MIS Quarterly

RESEARCH ARTICLE

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 = SubscribersLowestIE R(H , A)
            H = Subgraph(H ,V \  {v ∈ V chosen at random })
            H = Com ponentH ighestIE R(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 = Com ponents(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.

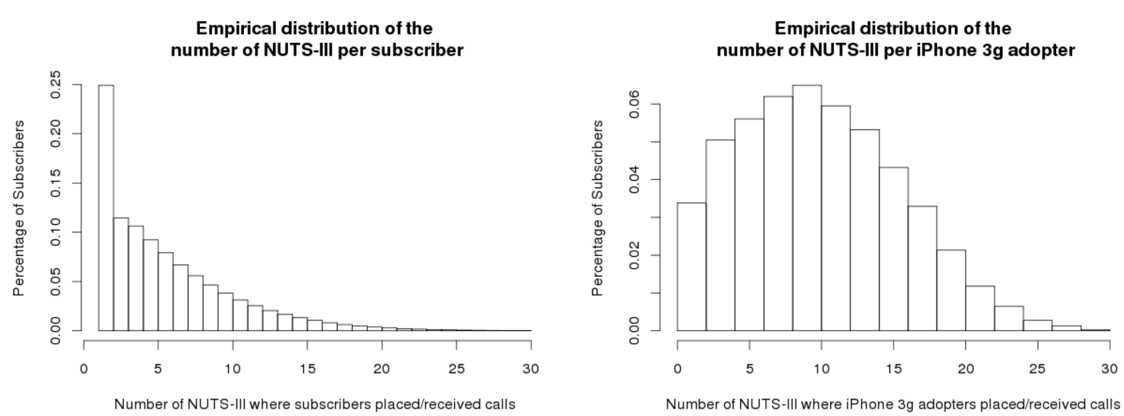

Figure B1. Number of NUTS-III Codes Where Subscribers Received or Placed Calls from August 2008 until July 2009

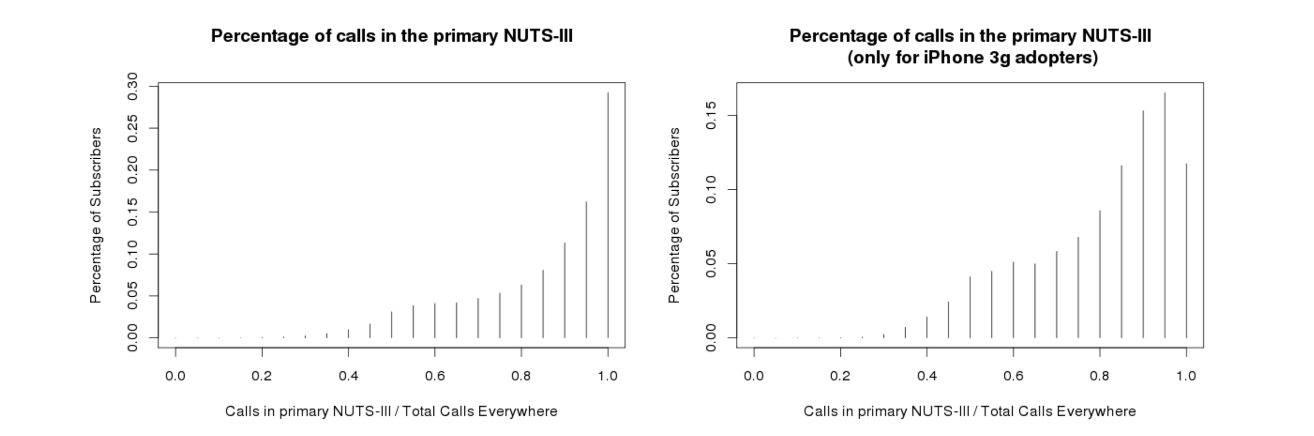

Figure B2. Proportion of Phone Calls Within the Primary NUTS-III Region from August 2008 until July 2009

