Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network

Ravi Bapna
University of Minnesota, Carlson School of Management, rbapna@umn.edu

Akhmed Umyarov
University of Minnesota, Carlson School of Management, aumyarov@umn.edu

Demonstrating compelling causal evidence of the existence and strength of peer to peer influence has become the holy grail of modern research in online social networks. In these networks, it has been consistently demonstrated that user characteristics and behavior tend to cluster both in space and in time. There are two well known rival mechanisms that compete to be the explanation for this observed clustering: peer influence and homophily. Both mechanisms lead to similar observational data, yet have tremendously different policy implications. In this paper, we present a novel randomized experiment that tests the existence of causal peer influence in the general population of a particular large-scale online social network and quantifies its strength as compared to homophily. We utilize a unique social feature to exogenously induce adoption of a paid product amongst a group of randomly selected users, and in the process develop a clean exogenous randomization of treatment and control groups. Our estimates show that peer influence causes more than 50% increase in odds of buying the product due to the influence coming from an adopting friend. In addition, we find that users with smaller number of friends are significantly more susceptible to be influenced by their peers as compared to the ones with larger number of friends. Finally, our experimental apparatus allows us to compare the strength of peer influence versus homophily as well as to compare our randomized trial with a standard matching-based quasi-experiment on the same data. We find that peer influence is a stronger force in the setting of our experiment as compared to homophily and that the quasi-experiment tends to produce results similar to the randomized trial, but over-estimating the effect on users with larger number of friends and under-estimating it for users with smaller number of friends.

Key words: Peer-effects, randomized experiment, social contagion, matching models, freemium communities, online social networks

---

1. Introduction and Background

The general challenge of demonstrating causal inference from observational data has been immortalized in Manski (1995) reference to the simultaneous movements of a man and his image in the mirror. He asks, "Does the mirror cause the man's movement or reflect them?" and concludes that without understanding optics and human behavior we cannot really tell. Interestingly, this quote from pre-Facebook era is extremely relevant to the causality questions that arise in today's digital age featuring massive online social networks such as Facebook and Twitter as well as niche networks such as Last.fm, Spotify, LinkedIn and others[1]. These online social networks are credited with playing roles that range from inspiring political action to driving viral and word-of-mouth spread of products and services (Aral and Walker 2011, Hill, Provost, and Volinsky 2006, Iyengar, Van den Bulte, and Valente 2011, Manchanda, Xie, and Youn 2008, Mayzlin 2006), and as such, represent a vast reservoir of social and economic influence. Central to the tapping into this reservoir is the understanding of causal relationships that drive the spread of products, services and information over these social networks, the focus of this paper.

It has been consistently demonstrated in the literature that in online social networks user characteristics and behavior tend to cluster both in space and in time, with users generally being similar to their online friends and acting similar to their online friends (Aral and Walker 2011). Interestingly, there are several different underlying causal mechanisms that can lead to this observed clustering with the most frequently cited ones being *peer influence* and *homophily*. Under the mechanism of *peer influence*, an individual causes her online friends to undertake a certain action, which in turn, leads to the observed correlation of the behavior of online friends.

On the other hand, under the mechanism of *homophily*, an individual tends to befriend peers that are similar to her on observed and unobserved characteristics and possibly the environment that they face. In this case, it is not surprising that behavior of an individual is correlated with the behavior of her friends: they may not influence each other at all, but the observed correlation of their actions comes from their intrinsic similarity. This underlying similarity is what forces them to make similar choices independently again and again, and therefore, we may observe the correlation between actions of online friends in case of homophily as well.

The importance of *disentangling* peer influence and homophily mechanisms stems from the fact that despite leading to very similar observational data, the policy implications of each of these mechanisms are vastly different. Under peer influence an effective policy may be to identify the "most influential" people and induce the desired behavior among them so that it would propagate through the social contagion, while under homophily mechanism this policy may have little effect (Aral 2011). Instead, under homophily, a careful segmentation based targeting strategy might be preferred. Moreover, the mechanisms of peer influence and homophily are not necessarily mutually exclusive and may complement each other, therefore social contagion processes in real online networks may contain a complex mixture of peer-influence and homophily.

Importantly, peer influence has the added bonus of bringing with it *the social multiplier effect*. Manski (1995) provides an intuitive example of that effect describing a potential positive feedback

[1] Online social networks such as Facebook, with a billion users, and Twitter, with more than 100 MM users, are consuming an increasingly significant portion of our time and attention. A 2010 Nielsen study estimated that the amount of time the average user spent on Facebook was about seven hours per month, and more importantly, was growing at the rate of 10% per month.

loop of peer influence in the context of academic performance of high school students. Manski (1995) posits that if an increase in individual student's academic performance causes the increase in the performance of the reference group of her peers, then this reference group may in turn increase the performance of that individual even further, and so on, leading to a positive self-reinforcing feedback loop with the social multiplier effect. On the other hand, homophily-based mechanisms that arise out of similarity of individual characteristics or contextual information do not typically exhibit this multiplier effect, perhaps explaining the fascination amongst researchers and practitioners about viral marketing of products and services.

All these factors make it critical, both for theory and for practice, to causally identify the presence of each of these mechanisms in the context of large-scale online social networks. It is fair to say that causal identification and measurement of peer influence in the general population of online social networks, or put simply, existence and strength of social contagion, has become the holy grail of modern research in online social networks.

In this paper, we present a novel randomized experiment that identifies the existence of peer influence in the general population of users of a particular online social network. Our work is inspired by Aral and Walker (2011) who demonstrate that significant social contagion can be created by embedding viral features into product design as well as by Aral and Walker (2012) who identify characteristics of influential and susceptible people by using the experimental data. This stream of work showcases the great potential of using randomized field experiments to study peer-effects in online social networks.

For our study, we selected the context of a *freemium* social network (Anderson 2008) called Last.fm.[2] A singular problem for the long-term viability of freemium communities is to convert free users into premium subscribers, the latter being far more profitable (Oestreicher-Singer and Zalmanson 2010). In the case of Last.fm, based on 2009 numbers, free users yield approximately 12 cents per registered user in the network per month, as opposed to paid subscribers who are almost 24 times more valuable, paying $3 per month. Interestingly, observational data that we collected from Last.fm website revealed that a) premium subscribers are extremely rare, accounting for only one percent of all users, and b) these premium subscribers are significantly more likely to be socially connected to other premium subscribers even controlling for the number of friends and other known covariates. However, as explained by Manski (1995), inferring the presence of peer-influence from this is not judicious. More specifically, there are several sorts of biases identified in making such an inference: these include simultaneity (Godes and Mayzlin 2004), unobserved heterogeneity (Van den

[2] Last.fm is a classical example of a freemium community featuring a large number of free users and a small number of premium subscribers. As is typical for a freemium community, the premium users bring in a disproportionately large share of company profits `http://www.guardian.co.uk/media/2010/dec/03/lastfm-protit-2009-figures`

Bulte and Lilien 2001), homophily (Aral, Muchnik, and Sundararajan 2009), and correlated effects (Manski 1995). While multiple attempts have been made at identifying peer effects using network structure based instrument variables (Bramoullé, Djebbari, and Fortin 2009, Oestreicher-Singer and Sundararajan 2010), natural experiments (Tucker 2008) and matched sample counterfactuals (Aral, Muchnik, and Sundararajan 2009, Susarla, Oh, and Tan 2012, Oestreicher-Singer and Zalmanson 2010), each method has its limitations (Aral 2011, Manski 1995) and the best we have in the absence of randomized or controlled exogenous variation are upper bounds of peer influence (Aral, Muchnik, and Sundararajan 2009).

Interestingly, Manski (1995) touches upon the possible reasons behind the lack of randomized trials involving general populations of different real-world networks. He reminds the reader that it is particularly harder to draw inference about general population from a self-selected sample of recruited subjects. In addition to self-selection bias, Manski (1995) argues that generalizable analysis is limited to the observations that are made without undue *intrusion*, since people behavior may change when they know they are being observed. In this study, we hope to demonstrate how our research pushes the frontier on these and other dimensions.

We present our findings starting with the insights gained from observational data followed by the analysis of randomized experiment. We also touch upon a series of simulated quasi-experiments that provide insights into the nature of the effect of self-selection bias. The randomized experiment demonstrated that new adoptions were significantly higher in the treatment group vs. control group. Moreover, both t-test and logistic regression estimates indicate that, on average, the odds of adopting the paid subscription by a user increase by more than 50% due to peer influence when her friend adopts it, indicating significant causal peer-effects in the monetization of social networks. In addition, we find that the peer influence can be significantly weakened by the size of the influenced user's friendship circle. Finally, we compare the strength of peer influence versus homophily in our setting and provide point estimates of each. In addition to that we briefly mention matching-based quasi-experiment. Quasi-experiment tends to produce results similar to the randomized trial, somewhat over-estimating the effect on users with larger number of friends and under-estimating it for the users with smaller number of friends, providing the first insights about the nature of bias in estimating peer effects by the models with self-selected populations.

To appreciate the economic significance of 50% increase in adoption rates due to peer influence, consider that while premium subscriptions are a rare event in the context of Last.fm, this 1% of paid users is very valuable, contributing to more than 18% of the site's total revenue (based on 2009 numbers that we had access to). With the modern social networks having over 1 billion users, a 50% increase in this kind of a "rare" event constitutes an absolute increase of *several million people*. Further, due to our pure randomization, the influencers in our study were just regular

1. *Novel identification strategy* that establishes peer influence for the general population:
   - We *avoid any voluntary recruitment procedures* by utilizing a unique "gifting" feature and selecting subjects completely at random from the full general population. Therefore, we eliminate the notorious subject self-selection bias when individuals who tend to respond to subject recruitment ads may be systematically different[a] from the target population (Camerer and Lovallo 1999, Harrison et al. 2009).
   - *Our manipulation is non-intrusive*, that is, subjects are completely unaware of being a part of the experiment at all, avoiding any observer bias.
   - In our study, *subjects cannot withdraw from the experiment and cannot escape our manipulation*, avoiding any possible subject mortality bias.
2. *Outcome is a real purchase decision with real money.* In our study, peer influence is established for economic transactions involving real money, because each observed outcome is a monetary transaction for $3 and subjects must actually pay their own money to buy the subscription. It is well established that customers may approach free products in a very different way than even the cheapest $0.01 products, basically acting under two different regimes: social norms versus market norms (Ariely 2010).
3. *Quantification of peer influence versus homophily.* Our experimental design allows us to provide a point estimate for the strength of peer influence as compared to the strength of homophily. We accomplish it by comparing natural adopters, who bring with them confounded signals of peer influence and homophily with "induced" adopters from our treatment that carry the pure peer influence signal only. We are also able to provide insights into the nature and extent of the bias that matching procedures may inflict when analyzed using quasi-experimental techniques such as propensity score matching (Rosenbaum and Rubin 1983) that attempt to match based only on the observables. This is achieved using simulated quasi-experiments prior to the experiment.

