



---

Peer Influence in the Diffusion of iPhone 3G over a Large Social Network

Author(s): Miguel Godinho de Matos, Pedro Ferreira and David Krackhardt

Source: *MIS Quarterly*, Vol. 38, No. 4 (December 2014), pp. 1103-1134

Published by: Management Information Systems Research Center, University of Minnesota

Stable URL: <https://www.jstor.org/stable/10.2307/26627964>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Management Information Systems Research Center, University of Minnesota is collaborating with JSTOR to digitize, preserve and extend access to *MIS Quarterly*

## PEER INFLUENCE IN THE DIFFUSION OF IPHONE 3G OVER A LARGE SOCIAL NETWORK<sup>1</sup>

**Miguel Godinho de Matos**

Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213-3890 U.S.A. {miguelgodinhomatos@cmu.edu} and  
Católica Lisbon School of Business and Economics, Palma de Cima, 1649-023 Lisbon, PORTUGAL

**Pedro Ferreira**

Heinz College and Department of Engineering and Public Policy, Carnegie Mellon University,  
Pittsburgh, PA 15213-3890 U.S.A. {pedrof@cmu.edu}

**David Krackhardt**

Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213-3890 U.S.A. {krack@cmu.edu}

*In this paper, we study the effect of peer influence in the diffusion of the iPhone 3G across a number of communities sampled from a large dataset provided by a major European Mobile carrier in one country. We identify tight communities of users in which peer influence may play a role and use instrumental variables to control for potential correlation between unobserved subscriber heterogeneity and friends' adoption. We provide evidence that the propensity of a subscriber to adopt increases with the percentage of friends who have already adopted. During a period of 11 months, we estimate that 14 percent of iPhone 3Gs sold by this carrier were due to peer influence. This result is obtained after controlling for social clustering, gender, previous adoption of mobile Internet data plans, ownership of technologically advanced handsets, and heterogeneity in the regions where subscribers move during the day and spend most of their evenings. This result remains qualitatively unchanged when we control for changes over time in the structure of the social network. We provide results from several policy experiments showing that, with this level of effect of peer influence, the carrier would have hardly benefitted from using traditional marketing strategies to seed the iPhone 3G to benefit from viral marketing.*

**Keywords:** Peer influence, homophily, diffusion, community identification, viral marketing

### Introduction

The pervasiveness of simple, small, and light handsets changed the way people communicate. The worldwide penetration of mobile handsets grew from 12 percent in 2000 to 87

percent in 2011 (ITU 2011). Smartphones accounted for a significant part of this growth. Their penetration was 65 percent in the United States, 50 percent in Europe, and 29 percent worldwide in 2011 (Vakulenko 2011). The mobile handset market is expected to reach \$340 billion in revenues by 2015. Smartphones are expected to account for 75 percent of this revenue, thus growing 24 percent per year on average between 2010 and 2015. In recent years, handsets have become small computers. Accordingly, their value has shifted toward the software and the data services they provide (*The Economist* 2011). Manufacturers increasingly are bundling

<sup>1</sup>Ravi Bapna was the accepting senior editor for this paper. Sanjukta Smith served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly*'s website (<http://www.misq.org>).

handsets with applications that generate positive network externalities. Examples include FaceTime and Google Talk, which allow consumers to video-call over the Internet at no cost. However, FaceTime requires an Apple device and Google Talk or a handset running Android. Hence, the utility derived from using these applications, and therefore the utility associated with the respective handsets, increases with the number of people that use them. The seminal works by Rohlfs (1974) and Katz and Shapiro (1986) show how the number of consumers buying a product may generate positive network externalities leading to the well-known S-shaped curve. This has been empirically demonstrated in several instances such as automatic teller machines (Saloner and Shepard 1995) and spreadsheet software (Brynjolfsson and Kemerer 1996). However, in the case of mobile communication products, and from the perspective of an individual, who adopts the product is likely to be as important, if not more, as how many people adopt because it is the exposure to these products through friends that may significantly reduce uncertainty and trigger adoption.<sup>2</sup> Yet, the role of peer influence in the diffusion of telecommunication products has not been largely explored. Exceptions include the study of video-conferencing (Tucker 2008), applications for personalized news (Aral et al., 2009) and caller ring back tones (Ma et al. 2010).

The penetration of handsets has been measured per brand and per country (Botelho and Pinto 2004; Doganoglu and Grzybowski 2007; Gruber 2001) but the specific mechanisms by which they diffuse have not been studied in detail. However, handsets are fundamental devices in the telecommunications industry. They are the gateways by which consumers talk to each other and obtain access to information. The whole business of cell phone operators relies on spreading handsets across consumers. If peer influence shapes handset sales, then carriers must place significant care and effort in targeting the right early adopters while reaching critical mass. If, on the other hand, the effect of peer influence is only small, then carriers are likely better off with mass marketing campaigns instead of targeted advertising.

Diffusion denotes the dissemination of a trait, product, or service within a social system (Strang and Soule 1998). The social system comprises the set of individuals, groups, and institutions that might adopt the innovation. The members of

the social system use communication channels, such as face-to-face meetings, phone conversations and email messages, to exchange information about the innovation being diffused (Mahajan 1985, p. 7). Diffusion can be caused by external influence, by mechanisms internal to the social system, or by both. External triggers include, for example, policy requirements (Mahajan 1985, p. 15; Valente 1996a, p. 95) and mass media (Valente 1996a, p. 81). Internal mechanisms relate mostly to how one's adoption depends on the behavior of the other members of the social system. In this context, peer influence arises as the dyadic process by which an individual shapes her behavior, beliefs, or attitudes according to what the other individuals in the social system think, express, or how they behave (Leenders 2002).<sup>3</sup>

A number of studies use randomized experiments to identify the effect of peer influence. Aral and Walker (2011) show evidence of contagious adoption of a Facebook application that allows users to share comments related to the movie industry. They seed the application to about 10,000 users and observe viral messaging. Bapna and Umyarov (2014) use Last.fm, a music-sharing website. They award free premium subscriptions to a random set of users and show that being connected to users that subscribe to the premium service increases the likelihood of acquiring this service. Most research on peer influence to date studies the diffusion of cheap, or even free, digital goods. However, it is not clear how the results found in these studies extend to settings in which consumers have to risk money to purchase physical products. While sometimes a good sold at a higher price is perceived as being better (McConnell 1968; Shiv et al. 2005; Stafford and Enis 1969), in other instances free goods may seem superior (Shampanier et al. 2007). It is, therefore, unclear how price determines consumption and, more importantly, how price mediates the effect of peer influence.

Empirical evidence on how information from friends may shape the consumption of costly goods is still lacking. This is not surprising because it is difficult to run randomized experiments with costly goods. On one hand, consumers would likely need to be compensated for the money they spend. On the other hand, it is hard to design such experiments in dense network settings in which treated and control

<sup>2</sup>Peer influence has been shown to play a significant role in many different contexts such as academic performance (Boucher et al. 2010; Carrell et al. 2009; Sacerdote 2001), smoking and drinking behavior (Case and Katz 1991; Mercken et al. 2010), sexual conduct (Romer et al. 1994), trade unions (Hedström 1994), vaccinations (Rao et al. 2007), new drugs (Burt 1987; Coleman et al. 1966; Strang and Tuma 1993; Valente 1996a) and electronic health records (Zheng et al. 2010).

<sup>3</sup>A broad range of mechanisms can drive peer influence (Leenders 2002; Strang and Soule 1998; Van den Bulte and Lilien 2001), such as information transfer (when individuals get to know about the innovation through previous adopters; Katz and Lazarsfeld 1955), competition (when individuals look at their rivals as frames of reference), conformity (when individuals comply with the behavior and attitudes of the groups to which they belong; Menzel 1960), network externalities (when the utility of adopting an innovation increases with the number of adopters; Katz and Shapiro 1994), and spatial proximity (the fact that geographical proximity facilitates all types of interactions across individuals; Strang and Soule 1998).

individuals connect, potentially contaminating results (Aral and Walker 2011). In these settings, a promising course of action might be to try to tease out the effect of peer influence from observational data (Aral 2010; Aral et al. 2009). This, however, is also hard to accomplish (Manski 1993; Van den Bulte and Lilien 2001). Difficulties arise, for example, when one cannot identify the group to which each individual belongs. Today's increased ability to collect data that captures the structure of the relationships among people facilitates identifying these groups (Wasserman and Faust 1994). Yet, collecting such data is expensive and only a handful of studies, such as Anagnostopoulos et al. (2008) and Aral et al. (2009), did so in large, real-world settings. In addition, these authors primarily look only at cheap digital goods.

Another difficulty in observational studies arises when unobserved heterogeneity across individuals that influences behavior also influences the formation of social ties (Aral et al. 2009). Peer effects models (Bramoulle et al. 2009; Oestreicher-Singer and Sundararajan 2012) and matching in high resolution panels (Aral et al. 2009) have been used to control for such unobservables. Other approaches include using structural models (Ma et al. 2010), propensity score matching (Aral et al. 2009), randomization (Anagnostopoulos et al. 2009), and instrumental variables (Tucker 2008). Interestingly, identifying groups of users and controlling for homophily are deeply interconnected. In fact, homophily suggests that social networks develop around groups of individuals that share similar unobserved traits, often called *communities* in the literature (Leskovec et al. 2009; Newman 2006). Therefore, controlling for community structure may help alleviate concerns with unobserved homophily (Shalizi and Thomas 2011). In particular, community membership may capture part of the unobservables that determine network structure, reducing the scope for unobserved characteristics that influence behavior and determine the formation of ties across individuals. Once communities of users have been identified, one can also use stochastic agent-based models to describe the coevolution of adoption and of the formation of social ties within and between communities (Snijders et al. 2010). These models can also help separate the contribution of homophily and peer influence for the diffusion of an innovation across a social system.

Our paper is an observational study aimed at characterizing the diffusion of the iPhone 3G in a large social network. We hypothesize that the likelihood of an individual adopting this handset is associated with her level of exposure to it (Valente 1996a) through her friends (Leenders 2010; Valente 1996b). This idea is in line with the arguments of Rogers (1995) whereby interactions between non-adopters and adopters mitigate the risk and uncertainty associated with the innovation, increasing the likelihood of adoption. As such, our

paper contributes to the set of IS literature that identifies peer effects in the adoption of technology in several new ways.

First, we study the effect of peer influence with a costly physical good. This complements the current literature looking at peer influence with cheap digital goods. Second, we do so in the context of a large social network from which we identify tight communities of individuals. In particular, we develop a new algorithm to identify communities with adopters when adoption is a rare event. This algorithm may inspire other researchers to study influence with costly goods, in which case adoption is likely to be limited. Third, we combine community identification with instrumental variables to identify the effect of peer influence. We also use agent-based modeling to measure this effect with a time-varying social network. Finally, we provide results from policy simulations characterizing the virality of this handset.<sup>4</sup>

We use a large dataset from a major European mobile carrier in one country, hereinafter called EuroMobile, to identify the effect of peer influence in the diffusion of the iPhone 3G. This is an expensive handset in the country analyzed and therefore it is unclear whether peer influence can affect its sales. We analyze call detailed records (CDRs) from thousands of users, their friends, and friends of their friends, between August 2008 and June 2009 to model their social network. We then explore the structure of this social network to identify the effect of peer influence. We combine community identification (Newman and Girvan 2004) with instrumental variables to alleviate endogeneity concerns. In particular, we instrument friends' adoption with the adoption of the friends of friends that are not friends of the ego and that live in a city different from that where the ego does to alleviate endogeneity concerns. We also use stochastic agent-based models (Snijders et al. 2010) to look at how the evolution of the social network over time correlates to the adoption of the iPhone 3G.

We find that if all friends adopt the iPhone 3G, then the probability of the ego to adopt increases by 15 percent on average. We obtain this result after controlling for heterogeneity across regions, across subscribers, and over time. This result is robust across specifications including different approaches to define our instruments as well as different ways to code the adjacency matrix across subscribers. We also show that peer influence accounted for roughly 14 percent of iPhone 3G sales during the first 11 months at EuroMobile. We provide results from several policy simulations showing

<sup>4</sup>In addition, our paper also complements the literature that measures the penetration of handsets in countries such as Germany (Singh 2008), India (Chu et al. 2009), Japan (Park and Ueda 2011), Korean (Park and Ueda 2011), Portugal (Botelho and Pinto 2004), and Taiwan (Chu et al. 2009),

that EuroMobile would have hardly benefitted from seeding the iPhone 3G to increase its penetration in light of the modest magnitude for the effect of peer influence that we find in this paper. This finding, however, depends highly on how much of the handset is subsidized to the seeds. In this regard, our paper contributes to improve our understanding of the potential of viral marketing (Leskovec et al. 2007; Watts et al. 2007) for costly products.

The reminder of this paper is structured as follows. The next section introduces EuroMobile and discusses the release of the iPhone 3G. The subsequent section describes our approach to identify communities of users. We then introduce our model and empirical strategy, after which we present our results and robustness checks. Following that, we show the results from policy simulations and, finally, present our conclusions.

## Mobile Carrier and the Release of iPhone 3G

Our industrial partner EuroMobile is the market leader of mobile services in this country. EuroMobile granted us access to all its detailed call records (CDRs) in one European country between August 2008 and June 2009. EuroMobile released the iPhone 3G in this country in July 2008. For every call placed or received by a EuroMobile subscriber, we have the anonymized phone numbers of the initiator and recipient, the GPS coordinates for the cell towers used, a timestamp, and the duration of the call. For every sent text message (sms) initiated or received by a EuroMobile subscriber, we have the anonymized phone numbers of the initiator and recipient and a timestamp. We also have a set of subscriber characteristics such as date of birth, gender, type of contract, tariff plan, handset, zip code for the account owner, and changes in all these covariates over time. Our dataset spans 5 million active phone numbers. We have about 3.7 billion calls and 13 billion sms in this dataset.

## The Social Network Graph

We use all CDRs in this dataset to define an undirected graph of communications across EuroMobile subscribers. We trimmed subscribers from other providers because we did not have sufficient information on them, such as whether they adopted the iPhone 3G. An edge between two EuroMobile subscribers is added to our graph if one of them called or sent an sms to the other and the latter answered back with a call or an sms within the same calendar month. This procedure disregards communications that are unlikely to proxy social

proximity such as those involving message bots, short numbers, and call centers.<sup>5</sup>

The resulting graph includes 4,986,313 subscribers and 57,069,798 edges. The undirected graph density is  $4.59 \times 10^{-6}$  and the mean degree is 22.9 with a standard deviation of 25.5. The median degree is 13. Figure 1 plots the empirical degree distribution for EuroMobile subscribers. We say that two EuroMobile subscribers are friends if they are connected through an edge in this graph. Hereafter, the number of friends of a EuroMobile subscriber is her *degree*.

## The Release of the iPhone 3G

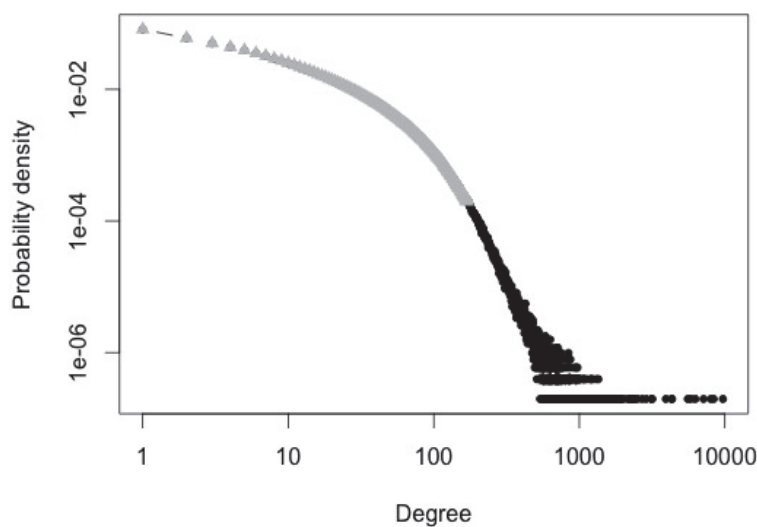
EuroMobile released the iPhone 3G in the country analyzed in July 2008. This was the first Apple smartphone commercialized in this country.<sup>6</sup> Smartphones accounted only for 11 percent of all handset sales during 2008. Figure 2 shows that ease of use, price, size, and weight were among the most important factors that consumers in this country considered when purchasing a handset. However, the iPhone 3G was a conspicuous luxury product in this market. The 8Gb model launched for 500 Euros without a contract with a carrier.<sup>7</sup> Furthermore, subscribers could only benefit from its full potential if they subscribed to a data plan. Figure 2 shows that in 2008 consumers in this country did not care much for Internet connectivity or network coverage on their handset. These facts suggest that there was significant uncertainty about the iPhone 3G, which likely motivated consumers to seek opinions from their friends. Therefore, the case of the iPhone 3G in this country in 2008 provides an appropriate setting to study the potential effect of peer influence. During the first 11 months in the market, less than 1 percent of EuroMobile subscribers adopted the iPhone 3G.

<sup>5</sup>We also removed from our analysis phone numbers that switched between pre-paid and post-paid because in these cases we are unable to track their history reliably. These operations eliminate less than 5 percent of the subscribers. We also trim from this graph subscribers whose degree is three standard deviations above the mean. This removes private branch machines (which only route calls among the internal telephones of an organization) and ensures that the size of our graph is computationally manageable. This eliminates less than 1.6 percent of the subscribers.

<sup>6</sup>At that time, Nokia was the preferred mobile phone brand with a market share above 40 percent, followed closely by Samsung, LG, and Sony, each with roughly a 10 percent market share.

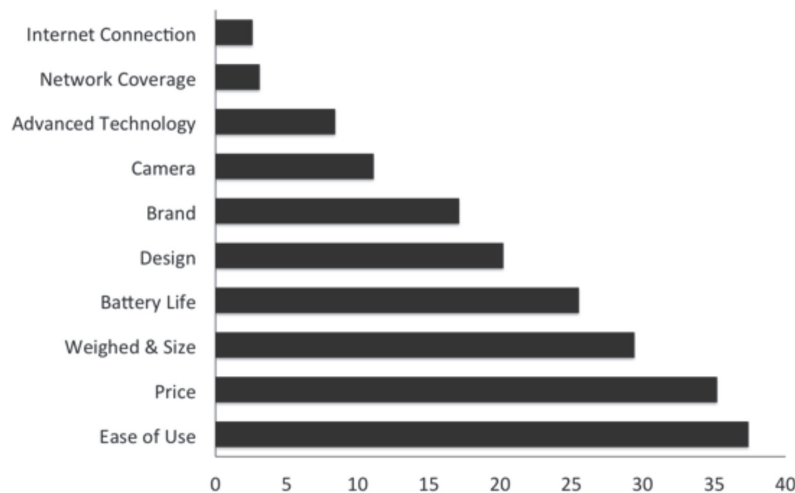
<sup>7</sup>Prices ranged between 130 and 400 Euros with contracts. Contracts would last for 24 months and significant penalties applied to terminate them before expiry. All iPhones were sold with a SW lock to the carrier selling the device. This could be unlocked at no charge after 24 months or before 24 months for a 200 Euro fee. Monthly contract fees ranged between 30 and 65 Euros across carriers. People in this country spent 20 Euros per month on average for mobile communications during 2008.





**Note:** Grey dots represent subscribers whose degree is three standard deviations below the mean.

**Figure 1. Degree Distribution across EuroMobile Subscribers**



**Note:** Percent of total respondents who checked each particular characteristic.

**Figure 2. Most Important Characteristics Considered When Purchasing a New Handset**

## Working with Communities

Peer influence is a local phenomenon. It measures whether peers play a part in someone's decision to undertake actions. For this reason, peer influence is likely to occur within communities of friends, that is, groups of people with many connections within the group and only a few connections to people outside the group. We argue that in order to study

peer influence one must identify tight communities of individuals in which peer influence can be meaningful instead of looking for average peer influence effects across large networks of people which are unlikely to embed real social meaning. In this paper, we aim at studying how friends influence one's decision to purchase the iPhone 3G but controlling for the fact that potential influence occurs within the scope of communities of friends. For this purpose, we focus

on communities with about 100 people, which according to Leskovec et al. (2009) is the size beyond which clusters tend to start blending into a giant component of the social graph and thus start to lose social meaning.

Furthermore, identifying communities of subscribers yields additional empirical advantages. For example, social networks tend to aggregate individuals that share common unobserved traits. These latent unobservables confound the identification of peer influence. Yet, community membership is likely to capture part of the unobservables that determine network structure and, at the same time, drive behavior. Thus, controlling for community membership is likely to help control for unobserved homophily (Shalizi and Thomas 2011).

Estimating peer effects on a community basis is also likely to reduce the downward bias that may arise from considering large networks of consumers. To see this, consider two communities, one completely immune to a disease and the other vulnerable to it. Introducing sick individuals into the former would lead us to think that the disease is not contagious. Introducing sick individuals into the latter would allow us to determine the rate of contagion. Studying contagion in this population as a whole would underestimate the virulence of the disease. Firms are typically interested in measuring peer effects in communities that are likely to purchase their products. In fact, they tailor their products to market segments separately in ways they believe might accelerate sales. From a managerial point of view, one is not interested in overall peer effects but rather in local influence effects. In the case of the expensive iPhone 3G, subscribers susceptible to adopting this handset might be very different from the average EuroMobile subscriber. Our interest is in identifying peer influence among adopters.

Finally, identifying communities also permits working with samples of independent, cohesive groups of consumers, which can be used to correct standard errors and avoid claiming effects when there might be none. Treating observations as independent and identically distributed when dealing with social network data is a major drawback of many empirical analyses which we try to limit in this paper.