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

	Adoption		
Time	No	Yes	Total
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 iPhone3G. 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 *adopted$_t$*	(2) IV Probit *adopted$_t$*
frd_adopters$_{t-1}$	2851*** (0258)	9.935*** (1.952) [2.087]
Log(*tenure*$_{t+1}$)	0.300*** (0.101)	0.329*** (0.0955) [0.103]
Log(*tenure*$_{t+1}$)^2	-0.0330** (0.0138)	-0.0370*** (0.0127) [0.0139]
prepaidY	-0.548*** (0.0283)	-0.508*** (0.0274) [0.0307]
genderF	-0.0616* (0.0343)	-0.0572 (0.0332) [0.0351]
genderM	0.144*** (0.0271)	0.1433*** (0.0261) [0.0275]
mobileNetY	0.425*** (0.0427)	0.398*** (0.0397) [0.0445]
phone2.5g	0.484*** (0.0547)	0.469*** (0.0536) [0.0581]
phone3.0g	0.637*** (0.0542)	0.628*** (0.0537) [0.0596]

Variables	(1) Probit $adopted_t$	(2) IV Probit $adopted_t$
phone3.5g	0.917*** (0.0658)	0.876*** (0.0650) [0.0735]
phoneOther	0.575*** (0.142)	0.503*** (0.140) [0.162]
phoneAge	-0.143*** (0.0551)	-0.152*** (0.0512) [0.0601]
phoneAge²	0.0256 (0.0251)	0.0279 (0.0240) [0.0273]
geoWageH	0.133*** (0.0431)	0.113*** (0.0381) [0.0420]
geoWageL	-0.146 (0.142)	-0.111 (0.150) [0.150]
geoWageVH	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^2	0.157	
Log Lik	-8721	

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

***p < 001; **p < 0.05; *p < 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	

	for *t* = 0 → *T* **do**
	EAI(*t*) = *N*(*t*) * (*m* * *AVG_FRD_ADP*(*t*) + *D*(*t*))
	N(*t* + 1) = *N*(*t*) – *EAI*(*t*)
	end for

Appendix F

Additional Time Partitions for the SIENA Analysis

<table>
<tr><th colspan="12">Table F1. Mean Jaccard Index Across the 263 Communities in the Sample</th></tr>
<tr><th>Time Span</th><th>AUG</th><th>SEP</th><th>OCT</th><th>NOV</th><th>DEC</th><th>JAN</th><th>FEB</th><th>MAR</th><th>APR</th><th>MAY</th><th>JUN</th></tr>
<tr><td>One Month</td><td></td><td>0.747
(0.046)</td><td>0.765
(0.041)</td><td>0.762
(0.042)</td><td>0.721
(0.047)</td><td>0.719
(0.051)</td><td>0.761
(0.045)</td><td>0.762
(0.043)</td><td>0.761
(0.044)</td><td>0.761
(0.045)</td><td>0.761
(0.046)</td></tr>
<tr><td>Two Months</td><td colspan="2"></td><td colspan="2">0.794
(0.043)</td><td colspan="2">0.798
(0.042)</td><td colspan="2">0.797
(0.039)</td><td colspan="2">0.800
(0.043)</td><td>0.762
(0.049)</td></tr>
<tr><td>Three Months</td><td colspan="3"></td><td colspan="3">0.815
(0.046)</td><td colspan="3">0.816
(0.041)</td><td colspan="2">0.804
(0.043)</td></tr>
<tr><td>[1; 2] – [3; 11]</td><td colspan="2"></td><td colspan="9">0.749
(0.056)</td></tr>
<tr><td>[1; 3] – [4; 11]</td><td colspan="3"></td><td colspan="8">0.794
(0.052)</td></tr>
<tr><td>[1; 4] – [5; 11]</td><td colspan="4"></td><td colspan="7">0.816
(0.050)</td></tr>
<tr><td>[1; 5] – [6; 11]</td><td colspan="5"></td><td colspan="6">0.826
(0.046)</td></tr>
<tr><td>[1; 6] – [7; 11]</td><td colspan="6"></td><td colspan="5">0.828
(0.045)</td></tr>
<tr><td>[1; 7] – [8; 11]</td><td colspan="7"></td><td colspan="4">0.821
(0.045)</td></tr>
<tr><td>[1; 8] – [9; 11]</td><td colspan="8"></td><td colspan="3">0.802
(0.048)</td></tr>
<tr><td>[1; 9] – [10; 11]</td><td colspan="9"></td><td colspan="2">0.765
(0.051)</td></tr>
<tr><td>[1; 10] – [11; 11]</td><td colspan="10"></td><td>0.681
(0.058)</td></tr>
</table>