[a] While the observed characteristics of these individuals can be accounted for by post-stratification, it is hard to account for unobserved characteristics that may be systematically biased and thus confound the study.

Figure 1 Summary of major contributions of our work

average users who we show exert significant peer influence on their friends. They are not some special elite influential users. It is a separate question for subsequent studies to discover how much more influence we would get if we were to use a biased sample to target influential people rather than an unbiased sample of average random people.

As summarized in Figure 1, our study contributes to the vast literature on peer influence in social networks in a variety of important directions. First, our experiment eliminates self-selection bias in the subject recruitment procedure and therefore allows us to make direct inference about general population. As explained in (Camerer and Lovallo 1999, Harrison et al. 2009, Aral and Walker 2011), self-selection of subjects into an experiment can potentially lead to biased results.

In addition to that, being a rare study that looks at economic transactions with real money rather than, say, re-tweeting of information, we contribute by testing whether peer effects span beyond free products or re-tweeting. As discussed by Shampanier, Mazar, and Ariely (2007) and

Ariely (2010), individuals act under two very different regimes when facing a free product versus a paid product: a regime of social norm versus a regime of market norm.

Lastly, our experimental design allows us to directly quantify the strength of peer influence versus the strength of homophily as well as to measure the bias of traditional observational "matching" methods in estimating peer influence as compared to the true experimental data.

To summarize, this study serves as a bridge in the scientific literature bringing the phenomenon of peer influence deep into the general population empirically and connecting peer influence to economic transactions with real money. This study reinforces and backs up the evidence from prior studies of self-selected experimental subjects and free products, supporting the idea that online peer influence constitutes a fundamental phenomenon rather than a peculiarity of socialization between a special set of experiment participants and their friends.

2. Institutional Details

The music industry today serves as a canonical example of how a long-established, growing and profitable industry can be disrupted and subsequently re-invented by the social machinery of Internet. One of the important emerging models of today's content consumption in the Internet is a *freemium* social community (Anderson 2008), as exemplified by sites such as Last.fm, Pandora, Spotify and many others. *Freemium* social communities typically operate based on a two-tiered business model that offers free access to the basic set of features and content while charging a fee for more advanced premium features. For example, free users of Last.fm[3] website can listen to the online music radio interrupted by commercials, while paid subscribers of Last.fm website enjoy continuous commercial-free music listening experience, a prestigious black "Subscriber" icon next to their user avatars that is visible to everyone on Last.fm as a sign of status, can listen to the online radio on a mobile phone and have access to additional colorful statistical charts.

Freemium communities often employ numerous social computing features (Parameswaran and Whinston 2007), such as, for example, a *friendship social network* feature that allows website users to become listed as *online friends* with another website user. Being an online friend with someone typically gives certain benefits: friends can easily share information among themselves and exert certain *peer influence* on each other. On Last.fm website, for instance, online friends can affect each other's music choices while sharing their own music listening experiences, they can listen to friend's "recommended radio", can review friend's "Loved songs" and so on. Appendix A provides a snapshot of a typical Last.fm user's page. Oestreicher-Singer and Zalmanson (2010) provide a nice overview of the institutional details of Last.fm website as a freemium social community.

[3] http://virtualmusic.tv/2011/02/2010-music-website-heat-map/ indicates that Last.fm, with reportedly 30 million subscribers, received 9.8 million hits per month in 2010.

Among the findings of their study is the fact that the music listening on Last.fm is socially driven which means it is based on what your friends are listening, and that a paid subscription appears as a distinct (ostensibly status) symbol visible to your friends. Also, as discussed in the studies of freemium communities (Oestreicher-Singer and Zalmanson 2010, Pauwels and Weiss 2008), a singular challenge for their long-term economic viability is discerning pathways and strategies for moving users *from-free-to-fee*, that is converting users from the large pool of free users to the elite set of premium paid subscribers.

In this paper, we present a randomized field experiment providing the evidence that making one person a premium subscriber can cause her online friends to pay for subscription and become premium subscribers as well. We chose Last.fm as a domain for conducting our experiment not only because it provides a typical example of a freemium community, a new and growing model of delivery of online services, but also because Last.fm makes for a unique experimental platform thanks to a unique social feature that allows gifting any random user in Last.fm social network with a premium subscription (paid by us). While this feature of Last.fm website has not yet been studied extensively in the social networks literature, it offers a great opportunity to create a "gold standard" randomized trial in an online social network. From an experimental design perspective *anyone in Last.fm social network has an equal chance of receiving a gift from us. Last.fm users cannot decline the gift or hide their subscription status from others. They cannot transfer the gift to anyone else, or postpone using it, or share it with someone else, or refund it.* This makes the unrestricted gifting social feature particularly valuable for online social networks in an experimental context, a fact this research is the first to bring forth.

3. Research Questions

The main research questions of this study are formulated as the following hypotheses:

HYPOTHESIS 1. *In an online social network there exists peer influence such that an individual's product adoption causes the adoption by her online friends.*

HYPOTHESIS 2. *The effect of peer influence is moderated by and is decreasing in the number of friends the influenced individual has.*

While the first hypothesis is the focal point of this paper and its rationale has been articulated at length already, it is worth dwelling a bit on the basis for the second hypothesis. Iyengar, Van den Bulte, and Valente (2011) make a compelling case for looking at moderating factors that may shape the nature and extent of social contagion at work. While the focus of many studies such as Godes and Mayzlin (2009) is on the influencer side of the equation: whether better connected adopters exert more influence than do less connected ones, we position ourselves on the susceptibility-to-influence-side of that equation, since peer influence also depends on the susceptibility of the individual being influenced. A user who has 1,000 friends on Last.fm may not even notice or care that one of her peers purchased a premium subscription. At the same time, a user who only has two friends may be more selective in befriending others and may pay closer attention to them. Therefore, this user is more likely to notice and follow the actions of just one manipulated friend. Similar distinctions between selective and non-selective tie forming behaviors in the context of trust have been observed in other online social networks such as Facebook (Bapna, Gupta, Rice, and Sundararajan 2012).

In order to address our research questions we first need to establish a causal link between person's B decision to subscribe and the influence from B's friend - person A. In this paper, our conceptualization of *peer influence* is due to Aral (2011). This conceptualization is rooted in utility theory in that the actions of one's peers change the utility one expects to receive from engaging in a certain behavior and thus the likelihood that one will engage in that behavior (Aral 2011). Such a conceptualization is flexible and encompassing with respect to the myriad influence mechanisms that could lead to social contagion. In other words, in order to demonstrate the presence of *peer influence* we do not seek to explain which influence mechanism from person A *causes* person B to subscribe: be it awareness raising, explicit or tacit persuasion, observational or social learning, imitation or any other mechanism. It is only required to demonstrate that person A causes person B to subscribe. Therefore, in this study, we do not raise the question of disentangling the general peer influence into the exact types of peer influence mechanisms as above. This disentanglement would definitely constitute a very different question than the one raised in this study and would require collecting a different type of data.

Our work relates to and builds upon the propensity score (Rosenbaum and Rubin 1983) matching based approaches of Aral, Muchnik, and Sundararajan (2009), Susarla, Oh, and Tan (2012) as well as Oestreicher-Singer and Zalmanson (2010). A key advancement of our work is that while propensity score matching accounts for observable user characteristics in crafting usable control groups, it is widely recognized (Aral, Muchnik, and Sundararajan 2009, Oestreicher-Singer and Zalmanson 2010) that other unobservable user characteristics (say amount of free-time an individual has, income level, sensitivity to commercials etc) or contextual effects such as marketing promotions (Van den Bulte and Stremersch 2004) could as well be influencing the propensity to be treated and be linked to homophily.

This limitation of not accounting for unobserved characteristics is overcome in our study through exogenous randomization such that there is no reason to believe that the treatment group and the control group (described in the next section) should have any systematic difference in observable and latent/unobservable characteristics. In the absence of randomization, the best we can get are upper bounds of the true estimate of contagion (Aral and Walker 2011).

4. Experimental Design

For illustrative purposes, we present the following intuitive explanation of our research approach before we describe the actual experimental setup using strict formalism. We will consistently rely on that illustration throughout the paper in order to convey abstract concepts more intuitively. Assume that the paid subscription in Last.fm social network is like a "disease" caused by a virus, albeit a benevolent one. We call this the U1B1-B virus[4]. Our observational data shows that people sick with this virus tend to be friends with other sick people, but this alone is not the evidence that the "disease" is contagious. This clustering could easily be explained by the fact that people tend to befriend people who are of similar "age" and in a similar "health" condition and therefore belong to the same "health" risk group and are equally likely to catch the U1B1-B virus from the environment (rather than from a peer), causing the observed clustering. Therefore, the question of our experiment would be: is the U1B1-B "subscription disease" contagious or is it just caught from the "environment" by cliques of people who are in "poor health"?

For the experiment, we will randomly select the *manipulated group* M of 1000 Last.fm users who will be chosen to receive the subscription gifts, which is akin to getting randomly infected by the U1B1-B virus, over which they have no control, ruling out any self-selection, and individual characteristics or contextual (observed or unobserved) homophily-based decisions that confound the analysis of observational data. We will also randomly select the *not-manipulated group* NM of 1000 random Last.fm users who do not get "infected" by us.

After a period of time, we compare the occurrence of the "disease" among the friends of group M and friends of group NM. Given the initial uniform randomization of groups M and NM, both observed and unobserved statistical properties of M and NM are expected to be statistically identical before the manipulation. Therefore, if any statistical difference is observed in the outcomes among friends of M and friends of NM groups, this difference should be attributed to our manipulation.

