Asymmetric Role of Social Influence in Smartphone Adoption in Large Mobile Networks

Qiwei Han[*†]
qiweih@cmu.edu

Pedro Ferreira[*‡]
pedrof@cmu.edu

João Paulo Costeira[†]
jpc@isr.ist.utl.pt

[*]Department of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, PA 15213
United States

[‡]Heinz College
Carnegie Mellon University
Pittsburgh, PA 15213
United States

[†]Instituto Superior Técnico
University of Lisbon
Lisbon, 1049-001
Portugal

1. INTRODUCTION

As the number of mobile subscriptions worldwide is reaching nearly 7 billion by the end of 2014 [10], the rapid growth of mobile handset adoption has evolved from communication devices (*a.k.a.* feature phones) with basic telephony service to advanced multifunction computing devices (*a.k.a.* smartphones) that provide one-stop solution to meet popular daily needs. Hence, smartphones have become platforms enabling a multitude of services and applications, through which the business of mobile market is delivered through these devices. Understanding the adoption pattern of this high-technology product is of great interest to not only handset manufacturers and mobile operators, but also to service providers and application developers, and it may also reveal significant managerial and policy implications [11].

Social influences have been acknowledged the important role in determining consumer's decision making process for various mobile communication products and services (e.g. [8, 9, 12, 13]. According to [5], the positive (negative) word-of-mouth message in the mobile phone market can increase (decrease) a company's market share by as much as 10 (20) percent. In particular, with regard to smartphone adoption, both [13] and [17] found positive effect, meaning the propensity for a focal customer to adopt smartphone increases with the number of adoptions in her ego network. However, identifying social influence in large networks, especially in observational studies, still remains as serious concerns to researchers [7], as it is known to confound with endogenous factors such as homophily - the tendency for individuals to choose friends with similar tastes, among others [14]. When correlated behavior between ego and her direct neighbors can be explained by both influence and their inherent similarities, misattribution of homophily to influence may lead to significant overestimation of the latter [3]. Moreover, [19] argued that social influence and homophily are generically confounded and can not be readily separated from each other, because latent homophily may still remain as a component of the estimated influence. Nevertheless, they proposed several constructive suggestions that may help alleviate the issue, one of which our work is motivated: to use community structure as the proxy to control for latent homophily. We intend to estimate the effect of social influence on adoption of iPhone 3G, an iconic smartphone, using dataset from a major wireless carrier operated in one European country (hereinafter EURMO). iPhone 3G is released in July 2008 and EURMO is the sole partner with Apple in this country, so we are able to capture the full cycle of its adoption.

2. DATA

EURMO dataset includes call detail records (CDR) for over 5 million subscribers between August 2008 and June 2009. Subscribers are identified by their anonymized phone numbers. For each call we know the caller and the callee, the timestamp, and the GPS coordinates of the connected cell tower. By aggregating GPS coordinates over the entire period, we can approximate subscriber's home location at municipal level as where they spend most of their days[1]. We further infer the socio-economic indicators (*e.g.* wage) by cross-referencing the latest census. We also have an (incomplete) set of subscriber characteristics such as gender and usage history since their subscription to EURMO, which includes tariff plan, handset and supplementary services (*e.g.* mobile broadband). In our period of analysis, there are 20,570 iPhone 3G adopters with complete profiles. Table 1 lists relevant variables and short descriptions that we extract from EURMO.

We use CDR to construct the social network as an undirected call graph. Specifically, we denote two subscribers to befriend each other if they exchange at least one call in the same calendar month. The mutual relationship between subscribers ensure that we preclude communications that are unlikely to represent the social ties, such as customer services and PBX machines. The resulting network consists of 5,535,388 subscribers and 66,717,468 edges with mean, standard deviation, and median of degree being 24.1, 25.7 and 16, respectively.

3. SUBPOPULATION EXTRACTION

Recent researches on statistical properties of real world social networks provide evidence of the existence of community structure within the network [6, 16]. These findings does not only validate the theoretic role of homophily and influence in tie formation, but also provide several important insights into the problem of community structure inference when we only observe the resulting network as follows. First, uncovered communities should exhibit real social meaning, as individuals in the same community have some natural affinity for each other or some fundamental characteristics in common. Meanwhile, they should be more likely to connect to each other than those who belong to different communities. Hence community discovery method should consider

[1]The municipal location is defined as Nomenclature of Units for Territorial Statistics (NUTS) III, which is a geocode standard across European countries by Eurostat for statistical purposes

Variable	Type	Description
gender	categorical	Self-reported gender (male, female, unknown)
wage	categorical	Inferred wage level (very low, low, average, high, very high)
prepaid	binary	prepaid tariff plan (yes, no)
phone_technology	categorical	handset technology (2G, 2.5G, 3G, 3.5G, other)
mobile_internet	binary	mobile broadband (yes, no)
phone_age	continuous	age of currently owned handset (year)
tenure	continuous	tenure since subscription (year)
region	categorical	home location at municipal level

Table 1: List of covariates extracted from EURMO

two different sources of information together, *i.e.* individual characteristics and social connections among them. Second, in many actual networks, individuals may belong to multiple overlapping communities [1, 15], *e.g.* families, coworkers and friends. This is also aligned with the notion of homophily across different social dimensions. Third, as also noted in [19], misspecification of community structure (*e.g.* simple modular and/or disjoint structure) may even worsen the problem and lead to biased model estimation. Fourth, as the complexity of network structure grows exponentially with the size, computational costs still remain challenging for the analyses on large scale networks. Therefore, extracting subpopulation via community discovery does not only significantly reduce group-level heterogeneity that may potentially confound the result but also help lessen the computational cost [22].