### Algorithms to Identify Communities

Table 1, based on the survey in Fortunato (2010), provides a list of algorithms used in the literature to identify communities in social graphs. Most of these algorithms cannot be applied to large social networks due to their computational complexity. The comparative studies in Pons and Latapy (2006), Lancichinetti et al. (2008), and Lancichinetti and Fortunato (2009) show clearly that the first 10 algorithms in

this table are unable to process a network with 5 million nodes and 57 million edges. For this reason, in this paper, we focus on algorithms 11 to 15.

Most of the algorithms in this table are based on greedy optimizations of the modularity index.<sup>8</sup> However, as Fortunato and Barthelemy (2007) show, this index has a resolution limit and fails to identify small communities in large social graphs. Therefore, we do not use modularity in our case. Not only is our network very large but our interest resides in identifying tight and small communities where peer influence can have real social meaning. Instead, we use the internal external ratio (IER) introduced in Krackhardt and Stern (1988) to measure the cohesiveness of communities. In our case,  $IER = (I - E) / (I + E)$ , where  $I$  denotes the number of calls and sms exchanged by individuals within the community and  $E$  denotes the number of calls and sms exchanged with individuals outside the community.<sup>9</sup> IER balances the effort that individuals place in maintaining friendships within versus outside the community ( $I - E$ ). In this regard, it is a good measure of the relative investment that individuals place in connecting with people within their community. Also, IER measures this effort relative to all of the effort that individuals place in maintaining friendships ( $I + E$ ). Therefore, IER also controls well for the heterogeneity in the effort to manage friendships across individuals.

Table 2 shows that the algorithms used to identify communities introduced above perform poorly when applied to our graph. Columns 4 to 7 show that more than half of the communities identified have at most 2 people. Column 6 shows that most individuals are placed in large communities, which are unlikely to entail social meaning. Columns 8 to 14 show that if one limits the analysis to communities with 25 to 200 people, then few of them have adopters. Columns 15 and 16 show that if one further limits the analysis to communities with positive IERs, then one is left with virtually no adopters. In sum, the communities identified by these algorithms can hardly be used to study peer influence with rare events.

<sup>8</sup> Modularity compares the number of edges within a given community to the number of edges that this community would have if network ties had been randomly established. Therefore, it assumes that any two people in the network can connect. This, however, is unlikely as networks grow.

<sup>9</sup> IER varies in  $[-1, +1]$ . The lower bound is attained when all communication involves individuals outside the community. The upper bound is attained when all communication remains within the community. The higher the IER, the more isolated the community. Negative IERs identify poor communities, in the sense that in such communities people talk more often to people outside the community rather than to people inside the community.

**Table 1. Community Detection Algorithms and Their Computational Complexity**

#	Authors	Computational Complexity
1	Palla et al. (2005)	$O(\exp(n))$
2	Newman and Girvan (2004)	$O(m^2n)$
3	Girvan and Newman (2002)	$O(mn^2)$
4	Furtunato et al. (2004)	$O(n^4)$
5	Bragow and Bollt (2004)	$O(n^3)$
6	Donetti and Munoz (2005)	$O(n^2)$
7	Zhou and Lipowsky (2005)	$O(n^2)$
8	Duch and Arenas (2005)	$O(n^2 \log(n))$
9	Radicchi et al. (2004)	$O(n^2)$
10	Newman (2003)	$O(n^2)$
11	Zhang et al. (2011)	$O(n_c^2 (n_c + n_b))$
12	Pons and Latapy (2006)	$O(n^2 \log(n))$
13	Clauset et al. (2004)	$O(n \log^2(n))$
14	Blondel et al. (2008)	$O(m)$
15	Rosvall and Bergstrom (2008)	$O(m)$

**Note 1:** Adapted from Fortunato (2010). **Note 2:**  $n$  denotes the number of nodes in the social network,  $m$  denotes the number of edges in the social network,  $n_c$  denotes the target number of nodes in the subpopulation of interest,  $n_b$  denotes the number of nodes in the boundary of the subpopulation of interest.

**Table 2. Summary of Results Obtained with Community Detection Algorithms**

Column Number	All Communities							Communities with Size between 25 and 200							IER > 0	
	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Community Size							Communities			Adopters per Community				Communities	
	Edge Weights	Execution Time	Communities	Min	Med	Max	Mean	All	With Adopters	All Adopters	Mean	Min	Med	Max	All	Adopters
Pons and Latapy (2006)	Yes	1 week	977,999	1	1	488,471	5.0	2,996	313	447	0.149	0	0	8	70	15
Blondel et al. (2008)	No	3 hours	8,417	2	2	886,764	592.4	47	3	8	0.179	0	0	6	46	8
Rosvall and Bergstrom (2008)	No	4 hours	8,113	2	2	1,564,858	614.6	24	2	2	0.018	0	0	1	20	2
Rosvall and Bergstrom (2008)	Yes	4 hours	8,081	2	2	2,490,201	617.6	14	0	10	0.000	0	0	0	10	0
Clauset et al. (2004)	Yes	1 month	25,373	2	2	317,426	179.3	232	17	37	0.159	0	0	3	21	0



**Table 3. Summary of Results Obtained Using the Modified T-CLAP Algorithms**

	All Communities							Communities with Size between 25 and 200							IER > 0	
Column Number	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Community Size							Communities			Adopters per Community				Communities	
	Edge Weights	Execution Time	Communities	Min	Med	Max	Mean	All	With Adopters	All Adopters	Mean	Min	Med	Max	All	Adopters
Zhang et al. (2011)	Yes	3 weeks	2,134	2	103	115	96.7	2,134	1,936	72,995	36.5	0	16	115	263	1,758

### A Modified Version of T-CLAP

We devise an alternative algorithm to extract communities from our large social network. This algorithm is inspired in T-CLAP (Zhang and Krackhardt 2010), which was previously used in the literature to identify small and tight communities. T-CLAP does not consider the entire social graph. Instead, it snowballs from a random seed and prunes nodes according to their IER. However, in our case, randomly sampling seeds is unlikely to ever hit an iPhone 3G adopter.<sup>10</sup>

Therefore, we change T-CLAP to (1) snowball starting from an iPhone 3G adopter, and (2) to discard non-adopters with higher probability. Appendix A shows the pseudocode for our modified version of T-CLAP. Note that in our version of this algorithm, if pruning a community results in two or more disjoint communities, then we proceed using the community with the highest IER. This helps us find communities with high IERs.

The computational complexity of T-CLAP is  $O(n_c^2(n_c + n_b))$ , where  $n_c$  denotes the number of nodes in the subpopulation of interest and  $n_b$  denotes the number of nodes in the boundary of this subpopulation. Our modified version of T-CLAP is simpler because it skips clustering. In this case it takes  $O(n_c)$  time to create the graph with the subpopulation of interest using snowball sampling. Computing IERs takes linear time on both  $n_c$  and the number of edges originating and reaching the subpopulation of interest, which we call  $m_c$ . Thus, our modified version of T-CLAP would take  $O(n(n_c + m_c))$  time to run in a network with  $n$  users. The majority of our code to

identify communities is developed in R, making use of iGraph (Csardi and Nepusz 2005).<sup>11</sup>

### Communities Identified Using the New T-CLAP

Table 3 shows the characteristics of the communities identified using our modified version of T-CLAP. We identify 2,134 communities comprising 72,995 adopters. More than 90 percent of these communities have adopters. We filter these communities and keep the ones that do not overlap, have at least one adopter, and positive IER. We are left with 263 communities comprising 1,758 adopters, which we use in the remainder of this paper. Figure 3 shows that there is no particular relationship between the size, the density, and the IER of these 263 communities.

Furthermore, these 263 communities are quite disconnected from each other. To see this, we select two communities at random and, also at random, we select one subscriber in each of these communities. We compute the length of the shortest path between these subscribers and repeat this process 500,000 times. Figure 4 shows the empirical distribution for the length of the shortest paths. The average length is 4.76 edges with a standard deviation of 0.66.

We also look at whether subscribers have friends in other communities. We perform a 1-wave snowball sample from all subscribers in these 263 communities and look for intersections. Figure 5 shows that more than 75 percent of the subscribers have no friends in other communities and 95 per-

<sup>10</sup>Recall that the adoption rate of the iPhone 3G was less than 1 percent in the country we analyze.

<sup>11</sup>The majority of our figures are generated using ggplot (Wickham 2009) and our tables are formatted using stargazer (Hlavac 2013).

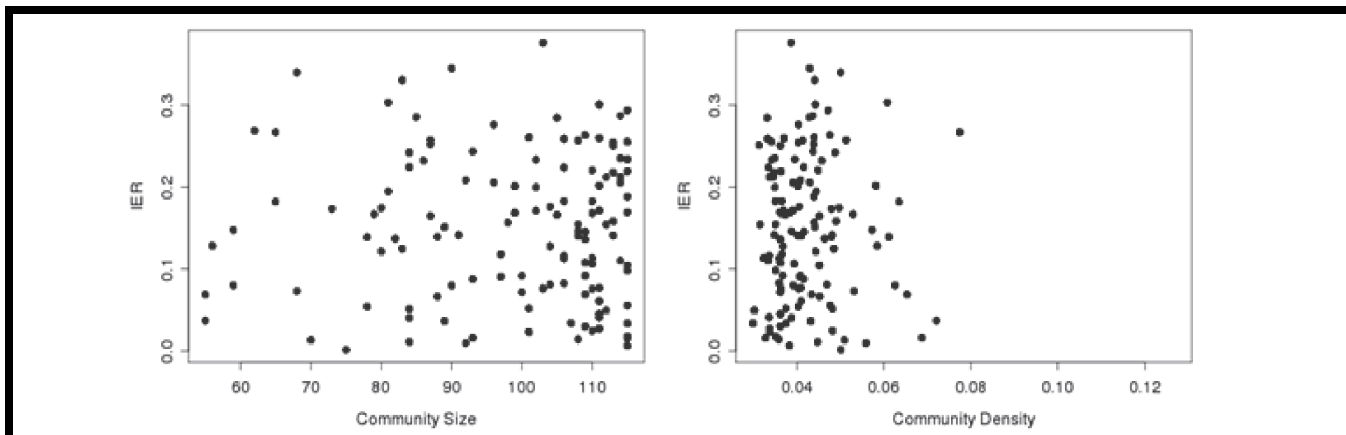


Figure 3. IER, Size, and Density for the Communities Extracted with the Modified T-CLAP

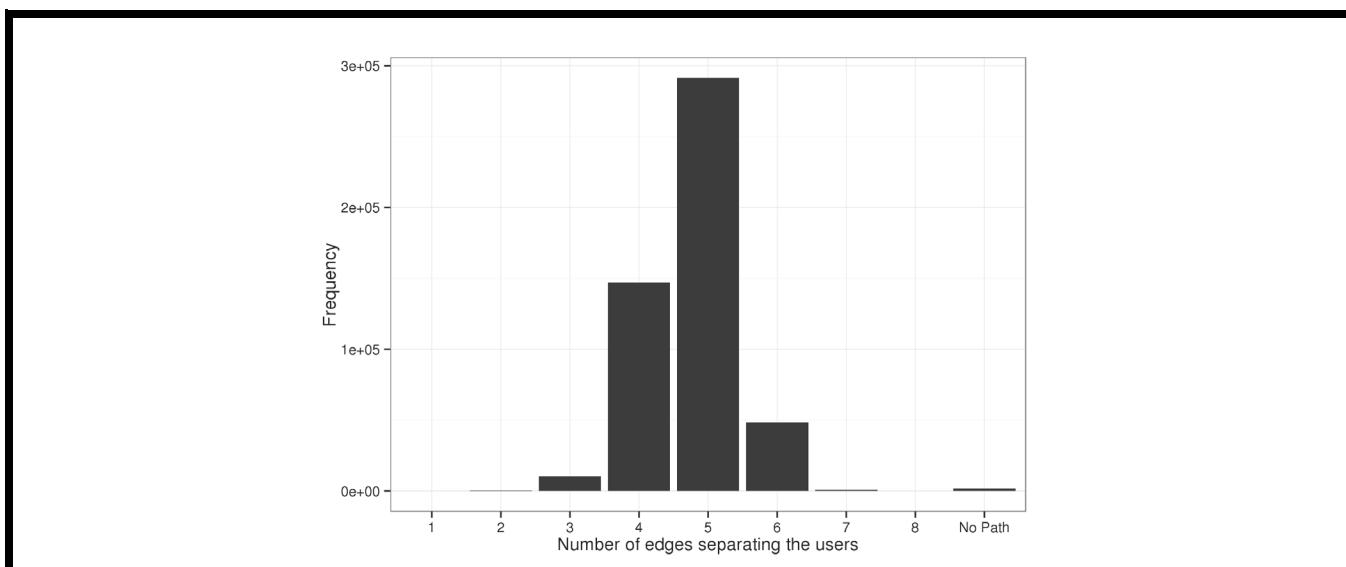


Figure 4. Length of the Shortest Path between Random Subscribers in Distinct Communities

cent have at most one friend in another community. On average, a community connects to 27.1 friends in other communities. The latter belong to 18.4 communities on average. Figure 6 shows the distribution of the average number of links that tie communities to each other. The overall average number of edges between two communities is 1.5 with a standard deviation of 0.6.<sup>12</sup> Taken together, these statistics show that these 263 communities are only sparsely connected.

<sup>12</sup>An undirected graph with 263 communities as nodes could have at most 34,453 edges. Only 7 percent of these edges are present in our case and 75 percent of them comprise only one edge between subscribers in the communities they connect.

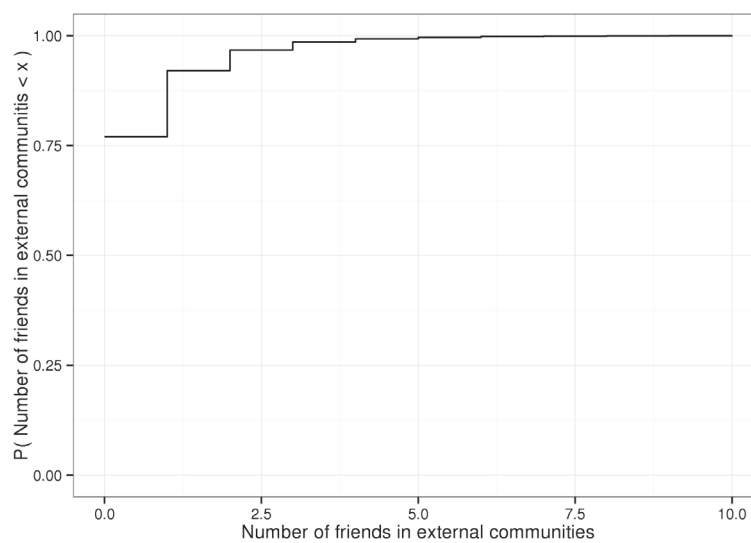
## Model and Empirical Strategy

### Consumer Utility Model

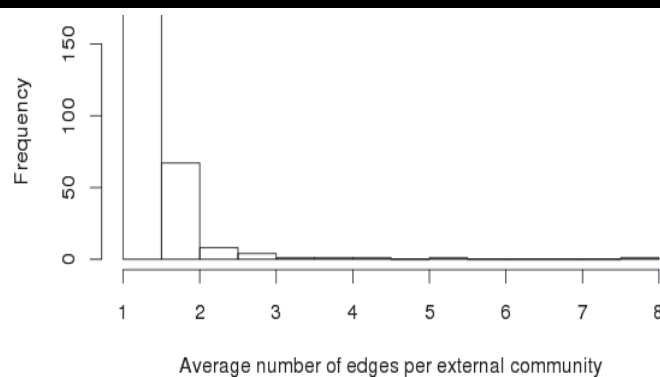
Let  $U_{it}$  represent the difference in utility for subscriber  $i$  between adopting and not adopting the iPhone 3G at time  $t$ . We assume that  $U_{it}$  is given by the following reduced form equation:

$$U_{it} = \alpha + X_i\beta + Z_{it}\gamma + \rho W_i Y_{t-1} + \varepsilon_{it} \quad (1)$$

$X_i$  and  $Z_{it}$  are time invariant and time variant subscriber-specific characteristics, respectively.  $W$  is an adjacency matrix. Entry  $w_{ij}$  is zero when subscribers  $i$  and  $j$  are not



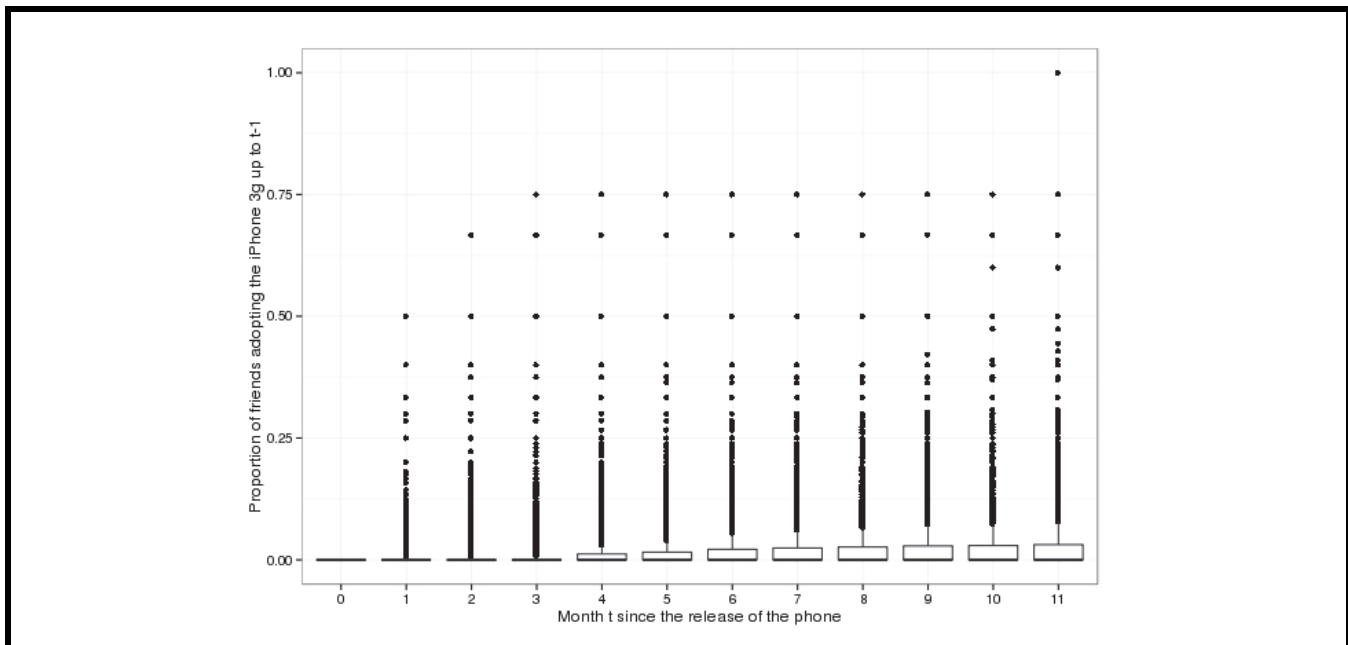
**Figure 5. Cumulative Distribution of the Number of Friends that Belong to Other Communities**



**Figure 6. Distribution of the Average Number of Links Tying Communities**

**Table 4. Descriptive Statistics of Time Invariant Covariates**

Variable	Description	Mean	Std. Dev.
<i>genderM</i>	1 for male	0.209	0.406
<i>genderF</i>	1 for female	0.185	0.388
<i>genderU</i>	1 for subscribers who did not report a gender	0.605	0.489
<i>prepaid</i>	1 if subscriber <i>i</i> had a prepaid tariff on June 30, 2008	0.606	0.489
<i>mobileNet</i>	1 if subscriber <i>i</i> had a mobile internet on June 30, 2008	0.023	0.150
<i>phone2.0g</i>	1 for ownership of a 2.0G handset on June 30, 2008	0.157	0.363
<i>phone2.5g</i>	1 for ownership of a 2.5G handset on June 30, 2008	0.478	0.499
<i>phone3.0g</i>	1 for ownership of a 3.0G handset on June 30, 2008	0.331	0.470
<i>phone3.5g</i>	1 for ownership of a 3.5G handset on June 30, 2008	0.028	0.166
<i>phoneOther</i>	1 for ownership of a handset with unknown range on June 30, 2008	0.004	0.066
<i>phoneAge</i>	Number of years a subscriber used the owned handset on June 30, 2008 (before the iPhone 3G release)	0.790	0.716



**Figure 7. Box Plot for the Proportion of Friend Adopters Over Time**

friends and the inverse of the degree of subscriber  $i$  otherwise.  $W_i$  is the  $i^{\text{th}}$  row of  $W$  and  $Y_{t-1}$  is a column vector that stacks the adoption decisions of all individuals up to time  $t-1$ .  $\varepsilon_{it}$  is the idiosyncratic error term. User  $i$  adopts the iPhone 3G at time  $t$  iff  $U_{it} > 0$ . Therefore,  $Y_{it} = 1\{U_{it} > 0\}$ . The term  $\alpha + X_{it}\beta + Z_{it}\gamma$  captures the intrinsic utility of subscriber  $i$  from adopting the iPhone 3G. Parameter  $\rho$  measures the effect of peer influence. If  $\rho > 0$ , adoption increases with the number of friends that adopt as well as with the number of friends of friends that adopt recursively, which in turn increases the propensity to adopt. In this case, this model originates the well-known s-shape diffusion curve as times goes by. If, however,  $\rho \leq 0$  and the intrinsic utility to adopt the iPhone 3G is small, few people will adopt and the handset will not spread largely across the population.