4.1. Formal Design

At the onset of the study we collected Last.fm social network data that consists of roughly 3.8 million users. Because there is a considerable number of inactive accounts in the network, we decided to direct our attention only to the *active* users for receiving subscription gifts, where a user is considered active if she listened to at least 1 song in the last 30 days before our manipulation. It turned out there were roughly 1.2 million active listeners in the Last.fm social network. Let's call this list L. We form the group G as a random sample of 2000 users drawn uniformly randomly from L with no replacement. This group G contains 2000 users who will then be randomly split

[4] The origin of this name is blinded for review process.

into manipulated and non-manipulated groups. Consequently, we form the manipulated group M as a random sample of 1000 users drawn uniformly randomly from G with no replacement. Finally, we form the non-manipulated group NM as $NM = G\backslash M$, that is the rest 1000 users that were left in G after we picked group M.

We define our *treatment group* T as all immediate friends of M who are not themselves in M or NM and who are not friends of someone in NM. Symmetrically, we define our *control group* C as all immediate friends of NM who are not themselves in M or NM and are not friends of someone in M. Figure 2 presents an intuitive Venn diagram for these sets of users. Given the real-world nature of our data, a small number of users will likely turn out to be friends of both M and NM groups simultaneously. These users cannot be unequivocally put either into the treatment or control group and hence were excluded[5] from the experiment.

As a robustness check, we repeated our analysis while keeping the intersection included in both T and C instead of excluding it. Our results remain strongly significant and almost identical to the "exclusion" case. This is not very surprising however, considering how small the intersection is compared to groups T and C. Also, in Section 6.7 we propose a new method to relax this assumption somewhat and demonstrate robustness of our results through bootstrapping.

Looking at Figure 2 it is easy to see why our randomization assigns users into treatment and control groups independently of observed and unobserved characteristics of a user. To see this consider group G of 2000 initial users *before* they were randomly split into M and NM. Also consider a particular person A who is in G and his online friend B as displayed in Figure 2.

Before our gifts are assigned, person A has absolutely equal chances of becoming a member of M or NM and these chances are completely independent of characteristics of person A, person B or anyone else. However, if person A is assigned to M, person B becomes a friend of M. Alternatively, if person A were assigned to NM, then person B would be a friend of NM. Just as person A has no way to predict whether he will end up in M or NM, his friend B has no way to predict whether B will end up as friend of M or friend of NM.

Therefore, person B has absolutely equal chances of becoming a friend of M or a friend of NM and these chances are independent of any characteristic of Person A, Person B or anyone else[6].

[5] In our empirical data, this intersection constitutes less than 5% of the treatment and control groups. Excluding the intersection seems to be the best way to proceed since keeping it in either T or C alone would break the symmetry and immediately throw T and C off-balance statistically. Alternatively, including this intersection in both T and C constitutes a downward bias: means become closer if exactly the same set of people is counted to be both in the treatment group and in the control group at the same time.

[6] This result is general and holds if Person B happened to have n different friends in group G: persons $A_1, A_2, \ldots, A_n$. It is easy to show that in this case Person B is as likely to end up in T as in C exactly the same way as it was shown in the example on Figure 2.

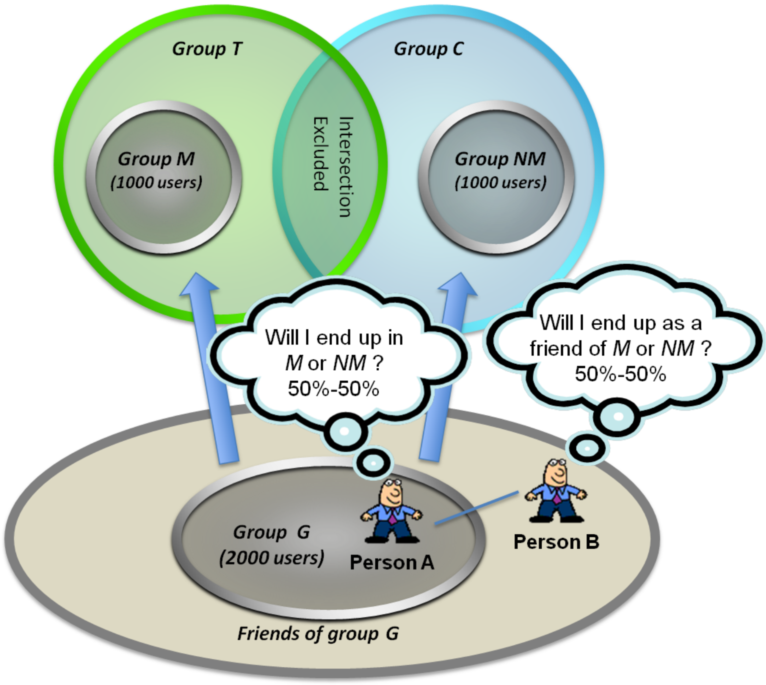

Figure 2 Venn Diagram of Experimental Design

This implies that assignment of Person B into group T versus C is independent of observed or unobserved characteristics of Person B.

Based on the experimental design explained above, our experimental procedure is composed of four stages. In the first stage we randomly assign users to groups M and NM, crawl their current friend network, and thus calculate groups T and C . In the second stage, we deploy a pre-treatment check and crawl the current status of M, NM, T and C groups immediately before the treatment. In the third stage, we deploy the U1B1-B virus by giving 1000 gifts to group M using our Paypal script[7]. Finally, we crawl the current status of M and NM groups immediately after manipulation to make sure our manipulation worked. Given Manski (1995) concern about subjects' behavior changing when they know they are being observed, we directed users to our Last.fm page (see Appendix A) where we took great care[8] in "explaining" to the users that these were expiring left-over funds from another project that we were simply giving away. We explicitly mentioned that we expected nothing in return and no action was needed from the user. The messaging worked, as can be gleaned by the comments of the gifted users left on our wall.

4.2. Strengths of the Experiment in Mitigating Threats to Validity

Our design has several intuitive benefits that help us overcome the myriad of challenges (Van den Bulte and Stremersch 2004, Aral 2011) in making causal detection of social contagion from observa-

[7] It only takes a couple of hours to distribute all 1000 gifts using our script

[8] Needless to say our protocol was approved by our IRB.

tional data, separating out homophily from peer influence. As mentioned above, one of the ways in which homophily manifests itself in observational data is through self-selection bias, when manipulations are not randomly assigned, which is not the case in our study. Also, since it is not possible to withdraw or refund a gift, we have no attrition or mortality bias. This is because users are selected randomly and they cannot escape, decline or remove themselves from the manipulation. It is also important to mention that each person's network will be collected immediately before the manipulation, immediately after the manipulation and with different levels of delay after the manipulation. Only "immediately before the manipulation" friend network is used to determine treatment group T and control group C . Clearly, if a person started self-selecting subscriber friends after the manipulation had occurred, it would not have any effect on our experiment. Further, because the subscriptions themselves are not transferrable and not refundable, we can rule out any direct treatment diffusion effect, suggesting that any effect that is observed must be through some kind of peer influence other than simple direct transfer of our gift. It is however possible given the real-life social network setting that our manipulation may "leak" from manipulated group M into control group C through 2nd degree friendship connections, i.e. there may be a possibility of an indirect treatment diffusion effect. This however would likely lead to an underestimation of the observed difference, not overestimation. Since the 2nd degree effect is probably slower and weaker than the 1st degree effect caused by the immediate friend, it can be mitigated by post-experimental controls on the shortest distances between the control group C and the treatment group T and on the time passed since manipulation. Finally, we can rule out any compensatory rivalry/resentful demoralization or experimenter bias, since neither treatment group T nor control group C know that they are being treated and watched.

It is important to highlight that only manipulated group M receives a gift from us and there are no other "disturbances" introduced into the network. Group M is told that the gift is given out of the expiring left-over funds from a prior survey and that a gift receiver is not required to do anything, thus group M itself is not even aware of being manipulated. Also, group M is not being tracked for the purposes of our experiment, it is their friends that we are interested in.

5. Data Description

5.1. Snapshot Data

Our dataset was collected by our custom multi-threaded Amazon Cloud-based web crawler and consists of panel data on approximately 3.8 million users that make up the largest connected component[9] of Last.fm network forming over 23 million friendship pairs . These users have been

[9] We employed multiple checks to ensure that we indeed got the largest connected component of the network and not just some smaller closed clique of users. Our checks ranged from looking for additional users in forums to crawling the lists of recommended music "neighbors" of each user. The total number of the extra users that we checked outside of our connected component amounts to the additional 0.5 million unique users. We have not discovered any other large connected component.

tracked consistently as a panel since May 2011 with updates roughly every 2 weeks. These dynamic updates provided us with fresh snapshots of the entire social network containing the list of friends and subscription status for every user. In addition to this information, we have been tracking self-reported demographic information and website-reported social activity information.

For every snapshot at time t, we have collected the following data for each user:

- $Age_{i,t}$. Self-reported age of user i. Age distribution was truncated to the interval between 8 and 79 in order to eliminate outlier data points that are likely fake.
- $Gender_{i,t}$. Self-reported gender of user i. Dummy variable.
- $FriendCnt_{i,t}$. Total count of number of friends of user i at time t.
- $SubscriberFriendCnt_{i,t}$. Total count of number of friends of user i who are paid subscribers.
- $SongsListened_{i,t}$. Total count of all songs ever listened and reported to Last.fm by user i. If a user listened to the same song twice, the song would be counted twice as well.
- $Playlists_{i,t}$. Total count of playlists ever made by user i on Last.fm.
- $Posts_{i,t}$. Total count of forum posts ever made by user i.
- $Shouts_{i,t}$. Total count of shouts (that is, wall posts) ever received by user i.
- $LovedTracks_{i,t}$. Total count of all tracks that were "loved" by user i.
- $RegDate_i$. User i original registration date on the website measured as the number of days since January 1, 1960 (standard date representation of SAS statistical package).
- $LastfmCountry_{i,t}$. Dummy variable. If user i's self-reported country is "USA", "Germany" or "UK", then $LastfmCountry$=1 for this user, otherwise 0. This variable is important because Last.FM subscription rules are slightly different[10] in the official Last.fm countries ("USA", "Germany", "UK") versus the rest of the world.
- $Subscriber_{i,t}$. Dummy variable indicating whether user i is a premium subscriber at time t.

