018) |
| $Phone\_age$ | -0.029* | -0.031* | -0.048*** | -0.45*** |
| | (0.012) | (0.01) | (0.006) | (0.006) |
| $Tenure_t$ | 0.08*** | 0.08*** | -0.05*** | -0.04*** |
| | (0.016) | (0.02) | (0.01) | (0.01) |
| Pseudo $R^2$ | 0.0751 | 0.0889 | 0.0657 | 0.0869 |
| Observations | 63,863 | 63,863 | 2,052,992 | 2,052,992 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table I
REGRESSION RESULTS OBTAINED FROM POOLED PROBIT (COLUMNS [1] AND [3]) AND 2SRI MODELS (COLUMNS [2] AND [4])


[13] S. Mogan and C. Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research), 2nd edition*. Cambridge University Press, New York NY, 2014.

[14] K. Hirano and G. W. Imbens, “The propensity score with continuous treatments,” in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, A. Gelman and X.-L. Meng, Eds. John Wiley & Sons, Ltd, 2005, pp. 73–84.

[15] R. Nielsen and J. Sheffield, “Matching with time-series cross-sectional data,” in *The 26th Annual Society for Political Methodology Summer Conference*, 2009.

[16] H. Risselada, P. C. Verhoef, and T. H. Bijmolt, “Dynamic effects of social influence and direct marketing on the adoption of high-technology products,” *Journal of Marketing*, vol. 78, no. 2, pp. 52–68, 2014.

[17] M. Matos, P. Ferreira, and D. Krackhardt, “Peer influence in the diffusion of the iPhone 3G over a very large social network,” *MIS Quarterly*, vol. 38, no. 4, pp. 1103–1133, 2014.

[18] M. Jackson, B. Rogers, and Y. Zenou, “The economic consequences of social network structure,” Centre for Economic Policy Research, Working Paper No. 10406, Feburary 2015.

[19] J. Park and A.-L. Barbasi, “Distribution of node characteristics in complex networks,” *PNAS*, vol. 106, no. 46, pp. 17 916–17 920, 2007.

[20] Y.-Y. Ahn, J. Bagrow, and S. Lehmann, “Link communities reveal multiscale complexity in networks,” *Nature*, vol. 466, pp. 761–764, 2010.

[21] J. Yang, J. McAuley, and J. Leskovec, “Community detection in networks with node attributes,” in *ICDM*, 2013, pp. 1151–1156.

[22] J. Yang and J. Leskovec, “Overlapping communities explain core-periphery organization of networks,” *Proceedings of the IEEE*, vol. 102, no. 12, pp. 1892–1902, December 2014.

[23] Y. Bramoullé, H. Djebbari, and B. Fortin, “Identification of peer effects through social networks,” *Journal of Econometrics*, vol. 150, no. 1, pp. 41–55, 2009.