Table 4 shows descriptive statistics for the covariates used in our model. We control for gender, for the sophistication of the handset owned prior to the release of the iPhone 3G, for how long the subscriber owned that handset, for the type of tariff plan subscribed, and for how long the subscriber has been with EuroMobile (tenure).<sup>13</sup> Figure 7 details the

evolution of the proportion of friends that adopted the iPhone 3G over the period of analysis.

### Identification Strategy

Allison (1982) and Tucker (2008) discuss how the model in equation 1 can be empirically estimated with pooled probit as long as the standard errors are adjusted to account for correlation in the error term using either a sandwich estimator or block bootstrap.<sup>14</sup> However, unobserved heterogeneity may bias our estimates of peer influence. This happens, in particular, if  $E_i[W_i Y_{t-1}, \varepsilon_{it}] \neq 0$ ; that is, if unobserved reasons that lead subscriber  $i$  to adopt the iPhone 3G relate to the reasons why she befriends other subscribers. We refer to this condition as 1-hop homophily. Our approach to deal with this problem is twofold. First, we use many controls to reduce the potential relevant unobserved effects in  $\varepsilon_{it}$ . We control for heterogeneity across regions, across subscribers, and over time. Subscribers living in large cities are likely to face lower

<sup>13</sup>Gender has been shown to play an important role in how people perceive the benefit of a new technology (Gefen and Straub 1997; Venkatesh and Morris 2000; Venkatesh et al. 2003). Prior studies have also shown that experience with technologies similar to the new product increases the likelihood of adoption (Karahanna et al. 2006; Venkatesh et al. 2003). Therefore, in our case, we would expect subscribers with more sophisticated handsets to be more likely to adopt the iPhone 3G. We also control for how

long the subscriber has that handset to account, for example, for the subscriber's propensity to be an early adopter. We control for tariff plan because post-paid subscribers are typically wealthier. Finally, tenure might proxy how much the subscriber trusts EuroMobile in order to acquire a new and expensive handset from it.

<sup>14</sup>To do so, the data must be organized into a panel and observations after the first adoption need to be removed from the sample.

**Table 5. Descriptive Statistics for All Dummy Variables**

Variable	Description	Mean	Std. Dev.
<i>geoWageV L</i>	1 if the user spends most days in regions where salaries are 2 standard deviations below the national average	0.001	0.065
<i>geoWageL</i>	1 if the user spends most days in regions where salaries are between 1 and 2 standard deviations below the national average	0.009	0.095
<i>geoWageA</i>	1 if the user spends most days in regions within 1 standard deviation of the national average	0.312	0.463
<i>geoWageH</i>	1 if the user spends most days in regions where salaries are between 1 and 2 standard deviations above the national average	0.439	0.496
<i>geoWageVH</i>	1 if the user spends most days in regions 2 standard deviations above the national average	0.237	0.425
		<b>Subscriber/Dummy</b>	
		<b>N Dummies</b>	<b>Mean</b>
<i>Zio_Code_FE</i>	Dummy variables for home zip code	70	342.3
<i>Community_FE</i>	Dummy variables for community membership	263	91.7
<i>Month_FE</i>	Dummy variables for each month	12	23,064
			<b>Std. Dev</b>
			521.8
			17.2
			532.8

**Note 1:** Regions are studied at municipal level and daytime is 8 a.m. – 8 p.m.

search costs because the iPhone 3G was mostly available from EuroMobile franchises located in major shopping malls. Subscribers in these cities may also find it more beneficial to use the iPhone 3G because network coverage is likely to be better in urban areas to support Internet access. We use regional dummies to control for these effects. We account for where subscribers live, which we identify from the zip code of the account holder, and for where they spend most of their daytime, which we identify from the GPS coordinates of the cell towers used to route calls between 8 a.m. and 8 p.m. that we match to wage information obtained from the last Census. Yet, and as described by Manski (1993) and Shalizi and Thomas (2011), unobservable group effects may still lead to the adoption of the iPhone 3G and confound our estimates of peer influence.<sup>15</sup> We include community dummies to help control for these effects. These dummies capture some of the unobserved homophily reducing the potential for unobservables that drive both network formation and adoption of the iPhone 3G. Finally, we add time dummies to control for the fact that the net benefit from adopting the iPhone 3G changes over time with seasonal promotions and declining price trends. Table 5 summarizes the dummy variables mentioned earlier.

Second, we explore the structure of the social network to derive an instrumental variable at the subscriber level and

thus further alleviate endogeneity concerns. Consider two subscribers,  $i$  and  $j$ , who are friends. Consider a third subscriber,  $k$ , who is a friend of subscriber  $j$  but not of subscriber  $i$ . We argue that subscriber  $k$ 's decision to adopt the iPhone 3G relates to subscriber  $j$ 's decision to do so and to subscriber  $i$ 's decision to do so only through subscriber  $j$ 's decision. Subscriber  $k$ 's decision to adopt the iPhone 3G is correlated to subscriber  $j$ 's decision to do so. Subscriber  $j$  becomes exposed to the iPhone 3G through her friends, which reduces her uncertainty about the utility of the new handset and thus increases her likelihood to adopt it. Furthermore, subscriber  $k$ 's and subscriber  $j$ 's decisions to adopt are correlated due to 1-hop homophily, which has been widely reported to arise in social networks (McPherson et al. 2001). Finally, even if our communities fail to exhibit 1-hop homophily, our instrument is algebraically related to our dependent endogenous variable by construction.<sup>16</sup>

<sup>15</sup>One example is a set of subscribers that communicates frequently within the group because they work for the same company and the company adopts the iPhone 3G (or another handset, for that matter) as the primary handset for its employees.

<sup>16</sup>To see this, note that  $W^2$  codifies the friends of friends and  $(1 - W)W^2$  the friends of friends not friends of the ego. Therefore, both the instrument and the endogenous variable depend on  $W$ . According to Bramoulle et al. (2009), this type of instrument can provide identification as long as  $I$ ,  $W$ , and  $W^2$  are linearly independent, which is the case in our data.

**Table 6. Descriptive Statistics for the Instrumental Variable**

Variable	Description	Mean	Std. Dev.
$ffnf\_adopters_{t-1}$	Proportion of friends not friends of the ego who adopted the iPhone 3G	0.009	0.011
$ffnfdc\_adopters_{t-1}$	Proportion of friends not friends of the ego who live in a city different from where the ego who adopted the iPhone 3G does	0.010	0.009
$N\_ff$	Number of friends of friends	1333.279	1081.448
$N\_ffnf$	Number of friends of friends not friends	1311.574	1064.555
$N\_ffnfdc$	Number of friends of friends not friends of the ego living in a city different from where the ego does	566.580	586.7173

**Note 1:** Mean and standard deviation for the time varying instruments were calculated considering the entire panel. **Note 2:** Mean and standard deviation for time invariant covariates is calculated over the subscribers in the sample.

However, our instrument only works properly if, in addition, we assume that  $E_i[W_i^2 Y_{t-1}, \varepsilon_{it}] = 0$ . We refer to this condition as absence of 2-hop homophily. We note that 1-hop homophily does not necessarily entail 2-hop homophily.<sup>17</sup> Yet, 2-hop homophily may still be present in our data. We address this concern in several ways. We only use the decision to adopt the iPhone 3G of the friends of friends of the ego not friends of the ego. This provides additional separation between our instrument and our dependent variable, which helps reduce the potential correlation between subscriber  $i$ 's and subscriber  $k$ 's decisions to adopt the iPhone 3G to that arising only through subscriber  $j$ 's decision. We also introduce geographical separation between subscriber  $i$  and subscriber  $k$  to further lessen the likelihood of shared unobserved homophily. We used the GPS coordinates of the cell towers used to route calls to identify the cities where subscribers live. Appendix B provides additional details of how this has been accomplished. We used only the adoption decisions of friends of friends of the ego that are not friends of the ego and that live in a city different from where the ego does as instruments. Previous literature has shown that geographic proximity facilitates interactions (Strang and Soule 1998) and that the frequency of both face-to-face and electronic interactions reduce with geographic separation (Tillema et al. 2010). Clearly, using instruments that are not friends of the ego and that are geographically far apart from the latter lessens significantly the opportunities for correlated unobserved homophily. Combining social network data with geography is an innovative approach to identification unique to our paper that we could pursue given the richness of our dataset. In our section on robustness checks, we discuss other

strategies to find additional instruments and use them to provide additional results.

Table 6 provides descriptive statistics for our instrumental variable, which is computed before applying T-CLAP.

Finally, we note that our identification strategy with IV introduces limitations for inference purposes. In short, we need to restrict the population in which we measure the effect of peer influence because we have no evidence that this effect is the same across all subscribers in our communities. In our case, we can look at treated egos as subscribers  $i$  with friends, subscribers  $j$ , that adopted the iPhone 3G and we can think of compliers as subscribers  $j$  that do not adopt the iPhone 3G when their friends, subscribers  $k$ , do not and that adopt the iPhone 3G when their friends (subscribers  $k$ ) do. Refer to Imbens and Angrist (1994), Angrist et al. (1996), and Wooldridge and Imbens (2007) for more details on these concepts. With one-side non-compliance (Bloom 1984), that is, when subscribers  $j$  do not adopt the iPhone 3G if their friends (subscribers  $k$ ) do not, we can identify the effect of peer influence in the subpopulation of friends of compliers. Subscribers  $j$  and subscribers  $k$  do not adopt the iPhone 3G in 15 percent of our observations. Only in 0.14 percent of observations did we see subscribers  $j$  adopt the iPhone 3G when subscribers  $k$  did not. Therefore, we find strong evidence of one-side non-compliance in our case. Now, note that the subpopulation of friends of compliers, in which we can identify the effect of peer influence, is by definition the subpopulation receiving treatment. Therefore, in our case, the measure of the average treatment effect on the treated (ATT) and the average partial effect (APE), reported in the next sections, should be interpreted along these lines. We note that reporting treatment effects in a subpopulation of interest is not unique to our paper. For example, Aral and Walker (2011) report the ATT while Bapna and Umyarov (2014) study the ATE on the non-treated.

<sup>17</sup>If it did, one could possibly use the decision to adopt the iPhone 3G of the friends of friends of friends not friends of the ego as an instrument. One could possibly use this idea recursively to further reduce the potential bias in the effect of peer influence. Yet, too many hops between egos and instruments may result in weak instruments.



**Table 7. Marginal Effects for Peer Influence**

	(1) Probit $P(\text{adopted}_{it})$	(2) Probit $P(\text{adopted}_{it})$	(3) Probit $P(\text{adopted}_{it})$
$\text{frd\_adopters}_{t-1} (\rho)$	3.241*** (0.269)	3.247*** (0.256)	2.851*** (0.258)
Observations	265,462	265,462	265,462
Community FE	No	No	Yes
Zip Code FE	No	Yes	Yes
Month FE	Yes	Yes	Yes
Other Covariates	Yes	Yes	Yes
Pseudo R <sup>2</sup>	0.112	0.117	0.157
Log Lik	-9192	-9133	-8721
APE	0.053*** (0.005)	0.053*** (0.005)	0.045*** (0.004)

\*\*\*p &lt; 0.01, \*\*p &lt; 0.05, \*p &lt; 0.1

**Note 1:** Cluster robust standard errors in parenthesis.

## A Results for Peer Influence

### Probit Results

Table 7 shows the results obtained without instruments.<sup>18</sup> We only show results for the covariate of interest. All other covariates behave as expected, as shown in Appendix D. All specifications in this table show a positive correlation between friends' adoption and the ego's adoption, as expected. The Average Partial Effect (APE) varies between 4.5 percent and 5.3 percent.<sup>19</sup> The magnitude of the coefficient reduces from Column (2) to Column (3) when we include community dummies. This provides evidence that using communities is potentially capturing some homophily that could otherwise be interpreted as influence. As such, this result provides support in favor of using communities to measure peer influence in social networks.

### IV Probit Results

Table 8 shows the results obtained after instrumentation.<sup>20</sup> The results in this table provide clear evidence of contagious adoption. If all friends of the ego adopt the iPhone 3G then the ego's probability to adopt increases by 15 percent. Figure 8 shows that this level of peer influence must have accounted for approximately 14 percent of all iPhone 3G adoption during our period of analysis. Table E1 in Appendix E shows the pseudocode of the algorithm used to compute this statistic. We note that the marginal effect of peer influence increases significantly after instrumentation. Budget constraints might explain this increase. People in the same household tend to call each other often and therefore will appear as friends in our social graph.

However, budget constraints might preclude all family members from adopting the new and expensive iPhone 3G. Family members are friends that will not adopt the iPhone 3G even if they want to. This artifact may downplay the estimates of peer influence before instrumentation because it leads the error term to be negatively correlated to our endogenous variable.

<sup>18</sup>We use 265,462 observations over 258 communities out of the 263 identified previously. Appendix C discusses why five communities were dropped from the analysis.

<sup>19</sup>For the probit estimator, APE is given by the partial derivative of the expected value function,  $E[\Phi(\alpha + Z\delta)]$ . The standard errors for the APE are computed using the Delta method as suggested in Wooldridge (2002).

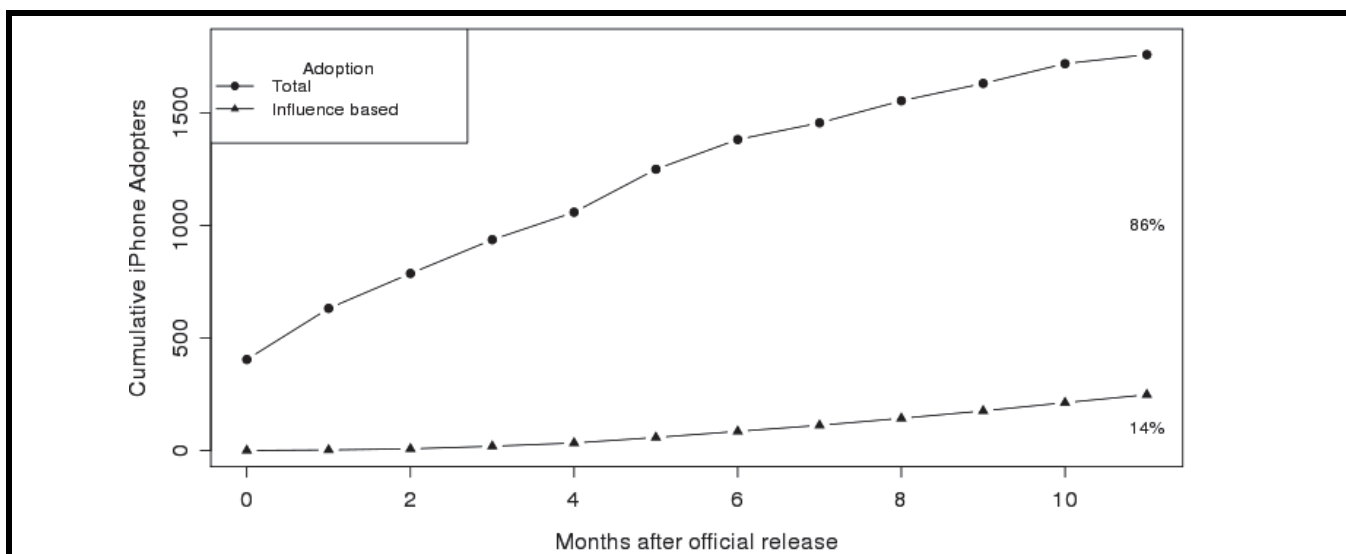
<sup>20</sup>In the case of IV probit, and following Wooldridge (2002, p. 475), the APE is obtained from the partial derivative of  $E_p[\alpha_p + Z\delta_p + \theta_p v]$ . Subscript p indicates that IV probit estimates parameters up to a scale factor. Subscript v denotes the residuals of the first stage regression. For more details on the Newey IV estimator and how to compute APE refer to Wooldridge and Imbens (2007), which also describes in detail how to average out effects over time.

**Table 8. Marginal Effects for Peer Influence with IV**

Instrument Variables	(1) IV Probit $\text{ffnf adopted}_{it}$	(2) IV Probit $\text{ffnfdc adopted}_{it}$
$\text{frd\_adopters}_{t-1} (\rho)$	7.864*** (1.171) [1.185]	9.935*** (1.952) [2.087]
Observations	265,462	265,462
Community FE	Yes	Yes
Zip Code FE	Yes	Yes
Month FE	Yes	Yes
Other Covariates	Yes	Yes
$\chi^2$ p-value	0.000	0.000
APE	0.120*** [0.018]	0.150*** [0.032]

\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1

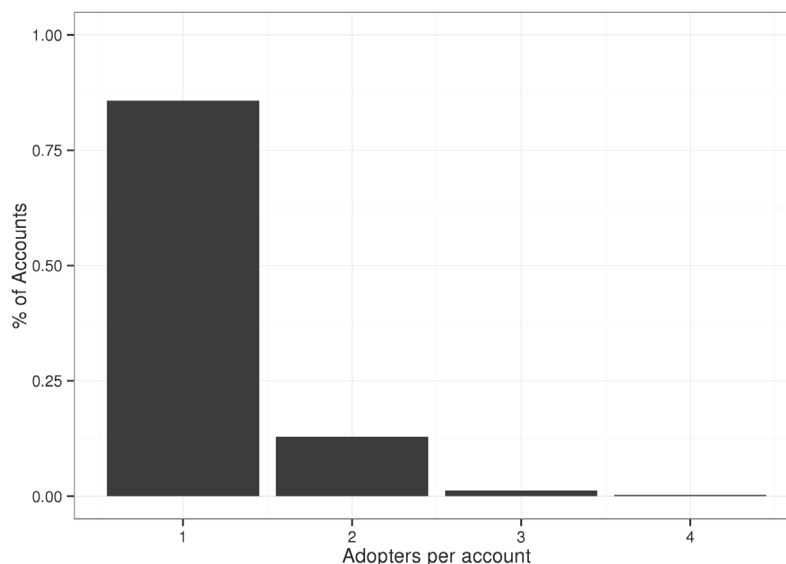
**Note 1:** Newey estimator standard errors in ( ) for IV probit. **Note 2:** Community block-bootstrap standard errors in [ ] for IV probit (200 replications are used to calculate the significance level).

**Figure 8. The Role of Peer Influence in the Adoption of the iPhone 3G**

To better understand the potential effect of budget constraints, we assume that all phone numbers in the same billing account belong to the same person, to people in the same family, or to employees of the same firm.<sup>21</sup>

<sup>21</sup>Nothing prevents unrelated subscribers from sharing a common phone bill, but these cases are likely exceptions.

On average, more than 20 percent of one's friends belong to the same account and therefore intra-account communication is significant. Figure 9 plots the number of iPhone 3G adopters in accounts with more than one and less than six phone numbers, most likely associated with families. Most accounts have only one adopter, which is consistent with the hypothesis of budget constraints. A similar plot is obtained for accounts with more than five phone numbers, most likely associated with firms. This reinforces the hypothesis of budget get constraints and shows that only a few firms, if at all, adopted the iPhone 3G.



**Figure 9. Adopters per Account for Accounts with Two to Five People**

## Robustness Checks

### Functional Form

Instrumental variables with linear probability models (LPM) rely on less restrictive assumptions on the distribution of the error term than nonlinear formulations. According to Allison (1982) and Wooldridge (2002), LPM estimates are often good approximations to the average partial effect in the population. Therefore, we estimate LPM versions of the model in equation (1).

Table 9 shows the results obtained. Both the OLS result, in Column (1), and the 2SLS results, in Columns (2) and (3), remain positive and highly statistically significant. Furthermore, they are not statistically different from their probit counterparts. Therefore, our results for the effect of peer influence do not seem to be an artifact of functional form.

### Effect of the Period of Analysis

The iPhone 3G is a big ticket item, in particular in the market studied in our paper, which can introduce friction into the process of peer influence. Therefore, a month-long window seems appropriate to analyze the effect of peer influence. In any case, this might bias our estimates downward, for example, by increasing the potential for contemporaneous peer influence, for which we may fail to account. We reproduce our results with weekly binning to check whether our

choice of period of analysis can be driving our results. Table 10 shows the results obtained for OLS, 2SLS, probit, and IV probit using weekly binning. These effects are four times smaller than the effects obtained with monthly binning, which suggests that our choice of period of analysis does not seem to drive our findings.