The descriptive summary statistics for approximately 1.2 million active[11] Last.fm users are displayed in Table 1. This table provides a breakdown of statistics for active subscribers and active non-subscribers for one particular snapshot of data collected around September 8, 2011 before our manipulation was done. From this data, we find that active subscribers are consistently different from active non-subscribers in a variety of metrics: they are older, tend to have more

[10] Even though the premium subscription costs the same amount for every country, the subscription is more valuable for people outside USA, Germany and UK. Several Last.FM services that are normally free for US/Germany/UK users require premium subscription for the rest of the world because of music licensing contracts.

[11] Active user means a user who listened to at least 1 song within 30 days prior to the collection of that particular snapshot of data.

Subscriber	N Obs	Variable	Mean	Std Dev	Missing	Median	Min	Max
0	1214303	Age	23.21	6.18	385200	22	8	79
		Gender (Male=1)	0.66	0.48	234278	1	0	1
		FriendCnt	24.18	70.65	0	10	1	11780
		SubscriberFriendCnt	**0.65**	2.85	0	0	0	541
		SongsListened	24913.30	32365.72	1	15022	0	1000472.00
		Playlists	0.53	3.32	0	0	0	2291
		Posts	7.67	141.70	0	0	0	64108
		Shouts	42.19	271.02	27717	5	0	131765
		LovedTracks	128.15	406.44	0	35	0	99109
		RegDate	17838.23	636.71	584	17902	15642	18877
		LastfmCountry	0.30	0.46	0	0	0	1
1	37161	Age	30.26	9.25	14165	28	8	78
		Gender (Male=1)	0.76	0.43	8449	1	0	1
		FriendCnt	33.73	116.62	0	10	1	9788
		SubscriberFriendCnt	**2.85**	10.35	0	1	0	709
		SongsListened	31996.64	43938.95	0	18139	0	1000070
		Playlists	1.44	5.38	0	1	0	496
		Posts	27.74	465.16	0	0	0	50740
		Shouts	85.31	531.56	1275	5	0	36508
		LovedTracks	370.05	1104.95	0	149	0	63595
		RegDate	17678.54	628.82	1	17735	15642	18868
		LastfmCountry	0.28	0.45	0	0.00	0.00	1.00

Table 1 Summary Statistics of Historical Data for Active Users

friends (approximately, 40% increase as compared to non-subscribers) and disproportionally more subscriber-friends (over 300% increase), more playlists, loved tracks and registered earlier than non-subscribers. These empirical observations confirm the observed clustering of subscription behavior indicating the underlying homophily or peer influence. Our summary data are remarkably in line with 2009 Last.fm data reported by Oestreicher-Singer and Zalmanson (2010) suggesting a stable long-term pattern.

5.2. Dynamic Data

The collection of snapshots allows us to look into the dynamics of user characteristics in the social network as well as the dynamics of the social network itself. The following network dynamic variable is the variable of interest in this particular study:

- $Adopter_{i,[t,t+1]}$. Dummy variable indicating whether user i who had not been a paid subscriber before time t adopted subscription and became a paid subscriber in the interval of time $[t, t+1]$. Since the minimum possible unit of a premium subscription is 1 month and we collected our data with the intervals of 2-3 weeks, our data collection process has not missed any single subscription event for any user in the network beginning from May 2011 and till the moment this paper is being read by our readers. Therefore, $Adopter_{i,[t,t+1]}$ variable is an objective and

Adopter	N Obs	Variable	Mean	Std Dev	Missing	Median	Min	Max
0	1211366	Age	23.20	6.18	384294	22	8	79
		Gender (Male=1)	0.66	0.48	233726	1	0	1
		FriendCnt	24.16	70.43	0	10	1	11780
		SubscriberFriendCnt	**0.65**	2.80	0	0	0	465
		SongsListened	24912.04	32363.37	1	15024	0	1000472
		Playlists	0.53	3.32	0	0	0	2291
		Posts	7.67	141.83	0	0	0	64108
		Shouts	42.14	271.01	27602	5	0	131765
		LovedTracks	127.97	406.32	0	35	0	99109
		RegDate	17838.13	636.65	584	17902	15642	18877
		LastfmCountry	0.30	0.46	0	0	0	1
1	1099	Age	26.31	7.13	346	25	11	74
		Gender (Male=1)	0.70	0.46	204	1	0	1
		FriendCnt	42.70	196.79	0	14	1	4730
		SubscriberFriendCnt	**2.76**	17.58	0	1	0	541
		SongsListened	31984.12	38619.43	0	18991	0	423529
		Playlists	1.05	1.98	0	1	0	27
		Posts	13.08	96.25	0	0	0	2266
		Shouts	93.17	381.14	43	7	0	6247
		LovedTracks	310.65	542.01	0	133	0	6143
		RegDate	17712.48	651.39	0	17734	15642	18877
		LastfmCountry	0.24	0.43	0	0	0	1

Table 2 Summary Statistics of Data for Recent Adopters over 2-3 weeks.

guaranteed indicator of adoption or absence of adoption in time period $[t, t+1]$ for every user among 3.8 million users.

Similarly to Table 1, Table 2 displays the summary statistics for the dynamic data of recent adopters vs. recent non-adopters. Please note that there is a subtle, but very important difference between the types of information displayed by Table 1 and Table 2:

- Table 1 compares the current subscribers versus current free active users. This is the information about the current state that the network has achieved over the years.
- Table 2 displays the information on the recent adopters. This is the information about the change in the current state: a change in the network over 2-3 week period.

The difference between Table 1 and Table 2 can be explained better if we mention that many people who are currently subscribers have been premium subscribers for very long time. Clearly, these "mature subscribers" should not be considered as either recent adopters or recent non-adopters and should not be counted in Table 2, but they are still subscribers and therefore, should be counted in Table 1[12].

Despite the differences, we observe that there is a similarity between Tables 1 and 2 suggesting a remarkable consistency in the data generation process over the years: recent adopters resemble the

[12] For this reason, the total number of users in Table 1 and Table 2 may not add up to exactly the same number.

large mass of existing premium subscribers based on the observed characteristics. More specifically, both tables demonstrate that subscribers and recent adopters tend to have disproportionately large count of subscriber friends: over 300% more as compared to non-subscribers and non-adopters while the total number of friends is only 40-70% larger.

We used this dynamic data to simulate and calibrate our experiment before actually running it. In particular, because new adoption is a rare event in our network, a key experimental challenge for us was to decide on the sample size for the manipulation so as to be able to pick up statistically any peer effect that may be there.

5.3. Quasi-Experiment to Calibrate the Randomized Trial

As Tables 1 and 2 demonstrate, paid subscriptions are rare events[13] in our network. Thus, before conducting the actual randomized experiment, we first constructed a quasi-experiment that simulated our randomized experiment using only observational panel data.

For our quasi-experiment, we looked at the users who had recently purchased subscription on their own and treated them as if they were gifted by us. We called these users group "M" (M in quotes). We then matched each user in "M" to his "mirror image": a corresponding user who has the same observed characteristics, but did not purchase a subscription. We called this "mirror" group as "NM". Subsequently we studied adoption behavior in groups "T" and "C" that correspond to these "M" and "NM". Methodologically, this approach is similar to the matching-based quasi-causal techniques seen in Aral, Muchnik, and Sundararajan (2009), Susarla, Oh, and Tan (2012) as well as Oestreicher-Singer and Zalmanson (2010). As expected for a purely observational matching technique, the influence of unobserved characteristics cannot be ruled out by the quasi-experiment, but we could still control for the observed characteristics of users as a "first-order approximation."

The quasi-experiment study was conducted in order to:

1. Do a sanity check: see whether the effect can be observed in observational-only data
2. Determine the appropriate sample size for the real experiment
3. Compare the quasi-experimental results against the real experiment and determine the type of a bias that results from quasi-experimental studies.

Based on our quasi-experimental trials, we determined that the effect is indeed observed in observational data even after controlling for individual characteristics. We also discovered that the sample size of 1000 manipulated users is adequate for observing a statistically significant effect given the rare-event nature of premium subscriptions and is also not too wasteful of our resources as each gift costs us $3.

[13] For example, only 3% of active users are currently subscribers and 0.2% of users are recent adopters in 1 month period. Yet the magnitude of these numbers should be considered in the context of the vast scale of real life online social networks. For example, in a network of size of Facebook these 0.2% would correspond to more than 1.5 million unique users (this is not even counting the social multiplier effect).

Variable	Friend of	Mean	Std Err	Minimum	Median	Maximum	t-value	Pr> $\|t\|$
Age	NM	22.72	0.261	8	21	79	−1.18	0.2387
	M	22.11	0.260	8	21	77		
Gender (Male=1)	NM	0.62	0.013	0	1	1	−1.16	0.2475
	M	0.59	0.013	0	1	1		
FriendCnt	NM	100.23	8.829	1	40	7800	0.52	0.6022
	M	109.35	8.754	1	41	4700		
SubscriberFriendCnt	NM	2.76	0.146	0	1	337	0.24	0.8129
	M	2.83	0.144	0	1	352		
LovedTracks	NM	238.76	9.208	0	61	32387	−1.24	0.2152
	M	216.04	9.252	0	59	23466		
Playlists	NM	0.67	0.037	0	0	204	−0.02	0.9853
	M	0.67	0.035	0	0	465		
RegDate	NM	17747.04	18.607	15742	17803	18844	0.09	0.9257
	M	17750.51	18.688	15746	17807	18762		
Shouts	NM	141.58	14.683	0	27	17699	0.67	0.5008
	M	161.32	14.730	0	27	25343		
SongsListened	NM	31970.58	696.408	0	19480	928318	0.20	0.8440
	M	32243.68	695.810	0	19873	1000000		

Table 3 Groups T and C have Similar Observed Statistical Properties

In addition to that, we subsequently used the same data and compared results of the quasi-experiment against the real experiment. We discovered that quasi-experiment tends to over-estimate the effect of the treatment on users with larger number of friends and tends to under-estimate it for the users with smaller number of friends. This pattern of over-estimation and under-estimation was robust across multiple randomization seeds and different runs of the quasi-experiment. To the best of our knowledge this is the first indication of the nature of the quasi-experimental bias in estimating peer-effects in online social networks. Because of space limitations, the exact procedure and details of the quasi-experiment are not included in this paper and are available upon request from authors.