We employ the method of discovering Communities from Edge Structure and Node Attributes (CESNA) for our purpose [21]. CESNA is a community discovery algorithm that consider both node attributes and network structure as well as the interactions between these two sources of information. It can detect overlapping communities with high accuracy and scalability over many existing community detection algorithms, particularly on large scale networks. For sake of space, details beyond the mechanics of CESNA can be found in [21], and we only note the following implementation procedures: i) for each iPhone 3G adopter, we construct the ego-network that contains adopter and their direct neighbors, similarly as [17]; ii) for each subscriber in the ego-network, we extract a list of 0-1 valued covariates specified in table 1 that 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 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 communities that contain iPhone 3G adopters. As a result, we obtain 11,454 communities with 202,743 subscribers, 14,685 of which are iPhone 3G adopters. After detecting community structure from networks, we are able to "naturally" extract subpopulations with half of the original network size which include over 70% of adopters and their cohesive groups of neighbors with whom are both similar and strongly connected.

4. CORE-PERIPHERY STRUCTURE

From the extracted subpopulation, we observe that 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. This is consistent with the findings suggested in [18, 20] that the intersection of overlapping communities may reveal *core-periphery structure* which complements current views of network organizations. In general, core nodes refer to set of central nodes that are connected to other core nodes as well as peripheral nodes, while peripheral nodes, by contrast, are only loosely connected to the core nodes but not to each other [4]. In this sense, following the measure proposed in [20] that unify both organizing principles of the network, we validate the existence of core-periphery structure in our subpopulation (see Fig. 1).

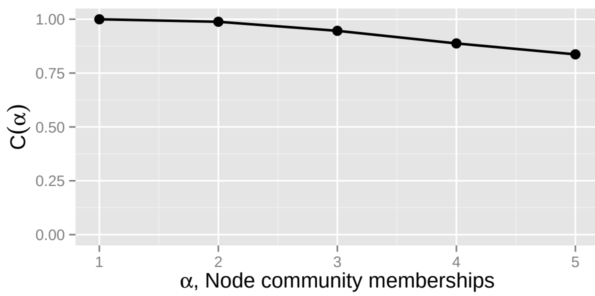

Figure 1: The fraction of nodes $C(\alpha)$ in the largest connected component of the induced subgraph on the nodes who belong to at least a communities. A high $C(\alpha)$ means that there is a single dominant core.

We then define a subscriber as *core* node if she belongs to at least five communities and as *periphery* node if otherwise. Among 9,194 core nodes, 5,548 (60%) are iPhone 3G adopters, whereas only 5% of peripheral nodes are adopters. This provides us with extra implications that iPhone 3G adopters tend to form densely connected groups that have the most shared properties with others thorough overlapping communities.

5. MODEL AND RESULTS

We describe our reduced form latent utility model as follows (index i are dropped for simplicity):

$$\begin{aligned} Y_t &= 1\{U_t > 0\} \\ U_t &= \alpha + \beta \cdot X + \gamma \cdot Z_t + \delta \cdot Core \\ &\quad + \mu_1 \cdot Core\_Friend\_Adopt_{t-1} \\ &\quad + \rho_1 \cdot Core \cdot Core\_Friend\_Adopt_{t-1} \\ &\quad + \mu_2 \cdot Peri\_Friend\_Adopt_{t-1} \\ &\quad + \rho_2 \cdot Core \cdot Peri\_Friend\_Adopt_{t-1} + \epsilon_t \end{aligned} \tag{1}$$

where X and Z_t are time invariant and time variant subscriber-specific characteristics listed in table 1. Dummy variable $Core$ indicates the incremental *network position effect* when subscriber is deemed as core. For social influence, we include the *main effects* of number of core friend adopters $Core\_Friend\_Adopt_{t-1}$ and number of peripheral friend adopters $Peri\_Friend\_Adopt_{t-1}$ and the *interaction effects* between number of friend adopters and subscriber's network position. A significant coefficient for the latter terms captures the social influence from core friend adopters and peripheral friend adopters, relative to focal subscriber's network position. Besides, we also introduce month and location fixed effects to control for heterogeneity across time and region.

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 need to be removed from the sample. We use the resulting sample (size=2,116,855) to empirically estimate the social influence on iPhone 3G adoption. Table 2 summarizes the estimates of covariates of interest from equation 1 using Probit model. The full regression table is available upon request.