Note 1: Jaccard Index defined as $Jaccard(g0, g1) = \frac{e_{11}}{e_{11}+e_{10}+e_{01}}$ where e_{11} denotes the edges that are present in both graphs $g0$ and $g1$, e_{10} denotes the edges that are only present in $g0$, and e_{10} denotes the edges that are only present in $g1$. **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	τ^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
	Behavior ➔Network	0.011	0.069	0.875	34.931	1.537	0.134	164.549	0.000	104
	Network ➔ Behavior	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
	Behavior ➔Network	0.056	0.051	0.269	26.986	1.370	0.084	213.184	0.000	150
	Network ➔ Behavior	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
	Behavior ➔Network	0.037	0.054	0.498	29.333	1.415	0.077	167.707	0.000	112
	Network ➔ Behavior	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
	Behavior ➔Network	0.069	0.054	0.199	26.291	1.357	0.065	144.909	0.003	102
	Network ➔ Behavior	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
	Behavior ➔Network	0.054	0.060	0.374	26.053	1.352	0.061	147.948	0.000	92
	Network ➔ Behavior	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
	Behavior ➔Network	0.204	0.089	0.026	32.069	1.472	0.110	94.797	0.001	58
	Network ➔ Behavior	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
	Behavior ➔Network	0.138	0.093	0.148	0.001	1.000	0.000	29.795	0.424	30
	Network ➔ Behavior	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
	Behavior ➔Network	0.319	0.167	0.093	0.001	1.000	0.000	7.541	0.480	9
	Network ➔ Behavior	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**: τ^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 $\frac{totalavailability}{samplingvariabiity}$.

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

Partition	Name	coeff	stderr	pval	I^2	H^2	τ^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
	Behavior ➔Network	0.012	0.037	0.748	37.558	1.601	0.052	163.277	0.000	104
	Network ➔ Behavior	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
	Behavior ➔Network	0.067	0.042	0.113	25.463	1.342	0.031	97.622	0.024	73
	Network ➔ Behavior	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**: τ^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 $\frac{totalavailability}{samplingvariabiity}$.

Appendix G

Robustness Checks on SIENA Analysis

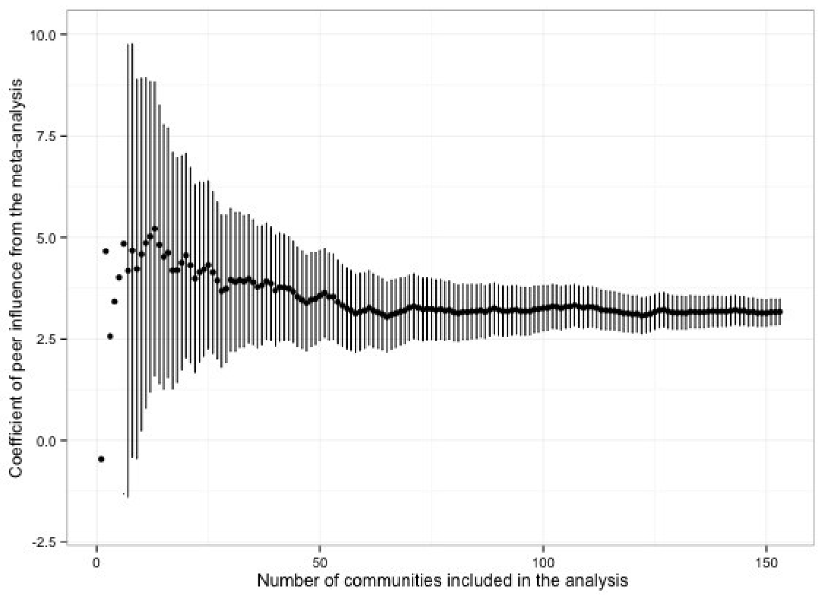

Figure G1. Cumulative Inclusion of Communities in Decreasing Order of the Standard Error Associated with the Effect of Peer Influence