6. Analysis and Results

6.1. Testing for Causal Social Contagion using the t-test

We conducted our randomized field experiment by sampling groups M and NM and computing their friends T and C respectively who are summarized in Table 3. Each person in group M subsequently received a 1 month subscription gift from us, with the 1000 gifts being distributed over the period of several hours by a GreaseMonkey script. The users from group NM did not receive any gift or any other communication from us and were only being silently tracked.[14]

[14] While clinical trials frequently give a placebo pill to the control group instead of just not giving anything at all, in our study we do not need it. Clinical trials deal with special circumstances of mind-body connection: it is well-known

Friend of	N	Mean	Std Dev	Std Err	t-value	Pr> $\|t\|$
NM	21534	0.00190	0.0436	0.000297	2.26	0.0237
M	22200	0.00297	0.0544	0.000365		

Table 4 Experimental results: t-test

A manipulation check was done immediately after distributing the gifts. This check demonstrated that all 1000 users in group M received the gift and became premium subscribers immediately. In one month after the manipulation was done, we collected a new snapshot of the social network and compared adoption behavior among all friends of group M versus all friends of group NM as described in the experiment design.

Given exogenous and independent randomization of our manipulation, the assignment of user i as a friend of M or NM is independent of her observed or unobserved characteristics as explained in Section 4 and confirmed in Table 3. Therefore we can compare the distributions of outcomes among friends of M and friends of NM without any need for controls.

As shown by the results of the t-test in Table 4, friends of group M demonstrated statistically and economically significant difference against friends of NM: there were approximately 55% more adoptions in the treatment group as compared to the control group. This offers valid support for the existence and importance of causal peer-effects for premium subscription adoption in the general population of Last.fm social network.

In order to demonstrate the *economic significance* of this effect, it should be noted that groups M and NM were selected as a random sample from the general population and *not from the population of very influential people*. It is remarkable that even 1000 *average* social network users have been able to exert that much peer influence on their friends. It is a part of a separate study to explore how much stronger the influence could have been had we focused ourselves on manipulating the sample of 1000 highly influential people rather than 1000 average people.

In addition to that, it is important to point out that we only look at the effect on immediate friends of M and NM in this paper. As was mentioned in Section 1, peer influence is subject to *social multiplier effect* such that once influenced the immediate friends of M and NM may themselves start influencing their own friends, possibly increasing economic significance of the original first-degree effect dramatically.

(Ariely 2010) that a placebo pill itself can demonstrate significant improvements in patient health as compared to no treatment at all. Therefore, clinical trials have to demonstrate not that the pill works in general, but that the pill works stronger than placebo works. Therefore, for clinical trials, it is typically a comparison of two alternative mechanisms both of which work. In our case, we do not intend to show that our manipulation works stronger than some other alternative manipulation. Instead, we plan to demonstrate that our manipulation works stronger than having no manipulation and simply "going with the flow".

Variable	Estimate	Std Err	t-value	Pr> $\|t\|$
Intercept: adopter=0	−3.5327	3.9038	−0.90	0.3655
OurTreatment	**0.4838**	**0.2006**	**2.41**	**0.0159**
log(FriendCnt)	−0.2480	0.1458	−1.70	0.0889
log(SubscriberFriendCnt)	0.4833	0.1595	3.03	0.0024
Age	0.0245	0.0148	1.65	0.0992
Gender (Male=1)	−0.3465	0.2135	−1.62	0.1050
LastfmCountry	−0.4339	0.2370	−1.83	0.0674
RegDate	−0.0003	0.0002	−1.52	0.1288
log(SongsListened)	0.1477	0.0878	1.68	0.0926
log(Posts)	0.0669	0.0652	1.03	0.3049
log(Playlists)	0.3688	0.1586	2.33	0.0200
log(Shouts)	−0.0049	0.0787	−0.06	0.9508
log(LovedTracks)	0.2580	0.0636	4.06	<0.0001

Table 5 Experimental results: logistic regression

6.2. Logistic Regression

As explained in Section 5, we were able to collect considerable data about individual users. Given the exogenously randomized nature of our experimental design, this data is not required for testing Hypothesis 1. Nevertheless, this data is useful in explaining the individual adoption decisions and we can utilize it to introduce control variables in order to increase statistical efficiency of our model as well as to set up the stage for testing Hypothesis 2. Since our outcome variable $Adopter_{i,[t,t+1]}$ is a binary variable we decided to use the standard logistic regression as the apparatus to control for the observed covariates and determine causality in this scenario. The formula below depicts our logistic regression model[15], the treatment variable and controls[16]:

$$\begin{aligned}\text{logit}\left(Pr\{\text{Adopter}=1\}\right) = \alpha &+ \beta_1 \cdot \mathit{OurTreatment} + \beta_2 \cdot \log(\mathit{FriendCnt}) + \beta_3 \cdot \mathit{RegDate} + \\ &+ \beta_4 \cdot \log(\mathit{SubscriberFriendCnt}) + \beta_5 \cdot \mathit{Age} + \beta_6 \cdot \mathit{Gender} + \\ &+ \beta_7 \cdot \mathit{LastfmCountry} + \beta_8 \cdot \log(\mathit{SongsListened}) + \\ &+ \beta_9 \cdot \log(\mathit{Posts}) + \beta_{10} \cdot \log(\mathit{Playlists}) + \beta_{11} \cdot \log(\mathit{Shouts}) + \\ &+ \beta_{12} \cdot \log(\mathit{LovedTracks})\end{aligned}$$

The following variable is used as a manipulation variable in this particular study:

- $OurTreatment_i$. This manipulation variable represents the dummy variable that indicates whether user i is a friend of group M or group NM[17].

[15] We also recognize that by specifying the logistic regression formula we assume a specific model

[16] Missing demographic variables are accounted for with a standard multiple imputation method (Schafer 1997)

[17] Since intersection was excluded, no user in our dataset is a friend of M and NM simultaneously.

Variable	*small* **FriendCnt** Estimate	Pr > $\lvert t \rvert$	*large* **FriendCnt** Estimate	Pr > $\lvert t \rvert$
Intercept: adopter=0	−3.5809	0.5866	−2.9245	0.5642
OurTreatment	**0.8662**	**0.0197**	**0.3315**	**0.1698**
log(FriendCnt)	0.3762	0.3108	−0.6427	0.0039
log(SubscriberFriendCnt)	0.6184	0.0645	0.7334	0.0003
Age	0.0646	0.0020	−0.0094	0.6694
Gender (Male=1)	0.0704	0.8701	−0.4647	0.0643
LastfmCountry	−0.5816	0.1444	−0.3923	0.1788
RegDate	−0.0005	0.1292	−0.0001	0.5608
log(SongsListened)	0.1721	0.2610	0.0792	0.4742
log(Posts)	0.0833	0.5779	0.0763	0.2933
log(Playlists)	0.0986	0.7807	0.4559	0.0099
log(Shouts)	−0.0268	0.8599	−0.0075	0.9354
log(LovedTracks)	0.2809	0.0076	0.2273	0.0049
Number of observations	21914		21820	

Table 6 Experimental results: separate logistic regressions for two groups split based on FriendCnt

As evident from Table 5, *OurTreatment* variable is statistically significant even after controlling for observed individual user characteristics. Moreover, since *OurTreatment* is assigned independently of characteristics of user i, this coefficient has causal interpretation: *OurTreatment causes* the adoption of subscription, thus providing additional evidence for Hypothesis 1.

Since *OurTreatment* is a dummy variable, it is easy to estimate the average marginal effect of *OurTreatment* on odds of adopting the subscription: if *OurTreatment* changes from 0 to 1, the odds of adoption increase by $e^{0.4838}$ that is by a factor of 1.62. It can be noted that this figure is in line with the results of the t-test from the previous section that demonstrated an increase of approximately 50% to 60% in adoptions in the treatment group.

It is also important to note that the estimated coefficient of log(*SubscriberFriendCnt*) is also statistically significant and positively associated with the likelihood of adoption of subscription: the effect that is likely to be observed if Hypothesis 1 is true.

6.3. Examining Susceptibility

In addition to testing for causal peer effects, we are also interested in examining whether certain characteristics of users are associated with more or less susceptibility to be influenced by their peers, as articulated in Hypothesis 2. It's worth emphasizing that while Hypothesis 1 is a causal claim, Hypothesis 2 is a correlation claim explaining the strength of the causal effect. More specifically, in Hypothesis 2, we claim that a random friend F is more susceptible to be influenced if F has few friends, but *we do not claim that we can actually force F to become even more susceptible by taking an additional friend away from her.* As a starting point for providing evidence for Hypothesis 2, we compare the strength of the effect of *OurTreatment* for users who have small (below median)

Variable	Estimate	Std Err	t-value	Pr> $\|t\|$
Intercept: adopter=0	−4.2204	3.9363	−1.07	0.2836
OurTreatment	**1.4857**	**0.6170**	**2.41**	**0.0160**
OurTreatment * log(FriendCnt)	**−0.2389**	**0.1266**	**−1.89**	**0.0591**
log(FriendCnt)	−0.0910	0.1706	−0.53	0.5937
log(SubscriberFriendCnt)	0.4765	0.1597	2.98	0.0028
Age	0.0249	0.0149	1.68	0.0939
Gender (Male=1)	−0.3548	0.2137	−1.66	0.0971
LastfmCountry	−0.4354	0.2372	−1.84	0.0666
RegDate	−0.0003	0.0002	−1.51	0.1320
log(SongsListened)	0.1461	0.0878	1.66	0.0960
log(Posts)	0.0704	0.0654	1.08	0.2815
log(Playlists)	0.3681	0.1588	2.32	0.0205
log(Shouts)	−0.0053	0.0786	−0.07	0.9461
log(LovedTracks)	0.2574	0.0635	4.05	0.0001

Table 7 Peer Effects are Moderated by the Number of Friends of Influencee

number of friends and users who have large number of friends (above median). To accomplish that we split both the treatment and control groups into two subgroups:

- Small *FriendCnt* subgroup that consists of all users i from T and C who have $FriendCnt_i \leq m$
- Large *FriendCnt* subgroup that consists of all users j from T and C who have $FriendCnt_j \geq m$

where m is the overall median *FriendCnt* that users from T and C have. By splitting our treatment and control group this way we ensure that both of these subgroups are of equivalent sample sizes and therefore the groups could be compared more directly. Consequently, we run separate logistic regressions for each of these two subgroups and compare the results side-by-side in Table 6.