Dependent Variable: $Adopted_t$	Probit
$Core(\delta)$	1.522^{***}
	(0.021)
$Core\_Friend\_Adopt_{t-1}(\mu_1)$	0.363^{***}
	(0.008)
$Peri\_Friend\_Adopt_{t-1}(\mu_2)$	0.187^{***}
	(0.008)
$Core \cdot Core\_Friend\_Adopt_{t-1}(\rho_1)$	0.092^{***}
	(0.01)
$Core \cdot Peri\_Friend\_Adopt_{t-1}(\rho_2)$	0.244^{***}
	(0.013)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Regression results of Probit Model

By computing the average partial effect using delta method, we find that for core subscribers, their propensity to adopt iPhone 3G is over 10% higher than peripheral subscribers, indicating that subscribers who are in the higher network position (maybe also social status) are more likely to adopt the smartphone. Meanwhile, having one more core friend adopter and one more peripheral friend adopter will increase the focal subscriber's adoption probability by 0.46% and 0.25%, respectively. However, interpreting the coefficients of interaction terms are far more straightforward [2], as the interaction effect needs to be calculated as the cross-partial derivatives. We find *asymmetric* interaction effect on core and peripheral subscribers when having core and peripheral adopter friends. Specifically, on average with one more core (peripheral) adopter friend, the changes between core and peripheral subscribers is about -1.52% (-1.76%). The implications of this finding are twofold: 1) peripheral subscribers are more likely to get influenced than core subscribers; 2) core subscribers are slightly more likely to be get influenced by a core adopter friend rather than a peripheral adopter friend.

6. REFERENCES

[1] Ahn, Y.-Y., Bagrow, J., and Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature 466* (2010), 761–764.

[2] Ai, C., and Norton, E. Interaction terms in logit and probit models. *Economic Letters 80*, 1 (2003), 123–129.

[3] Aral, S., Muchnik, L., and Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of National Academy of Science 106*, 51 (2009), 21544–21549.

[4] Borgatti, S., and Everett, M. Models of core/periphery structures. *Social Networks 21*, 4 (2000), 375–395.

[5] Bughin, J., Doogan, J., and Vetvik, O. A new way to measure word-of-mouth marketing. *Mckinsey Quarterly* (April 2010), 1–9.

[6] Girvan, M., and Newman, M. Community structure in social and biological networks. *Proceedings of National Academy of Sciences 99* (2002), 7821–7826.

[7] Goldsmith-Pinkham, P., and Imbens, G. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*, 3 (2013), 253–264.

[8] Han, Q., and Ferreira, P. Role of peer influence in churn in wireless networks. In *Proceedings of 7th International Conference on Social Computing, SocialCom 2014, Beijing, China* (2014), pp. 176–183.

[9] Hill, S., Provost, F., and Volinsky, C. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science 21*, 2 (2006), 256–276.

[10] ITU. The World in 2014: ICT Facts and Figures, 2014.

[11] Kauffman, R., and Techatassanasoontornb, A. Understanding early diffusion of digital wireless phones. *Telecommunications Policy 33*, 8 (2009), 432–450.

[12] Ma, L., Krishnan, R., and Montgomery, A. Latent homophily or social influence? an empirical analysis of purchase within a social network. *Management Science 57*, 9 (2014), 1623–1639.

[13] Matos, M., Ferreira, P., and Krackhardt, D. Peer influence in the diffusion of the iPhone 3G over a very large social network. *MIS Quarterly 38*, 4 (2014), 1103–1133.

[14] McPherson, M., Smith-Lovin, L., and Cook, J. M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27* (2001), 415–444.

[15] Palla, G., Derenyi, I., Farkas, I., and Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature 435* (2005), 814–818.

[16] Park, J., and Barbasi, A.-L. Distribution of node characteristics in complex networks. *Proceedings of National Academy of Sciences 106*, 46 (2007), 17916–17920.

[17] Risselada, H., Verhoef, P., and Bijmolt, T. Dynamic effects of social influence and direct marketing on the adoption of high-technology products. *Journal of Marketing 78* (March 2014), 52–68.

[18] Rombach, M. P., Porter, M., Fowler, J., and Mucha, P. Core-periphery structure in networks. *SIAM Journal of Applied Math 74*, 1 (2014), 167–190.

[19] Shalizi, C. R., and Thomas, A. C. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research 40*, 2 (May 2011), 211–239.

[20] Yang, J., and Leskovec, J. Overlapping communities explain core-periphery organization of networks. *Proceedings of the IEEE 102*, 12 (December 2014), 1892–1902.

[21] Yang, J., McAuley, J., and Leskovec, J. Community detection in networks with node attributes. In *Proceedings of the IEEE 13rd International Conference on Data Mining, Dallas TX, USA* (2013), pp. 1151–1156.

[22] Zhang, B., Krackhardt, D., Krishnan, R., and Doreian, P. An effective and efficient subpopulation extraction method in large social networks. In *Proceedings of the 32rd International Conference on Information Systems, Shanghai, China* (2011).