### Alternative Weighting Matrices

In the preceding sections,  $W$  was a row standardized version of the adjacency matrix coding the social graph giving the same weight to every friend. However, it is likely that subscribers weight differently the information they obtain about the iPhone 3G from their friends. Below, we introduce five new ways to weight friends and show that our estimates of peer influence remain unchanged. Table 11 shows the results obtained using the friends of friends not friends of the ego who live in a city different from where the ego does as our instrumental variable. We also center on the mean and standardize the proportion of friends that adopted the iPhone 3G to facilitate the comparison of the magnitude of the effect of peer influence across columns. Column (1) displays the results obtained before. Column (2) weights friends according to the number of friends they have. This specification captures the fact that subscribers may place more weight on friends that are more connected, which therefore may convey more information about the new handset. Column (3) weights friends according to air time over the entire period of analysis, which captures the fact that subscribers may give more weight

**Table 9. Coefficient Estimates for the Linear Probability Model**

Instruments Variables	(1) OLS <i>adopted<sub>it</sub></i>	(2) 2SLS <i>ffnfdc adopted<sub>it</sub></i>	(3) 2SLS <i>ffnfdc adopted<sub>it</sub></i>
<i>frd_adopters<sub>t-1</sub></i> ( $\rho$ )	0.058*** (0.007)	0.094*** (0.018)	0.128*** (0.029)
Observations	265,462	265,462	265,462
Community FE	Yes	Yes	Yes
Zip Code FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Other Covariates	Yes	Yes	Yes
R <sup>2</sup>	0.015	0.015	0.015

\*\*\*p &lt; 0.01, \*\*p &lt; 0.05, \*p &lt; 0.1

**Note 1:** Community cluster robust standard errors in parentheses.**Table 10. Coefficient Estimates for IV Probit and the Linear Probability Model Using Weekly Binning**

Instrument Variables	(1) OLS <i>adopted<sub>it</sub></i>	(2) 2SLS <i>ffnfdc_adopted<sub>it</sub></i>	(3) Probit <i>adopted<sub>it</sub></i>	(4) IV Probit <i>ffnfdc_adopted<sub>it</sub></i>
<i>frd_adopters<sub>t-1</sub></i> ( $\rho$ )	0.017*** (0.001)	0.041*** (0.007)	2.444*** (0.149)	8.526*** (1.848)
Observations	1,169,364	1,169,364	1,169,364	1,169,364
Community FE	Yes	Yes	Yes	Yes
Zip Code FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Other Covariates	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.004	0.004		
APE			0.011*** (0.002)	0.039*** (0.007)

\*\*\*p &lt; 0.01, \*\*p &lt; 0.05, \*p &lt; 0.1

**Note 1:** Community cluster robust standard errors in parentheses.**Table 11. Peer Effects with Different Weighting Matrices**

	(1) IV Probit <i>adopted<sub>it</sub></i>	(2) IV Probit <i>adopted<sub>it</sub></i>	(3) IV Probit <i>adopted<sub>it</sub></i>	(4) IV Probit <i>adopted<sub>it</sub></i>	(5) IV Probit <i>adopted<sub>it</sub></i>	(6) IV Probit <i>adopted<sub>it</sub></i>
<i>frd_adopters<sub>t-1</sub></i> weighed by...		Degree	Calls + SMS	Calls	SMS	Airtime
<i>std_frd_adopters<sub>t-1</sub></i>	0.327*** (0.0643)	0.358*** (0.0660)	0.814*** (0.144)	0.798*** (0.141)	0.756*** (0.134)	0.747*** (0.133)
Observations	265,462	265,462	265,462	265,462	265,462	265,462
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Zip Code FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Covariates	Yes	Yes	Yes	Yes	Yes	Yes

\*\*\*p &lt; 0.01, \*\*p &lt; 0.05, \*p &lt; 0.1

**Note 1:** Network tenure, previous handset age, as well as dummies for gender, previous handset technology, regional wage level, type of contract (pre/post paid), and subscription to mobile internet data plans prior to the release of the iPhone 3G included in all regressions. *std\_frd\_adopters<sub>t-1</sub>* is the same as variable *frd\_adopters<sub>t-1</sub>*, but mean centered and standardized to facilitate comparison across regressions. The instrument for *std\_frd\_adopters<sub>t-1</sub>* is the mean centered and standardized proportion of i's friends of friends that are not simultaneously i's friends and that spend most of their time in regions with different NUTS-III code than i himself. **Note 2:** Newey estimator standard errors in parentheses.

**Table 12. First Stage Regression for IV Analysis**

Instrumental Variables	(1) OLS wfwfnf <i>frd_adopters</i> <sub><i>t-1</i></sub>	(2) OLS ffnf 1path <i>frd_adopters</i> <sub><i>t-1</i></sub>	(3) OLS wfwfnf 1path <i>frd_adopters</i> <sub><i>t-1</i></sub>
<i>X_adopters</i> <sub><i>t-1</i></sub>	0.184*** (0.028)	1.067*** (0.0672)	0.183*** (0.028)
Observations	265,462	265,462	265,462
Community FE	Yes	Yes	Yes
Zip Code FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Other Covariates	Yes	Yes	Yes
Underidentification and Weak Instrument Tests			
Kleibergen-Paap rK LM $\chi^2$ p-value	0.000	0.000	0.000
Kleibergen-Paap rK Wald F-stat	42.927	252.323	42.362

\*\*\*p &lt; 0.01, \*\*p &lt; 0.05, \*p &lt; 0.1

**Note 1:** Community cluster robust standard errors in parentheses.

to friends with whom they communicate more often. Finally, Columns (4), (5), and (6) weight friends according to the number of calls and sms exchanged over the entire period of analysis. These specifications capture the fact that subscribers may give more weight to friends with whom they interact more often. The effect of peer influence remains positive and highly statistically significant in all specifications. Therefore, the effect of peer influence is robust to how subscribers might weight their friends' decisions.

### Alternative Instrumental Variables

We test the sensitivity of our results to the definition of our instrumental variable by further exploring the structure of the social graph. The first alternative instrument we consider is the friends of the friends of the ego that are not her friends and that connect to her only through a single path in the social graph (ffnf\_1path). Formally, when subscriber *i* is the ego, subscriber *j* is a friend of subscriber *k* and subscriber *k* is a friend of subscriber *j* not a friend of subscriber *k*, subscriber *k*'s decision to adopt the iPhone 3G is used as an instrument for subscriber *j*'s decision to adopt only if there is no subscriber *j'* ≠ *j* that is a friend of both subscriber *k* and subscriber *i*, nor does there exists a subscriber *j''* that is a friend of both subscriber *k* and a friend of a friend of subscriber *i* recursively. Using only one path between subscriber *i* and subscriber *k* introduces additional separation between them, which further reduces the potential for unobserved 2-hop homophily between egos and instruments. The only connection between subscriber *i* and subscriber *k* is

through subscriber *j*, whose decision to adopt the iPhone 3G is our endogenous variable.

The second alternative instrument we use is the decision to adopt the iPhone 3G of the weak friends of the weak friends of the ego that are not her friends (wfwfnf). Using only paths in the social graph through weak friends introduces additional social separation between the instrument and the ego. Subscriber *j* is a weak friend of subscriber *i* if she is in the bottom quartile of the distribution of the number of calls and sms that subscriber *i* exchanges with her friends throughout the whole period of analysis. Using only weak friends helps further reduce the potential for unobserved 2-hop homophily between subscriber *i* and subscriber *k*. Table 12 shows the first stage regressions and the Stock and Yogo tests for weak instruments. Despite the separation that these new instruments introduce, they are not weak instruments. Table 13 shows the second stage results. Column (1) uses the former alternative, Column (2) uses the latter alternative, and Column (3) uses both of them combined. It is clear that the effect of peer influence remains positive and highly statistically significant with similar magnitude across models.

These results provide additional evidence of the robustness of our identification approach. In particular, the alternative instruments used in this section rely on mechanisms other than geographical distance to introduce separation between instruments and the ego. Yet, all our results agree in terms of statistical significance and magnitude of the observed effect of peer influence.

**Table 13. Peer Effects with Alternative Instrumental Variables**

Instrumental Variables	(1) IV Probit wfwfnf adopted <sub>it</sub>	(2) IV Probit ffnf_1path adopted <sub>it</sub>	(3) IV Probit wfwfnf_1path adopted <sub>it</sub>
<i>frd_adopters<sub>t-1</sub></i>	10.347*** (3.183) [2.761]	9.037*** (1.350) [1.228]	10.200*** (3.211) [3.116]
Observations	265,462	265,462	265,462
Community FE	Yes	Yes	Yes
Zip Code FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Other Covariates	Yes	Yes	Yes
$\chi^2$ p-value	0.000	0.000	0.000
APE	0.160*** [0.041]	0.142*** [0.018]	0.154*** [0.046]

\*\*\*p &lt; 0.01, \*\*p &lt; 0.05, \*p &lt; 0.1

**Note 1:** Community clusters robust standard errors in ( ) for probit. **Note 2:** Newey estimator standard errors in ( ) for IV probit. **Note 3:** Community block-bootstrap standard errors in [] for IV probit (200 replications) are used to calculate the significance level.

### A Time-Varying Social Network

A central hypothesis of our analysis so far is that social ties inferred from the CDR data are stable over the whole period of analysis. We use stochastic actor-based models for the co-evolution of network dynamics and behavior (SAMCNDB) to relax this assumption. SAMCNDBs are used to model the dynamics of link formation across the members of a social network and the relationship between ties and behavior. SIENA (Snijders et al. 2010) provides an implementation of such a model. It requires at least two snapshots of the social network and of the behavior of interest. Actors play a game that establishes a path from one snapshot to the next. At each stage, an actor is selected to take action. The time between actions from the same actor is modeled as a Poisson process with a constant rate,  $\lambda$ , common to all actors.

Two types of actions can occur:

- (1) Actors can create or eliminate a tie in the social network. They do so to maximize the utility given by a function of the form  $f_k(B, x) = \sum_k B_k S_k(x)$ , which is the sum over all effects  $k$  of the utility associated with each effect,  $S_k(x)$ , weighted by a scaling factor,  $B_k$ . The utility of an effect depends on the state of the network,  $x$ .
- (2) Actors can change their behavior. In our case, they can choose to adopt the iPhone 3G. They do so to maximize the utility given by a function of the form  $f_i^Z(B^Z, x, z) = \sum_{\tilde{k}} B_{\tilde{k}}^Z S_{\tilde{k}}(x, z)$ , which is the sum over

all effects  $\tilde{k}$  of the utility associated with each effect,  $S_{\tilde{k}}^Z(x, z)$ . In this case, utility depends on the state of the network and the behavior of all actors,  $z$ .

Functions  $S$  and  $S^Z$  represent effects that, according to the theory of social networks, influence the formation of ties and behavior, respectively, such as transitivity and homophily. Borrowing from the econometric theory of random utility models, SIENA introduces a separable unobservable in the utility of each actor. This error term follows a type I extreme value distribution with mean 0 and a scale parameter normalized to 1. Consequently, the probability with which an actor chooses a particular action follows a multinomial logit function.  $B_k$ ,  $B_k^Z$ , and  $\lambda$  parameterize the probability distribution of a Markov process that captures the dynamics of this game. The likelihood function of this process is often impossible to determine analytically (Snijders, 1996). For this reason, SIENA estimates parameters using a method of moments, whose solutions are approximated using Monte Carlo simulation.

In SIENA, a snapshot of the form  $[t_1, t_2]$  accumulates the effect of all actions by all actors between the first day of  $t_1$  and the last day of  $t_2$ .<sup>22</sup> We partitioned our data into two snapshots of the form  $[1, t] - [t + 1, 11]$ . Appendix F shows

<sup>22</sup>SIENA requires changes in both network and behavior between consecutive snapshots to embody enough variance allowing for estimating the parameters of interest. Still, these changes cannot be exaggeratedly large in order to challenge the assumption that they pertain to a network that evolves sufficiently smoothly over time. For this reason, it is fundamental to carefully prepare our data for a SIENA analysis.



results for all  $t = 2, 3, \dots, 9$ , which are all aligned with the ones presented below for  $t = 4$ , which is the partition for which SIENA converges in most communities. This partition provides a good balance between the two non-overlapping periods: 1,058 subscribers adopt the iPhone 3G in [1, 4] and 700 do so in [5, 11]. Furthermore, the Jaccard Index for the networks at  $t = 4$  and  $t = 11$  is 0.816 (with a standard error of 0.050), which provides evidence that a number of social ties are stable from the first period to the second and thus likely to entail real social meaning.<sup>23</sup>

An adjacency matrix,  $A^t$  is computed for each and every community,  $c$ , at each and every time period,  $t$ .  $a_{ij}^t$  indicates whether subscriber  $i$  called or sent an sms to subscriber  $j$  during time period  $t$ . Vector  $y^t$  stacks the behavior of interest.  $y_i^t$  indicates whether subscriber  $i$  owned the iPhone 3G during time period  $t$ . We use the following covariates to explain the evolution of the social network: (1) out-degree,  $\sum_j a_{ij}^t$ , measures the overall tendency to form ties that are costly to establish and maintain; (2) reciprocity,  $\sum_j a_{ij} a_{ji}$ , controls for the fact that people tend to return calls and sms from their friends; (3) transitivity,  $\sum_j a_{ij} \max_h(a_{ih}, a_{hj})$ , controls for triad closure bias—people tend to communicate more with the friends of their friends; (4) adoption similarity,  $\sum_j a_{ij} (sim_{ij} - \bar{sim}_i)$  with  $sim_{ij} = 1 \{y_i = y_j\}$ , which measures the subscribers' tendency to befriend subscribers with similar behavior and thus captures homophily (Snijders et al. 2010). Finally, we use  $\sum_j a_{ij} sim_{ij} / \sum_j a_{ij}$  to capture the effect of exposure to the new handset through friends, which measures the effect of peer influence.

We use meta-analysis to combine the effects obtained from running SIENA in each community (DerSimonian and Laird 1986; Hedges and Olkin 1985; Viechtbauer 2010). A summary effect for each parameter is obtained using a random-effects model to allow parameters to vary across communities (DerSimonian and Laird 1986). We use a restricted maximum likelihood estimator and the standard error bias correction suggested by Paule and Mandel (1982), Rukhin et al. (2000), and Knapp and Hartung (2003). Table 14 provides the results obtained. The effect of peer influence is positive and highly statistically significant. Therefore, we find strong evidence of peer influence even in the presence of a social network that changes over time. The summary effect of homophily is positive but not statistically significant. Therefore, we do not find evidence that homophily played a significant role in the structure of the diffusion network of the iPhone 3G, that is,

<sup>23</sup>Still, in 30 communities out of the 263 identified earlier, there is no variation in behavior from the first to the second period. In the 80 other communities, the variation in behavior is too little for SIENA to converge. We are left with 153 communities, which we analyze.

subscribers do not seem to have become friends because they bought the new handset. Figure 10 shows the parameters and the standard errors for influence and homophily in panels (a) and (b), respectively, for each community analyzed. The effect of peer influence is positive in all communities and statistically different from zero in most of them. The effect of homophily is negative and statistically significant in some communities, positive and statistically significant in other communities but close to zero and not statistically significant in most of them.<sup>24</sup>

Appendix G shows the results for the leave-one-out test (Viechtbauer 2010) and the cumulative inclusion test (Lau et al. 1992) for the previous results. The first test consists of re-estimating the meta-analysis coefficient repeatedly including one community at time in the analysis. The test is sensitive to the order in which communities are included. Therefore, we report two limiting cases, one in which communities are included in decreasing order of the standard error associated with the estimate of the effect of peer influence and another one in which communities are included in increasing order of the standard error associated with the estimate of the effect of peer influence. In both cases the results converge quickly to the effect of peer influence reported before. The second test shows that this result is not sensitive to removing any one community from the analysis.

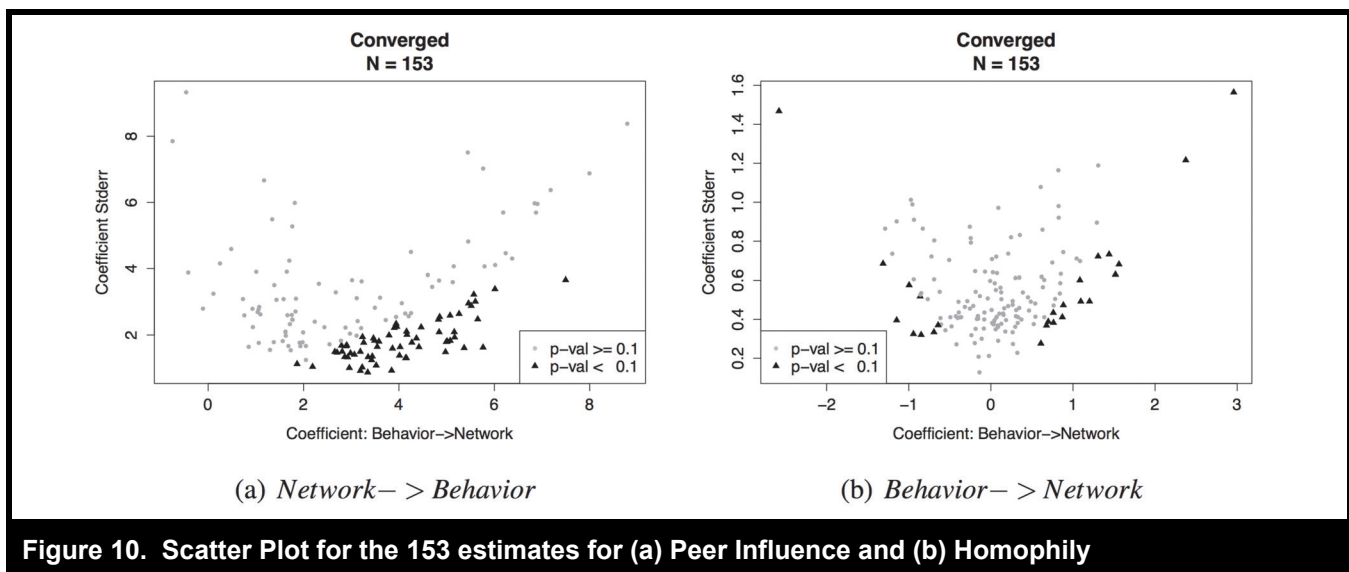
Interpreting the economic meaning of parameter estimates in SIENA is far from straightforward (Ripley and Snijders 2010). Still, the log-odds ratio of adoption versus non-adoption for the effect of peer influence is given by  $\beta(A - \bar{A})/(A + \bar{A})$ , where  $A$  and  $\bar{A}$  denote the number of friends that adopted and did not adopt the iPhone 3G, respectively, and  $\beta$  denotes the coefficient obtained. For a subscriber whose friends all adopted the iPhone 3G, the odds ratio of adoption versus no adoption is given by  $\exp(\beta)$ , which in our case is

<sup>24</sup>We perform a Q-test of heterogeneity (Huedo-Medina et al. 2006) for each parameter in our model. The null hypothesis for this test states that the true value of a parameter is the same across communities. If this hypothesis is rejected, then there is evidence of unobserved heterogeneity that explains differences in parameter estimates across communities that cannot be attributed to sampling error. If this is not the case, then the summary effect can be interpreted as the true value for the parameter in the population. We also report the  $I^2$  index, which measures the percentage of the total variation in the effect that is explained by unobserved heterogeneity across communities. The Q-test cannot reject the null hypothesis for the effect of peer influence and the associated  $I^2$  is essentially zero. The Q-test shows evidence of unexplained heterogeneity in homophily across communities. Finally, the Q-test rejects the null hypothesis and the  $I^2$  is large for out-degree, reciprocity, and transitivity and thus there are unobservables driving the variation of these three parameters across communities. Nevertheless, the coefficients obtained exhibit the expected signs and are statistically significant.

**Table 14. Results of the Meta-Analysis**

Variable	Coeff	Stderr	pval	I <sup>2</sup>	H <sup>2</sup>	$\tau^2$	Q-Test	Q-Test (pval)	N Obs
outdegree (density)	-4.414	0.043	0.000	62.851	2.692	0.182	543.708	0.000	153
reciprocity	4.398	0.067	0.000	75.399	4.065	0.448	782.592	0.000	153
transitive ties	2.174	0.034	0.000	51.141	2.047	0.086	345.009	0.000	153
<i>Behavior</i> → <i>Network</i>	0.043	0.044	0.334	24.219	1.320	0.061	208.778	0.002	153
<i>Network</i> → <i>Behavior</i>	3.168	0.102	0.000	0.000	1.000	0.000	62.002	1.000	153

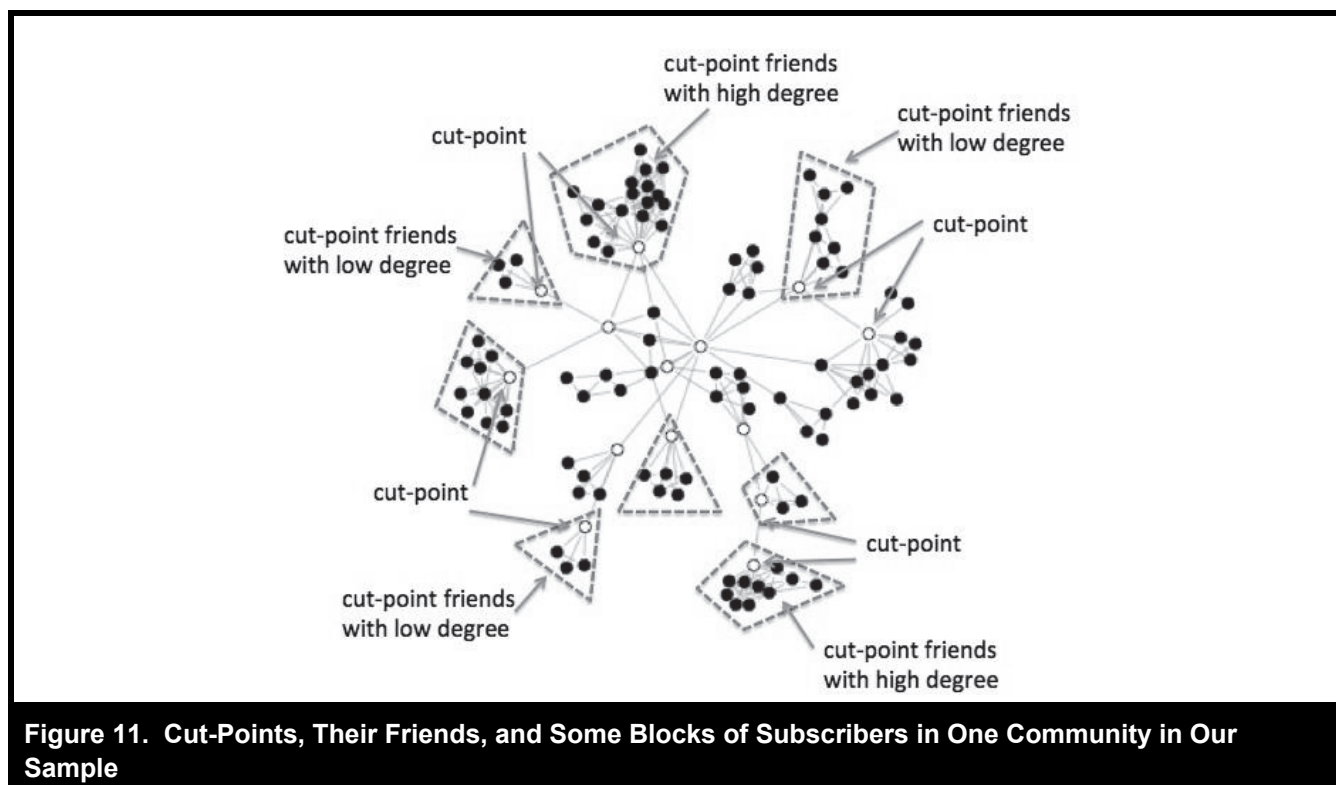
**Note 1:** Variable *Behavior* → *Network* is captured by the behavior similarity SIENA effect. **Note 2:** Variable *Network* → *Behavior* is implemented through the behavior average similarity SIENA effect. **Note 3:** Meta analysis estimated through maximum likelihood assuming a random effects model with Knapp and Hartung standard error correction. **Note 4:**  $\tau^2$  is the estimate of the total amount of heterogeneity,  $I^2$  is the percentage of total variability due to heterogeneity,  $H^2$  is  $\frac{\text{total variability}}{\text{sampling variability}}$ .