As demonstrated by the results in Table 6, *OurTreatment* variable is statistically significant for users who have small number of friends while being statistically insignificant for users who have large number of friends. This result suggests that the strength of the effect of *OurTreatment* on the target user actually depends on the total number of friends the target user has.

6.4. Logistic Regression with Interaction Term

Given the results of Table 6, it is natural to refine the model such as it becomes capable of learning the decreasing strength of *OurTreatment* from the data itself. To achieve this, we introduce the interaction term between *OurTreatment* and *FriendCnt*.

In addition to that we refine the model further and notice that some of our users in group T naturally received “multiple treatments”: just by chance they happened to have two or more friends that were manipulated by us. For these users, we assume the linear response to additional treatment and assume our model as follows:

- $OurTreatment_{i,t}$. This manipulation variable represents the count of how many friends of user i were manipulated by us at time t.

Table 7 presents the results of fitting the model with interaction term to our data. As we can observe, we obtain a negative coefficient for the interaction term. This implies that the larger *FriendCnt* is, the weaker is the response to *OurTreatment* other things being equal, thus supporting Hypothesis 2.

6.5. Some Insights into the Homophily and Peer Influence Rivalry

An interesting aspect of our real-world experimental setup is that while our manipulation was in progress, users in groups T and C still had some other friends from the other parts of the network who decided to purchase the subscription on their own as demonstrated in Figure 3. These adopter friends from the "outer" network[18] may have also exerted some peer influence on the treatment and control groups.

Because of the exogenous random assignment of *OurTreatment*, these "other treatments" do not introduce any statistical bias since the friends of M and friends of NM should be exposed to statistically equivalent levels of this background "other treatment". Nevertheless, by creating a new control variable called *OtherTreatment* and controlling for this "other treatment" administered by the rest of the network, we are able to learn the strength of peer influence as compared to homophily.

Considering Figure 3, the key difference between *OurTreatment* and *OtherTreatment* lies in the fact that user A did not choose to be a subscriber on her own. It is we, the experimenters, who chose to subscribe her. At the same time, user C *is* a self-selected subscriber and therefore, is more likely than average to come from the "premium subscription risk-group".

Rephrasing the same idea in U1B1-B virus terms, user A is an average user artificially infected by us with the virus. At the same time, user C got infected on her own, so likely belongs to the "poor health" group. Looking at the network from the perspective of an external observer, we see that user A's infection sends us, the external observers, just 1 new signal about user B:

- "I, user A, may personally infect user B" (peer influence). No other signals about user B's health are received by us from user A's "infection".

On the other hand, user C sends us, the external observers, two signals at the same time:

- "I, user C, may personally infect user B" (peer influence)
- "I, user C, am likely in poor health group. User B is my friend, so she is likely in the same poor health group as I am, so user B is quite likely to get infected on her own too" (homophily)

[18] That is, these naturally occurring adopter friends are not from groups M, NM, T or C, but from the rest of the 3.8 million network.

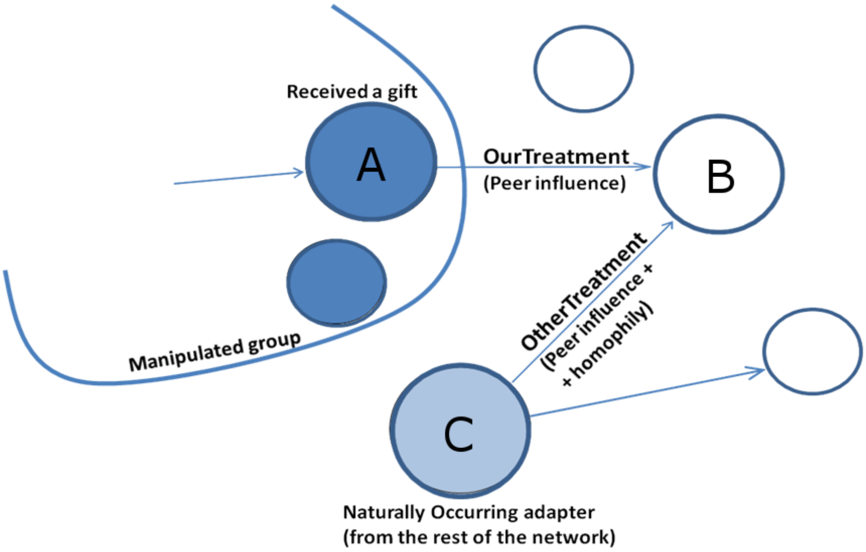

Figure 3 Randomized Experiment Design:
Naturally Occurring "OtherTreatment" Offers an Interesting Contrast to OurTreatment

Variable	Estimate	Std Err	t-value	Pr> $\|t\|$
Intercept: adopter=0	−4.5518	3.9264	−1.16	0.2463
OurTreatment	**1.5782**	**0.6955**	**2.27**	**0.0233**
OurTreatment * log(FriendCnt)	**−0.2580**	**0.1465**	**−1.76**	**0.0783**
OtherTreatment	**2.0655**	**0.5521**	**3.74**	**0.0002**
OtherTreatment * log(FriendCnt)	**−0.3536**	**0.1019**	**−3.47**	**0.0005**
log(FriendCnt)	0.0433	0.1837	0.24	0.8136
log(SubscriberFriendCnt)	0.4414	0.1769	2.50	0.0126
Age	0.0239	0.0150	1.60	0.1108
Gender (Male=1)	−0.3332	0.2137	−1.56	0.1193
LastfmCountry	−0.4025	0.2382	−1.69	0.0914
RegDate	−0.0003	0.0002	−1.51	0.1306
log(SongsListened)	0.1272	0.0884	1.44	0.1501
log(Posts)	0.0690	0.0650	1.06	0.2884
log(Playlists)	0.3743	0.1600	2.34	0.0193
log(Shouts)	−0.0170	0.0777	−0.22	0.8265
log(LovedTracks)	0.2474	0.0640	3.87	0.0001

Table 8 Peer Effects hold for OurTreatment and OtherTreatment

Given this, comparing the strength of *OurTreatment* vs. *OtherTreatment* becomes comparing the strength of "peer influence" vs. "peer influence + homophily". Therefore, our research gives us an opportunity to provide the point estimates of the strength of peer influence vs. homophily. We believe this is a unique feature of our design that gives us an insight on the extent of the homophily strength that has been so hard to quantify earlier.

Given the model described by Figure 3 and the insights learned from Section 6.4, we extend our logistic model by including *OtherTreatment* and its interaction with *FriendCnt* as new variables into our model:

- $OtherTreatment_{i,[t-1,t]}$. This variable represents the count of how many friends of user i adopted the subscription on their own in the time interval $[t-1,t]$ independently from our manipulation.

Table 8 presents the results of fitting this augmented model to our experimental data. As Table 8 demonstrates, *OurTreatment* variable and its interaction are statistically significant even after accounting for *OtherTreatment*. Interestingly, *OtherTreatment* and its interaction term are also significant with negative sign and follow the pattern that resembles *OurTreatment*: the effect that Hypothesis 2 would imply. Note that the *OtherTreatment* coefficient is larger in magnitude than *OurTreatment*, as would be expected given insights from Section 6.5.

6.6. Strength of Peer Influence versus Homophily

As explained in Section 6.5, variable *OurTreatment* represents the effect of pure peer influence, while the effect of *OtherTreatment* represents the effect of peer influence and homophily combined. As outlined in Table 8, both *OurTreatment* and *OtherTreatment* enter our model with interactions terms, therefore, the marginal effects of either of these variables is not constant and depends on the exact characteristics of the influenced user. In order to provide a point estimate of the marginal effect of each of these variables, we take a median Last.fm user from the social network and apply *OurTreatment* to that user varying only her number of friends and holding all other variables constant (assuming also $OtherTreatment = 0$). We repeat the same procedure for *OtherTreatment* by varying only the median user's number of friends and holding all other variables constant (assuming also $OurTreatment = 0$).

The two resulting curves are displayed in Figure 4 where the solid line represents the marginal effect of *OurTreatment* varying with the number of friends and the dashed line represents the marginal effect of *OtherTreatment* respectively: the horizontal axis represents the number of friends the influenced user has, while the vertical axis represents the ratio of increase in the user's odds of adopting the subscription. For example, for a median user with 10 friends a unit of *OurTreatment* increases the user's odds of buying subscription 2.72 times (that is, 172% increase in odds).

As Figure 4 demonstrates, users with small number of friends are the ones who are the most susceptible to peer influence demonstrating hundreds of percents increases in odds. Moreover, given that users with small number of friends represent the vast majority of our social network, this finding shows not only the statistical significance of our result, but also its economic significance.