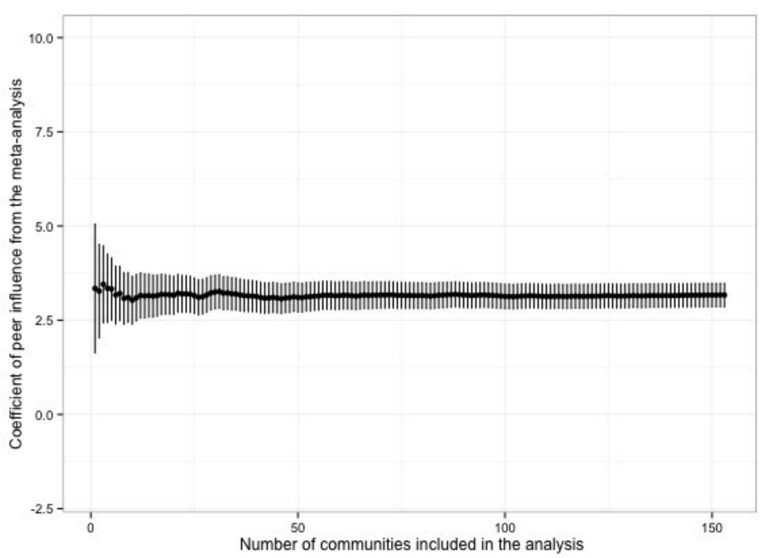

Figure G2. Cumulative Inclusion of Communities in Increasing Order of the Standard Error Associated with the Effect of Peer Influence

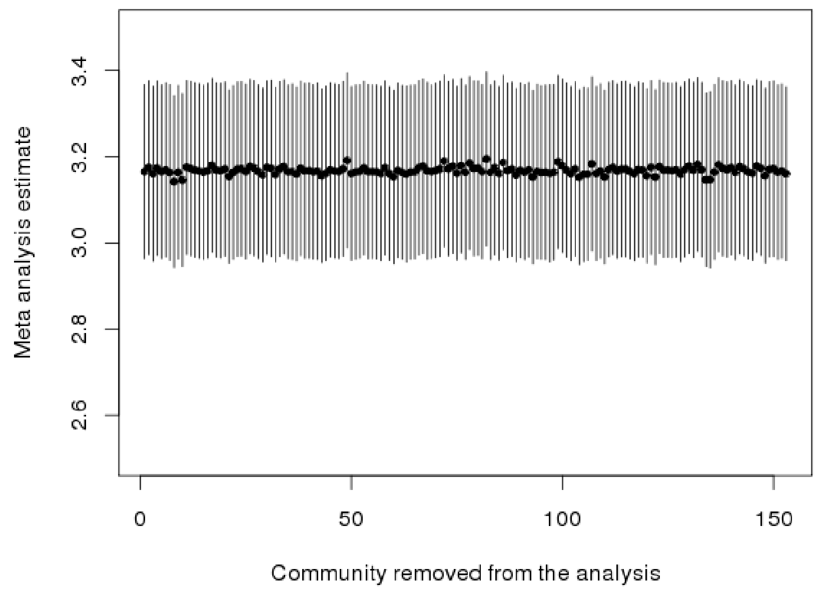

Figure G3. Effect of Peer Influence Leaving One Community Out of the Analysis at a Time

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.
cengralization.deg	0.12	0.04	0.04	0.28	Measures how central the most central subscriber is relative to all other subscribers $C_x = \frac{\sum_{i=1}^{N} C_x(p_*)-C_x(p_i)}{\max \sum_{i=1}^{N} C_x(p_*)-C_x(p_i)}$. $C_x(p_i)$ 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	-005	-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	-038	-0.50	0.15	-0.21	-0.07	00.08	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 $S(t, i) = (X_i, Z_{i,t}, BPA_{t,i}, PA_{t,i}, AO_{t,i}, A_{t,i})$ is a data structure for subscriber i a t time t. $BPA(t, i)$ represents the baseline propensity for subscriber i to adopt the iPhone 3G at time t. This is given by $\Phi\left(\hat{\alpha} + X_i\hat{\beta} + Z_{i,t}\hat{\gamma} + \hat{\rho}W_iG(t-1).S(i \in S).AO\right)$ 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 I1. 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 $\mathrm{AA}(n, m, G)$ – Additional adopters
  $G(0).S(i \in I).A$ = *seedingpolicy*$(n, G(0).S(i \in I))$
  for $t = 0 \rightarrow T, i \in S$ **do**
    $G(t).S(i \in I).BPA = \Phi\left(\hat{\alpha} + X_i\hat{\beta} + \hat{\gamma}Z_{it} + \hat{\rho}W_iG(t-1).S(i \in I).AO\right)$
  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.