**Figure 10. Scatter Plot for the 153 estimates for (a) Peer Influence and (b) Homophily**

$\exp(3.168) = 23.67$ . The standard error for this effect is 0.102 and therefore a 95 percent confidence interval in this case is [19.4, 29.1]. Let  $p_0$  represent the probability of adoption of a subscriber without friends that have adopted the iPhone 3G. Let  $p_1$  represent the additional probability of adoption of a subscriber whose friends all adopted the iPhone 3G. According to the results the “IV Probit Results” section presented earlier,  $p_0 = 0.006$  and  $p_1 = 0.150$ . Therefore, the odds ratio in this case is given by  $[(p_0 + p_1)/(1 - (p_0 + p_1))]/[p_0/(1 - p_0)] = 30.6$ . The standard error for the estimate of  $p_1$  is 0.032, therefore a 95 percent confidence interval for this odds ratio is [16.9, 44.3]. This analysis shows that the effect of peer influence obtained with SIENA is similar to that obtained the “IV Probit Results” section using instrumental variables, which increases our confidence in our results.

### The Effect of Cut-Points

We investigate whether variance in the structural properties of our communities relates to the heterogeneity in adoption and peer influence across our communities. Table H1 in Appendix H shows the descriptive statistics for a number of well-known structural measures and Table H2 shows that a number of them are highly correlated. The most striking relationship between these measures and adoption arises for cut-points. Cut-points are small world bridges that connect blocks of subscribers that would otherwise be separated. Figure 11 highlights the role of cut-points in one of our communities.



**Figure 11. Cut-Points, Their Friends, and Some Blocks of Subscribers in One Community in Our Sample**

Cut-points serve as pathways for information to flow to these blocks of subscribers that have no other connection to the “core” of the community. Communities with many cut-points may be more fragile but also less redundant. Dense networks with only a few cut-points may trigger inertia and be more likely to exhibit the status quo, which, in our case, is little adoption of the iPhone 3G.

Table 15 shows the results obtained from regressing adoption on a subset of the measures identified in Table H1. Columns (1), (2), and (3) in this table show that communities with more cut-points have higher (monthly) adoption rates, more months with adoption, and adoption over longer periods of time, respectively. The fact that more cut-points is associated with more adoption over a longer period of time shows the potentially important role of cut-points in spreading information about the iPhone 3G in these communities. These results are obtained after controlling for density, average path length, and average betweenness. Together, these covariates measure how big a community is and how connected its members are. These covariates exhibit the expected signs in these regressions but only average path length is statistically significant in the last specification.

We also focus on the relationship between cut-points and peer influence. Cut-points can exert more influence over their

friends when the degree of the latter is low. Otherwise, cut-points are just one friend among many. Therefore, we look at whether the effect of peer influence is different in communities with many cut-points whose friends have low degree. Column (4) in Table 15 shows the results obtained using the coefficient from the SIENA analysis as a measure of peer influence. Communities with many cut-points, namely above the median, exhibit more peer influence. However, this effect is attenuated by the degree of the friends of the cut-points. Therefore, we find suggestive evidence that cut-points play an important role in adoption and shape peer influence when their friends have few friends.

## Policy Experiments

We test whether EuroMobile could have increased its profit by seeding the iPhone 3G to appropriate subscribers. We code the simulator described in Table I1 in the Appendix I for this purpose. This simulator computes the expected number of additional adopters (AA) as a function of the seeding strategy ( $k$ ), the number of seeds ( $n$ ), an exogenous marginal effect for peer influence ( $m$ ) and a model for how consumers

**Table 15. Regression Results**

	log(adoption_rate) (1)	months_with_adoption (2)	adoption_span (3)	peer-influence <sup>1</sup> (4)
Density	1.560 (5.806)	-5.168 (19.627)	22.947 (21.188)	22.266 (17.293)
Average path length	-0.120 (0.119)	-0.534 (0.401)	-0.759* (0.433)	-0.869* (0.477)
Average betweenness	0.001 (0.002)	0.006 (0.006)	0.004 (0.007)	0.019** (0.007)
Number of cut-points	0.037*** (0.014)	0.078* (0.047)	0.120** (0.050)	
Average degree of cut-points friends				-0.042 (0.288)
High number of cut-points <sup>2</sup>				4.730** (2.320)
High number of cut-points * Average degree of cut-points friends				-0.955** (0.446)
Constant	-0.897* (0.475)	5.679*** (1.607)	7.87*** (1.735)	3.233 (2.178)
Observations	153	153	153	153
R <sup>2</sup>	0.096	0.061	0.047	0.078
Adjusted R <sup>2</sup>	0.071	0.036	0.022	0.040
Residual Standard Error	0.567 (df = 148)	1.915 (df = 148)	2.068 (df = 148)	1.806 (df = 146)
F Statistic	3.922*** (df = 4; 148)	2.423* (df = 4; 148)	1.843 (df = 4; 148)	2.057* (df = 6; 146)

**Note 1:** Peer influence measured by the coefficient obtained in the SIENA analysis. **Note 2:** High number of cut-points indicates a community with more cut-points than the median. \*p < 0.1; \*\*p < 0.05; \*\*\*p < 0.01

decide whether to purchase the iPhone 3G. Table 16 defines the five seeding strategies considered. These are common heuristics used in marketing as described in Watts et al. (2007) and Hinz et al. (2011). We use the estimate for the marginal effect of peer influence obtained in the “IV Probit Results” section in our simulations. Therefore, they apply only to treated subscribers as discussed in the “Identification Strategy” section.

For each set of parameters, we present results for month 11 (T) in our panel. These results are averages over 1,000 simulations. We provide three types of results. Figure 12 shows the number of additional adopters as a function of the seeding policy, the number of seeds, and the exogenous marginal effect of peer influence. We observe that local and global degree yield the largest number of additional adopters. Random seeding performs poorly but global low degree performs even worse.

Figure 13 shows the cost to revenue ratio  $C/R$  that allows EuroMobile to breakeven as a function of the seeding strategy,

the number of seeds and the exogenous marginal effect of peer influence when EuroMobile subsidizes 50 percent of the handset to the seeds ( $v = 0.5$ ).<sup>25</sup>

We observe again that local and global degree perform the best. However, the former strategy allows for significantly higher  $C/R$  ratios for small sized campaigns.

Finally, Figure 14 shows the ratio of EuroMobile profits with seeding relative to no seeding for  $C/R = 0.5$ <sup>26</sup> and  $v = 0.5$ .<sup>27</sup>

<sup>25</sup> $C$  is the cost of iPhone 3G to EuroMobile and  $R$  is the price charged by EuroMobile per handset. The revenue generated by EuroMobile from seeding the iPhone 3G is given by  $AA(R - C)$ . The cost of seeding is given by  $nR(1 - v)$ . Therefore, EuroMobile breaks even when  $C/R = 1 - (1 - v)n/AA$ .

<sup>26</sup>According to iSuppli (<http://www.isuppli.com/>), the marginal cost of the iPhone 3G is roughly 50 percent of the revenue.

<sup>27</sup>The profit with seeding is given by  $AA(R - C) - nR(1 - v)$  whereas the profit without seeding must have been  $AO(R - C)$ .

Table 16. Seeding Policy Interventions Tested	
Policy $t = 0$	Policy Description
Random	Award the iPhone 3G to a random sample of $n$ subscribers in $S$
Global Degree	Award the iPhone 3G to the $n$ subscribers in $S$ with highest degree
Global Betweenness	Award the iPhone 3G to $n$ subscribers in $S$ with highest betweenness
Local Degree	Randomly select $n$ communities with replacement and at each draw award the iPhone 3G to the subscriber with the highest degree in that community without and iPhone 3G
Low Degree	Award the iPhone 3G to the $n$ subscribers in $S$ with lowest degree

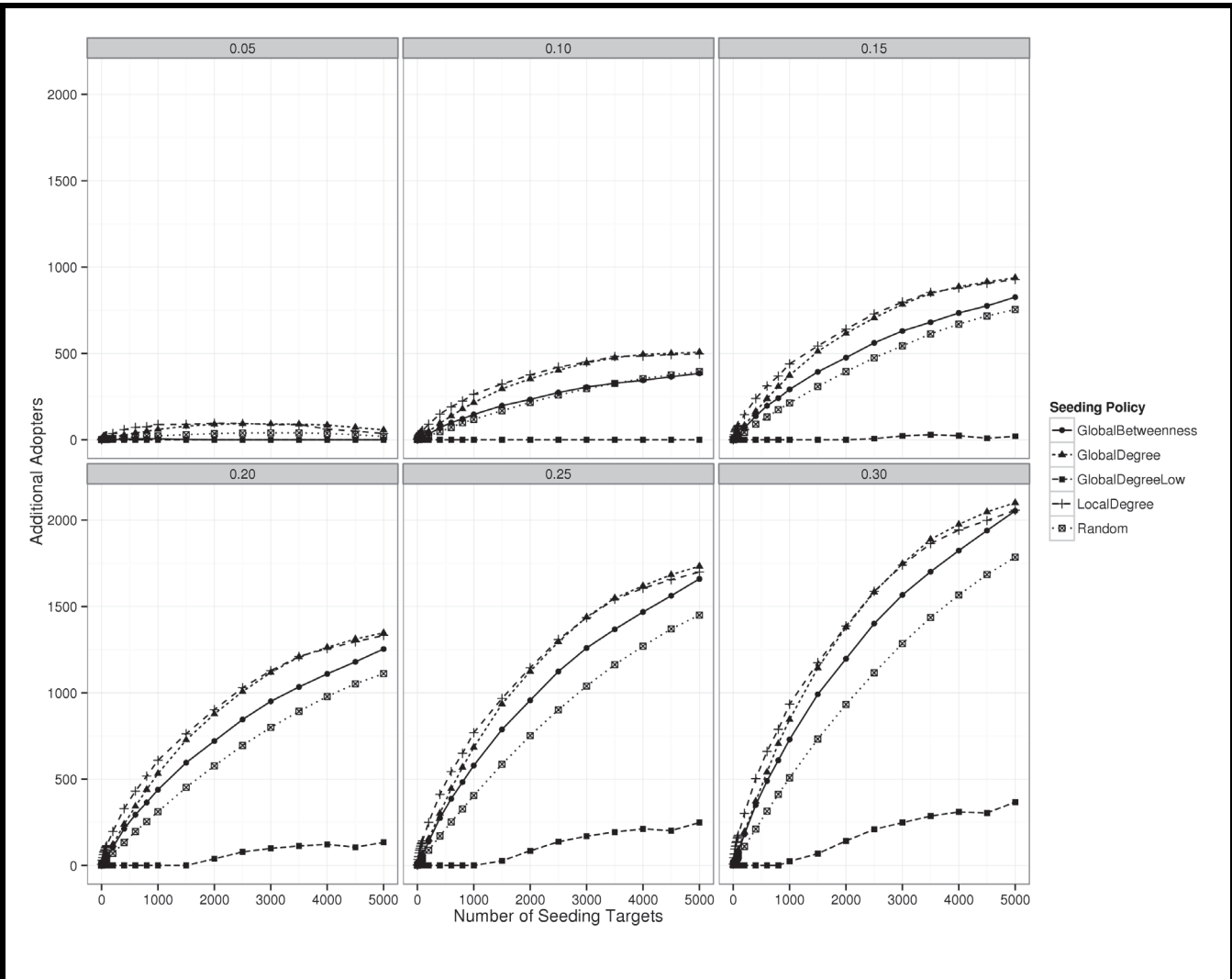
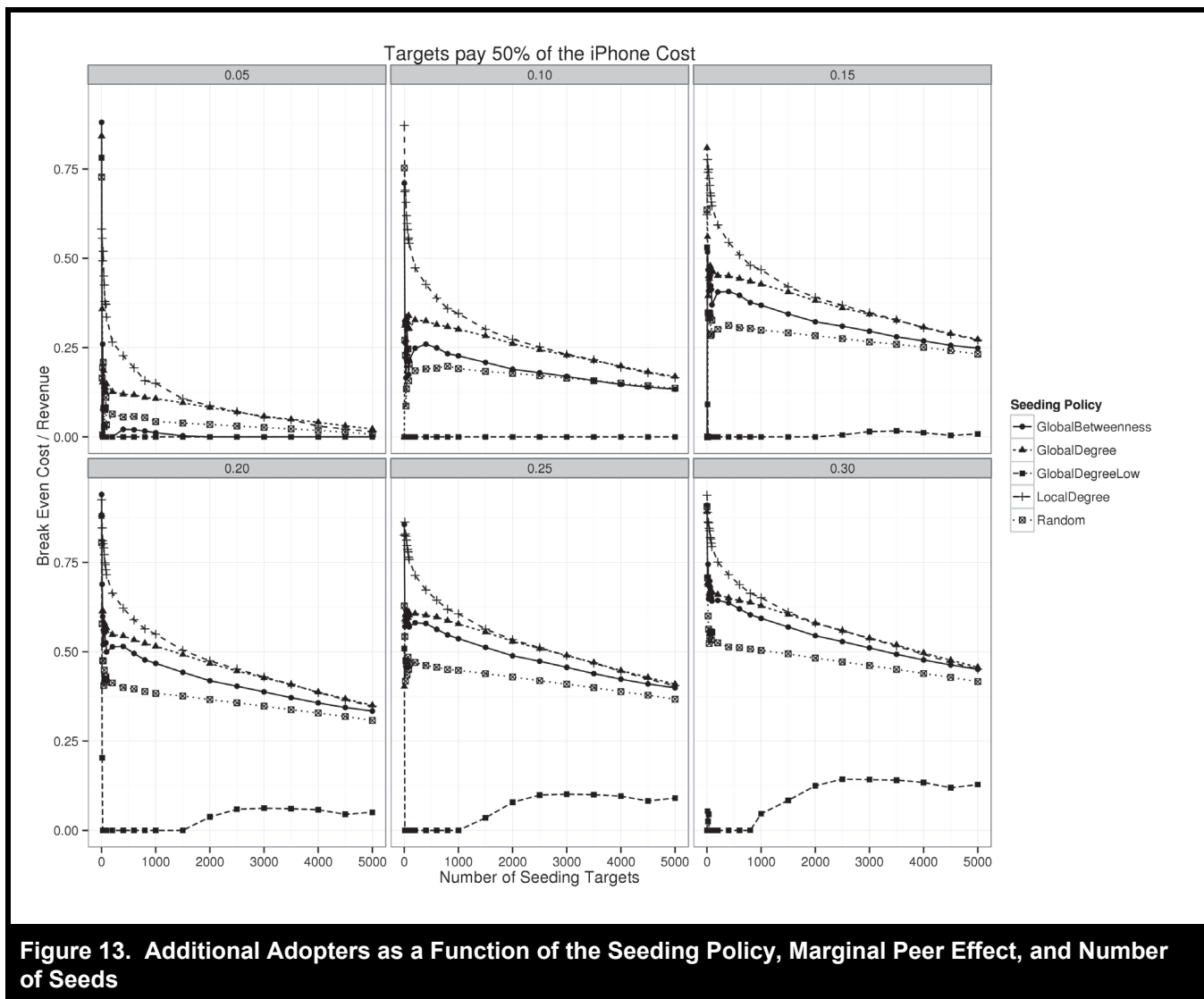


Figure 12. Additional Adopters as a Function of the Seeding Policy, Marginal Peer Effect, and Number of Seeds



We observe that EuroMobile can hardly benefit from seeding the iPhone 3G given the modest marginal effects of peer influence found in this paper. However, these results are very sensitive to how much EuroMobile subsidizes the iPhone 3G to the seeds. Figure 15 shows that EuroMobile can actually double its profit by using local or global degree to seed 5,000 subscribers when the marginal effect of peer influence is twice as much as that observed in our data and the subsidy is relatively small.

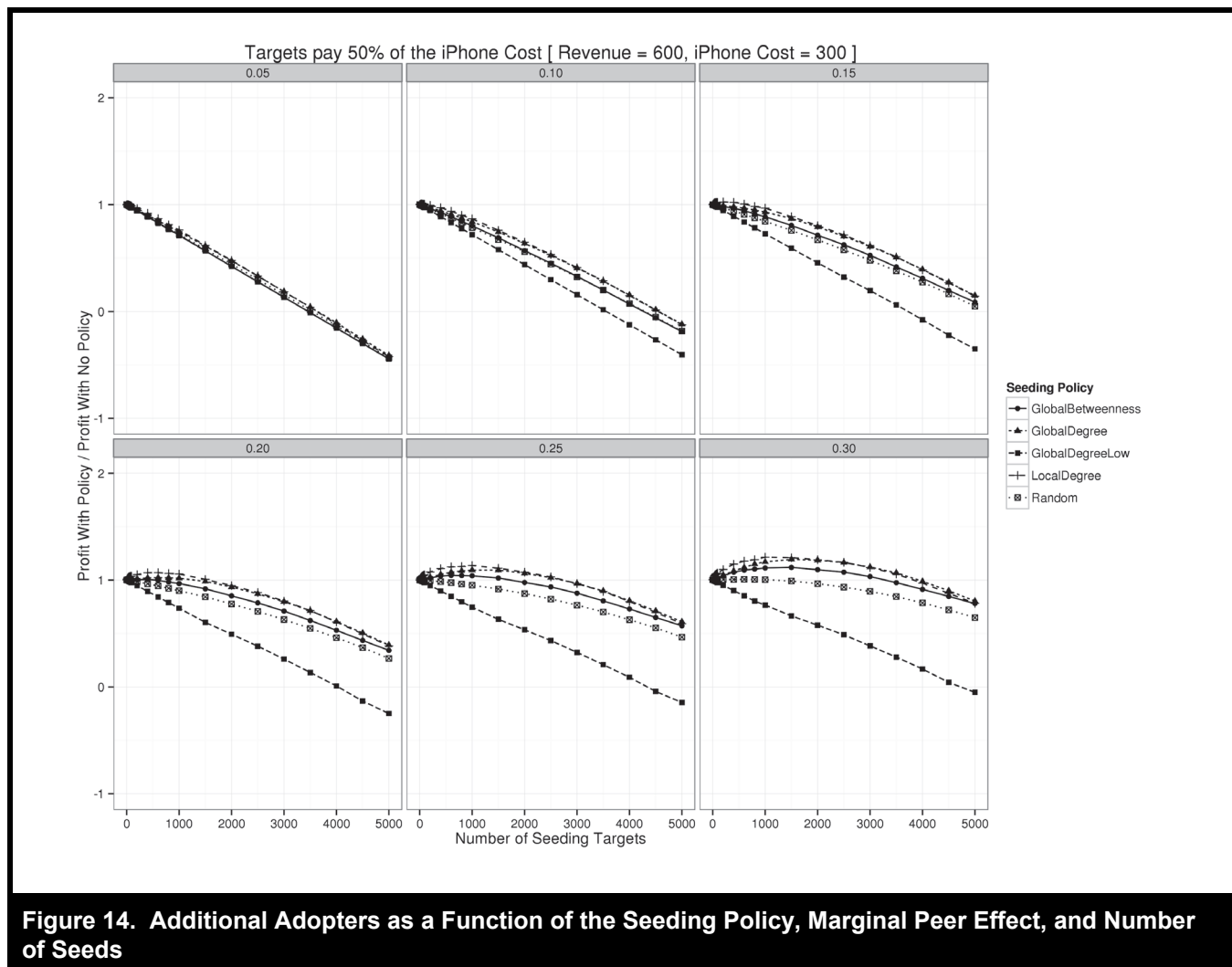
In sum, the effect of peer influence found in this paper is smaller than other peer effects reported in the literature such as in Aral and Walker (2011) and Bapna and Umyarov (2014). However, the settings in these papers are not directly comparable to ours. In particular, we study the effect of peer influence for an expensive new handset. It is, therefore,

reasonable to observe a smaller effect of peer influence in our case. In fact, our paper shows that the potential of viral marketing strategies for expensive physical products might not be the same as for inexpensive digital products. While there might still be peer influence in the former, ultimately the decision to adopt lies with each consumer individually and purchasing expensive products entails significant financial risks that previous studies have not been able to capture.