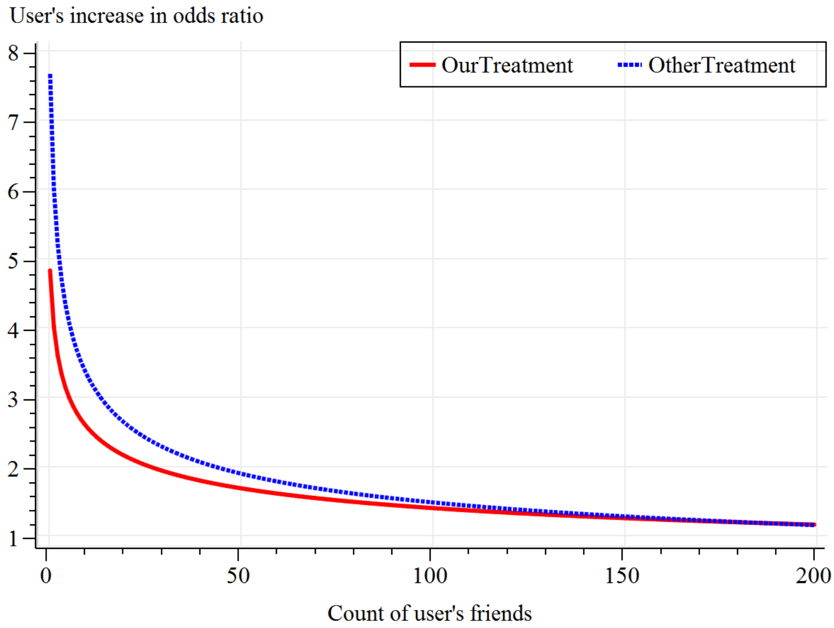

Figure 4 OtherTreatment Estimates are Upwardly Biased

Notably, *OtherTreatment* is estimated to be stronger than *OurTreatment* across the large spectrum of user friend counts. This result is not surprising, however, since *OtherTreatment* contains both the homophily signal and peer influence, while *OurTreatment* is pure peer influence. In an attempt to warn against over-generalizing the observed difference between the two, we would hypothesize that homophily in this setting can be a weaker force that acts continuously, while peer influence is a sudden and stronger force, but more short-lived, therefore homophily may not manifest itself enough over the short periods of time such as our experiment, while it can potentially manifest itself considerably over longer periods of time. We also acknowledge that peer influence of a person who received a gift subscription as in *OurTreatment* can potentially be weaker than peer influence of a person who paid for subscription herself as in *OtherTreatment*, however that would only make our estimates of the strength of true peer influence more conservative.

6.7. Robustness

In addition to the analysis reported above we performed a number of additional statistical tests ranging from parametric tests (logistic regression, survival model analysis) to semiparametric tests (binomial test) to non-parametric tests (reshuffling, bootstrapping) that all demonstrate significance and robustness of our results under different model assumptions.

Binomial test. Assume our manipulation did not cause any effect. We note that in terms of absolute numbers we had 107 total new adopters with the treatment group having 66 adopters versus 41 adopters observed in the control group. As we will see, the difference that large is very unlikely to occur just by chance.

If manipulation did not work, then we would have had 66+41=107 adopters in total anyway, irrespective of had we gifted group M or had we not. These 107 adopters would have been simply "scattered around" by our randomization procedure described in Figure 2. If manipulation had no effect, each "future adopter" had equal chances of ending up in either T or C ($p = 0.5$), however we observed that out of 107 adopters, only 41 adopter ultimately ended up being in C and 66 ended up being in T.

The chance of the difference that large is:

$$P(\text{Observing 41 adopter in one group and 66 in another}|p = 0.5, n = 107) =$$

$$= \text{Binomial}(k = 41, p = 0.5, n = 107) + (1 - \text{Binomial}(k = \underbrace{(66-1)}_{65}, p = 0.5, n = 107)) < 0.02$$

According to this test, probability that the difference that large can occur by chance is less than 2%, so there is more than 98% probability that it did not happen by chance, giving us confidence in our result.

Reshuffling test. Assume our manipulation did not cause any effect, so any difference we see is just a chance. This means we can randomly reshuffle users between groups M and NM, repeat our complete analysis with those reshuffled "M" and "NM" and still see the same magnitude of difference in adoption fairly often. As we will see, it is actually quite rare to see this magnitude of difference if we were to reshuffle M and NM randomly.

For our test we used the following procedure to construct reshuffled "M" and "NM":

1. Pick random 500 users from the true M + random 500 users from the true NM
2. Put these 1000 users into a new group called "M" (in quotes)
3. Pick the leftover 500 users from the true M + leftover 500 users from the true NM
4. Put these 1000 users into a new group called "NM" (in quotes)
5. For these reshuffled "M" and "NM", find the corresponding "T" and "C" and repeat all the analysis we did in exactly the same way.
6. Compute the difference in adopters between "T" and "C" and save it
7. Repeat from Step 1

After running 800 iterations of this simulation, we find that only in 3.2% of random reshufflings does group "T" beat group "C" with the difference as large as we see in our data. If we were to believe that the difference we observe is just a chance, it means we would need to believe we hit a 3.2% probability event. According to this test, there is more than 96% probability that the difference that large did not happen by chance.

Bootstrap test. Assume our manipulation did not cause any effect, so any difference we see is just a chance. We may remember that group NM is just an arbitrary random sample of 1000 users

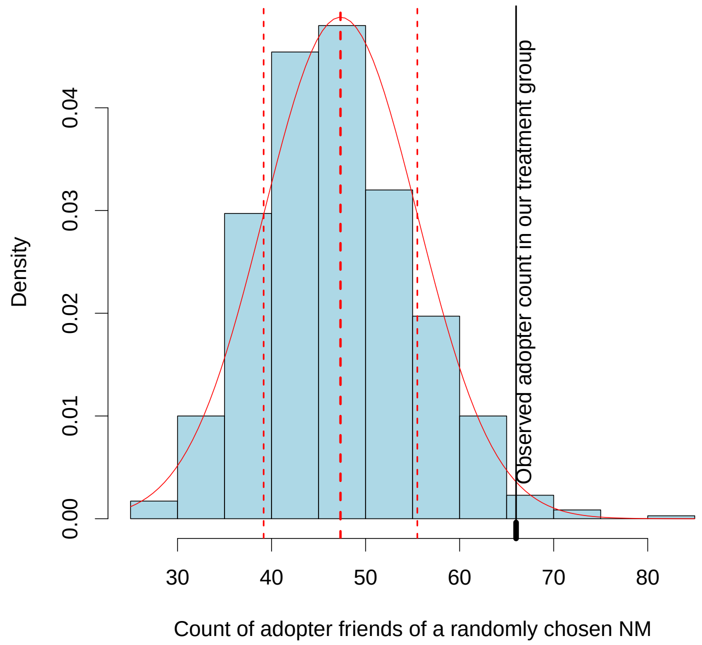

Figure 5 **Bootstrap test reveals how unlikely it is to see 66 adopters just by chance**

that was sampled from the population of 1.2 million eligible ones. Therefore, we may easily generate many other random samples $NM_1, \ldots, NM_k$, each with 1000 users, from 1.2 million eligible users and compare the friends of each NM_i against the friends of our true group M.

If our manipulation did not work, we would see that friends of many NMs achieve adoption levels similar to the one achieved by friends of M. However, as we see in Figure 5, the friends of M achieved such a high level that is rarely seen with a typical NM sample of 1000 eligible users.

For our test we used the following procedure to conduct bootstrapping:

1. Create list A of all users who were eligible to receive a gift but did not.
2. Initialize i=1
3. Pick random 1000 users from A and form group NM_i
4. Based on the given groups M and NM_i compute groups T_i and C_i
5. Compute the difference in adopters between T_i and C_i and save it
6. Increase i by 1
7. Repeat from Step 3

It is important to note that:

- this technique statistically utilizes *the entire 1.2 million sample of eligible users* to compare against 1,000 manipulated users, while at the same time keeping intact the perfect intrinsic symmetry of our experiment

- this technique allows us to include users from the "intersection" back into sample. For example, if a particular friend of M happened to be in the intersection at iteration $i = 1$ and thus was excluded, she will not necessarily be excluded at iteration $i = 2$ once different NM_i is sampled. Moreover, it is guaranteed that if we run this simulation long enough, no user will remain in the intersection forever.

After running 700 iterations of this simulation, we plot the distribution of typical counts of adopter friends of each NM_i and contrast it with the observed count of adopter friends of group M that received a gift from us. As is evident from Figure 5, friends of group M really stand out and it is very unlikely to see the level of adoption that high just by sampling a random NM_i. Only 1.7% of random NMs were able to reach the level achieved by our group M.

According to this test, there is more than 98% probability that it did not happen by chance.

Other models. In addition to logistic regression model presented in Section 6, we tried a number of different models including survival models. Survival models offer a very interesting approach that can take advantage of longitudinal nature of our data. Unfortunately our data does not provide us with precise timing of adoption required by survival models and only contains a few time periods. Nevertheless we tried a number of survival models that allow for ties and imprecise timing. The results of these models were remarkably similar to the results of our logistic regression. We decided to mention survival models only briefly because of space limitations and taking into account that our adoption timing might be too imprecise to motivate the use of a survival model.

Related phenomena. During our experiment, we also independently observed several effects that are intuitive and confirmed by already published results. More specifically:

- We observed that even after our gift manipulation had expired in group M, some people in group M decided to renew subscription on their own. The count of "renew-ers" in group M was statistically larger than the count of "new adopters" in group NM despite the fact that the groups were chosen initially at random, confirming the well-known effect of free promotions.
- The estimation results suggest that subscribers and adopters tend to be older and registered earlier than general population confirming the earlier findings of Oestreicher-Singer and Zalmanson (2010) who collected Last.fm data for a different study several years before us.
- We discovered that being in non-*LastfmCountry* provides a significant increase in the likelihood of adopting: a finding that is consistent with the fact that premium subscription gives much more features to people outside of the US, UK and Germany even though it costs the same.

While these findings are not the main research question of this study, they serve as additional evidence that Last.fm social network is a domain that is subject to traditional economic laws and therefore the insights learned from Last.fm domain can be a manifestation of more fundamental laws that are applicable across other domains as well.

7. Conclusions and Discussion

In this paper, we present a novel randomized experiment that allows us to make a *causal inference about the presence of economic social contagion* and peer-effects in the *general population* of an online social network without any subject recruitment procedures. More specifically, we conduct the experiment in the context of purchasing premium subscriptions of a "freemium" social network using a unique website feature that allows us to buy a premium subscription gift for any user in the network and thus creates a perfect "seeding tool". This unique feature induces the proverbial "helicopter drop," an exogenous random assignment of a treatment to a subset of the population, which can be compared against a statistically identical control group. We believe that this research is at the frontier of what IS can do - an "economic experiment in the wild" with real subjects but without a self-selection based subject recruitment procedure.

In this study, we demonstrate that there exists statistically and economically significant causal peer influence in the general population of a freemium social network. In addition to that, we quantify the strength of this peer influence and discover that the strength of peer influence decreases with the size of the friendship circle of the influenced user.

Moreover, in our study, each individual outcome is a purchase of the paid product with well defined monetary cost as compared to prior research that looked at the adoption of free products. Therefore, product adoption requires subjects to make *explicit economic decision with their own money* in our setting.

In addition to that, we compare the *point estimates of pure peer influence effect vs. homophily effect*. While these estimates provide a way of quantifying the strength of homophily vs. peer influence in a social network, this study suggests to look at peer influence and homophily as forces of nature acting over different time horizons and suggests that a separate study is needed to identify the longitudinal effects of both of these forces. We also separately compare the results of observational quasi-experiment with the randomized experiment on the same data and conclude that quasi-experiments tend to overestimate the strength of peer influence for users with large number of friends, while underestimating it for users with small number of friends.

Our work does not concern the exact peer influence mechanisms that are at work in the ongoing social contagion process: we do not distinguish between tactics like persuading a friend to subscribe versus imitation of a friend etc, as we combine all of them under the umbrella of peer influence mechanism that is contrasted with the umbrella of homophily mechanism. We do have to mention however that premium subscription is definitely not a hidden secret inside freemium communities: it is being constantly advertised and free users are generally aware of it. The fact that peer influence is clearly observed despite the target population being aware of the product is an interesting aspect of peer influence in our context suggesting complex underlying mechanisms that span beyond "awareness".

In this paper, we also do not study whether the influence comes from a few elite highly influential users or a large number of low influential users: our major goal for this paper is to demonstrate that significant economic social contagion is at work on average in the general population of a freemium social network such as Last.fm. In addition to that we only limit our attention to the influence on first-degree friends and do not look at second-degree effects and at importance of strength of ties (Centola 2010). Both of these issues are part of our ongoing research.

Finally, we believe our experimental design can be practically carried out by both researchers and practitioners. Practitioners may build a similar gift artifact into their products and use it as a dipstick to examine the nature and strength of social contagion in their setting. We expect to see more such random acts of kindness to solve interesting problems facing businesses and society.

References

Anderson, C. 2008. *The long tail: Why the future of business is selling less of more*. Hyperion Books.

Aral, S. 2011. Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science* **30** 217–223.

Aral, S., L. Muchnik, A. Sundararajan. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Procs of the National Academy of Sciences* **106**(51) 21544–21549.

Aral, S., D. Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* **57**(9) 1623.

Aral, S., D. Walker. 2012. Identifying influential and susceptible members of social networks. *Science* **337**(6092) 337–341.

Ariely, D. 2010. *Predictably irrational: The hidden forces that shape our decisions*. HarperCollins.

Bapna, R., A. Gupta, S. Rice, A. Sundararajan. 2012. Trust, reciprocity and the strength of social ties: An online social network based field experiment. Working Paper. University of Minnesota.

Bramoullé, Y., H. Djebbari, B. Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**(1) 41–55.

Camerer, C., D. Lovallo. 1999. Overconfidence and excess entry: An experimental approach. *The American Economic Review* **89**(1) 306–318.

Centola, D. 2010. The spread of behavior in an online social network experiment. *Science* **329**(5996) 1194–1197.

Godes, D., D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Science* 545–560.

Godes, D., D. Mayzlin. 2009. Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science* **28**(4) 721–739.

Harrison, G.W., M.I. Lau, E. Elisabet Rutström. 2009. Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization* **70**(3) 498–507.

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

Iyengar, R., C. Van den Bulte, T.W. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 195–212.

Manchanda, P., Y. Xie, N. Youn. 2008. The role of targeted communication and contagion in product adoption. *Marketing Science* **27**(6) 961–976.

Manski, C.F. 1995. *Identification problems in the social sciences*. Harvard University Press.

Mayzlin, D. 2006. Promotional chat on the internet. *Marketing Science* 155–163.

Oestreicher-Singer, G., A. Sundararajan. 2010. The visible hand of peer networks in electronic markets. URL `http://ssrn.com/abstract=1268516`. Working paper, New York University, New York.

Oestreicher-Singer, G., L. Zalmanson. 2010. Paying for content or paying for community ? the effect of social computing platforms on willingness to pay in content websites. Working Paper. Tel-Aviv University.

Parameswaran, M., A.B. Whinston. 2007. Research issues in social computing. *Journal of the Association for Information Systems* **8**(6) 336–350.

Pauwels, K., A. Weiss. 2008. Moving from free to fee: How online firms market to change their business model successfully. *Journal of Marketing* **72**(3) 14–31.

Rosenbaum, P.R., D.B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1) 41–55.

Schafer, J.L. 1997. *Analysis of incomplete multivariate data*, vol. 72. Chapman & Hall/CRC.

Shampanier, K., N. Mazar, D. Ariely. 2007. Zero as a special price: The true value of free products. *Marketing Science* **26**(6) 742–757.

Susarla, A., J. Oh, Y. Tan. 2012. Influentials or susceptibles? analyzing cascades of word-of-mouth conversations in online social networks. *forthcoming in Information Systems Research* .

Tucker, C. 2008. Identifying formal and informal influence in technology adoption with network externalities. *Management Science* **54**(12) 2024–2038.

Van den Bulte, C., G.L. Lilien. 2001. Medical innovation revisited: Social contagion versus marketing effort. *American Journal of Sociology* **106**(5) 1409–1435.

Van den Bulte, C., S. Stremersch. 2004. Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science* **23**(4) 530–544.

Appendix A: Sample Last.fm web pages

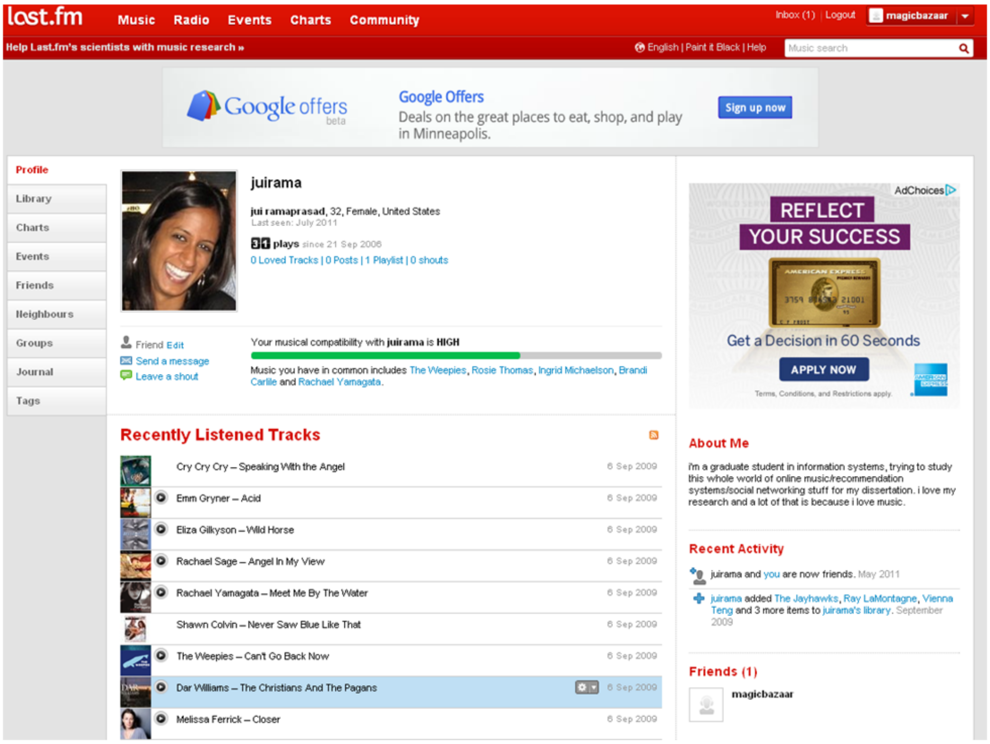

Figure 6 Sample Last.fm user page

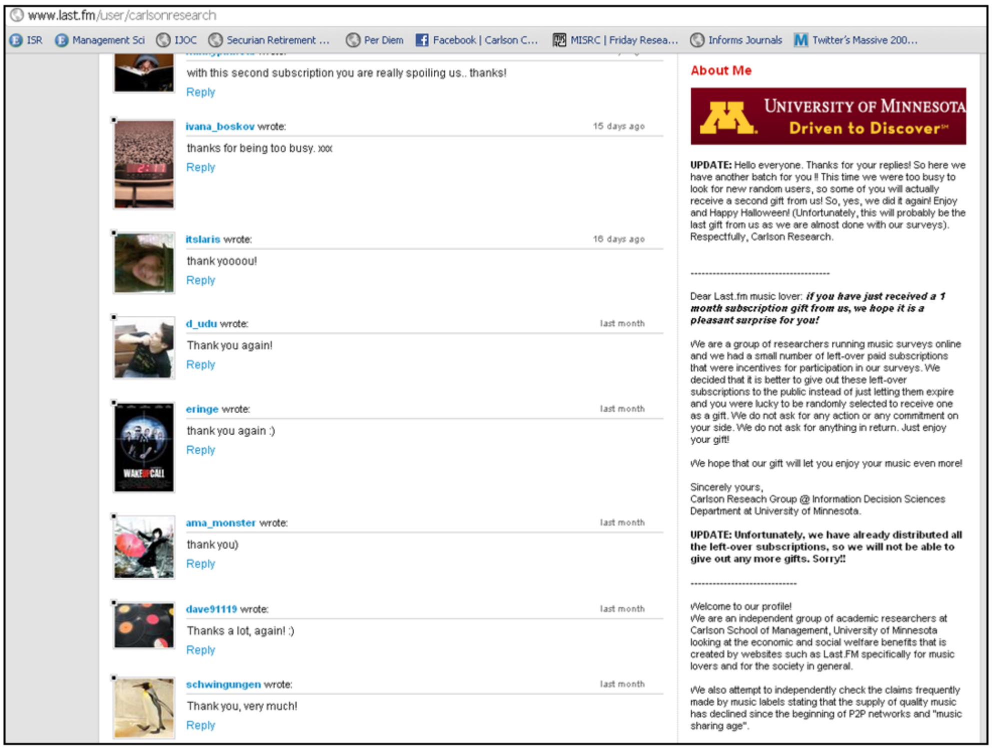

Figure 7 Gift landing page