## Conclusions

This paper complements the literature that looks at contagion in the case of cheap and free digital goods by studying the effect of peer influence in the diffusion of a costly physical

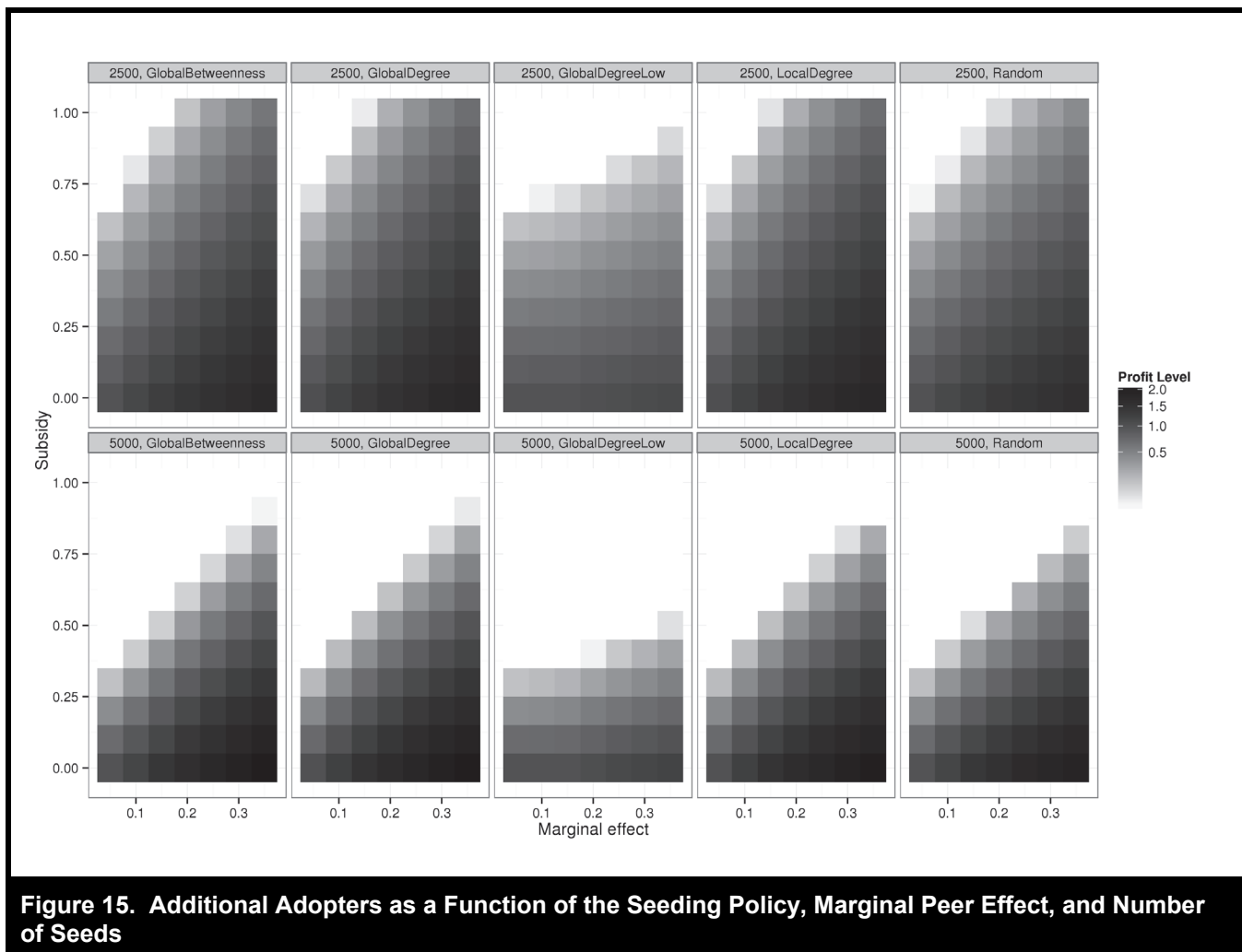




product. We argue that the results previously reported in the literature do not extend to costly products but studying the latter type of products provides for more realistic settings. In particular, it is hard for firms to breakeven when seeding costly goods to benefit from peer influence. In these cases, the effect of peer influence would need to be significant in magnitude to compensate for the high costs of seeding. In this paper we show that while there is peer influence in the adoption of the iPhone 3G, the magnitude of the effect will hardly justify seeding iPhones for free. Firms seeding other costly products are likely to face similar concerns and thus think that mass marketing is more appropriate in these cases. Yet, two additional alternatives might be considered. One is to reduce the cost of seeding. Instead of offering the product for free, firms can try to offer the product to seeds at a discount. This will facilitate breaking even. However, it may be hard to convince consumers to purchase costly products even when these are offered at discounted prices. Another

approach is to explicitly incentivize peer influence, for example, offering the product, for free or at a discount, to seeds that bring friends who may also purchase the product. This way, the firm actively rewards peer influence, which is thus likely to increase to a level that allows for breaking even.

We contribute with a new methodology to measure the effect of peer influence using observational data, which rests on identifying tight communities of likely adopters when adoption is a rare event and exploring the structure of the social network to derive instrumental variables at the individual level. We develop a new algorithm capable of identifying these communities in large network graphs. This algorithm can be easily applied in other settings where the product studied is expensive, which is likely to lead to little adoption. We work with a major European cell phone provider (EuroMobile) to study the role of peer influence in the diffusion of the iPhone 3G. We use call detailed records (CDRs)



from millions of EuroMobile users that we follow over a period of 11 months. We infer a social graph across users using these CDRs and identify tight communities of friends whose majority of calling behavior occurs inside the community. We work with 263 non-overlapping communities comprising 24,131 users of which 1,758 adopted the iPhone 3G. We posit that the subscribers' decision to adopt the iPhone 3G depends on whether their friends adopted this handset. Yet, unobserved heterogeneity correlated with friends' adoption may render our estimates of peer influence biased. We add a number of dummy variables to our model to control for heterogeneity across regions, across subscribers, and over time. Furthermore, we also examine adoption with the adoption of the friends of the friends of the ego not friends of the ego living in cities other than where the ego does. We provide a series of results showing the appropriateness and robustness of this instrument.

We show that if all friends adopt the iPhone 3G, then the ego's probability to adopt increases by 15 percent on average. Using this estimate, we show that peer influence might have accounted for roughly 14 percent of iPhone 3G sales at EuroMobile during the 11 months of data we analyze. We also use SIENA, which is an implementation of a stochastic actor-named model for the coevolution of network dynamics and behavior, to model a time-varying social network in which subscribers can adopt the iPhone 3G. Using SIENA we find an effect of peer influence consistent both in statistical significance and magnitude with our IV estimates, which increases our confidence in our findings. We use different snapshots of our data to parameterize these simulations and all of them yield similar results. Finally, we provide results from several policy simulations showing that, with the modest magnitude of the effect of peer influence found in this paper, EuroMobile would have hardly benefitted from using com-

mon marketing strategies to seed the iPhone 3G as a way to accelerate its diffusion. This finding, however, depends highly on how much the handset is subsidized to the seeds.

Our paper also comes with limitations. Our analysis relies on a single random sample of communities derived from the CDRs. While identifying communities is important to control for unobserved confounders that can drive both adoption and social proximity, it is still the case that the clustering criteria we used are not unique and even multiple runs of our sampling strategy could produce different communities which could, in principle, lead to additional variance in our results. Yet, we show that most of the algorithms previously used in the literature to identify communities cannot process real-world, large social networks. We use CDRs to proxy social proximity. However, we know that subscribers in our sample must have communicated over other channels, such as face-to-face meetings, e-mail, Twitter, and Facebook, which we cannot measure. In addition, EuroMobile was not the only wireless provider in the country analyzed and thus the subscribers we analyzed had friends that we could not observe. Still, the second and the third largest providers in this country did not launch the iPhone 3G until much later, which reduces significantly the likelihood of having friends with other carriers that adopted the iPhone 3G. Finally, we identify the effect of peer influence but we are unable to determine the underlying mechanism by which influence occurs, such as information transmission, competition, or conformity.

## Acknowledgments

This work was partially supported by the CMU-Portugal Program through Ph.D. fellowship SFRH/BD/33772/2009, and by the iLab and the Living Analytics Research Center at the Heinz College. We also thank our Industrial Partner for their support. We thank the senior editor, the associate editor, and two anonymous reviewers for their valuable comments and ideas throughout the review process of this paper. We thank Rahul Telang and Rodrigo Belo for valuable comments to earlier versions of this paper and for many brainstorming sessions. We also thank participants in the International Conference on Information Systems (2012) and in the Symposium on Statistical Challenges in e-Commerce Research (2012) for the valuable feedback received. All remaining errors are our own.

## References

- Allison, P. D. 1982. "Discrete-Time Methods for the Analysis of Event Histories," *Sociological Methodology* (13), pp. 61-98.
- Anagnostopoulos, A., Kumar, R., and Mahdian, M. 2008. "Influence and Correlation in Social Networks," in *Proceedings of the 14<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM, pp. 7-15.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. 1996. "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association* (91:434), pp. 444-455.
- Aral, S. 2010. "Commentary—Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science* (30:2), pp. 217-223.
- Aral, S., Muchnik, L., and Sundararajan, A. 2009. "Distinguishing Influence Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences* (106:51), pp. 21544-21549.
- Aral, S., and Walker, D. 2011. "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science* (57:9), pp. 1623-1639.
- Bagrow, J., and Boltt, E. 2004. "A Local Method for Detecting Communities," *Physical Review E* (72:4), p. 045108-1-045108-10.
- Bapna, R., and Umyarov, A. 2014. "Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network," *Management Science*, forthcoming.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. 2008. "Fast Unfolding of Communities in Large Networks," *Journal of Statistical Mechanics: Theory and Experiment* (8:10), P10008.
- Bloom, H. S. 1984. "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review* (8:2), pp. 225-246.
- Botelho, A., and Pinto, L. C. 2004. "The Diffusion of Cellular Phones in Portugal," *Telecommunications Policy* (28:5-6), pp. 427-437.
- Boucher, V., Bramoulle, Y., Djebbari, H., and Fortin, B. 2010. "Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation," *Journal of Applied Econometrics* (29:1), pp. 91-109.
- Bramoulle, Y., Djebbari, H., and Fortin, B. 2009. "Identification of Peer Effects through Social Networks," *Journal of Econometrics* (150:1), pp. 41-55.
- Brynjolfsson, E., and Kemerer, C. F. 1996. "Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market," *Management Science* (42:12), pp. 1627-1647.
- Burt, R. S. 1987. "Social Contagion and Innovation: Cohesion Versus Structural Equivalence," *American Journal of Sociology* (92:6), pp. 1287-1335.
- Carrell, S. E., Fullerton, R. L., and West, J. E. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics* (27:3), pp. 439-464.
- Case, A. C., and Katz, L. F. 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," National Bureau of Economic Research Working Paper Series No. 3705.
- Chu, W.-L., Wu, F.-S., Kao, K.-S., and Yen, D. C. 2009. "Diffusion of Mobile Telephony: An Empirical Study in Taiwan," *Telecommunications Policy* (33:9), pp. 506-520.
- Clauset, A., Newman, M. E. J., and Moore, C. 2004. "Finding Community Structure in Very Large Networks," *Physical Review* (70:6), 066111.

- Coleman, J. S., Katz, E., Menzel, H., and Bureau of Applied Social Research. 1966. *Medical Innovation: A Diffusion Study*, Indianapolis, IN: Bobbs-Merrill Co.
- Csardi, G., and Nepusz, T. 2005. *iGraph Reference Manual* (<http://igraph.org/c/doc/igraph-docs.pdf>).
- DerSimonian, R., and Laird, N. 1986. "Meta-Analysis in Clinical Trials," *Controlled Clinical Trials* (7:3), pp. 177-188.
- Doganoglu, T., and Grzybowski, L. 2007. "Estimating Network Effects in Mobile Telephony in Germany," *Information Economics and Policy* (19:1), pp. 65-79.
- Donetti, L., and Munoz, M. A. 2005. "Improved Spectral Algorithm for the Detection of Network Communities," *AIP Conference Proceedings* (779:1), pp. 104-107.
- Duch, J., and Arenas, A. 2005. "Community Detection in Complex Networks Using Extremal Optimization," *Physical Review E* (72:2), p. 027104.
- Economist, The. 2011. "Nokia at the Crossroads: Blazing Platforms," *The Economist*, February 10.
- Fortunato, S. 2010. "Community Detection in Graphs," *Physics Reports* (486:3-5), pp. 75-174.
- Fortunato, S., and Barthelemy, N. 2007. "Resolution Limit in Community Detection," *Proceedings of the National Academy of Sciences* (104:1), pp. 36-41.
- Fortunato, S., Latora, V., and Marchiori, M. 2004. "A Method to Find Community Structures Based on Information Centrality," *Physical Review E* (70), p. 056104.
- Gefen, D., and Straub, D. W. 1997. "Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model," *MIS Quarterly* (21:4), pp. 389-400.
- Girvan, M., and Newman, M. E. J. 2002. "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Sciences of the United States of America* (99:12), p. 7821.
- Gruber, H. 2001. "Competition and Innovation: The Diffusion of Mobile Telecommunications in Central and Eastern Europe," *Information Economics and Policy* (13:1), pp. 19-34.
- Hedges, L. V., and Olkin, I. 1985. *Statistical Methods for Meta-Analysis* (1<sup>st</sup> ed.), New York: Academic Press.
- Hedström, P. 1994. "Contagious Collectivities: On the Spatial Diffusion of Swedish Trade Unions, 1890-1940," *American Journal of Sociology* (99:5), pp. 1157-1179.
- Hinz, O., Skiera, B., Barrot, C., and Becker, J. U. 2011. "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing* (75:6), pp. 55-71.
- Hlavac, M. 2013. "Stargazer: LaTeX Code and ASCII Text for Well-Formatted Regression and Summary Statistics Tables," <https://sites.google.com/site/marekhlavac/stargazer>.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., and Botella, J. 2006. "Assessing Heterogeneity in Meta-Analysis: Q Statistic or  $i^2$  Index?," *Psychological Methods* (11:2), pp. 193-206.
- Imbens, G. W., and Angrist, J. D. 1994. "Identification and Estimation of Local Average Treatment Effects," *Econometrica* (62:2), pp. 467-475.
- ITU. 2011. "The World in 2011: ICT Facts and Figures," International Telecommunication Union, Geneva.
- Karahanna, E., Agarwal, A., and Angst, C. M. 2006. "Reconceptualizing Compatibility Beliefs in Technology Acceptance Research," *MIS Quarterly* (30:4), pp. 781-804.
- Katz, E., and Lazarfeld, P. F. 1955. *Personal Influence: The Part Played by People in the Flow of Mass Communications*, Glencoe, IL: The Free Press.
- Katz, M. L., and Shapiro, C. 1986. "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy* (94:4), pp. 822-841.
- Katz, M. L., and Shapiro, C. 1994. "Systems Competition and Network Effects," *The Journal of Economic Perspectives* (8:2), pp. 93-115.
- Knapp, G., and Hartung, J. 2003. "Improved Tests for a Random Effects Meta-Regression with a Single Covariate," *Statistics in Medicine* (22), pp. 2693-2710.
- Krackhardt, D., and Stern, R. N. 1988. "Informal Networks and Organizational Crises: An Experimental Simulation," *Social Psychology Quarterly* (51:2), pp. 123-140.
- Lancichinetti, A., and Fortunato, S. 2009. "Community Detection Algorithms: A Comparative Analysis," *Physical Review E* (80:5), p. 056117.
- Lancichinetti, A., Fortunato, S., and Radicchi, F. 2008. "Benchmark Graphs for Testing Community Detection Algorithms," *Physical Review E* (78:4), p. 046110.
- Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., and Chalmers, T. C. 1992. "Cumulative Meta-Analysis of Therapeutic Trials for Myocardial Infarction," *New England Journal of Medicine* (327:4), pp. 248-254.
- Leenders, R. T. A. J. 2002. "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix," *Social Networks* (24:1), pp. 21-47.
- Leskovec, J., Adamic, L. A., and Huberman, B. A. 2007. "The Dynamics of Viral Marketing," *ACM Transactions on the Web* (1:1), Article 5.
- Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. 2009. "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters," *Internet Mathematics* (6:1), pp. 29-123.
- Ma, L., Krishnan, R., and Montgomery, A. 2010. "Homophily or Influence? An Empirical Analysis of Purchase Within a Social Network," Working Paper, Carnegie Mellon University.
- Mahajan, V. 1985. *Models for Innovation Diffusion*, Newbury Park, CA: Sage Publications, Inc.
- Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies* (60:3), pp. 531-542.
- McConnell, J. D. 1968. "The Price-Quality Relationship in an Experimental Setting," *Journal of Marketing Research* (5:3), pp. 300-303.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* (27:1), pp. 415-444.
- Menzel, H. 1960. "Innovation, Integration, and Marginality: A Survey of Physicians," *American Sociological Review* (25:2), pp. 704-713.



- Mercken, L., Snijders, T. A. B., Steglich, C., Vartiainen, E., de Vries, H. 2010. "Dynamics of Adolescent Friendship Networks and Smoking Behavior," *Social Networks* (32:1), pp. 72-81.
- Newman, M. E. J. 2003. "Fast Algorithm for Detecting Community Structure in Networks," *Physical Review E* (69), p. 066133.
- Newman, M. E. J. 2006. "Modularity and Community Structure in Networks," *Proceedings of the National Academy of Sciences* (103:23), pp. 8577-8582.
- Newman, M. E. J., and Givran, M. 2004. "Finding and Evaluating Community Structure in Networks," *Physical Review E* (69:2), p. 026113.
- Oestreicher-Singer, G., and Sundararajan, A. 2012. "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management Science* (58:11), pp. 1963-1981.
- Palla, G., Derényi, I., Farkas, I., and Vicsek, T. 2005. "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society," *Nature* (435:7043), pp. 814-818.
- Park, Y., and Ueda, M. 2011. "A Comparative Study on the Diffusion of Smartphones in Korea and Japan," in *Proceedings of the IEEE/IPSJ 11<sup>th</sup> International Symposium on Applications and the Internet*, Los Alamitos, CA: IEEE Computer Society, pp. 545-549.
- Paule, R. C., and Mandel, J. 1982. "Consensus Values and Weighting Factors," *Journal of Research of the National Bureau of Standards* (87:5), pp. 377-385.
- Pons, P., and Latapy, M. 2006. "Computing Communities in Large Networks Using Random Walks," *Journal of Graph Algorithms and Applications* (10:2), pp. 191-218.
- Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., and Parisim, D. 2004. "Defining and Identifying Communities in Networks," *Proceedings of the National Academy of Sciences of the United States of America* (101:9), pp. 2658-2663.
- Rao, N., Mobius, M. M., and Rosenblat, T. 2007. "Social Networks and Vaccination Decisions," Working Paper No. 07-12, Federal Reserve Bank of Boston.
- Ripley, R. M., and Snijders, T. A. B. 2010. *Manual for SIENA Version 4.0*, University of Oxford, Department of Statistics & Nuffield College.
- Rogers, E. M. 1995. "Diffusion of Innovations," New York: Free Press.
- Rohlf, J. 1974. "A Theory of Interdependent Demand for a Communications Service," *The Bell Journal of Economics and Management Science* (5:1), pp. 16-37.
- Romer, D., Black, M., Ricardo, I., Feigelman, S., Kaljee, L., Galbraith, J., Nesbit, R., Hornik, R. C., and Stanton, B. 1994. "Social Influences on the Sexual Behavior of Youth at Risk for HIV Exposure," *American Journal of Public Health* (84:6), pp. 977-985.
- Rosvall, M., and Bergstrom, C. T. 2008. "Maps of Random Walks on Complex Networks Reveal Community Structure," *Proceedings of the National Academy of Sciences* (105:4), pp. 1118-1123.
- Rukhin, A. L., Biggerstaff, B. J., and Vangel, M. G. 2000. "Restricted Maximum Likelihood Estimation of a Common Mean and the Mandel-Paule Algorithm," *Journal of Statistical Planning and Inference* (83:2), pp. 319-330.
- Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics* (116:2), pp. 681-704.
- Saloner, G., and Shepard, A. 1995. "Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines," *The RAND Journal of Economics* (26:3), pp. 479-501.
- Shalizi, C. R., and Thomas, A. C. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods & Research* (40:2), pp. 211-239.
- Shampanier, K., Mazar, N., and Ariely, D. 2007. "Zero as a Special Price: The True Value of Free Products," *Marketing Science* (26:6), pp. 742-757.
- Shiv, B., Carmon, Z., and Ariely, D. 2005. "Placebo Effects of Marketing Actions: Consumers May Get What They Pay For," *Journal of Marketing Research* (42:4), pp. 383-393.
- Singh, S. K. 2008. "The Diffusion of Mobile Phones in India," *Telecommunications Policy* (32:9-10), pp. 642-651.
- Snijders, T. A. B. 1996. "Stochastic Actor-Oriented Models for Network Change," *Journal of Mathematical Sociology* (21:1-2), pp. 149-172.
- Snijders, T. A. B., van de Bunt, G. G., and Steglich, C. E. G. 2010. "Introduction to Stochastic Actor-Based Models for Network Dynamics," *Social Networks* (32:1), pp. 44-60.
- Stafford, J. E., and Enis, B. M. 1969. "The Price-Quality Relationship: An Extension," *Journal of Marketing Research* (6), pp. 456-458.
- Strang, D., and Soule, S. A. 1998. "Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills," *Annual Review of Sociology* (24), pp. 265-290.
- Strang, D., and Tuma, N. B. 1993. "Spatial and Temporal Heterogeneity in Diffusion," *American Journal of Sociology* (99:3), pp. 614-639.
- Tillema, T., Dijst, M., and Schwanen, T. 2010. "Face-to-Face and Electronic Communications in Maintaining Social Networks: The Influence of Geographical and Relational Distance and of Information Content," *New Media & Society* (12:6), pp. 965-983.
- Tucker, C. 2008. "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities," *Management Science* (54:12), pp. 2024-2038.
- Valente, T. W. 1996a. "Network Models of the Diffusion of Innovations," *Computational and Mathematical Organization Theory* (2:2), pp. 163-164.
- Valente, T. W. 1996b. "Social Network Thresholds in the Diffusion of Innovations," *Social Networks* (18:1), pp. 69-89.
- Vakulenko, M. 2011. "Mobile Platforms: The Clash of Ecosystems," Vision Mobile, November 7 (<http://www.visionmobile.com/blog/2011/11/new-report-mobile-platforms-the-clash-of-ecosystems/>).
- Van den Bulte, C., and Lilien, G. L. 2001. "Medical Innovation Revisited: Social Contagion Versus Marketing Effort," *The American Journal of Sociology* (106:5), pp. 1409-1435.
- Venkatesh, V., and Morris, M. G. 2000. "Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior," *MIS Quarterly* (24:1), pp. 115-139.

- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* (27:3), pp. 425-478.
- Viechtbauer, W. 2010. "Conducting Meta-Analyses in R with the Metafor Package," *Journal of Statistical Software* (36:3), pp. 1-48.
- Wasserman, S., and Faust, K. 1994. *Dynamic Network Analysis: Methods and Applications*, New York: Cambridge University Press.
- Watts, D. J., Peretti, J., and Harvard Business School. 2007. *Viral Marketing for the Real World*, Boston: Harvard Business School.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*, New York: Springer.
- Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
- Wooldridge, J. M., and Imbens, G. 2007. "What's New in Econometrics?," National Bureau of Economic Research ([http://www.nber.org/WNE/lect\\_13\\_weakmany\\_iv.pdf](http://www.nber.org/WNE/lect_13_weakmany_iv.pdf)).
- Zhang, B., and Krackhardt, D. 2010. "Keeping up with the Joneses: an Empirical Investigation of Contagion on CRBT Adoption," Working Paper, Carnegie Mellon University.
- Zhang, B., Krackhardt, D., Krishnan, R., and Doreian, P. 2011. "An Effective and Efficient Subpopulation Extraction Method in Very Large Social Networks," in *Proceedings of the 32<sup>nd</sup> International Conference on Information Systems*, Shanghai (<http://aisel.aisnet.org/ics2011/proceedings/breakthroughideas/7>).
- Zheng, K., Padman, R., Krackhardt, D., Johnson, M. P., and Diamond, H. S. 2010. "Social Networks and Physician Adoption of Electronic Health Records: Insights from an Empirical Study," *Journal of the American Medical Informatics* (17:3), pp. 328-336.
- Zhou, H., and Lipowsky, R. 2004. "Network Brownian Motion: A New Method to Measure Vertex-Vertex Proximity and to Identify Communities and Subcommunities," *Computational Science-ICCS*, Berlin: Springer-Verlag, pp. 1062-1069.

## About the Authors

**Miguel Godinho de Matos** is a post-doctoral researcher of Information Systems and Management at both Heinz College and

Católica Lisbon School of Business and Economics. He received a Ph.D. in Telecommunications Policy and Management and a M.Sc. in Engineering and Public Policy from Carnegie Mellon University. Miguel's research interests focus on the analysis of social networks and peer influence on consumer behavior and the impact of digitization on consumer search and choice. Miguel has published his work in top peer-reviewed research conferences such as the International Conference of Information Systems, IEEE Conference on Social Computing and the Economics of Digitization Seminar Series of the National Bureau of Economic Research.

**Pedro Ferreira** is an assistant professor of Information Systems and Management at the Heinz College and at the Department of Engineering and Public Policy, Carnegie Mellon University. He received a Ph.D. in Telecommunications Policy from CMU and a M.Sc. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology. Pedro's research interests lie in two major domains: identifying causal effects in dense network settings, with direct application to understanding the future of the digital media industry, and the evolving role of technology in the economics of education. Currently, he is working on a series of large-scale randomized experiments in network settings looking at identifying the role of peer influence in the consumption of media. Pedro has published in top journals and top peer-reviewed research conferences such as *Management Science* and the IEEE Conference on Social Computing.

**David Krackhardt** is a professor of organization at the Heinz College of Public Policy and the Tepper School of Business, Carnegie Mellon University. He received his Ph.D. in organizational behavior from the University of California, Irvine. His primary focus is the development of theories and methods for the study of social networks, especially within organizations. He has published in top journals in the fields of anthropology, sociology, psychology, management, information systems, medicine, and statistics. He has served on the editorial boards of several leading journals, including *Administrative Science Quarterly*, *Academy of Management Journal*, and *Computational and Mathematical Organizational Theory*. He is the founding editor of *Journal of Social Structure* and currently serves as an associate editor for *Network Science*.





# PEER INFLUENCE IN THE DIFFUSION OF IPHONE 3G OVER A LARGE SOCIAL NETWORK

**Miguel Godinho de Matos**

Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213-3890 U.S.A. {miguelgodinhomatos@cmu.edu} and  
Católica Lisbon School of Business and Economics, Palma de Cima, 1649-023 Lisbon, PORTUGAL

**Pedro Ferreira**

Heinz College and Department of Engineering and Public Policy, Carnegie Mellon University,  
Pittsburgh, PA 15213-3890 U.S.A. {pedrof@cmu.edu}

**David Krackhardt**

Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213-3890 U.S.A. {krack@cmu.edu}

## Appendix A

### Modified T-CLAP Algorithm

Our modified version of T-CLAP snowballs from a random adopter with depth  $d$  (three, in the case of our paper) and keeps pruning subscribers from this sample while the resulting community has more than  $s$  subscribers (115, in the case of our paper).  $H$  is the social network graph.  $H$ .  $S$  is the set of subscribers in this graph and  $A$  the set of adopters. The algorithm identifies the subscribers in  $H$  that are not adopters with the lowest IER. These subscribers are kept in set  $V$ . If  $V$  is empty, then the algorithm identifies the subscribers in  $H$  that are adopters with the lowest IER. These subscribers are kept in  $V$ . This allows us to prune non-adopters with higher probability. A subscriber from  $V$  is then selected at random and removed from  $H$ . This can, however, separate  $H$  into disjoint graphs. If this is the case, then  $H$  keeps only the subgraph with the highest IER (ties decided at random).

**Table A1. Pseudocode of the Modified Version of T-CLAP**
**Variables**

G Network graph  
A Subset of adopters  
d Maximum community  
depth s Target size for communities

**Auxiliary variables**

H Network graph  
V Set of subscribers  
C Set of network graphs

**Algorithm**

```

function MTCLAP(G,A,s) - Modified T-CLAP
    return NP(G, Snowball(G,random seed, d), A, s)
end function

function NP(G,H ,A,s) - Node Pruning
    Compute H .S.IE R
    while |H .S| > s do
        V = SubscribersLowestIER(H , A)
        H = Subgraph(H ,V \ {v ∈ V chosen at random })
        H = ComponentHighestIER(H )
    end while
    return H
end function

function SUBSCRIBERSLOWESTIER(H ,A)
    V = {H .S(i) \ A : H .S(i).IE R = min{H .S.IE R}}
    if V.S == 0/ then
        V = {H .S(i) : H .S(i).IE R = min{H .S.IE R}}
    end ifreturn V
end function

function COMPONENTHIGHESTIER(H )
    C = Components(H )
    Compute C.IE R
    C = {C(k) ∈ C : C(k).IE R = max{C.IE R}}
    return {c ∈ C chosen at random}
end function

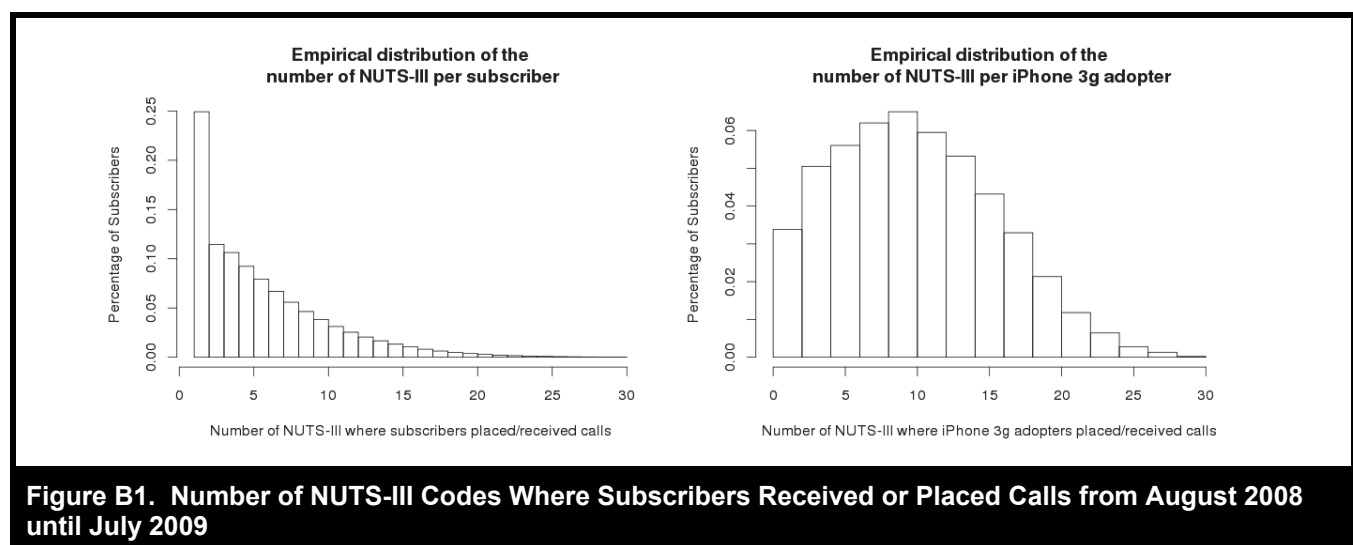
```

# Appendix B

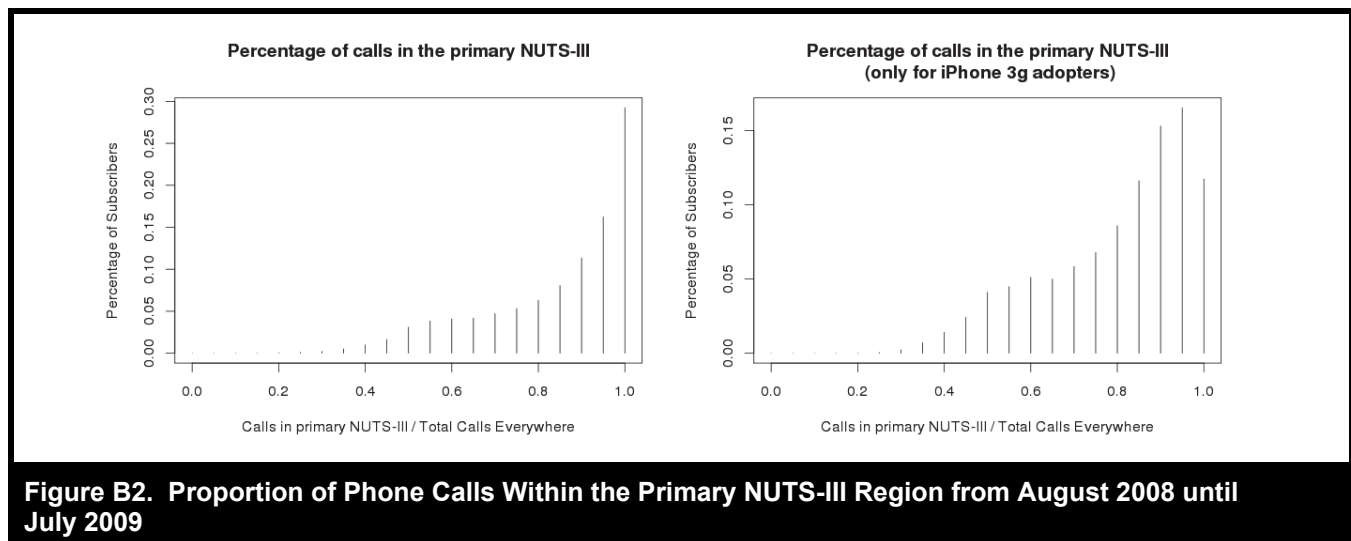
## Computing Our Instrumental Variables

Our dataset contains information that identifies the cellular towers that are used in each and every call placed or received by EuroMobile subscribers. Cellular towers are associated with GPS coordinates that allow us to track subscribers through space and time. For every call placed or received, we use the GPS coordinates of the cell towers to determine the NUTS-III region where the initiator and the recipient are located. We then use the mode of the NUTS-III regions obtained to determine the primary region for each individual (the one where he is most often observed). Cell tower ranges can cover from 1 km up to 30 km. Therefore, there is some uncertainty associated with the true location of each subscriber, particularly in regions with low population density where there are few cell towers with broad ranges. On average, we used 753.2 calls to identify the primary region of each subscriber.

NUTS codes (nomenclature of territorial units for statistics) are statistical divisions of the economic territory use throughout the European Union designed to develop consistent regional statistics across countries (<http://epp.eurostat.ec.europa.eu/>). We choose NUTS-III for our analysis because, by construction, NUTS-III represents contiguous municipalities that face similar economic and development challenges and within which people are likely to move substantially. We only know the location of a subscriber when she places or receives a call. Using NUTS-III ensures that we use large enough regions so that each subscriber moves most of the time within the region rather than between regions. Furthermore, we also try to capture the fact that the socio-economic challenges faced by subscribers within the same NUTS-III are similar. Figure B1 depicts the number of NUTS-III codes where each subscriber was seen receiving or placing calls. The average number of NUTS-III per subscriber, for the entire population of subscribers, is 5.9 with a standard deviation of 4.5. Still, about 25 percent of the subscribers were seen placing or receiving calls within a single NUTS-III region. iPhone 3G adopters were more mobile than the average user with 10.4 NUTS-III per subscriber on average and a standard deviation of 5.6. This fact does not affect our separation strategy because looking in detail at the number of calls that allowed for identifying each subscriber within each NUTS-III, an overwhelming majority of them were placed or received within the primary NUTS-III region. This is true both overall and for the iPhone 3G adopters in particular. This highlights that people do move around in their daily lives (particularly when considering a large span of time such as 11 months), but they tend to stay within their primary region most of the time. Therefore, people with distinct primary (NUTS-III) regions will be clearly separated geographically almost all of the time. Figure B2 shows that the average proportion of calls placed within the primary NUTS-III region across subscribers is 83 percent (median 88 percent) overall and 78 percent (median 82 percent) for iPhone 3G adopters.



**Figure B1. Number of NUTS-III Codes Where Subscribers Received or Placed Calls from August 2008 until July 2009**



## Appendix C

### Community Sample Details

Table C1 shows the number of observations included in the regressions reported in the “Results for Peer Influence” section of the paper. From the original 263 communities with 24,131 users, we discarded 44 adopters and 241 non-adopters. These users were removed from the sample because of missing data for tenure, mobileNet ,and/or the dummies that described the previous handset that they owned. As a consequence, 5 out of the 263 communities were removed from the sample because they had no adopters. Including them in the analysis would generate a problem of complete separation of the outcome (Albert and Anderson 1984). These 5 communities had 711 subscribers. An additional 269 subscribers were also removed from the sample because they lived in zip codes where no one else adopted the iPhone 3G. Again, including them would lead to a problem of perfect separation of the outcome. Therefore, we estimate our model using 258 communities with a total of 23,151 subscribers of which 1,714 adopted the iPhone 3G during the period of analysis.

**Table C1. Observations Used in the Estimation of Our Empirical Model**

Time	Adoption		Total
	No	Yes	
0	22,782	369	23,151
1	22,558	224	22,782
2	22,403	155	22,558
3	22,254	149	22,403
4	22,133	121	22,254
5	21,945	188	22,133
6	21,813	132	21,945
7	21,739	74	21,813
8	21,642	97	21,739
9	21,565	77	21,642
10	21,477	88	21,565
11	21,437	40	21,477
Total	263,748	1,714	265,462

# Appendix D

## Complete Regression Output

Columns (1) and (2) of Table D1 show the complete regression output for column (3) of Table 7 and column (2) of Table 8, respectively. The results for the other models presented in this paper are similar but were omitted due to lack of space. The signs of the control variables are consistent with what one would expect prior. This is true both before and after instrumentation. People with previous plans of mobile Internet were more likely to adopt the iPhone 3G. This is also true for subscribers using handsets 2G or above prior to the release of the iPhone 3G. EuroMobile subscribers that spent most of their daytime in regions with very high or high average wage levels were more likely to adopt than individuals spending most of their time in regions with wages close to the national average. The opposite was true for subscribers moving in low and very low wage regions. Users subscribing to prepaid tariff plans before the release of the iPhone 3G were less likely to adopt, also as expected. The explanation is one of price because in order to buy the iPhone 3G and still remain a prepaid subscriber, consumers needed to pay the full price of the handset up front. The alternative would be to change from prepaid to postpaid, but in the country analyzed, consumers have a clear preference toward prepaid plans (approximately 80 percent of all subscribers are prepaid).

Finally, up to a certain point, network tenure contributed positively to the probability of adoption. This is likely to indicate that subscribers required some experience with the services provided by EuroMobile prior to purchasing a phone that would bind them for at least 24 months. For the sake of readability, note that the dummies dropped from the regression below are genderU, phone2.0g, and geoWageVL.

**Table D1. Complete Regression Outputs for Probit and IV Probit**

Variables	(1) Probit <i>adopted<sub>t</sub></i>	(2) IV Probit <i>adopted<sub>t</sub></i>
<i>frd_adopters<sub>t-1</sub></i>	2851*** (0258)	9.935*** (1.952) [2.087]
<i>Log(tenure<sub>t+1</sub>)</i>	0.300*** (0.101)	0.329*** (0.0955) [0.103]
<i>Log(tenure<sub>t+1</sub>)<sup>2</sup></i>	-0.0330** (0.0138)	-0.0370*** (0.0127) [0.0139]
<i>prepaidY</i>	-0.548*** (0.0283)	-0.508*** (0.0274) [0.0307]
<i>genderF</i>	-0.0616* (0.0343)	-0.0572 (0.0332) [0.0351]
<i>genderM</i>	0.144*** (0.0271)	0.1433*** (0.0261) [0.0275]
<i>mobileNetY</i>	0.425*** (0.0427)	0.398*** (0.0397) [0.0445]
<i>phone2.5g</i>	0.484*** (0.0547)	0.469*** (0.0536) [0.0581]
<i>phone3.0g</i>	0.637*** (0.0542)	0.628*** (0.0537) [0.0596]



**Table D1. Complete Regression Outputs for Probit and IV Probit (continued)**

Variables	(1) Probit <i>adopted<sub>t</sub></i>	(2) IV Probit <i>adopted<sub>t</sub></i>
<i>phone3.5g</i>	0.917*** (0.0658)	0.876*** (0.0650) [0.0735]
<i>phoneOther</i>	0.575*** (0.142)	0.503*** (0.140) [0.162]
<i>phoneAge</i>	-0.143*** (0.0551)	-0.152*** (0.0512) [0.0601]
<i>phoneAge</i> <sup>2</sup>	0.0256 (0.0251)	0.0279 (0.0240) [0.0273]
<i>geoWageH</i>	0.133*** (0.0431)	0.113*** (0.0381) [0.0420]
<i>geoWageL</i>	-0.146 (0.142)	-0.111 (0.150) [0.150]
<i>geoWageVH</i>	0.198*** (0.0563)	0.150*** (0.0510) [0.0540]
Constant	-3.602*** (0.214)	-3.652*** (0.310) [0.249]
Observations	265,462	265,462
Community FE	Yes	Yes
Zip Code FE	Yes	Yes
Month FE	Yes	Yes
Pseudo R <sup>2</sup>	0.157	
Log Lik	-8721	

\*\*\*p &lt; 0.01; \*\*p &lt; 0.05; \*p &lt; 0.1

**Note 1:** Community clusters robust standard errors in ( ) for Probit. **Note 2:** Newey estimator standard errors in ( ) for IV Probit. **Note 3:** Community block-bootstrap standard errors in [] for IV Probit based on 200 replications.

# Appendix E

## Pseudocode to Estimate Adoption Due to Peer Influence

**Table E1. Pseudocode to Estimate Adoption Due to Peer Influence**

Key variables	
$m$	Marginal effect of peer (obtained from IV probit)
$D(t)$	Function that returns the marginal effect of time dummies
$N(t)$	Function that returns the number of people who did not adopt the iPhone 3G. For $t = 0$ it returns the sample size.
$AVG\_FRD\_ADP(t)$	The sample average for $frd\_adopters_{t-1}$
$EAI(t)$	Expected adoptions that occur due to peer influence
Algorithm	
<pre> <b>for</b> <math>t = 0 \rightarrow T</math> <b>do</b>   <math>EAI(t) = N(t) * (m * AVG\_FRD\_ADP(t) + D(t))</math>   <math>N(t + 1) = N(t) - EAI(t)</math> <b>end for</b> </pre>	

# Appendix F

## Additional Time Partitions for the SIENA Analysis

Table F1. Mean Jaccard Index Across the 263 Communities in the Sample											
Time Span	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN
One Month		0.747 (0.046)	0.765 (0.041)	0.762 (0.042)	0.721 (0.047)	0.719 (0.051)	0.761 (0.045)	0.762 (0.043)	0.761 (0.044)	0.761 (0.045)	0.761 (0.046)
Two Months			0.794 (0.043)		0.798 (0.042)		0.797 (0.039)		0.800 (0.043)		0.762 (0.049)
Three Months					0.815 (0.046)			0.816 (0.041)			0.804 (0.043)
[1; 2] – [3; 11]								0.749 (0.056)			
[1; 3] – [4; 11]								0.794 (0.052)			
[1; 4] – [5; 11]								0.816 (0.050)			
[1; 5] – [6; 11]								0.826 (0.046)			
[1; 6] – [7; 11]								0.828 (0.045)			
[1; 7] – [8; 11]								0.821 (0.045)			
[1; 8] – [9; 11]								0.802 (0.048)			
[1; 9] – [10; 11]										0.765 (0.051)	
[1; 10] – [11; 11]											0.681 (0.058)

**Note 1:** Jaccard Index defined as  $Jaccard(g_0, g_1) = \frac{e_{11}}{e_{11} + e_{10} + e_{01}}$  where  $e_{11}$  denotes the edges that are present in both graphs  $g_0$  and  $g_1$ ,  $e_{10}$  denotes the edges that are only present in  $g_0$ , and  $e_{01}$  denotes the edges that are only present in  $g_1$ . **Note 2:** Standard errors in ( ).

Table F2. Results of the Meta-Analysis Using SIENA for Different Time Partitions

Partition	Name	coeff	stderr	pval	$I^2$	$H^2$	$\tau^2$	Q-Test	Q-Test (p-val)	N obs
[1; 2] – [3; 11]	outdegree (density)	-4.214	0.052	0.000	62.458	2.665	0.169	325.644	0.000	104
	reciprocity	5.014	0.090	0.000	59.418	2.464	0.421	269.346	0.000	104
	transitive ties	2.538	0.045	0.000	56.025	2.274	0.113	263.453	0.000	104
	<i>Behavior</i> → <i>Network</i>	0.011	0.069	0.875	34.931	1.537	0.134	164.549	0.000	104
	<i>Network</i> → <i>Behavior</i>	3.347	0.107	0.000	0.000	1.000	0.000	43.340	0.000	104
[1; 3] – [4; 11]	outdegree (density)	-4.393	0.043	0.000	59.972	2.498	0.167	448.472	0.000	150
	reciprocity	4.772	0.079	0.000	72.366	3.619	0.530	595.384	0.000	150
	transitive ties	2.377	0.035	0.000	52.279	2.095	0.095	351.168	0.000	150
	<i>Behavior</i> → <i>Network</i>	0.056	0.051	0.269	26.986	1.370	0.084	213.184	0.000	150
	<i>Network</i> → <i>Behavior</i>	3.156	0.094	0.000	0.000	1.000	0.000	58.394	1.000	150
[1; 5] – [6; 11]	outdegree (density)	-4.370	0.053	0.000	66.291	2.967	0.204	374.884	0.000	112
	reciprocity	4.107	0.073	0.000	75.442	4.072	0.404	475.923	0.000	112
	transitive ties	1.939	0.036	0.000	47.098	1.890	0.070	230.529	0.000	112
	<i>Behavior</i> → <i>Network</i>	0.037	0.054	0.498	29.333	1.415	0.077	167.707	0.000	112
	<i>Network</i> → <i>Behavior</i>	2.857	0.232	0.000	27.452	1.378	2.014	121.619	0.231	112
[1; 6] – [7; 11]	outdegree (density)	-4.443	0.057	0.000	65.275	2.880	0.214	337.356	0.000	102
	reciprocity	3.896	0.070	0.000	67.843	3.110	0.295	324.878	0.000	102
	transitive ties	1.723	0.042	0.000	53.516	2.151	0.096	228.744	0.000	102
	<i>Behavior</i> → <i>Network</i>	0.069	0.054	0.199	26.291	1.357	0.065	144.909	0.003	102
	<i>Network</i> → <i>Behavior</i>	3.112	0.130	0.000	0.000	1.000	0.000	24.392	1.000	102
[1; 7] – [8; 11]	outdegree (density)	-4.504	0.061	0.000	65.212	2.875	0.226	299.468	0.000	92
	reciprocity	3.777	0.073	0.000	65.190	2.873	0.286	265.460	0.000	92
	transitive ties	1.495	0.042	0.000	48.672	1.948	0.079	186.580	0.000	92
	<i>Behavior</i> → <i>Network</i>	0.054	0.060	0.374	26.053	1.352	0.061	147.948	0.000	92
	<i>Network</i> → <i>Behavior</i>	3.173	0.129	0.000	0.000	1.000	0.000	16.742	1.000	92
[1; 8] – [9; 11]	outdegree (density)	-4.845	0.077	0.000	52.680	2.113	0.189	125.514	0.000	58
	reciprocity	3.935	0.104	0.000	62.745	2.684	0.394	175.865	0.000	58
	transitive ties	1.275	0.053	0.000	35.695	1.555	0.055	90.531	0.003	58
	<i>Behavior</i> → <i>Network</i>	0.204	0.089	0.026	32.069	1.472	0.110	94.797	0.001	58
	<i>Network</i> → <i>Behavior</i>	3.191	0.134	0.000	0.000	1.000	0.000	5.989	1.000	58
[1; 9] – [10; 11]	outdegree (density)	-5.118	0.133	0.000	62.162	2.653	0.339	103.865	0.000	30
	reciprocity	4.072	0.158	0.000	52.273	2.095	0.365	65.433	0.000	30
	transitive ties	1.082	0.084	0.000	51.543	2.064	0.096	64.265	0.000	30
	<i>Behavior</i> → <i>Network</i>	0.138	0.093	0.148	0.001	1.000	0.000	29.795	0.424	30
	<i>Network</i> → <i>Behavior</i>	2.664	0.183	0.000	0.000	1.000	0.000	2.410	1.000	30
[1; 10] – 11	outdegree (density)	-5.539	0.259	0.000	39.885	1.663	0.222	14.597	0.067	9
	reciprocity	4.071	0.329	0.000	31.055	1.450	0.241	13.649	0.091	9
	transitive ties	0.860	0.164	0.001	36.805	1.582	0.066	15.223	0.055	9
	<i>Behavior</i> → <i>Network</i>	0.319	0.167	0.093	0.001	1.000	0.000	7.541	0.480	9
	<i>Network</i> → <i>Behavior</i>	2.051	0.372	0.001	0.000	1.000	0.000	0.431	1.000	9

**Note 1:** *Behavior* → *Network* is captured by the *behaviorsimilarity*. **Note 2:** *Network* → *Behavior* is implemented through *behavioraveagesimilarity*.

**Note 3:** Meta-analysis estimated through maximum likelihood assuming a random effects model with Knapp and Hartung standard error correction.

**Note 4:**  $\tau^2$  is the estimate of the total amount of heterogeneity.  $I^2$  is the percentage of the total variability due to heterogeneity.  $H^2$  is  $\frac{\text{totalavailability}}{\text{samplingvariability}}$ .

**Table F3. Results of the Meta-Analysis Using SIENA with Three Snapshots**

Partition	Name	coeff	stderr	pval	$I^2$	$H^2$	$\tau^2$	Q-Test	Q-Test (p-val)	N obs
[1; 2] – [4; 6] – [7; 11]	outdegree (density)	-4.195	0.038	0.000	69.841	3.316	0.107	392.527	0.000	104
	reciprocity	4.029	0.059	0.000	82.368	5.672	0.290	708.874	0.000	104
	transitive ties	1.833	0.028	0.000	57.796	2.369	0.046	265.849	0.000	104
	<i>Behavior</i> → <i>Network</i>	0.012	0.037	0.748	37.558	1.601	0.052	163.277	0.000	104
	<i>Network</i> → <i>Behavior</i>	3.040	0.111	0.000	0.000	1.000	0.000	43.034	1.000	104
[1; 4] – [5; 8] – [9; 11]	outdegree (density)	-4.320	0.051	0.000	71.312	3.486	0.129	272.047	0.000	73
	reciprocity	3.884	0.072	0.000	81.336	5.358	0.288	419.031	0.000	73
	transitive ties	1.677	0.039	0.000	64.556	2.821	0.067	211.380	0.000	73
	<i>Behavior</i> → <i>Network</i>	0.067	0.042	0.113	25.463	1.342	0.031	97.622	0.024	73
	<i>Network</i> → <i>Behavior</i>	3.009	0.133	0.000	0.000	1.000	0.000	25.731	1.000	73

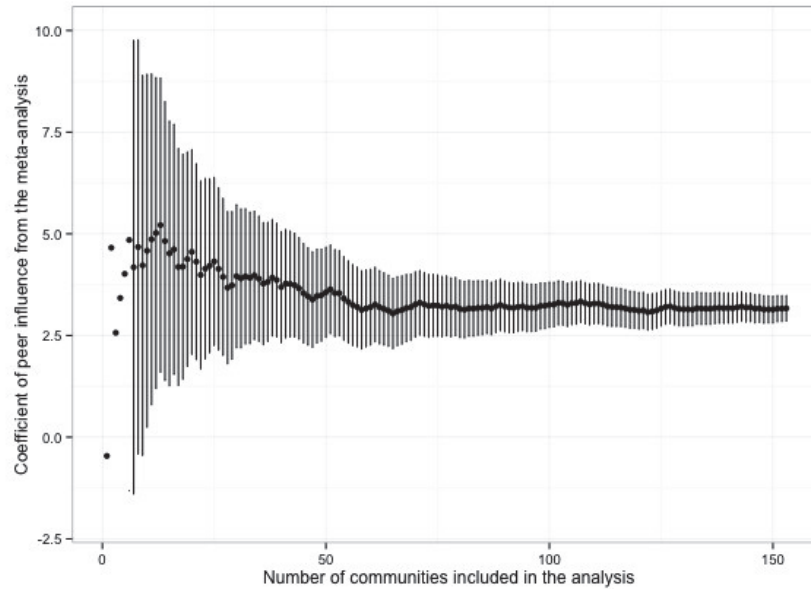
**Note 1:** *Behavior* → *Network* is captured by the *behaviorsimilarity*. **Note 2:** *Network* → *Behavior* is implemented through *behavioraveagesimilarity*.

**Note 3:** Meta-analysis estimated through maximum likelihood assuming a random effects model with Knapp and Hartung standard error correction.

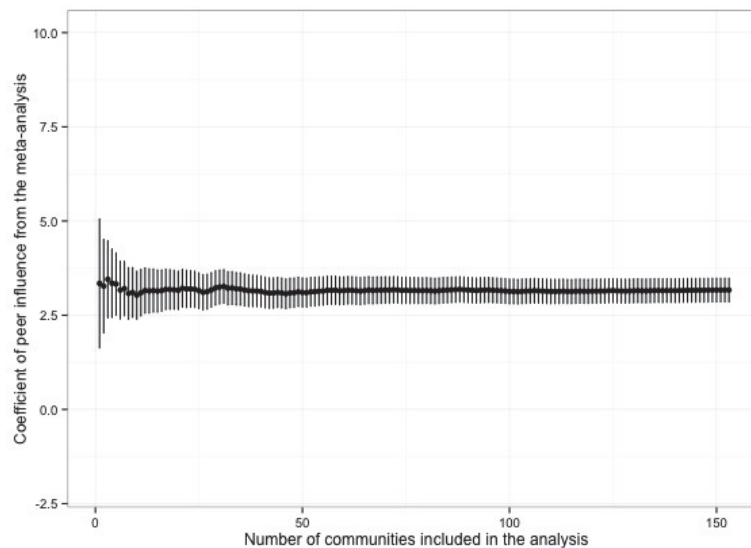
**Note 4:**  $\tau^2$  is the estimate of the total amount of heterogeneity.  $I^2$  is the percentage of the total variability due to heterogeneity.  $H^2$  is  $\frac{\text{totalavailability}}{\text{samplingvariability}}$ .

# Appendix G

## Robustness Checks on SIENA Analysis

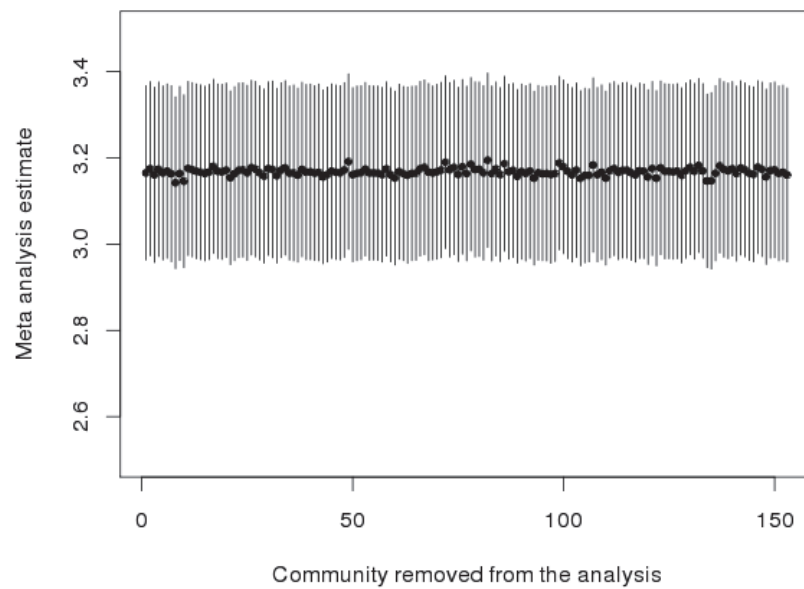


**Figure G1. Cumulative Inclusion of Communities in Decreasing Order of the Standard Error Associated with the Effect of Peer Influence**



**Figure G2. Cumulative Inclusion of Communities in Increasing Order of the Standard Error Associated with the Effect of Peer Influence**





**Figure G3. Effect of Peer Influence Leaving One Community Out of the Analysis at a Time**

# Appendix H

## Community-Level Descriptive Statistics

**Table H1. Definition and Descriptive Statistics for Community-Level Covariates Considered in Our Study**

	avg	sd	min	max	definition
adopters	8.81	6.78	1.00	49.00	Number of adopters.
adoption_rate	0.57	0.43	0.08	3.50	Number of adoptions per month.
months_with_adoption	4.33	2.25	1.00	10.00	Number of months in which adoption took place.
adoption_span	6.83	3.59	0.00	11.00	Time between first and last adoption.
size	91.88	17.14	5.00	115.00	Number of subscribers (N).
edges	211.98	55.72	91.00	430.00	Number of links between subscribers (E).
density	0.05	0.01	0.03	0.11	$2e/(n(N-1))$ .
diameter	9.25	2.82	5.00	19.00	Longest path between subscribers (no cycles).
avg_path_length	4.57	0.86	2.78	7.38	$\sum_i \sum_j p(i,j)/(N(N-1))$ ; $p(i,j)$ length of the shortest path between $i$ and $j$ .
av_degree	4.61	0.82	3.04	8.96	$\sum_i deg(i)/N$ ; $deg(i)$ degree of subscriber $i$ .
avg_betweenness	160.10	57.26	16.65	364.37	$\sum_i bet(i)/N$ ; $g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$ ; $\sigma_{st}$ number of shortest paths between $s$ and $t$ ; $\sigma_{st}(v)$ number of such paths through $v$ .
avg_closeness	0.21	0.06	0.02	0.32	$\sum_i clo(i)/N$ ; $g(v) = \sum_{t \in V \setminus v} 2^{-d_G(v,t)}$ ; $d_G(v,t)$ distance between $v$ and $t$ in $G$ .
centralization.deg	0.12	0.04	0.04	0.28	Measures how central the most central subscriber is relative to all other subscribers $C_x = \frac{\sum_{i=1}^N C_x(p_i) - C_x(p_i)}{\max \sum_{i=1}^N C_x(p_i) - C_x(p_i)}$ . $C_x(p_i)$ is any centrality measure of point $i$ and $C_x(p_*)$ is the largest such measure in the network. Apply this column to the three rows of centralization.
centralizatio.bet	0.54	0.15	0.06	0.85	
centralization.clo	0.25	0.11	0.00	0.46	
transitivity	0.54	0.08	0.32	0.90	Number of closed triplets relative to the number of connected triples to vertices.
transitivity.corr	0.68	0.04	0.45	0.81	Correlation between $W$ and $W^2$ ; $W$ is the adjacency matrix.
cut-points	17.11	5.43	4.00	32.00	Number of subscribers that, if removed, separate the community in more than one component.

**Table H2. Correlation Table Across the Community-Level Covariates Considered in Our Study**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
(1) adopters	1.00																				
(2) adoption_rate	1.00	1.00																			
(3) months_with_adoption	0.85	0.85	1.00																		
(4) adoption_span	0.57	0.57	0.74	1.00																	
(5) size	0.27	0.27	0.32	0.29	1.00																
(6) density	-0.29	-0.29	-0.32	-0.30	-0.75	1.00															
(7) avg_path_length	0.18	0.18	0.09	0.04	0.28	-0.40	1.00														
(8) diameter	0.13	0.13	0.06	-0.00	0.23	-0.29	0.88	1.00													
(9) avg_pagerank	-0.27	-0.27	-0.32	-0.29	-0.98	0.76	-0.28	-0.22	1.00												
(10) avg_coreness	-0.15	-0.15	-0.13	-0.12	-0.00	0.61	-0.24	-0.16	-0.01	1.00											
(11) avg_degree	-0.14	-0.14	-0.11	-0.11	0.06	0.59	-0.30	-0.19	-0.07	0.95	1.00										
(12) avg_betweenness	0.25	0.25	0.21	0.15	0.67	-0.61	0.86	0.73	-0.65	-0.15	-0.16	1.00									
(13) avg_closeness	-0.13	-0.13	-0.12	-0.09	-0.29	0.40	-0.47	-0.47	0.30	0.22	0.27	-0.30	1.00								
(14) centralization_deg	-0.27	-0.27	-0.29	-0.28	-0.42	0.63	-0.45	-0.33	0.43	0.39	0.44	-0.48	0.41	1.00							
(15) centralization_deg	-0.09	-0.09	-0.09	-0.05	-0.02	-0.11	0.09	-0.10	0.01	-0.09	-0.21	.13	0.18	-0.04	1.00						
(16) centralization_clo	-0.09	-0.09	-0.05	0.01	-0.16	0.20	-0.47	-0.52	0.16	0.12	0.13	-0.26	0.92	0.27	0.39	1.00					
(17) transitivity	-0.04	-0.04	-0.05	-0.03	-0.12	0.41	0.08	0.04	0.10	0.68	0.48	-0.01	-0.06	0.02	0.19	-0.04	1.00				
(18) transitivity_cor	-0.06	-0.06	-0.05	-0.04	-0.00	0.22	0.04	0.01	-0.01	0.50	0.32	0.02	-0.08	-0.14	0.16	-0.03	0.78	1.00			
(19) cut_points	0.41	0.41	0.35	0.33	0.53	-0.71	0.56	0.44	-0.52	-0.41	-0.48	0.65	-0.38	-0.50	0.15	-0.21	-0.07	0.08	1.00		
(20) cut_points_ff_avg_deg	-0.11	-0.11	-0.10	-0.11	0.13	0.31	-0.45	-0.38	-0.13	0.56	0.67	-0.20	0.42	0.49	-0.24	0.29	0.04	0.06	-0.38	1.00	
(21) logic(cliques_min_size_3)	-0.03	-0.03	-0.01	-0.01	0.21	0.38	-0.12	-0.08	-0.23	0.91	0.87	0.04	0.11	0.31	-0.05	0.05	0.59	0.39	-0.19	0.50	1.00

# Appendix I

## Policy Simulator Code

Our simulator evolves a graph object over time of the form  $G(t) = (S(t), W)$  where  $S(t)$  is a set of subscribers and their characteristics and  $W$  is a fixed adjacency matrix. In our setting  $S(t, i) = (X_p, Z_{i,t}, BPA_{i,p}, PA_{i,p}, AO_{i,p}, A_{i,i})$  is a data structure for subscriber  $i$  at time  $t$ .  $BPA(t, i)$  represents the baseline propensity for subscriber  $i$  to adopt the iPhone 3G at time  $t$ . This is given by  $\Phi(\hat{\alpha} + X_i\hat{\beta} + Z_{i,t}\hat{\gamma} + \hat{\rho}W_iG(t-1).S(i \in S).AO)$  and thus introduces heterogeneity across consumers and evolves over time.  $G(t)$  is the graph at time  $t$ ,  $S(i)$  indexes subscriber  $i$  in that graph and  $AO$  is the adoption observed for that subscriber in our dataset.  $S(i \in S)$  refers to a vector of all subscribers in  $S$ . The probability of subscriber  $i$  adopting the iPhone 3G at time  $t$  is given by the sum of  $BPA(t, i)$  and the effect that additional adopters exert on her, computed as  $mW_i(G(t-1).S(i \in S).A - G(t-1).S(i \in S).AO)$ , where  $A$  represents whether the subscriber adopts the iPhone 3G in our simulation.  $A$  is determined using a random draw from a uniform distribution in  $[0, 1]$ .

**Table 11. Pseudocode for Policy Simulator****Variables**

$n$	number of seeds
$m$	marginal peer influence effect
$S$	set of subscribers
$\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\rho}$	parameters from probit estimation
$W, X_i, Z_{it}$	covariates in equation 1
$AO(i, t)$	adoptions generated in our dataset
$A(i, t)$	adoptions generated in simulation
$BPA(i, t)$	baseline probability of adoption
$PA(i, t)$	probability of adoption

**Algorithm**

```

function AA( $n, m, G$ ) – Additional adopters
   $G(0).S(i \in I).A = \text{seedingpolicy}(n, G(0).S(i \in I))$ 
  for  $t = 0 \rightarrow T, i \in S$  do
     $G(t).S(i \in I).BPA = \Phi(\hat{\alpha} + X_i\hat{\beta} + \hat{\gamma}Z_{it} + \hat{\rho}W_iG(t-1).S(i \in I).AO)$ 
  end for
  for  $t = 0 \rightarrow T, i \in S: A(i, t) == ()$  do
     $G(t).S(i).PA = G(i).S(i).BPA + mW_i(G(t-1).S(i \in I).A - G(t-1).S(i \in I).AO)$ 
     $G(t).S(i).A = 1 \{draw U(0,1) > G(t).S(i).PA\}$ 
  end for
  return  $\sum_{i \in S} G(T).S(i).A$ 
end function

```

**Reference**

Albert, A., and Anderson, J. A. 1984. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika* (71:1), pp. 1-10.