Forthcoming in *Management Science*
2014

Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks Online Appendix

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 multiple well-known rival mechanisms that compete to be the explanation for this observed clustering. These range from peer influence to homophily to other unobservable external stimuli. These multiple mechanisms lead to similar observational data, yet have vastly different policy implications. In this paper, we present a novel randomized experiment that tests the existence of causal peer influence in the general population — one that did not involve subject recruitment for experimentation — of a particular large-scale online social network. We utilize a unique social feature to exogenously induce adoption of a paid service amongst a group of randomly selected users, and in the process develop a clean exogenous randomization of treatment and control groups. A variety of nonparametric, semiparametric and parametric approaches, ranging from resampling-based inference to ego-level random effects to logistic regression to survival models, yield close to identical, statistically and economically significant estimates of peer influence in the general population of a freemium social network. Our estimates show that peer influence causes more than a 60% increase in odds of buying the service due to the influence coming from an adopting friend. In addition, we find that users with a smaller number of friends experience stronger relative increase in the adoption likelihood due to influence from their peers as compared to the users with a larger number of friends. Our nonparametric resampling procedure based estimates are helpful in situations of networked data that violate independence assumptions. We establish that peer influence is a powerful force in getting users from free to premium levels, a known challenge in freemium communities.

Key words: Peer effects, randomized experiment, social contagion, nonparametric inference, freemium communities, online social networks

---

Appendix A: Sample Last.fm web pages

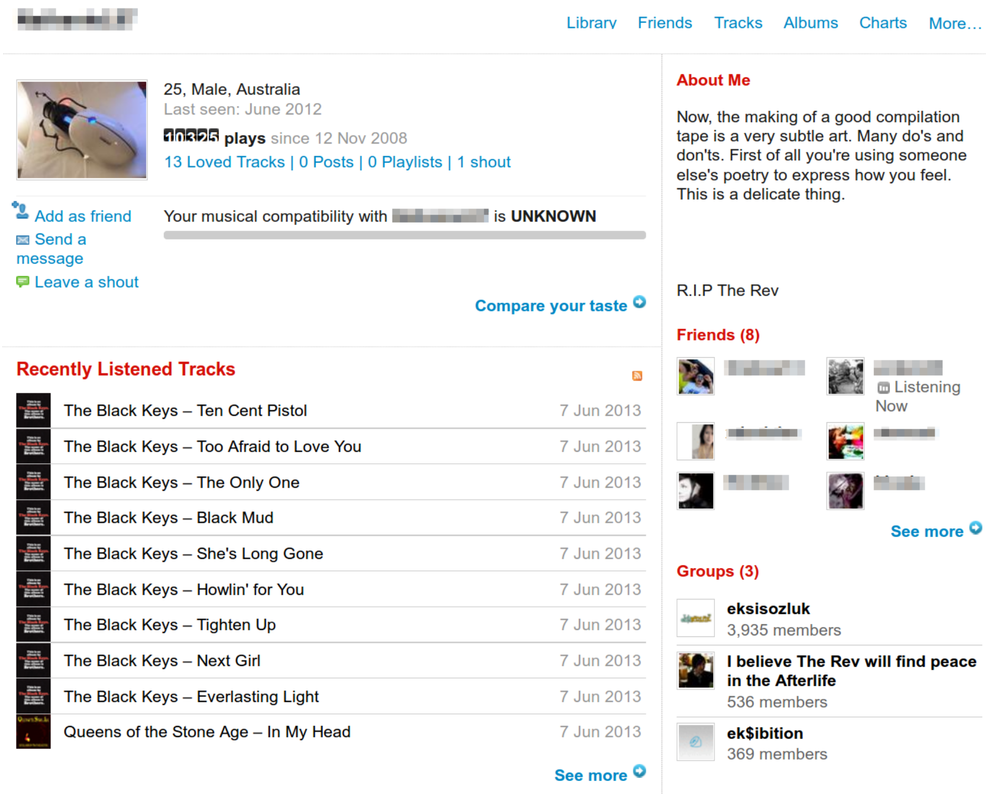

Figure 1 **Sample Last.fm user page**

Dear Last.fm music lover: ***if you have just received a 1 month subscription gift from us, we hope it is a pleasant surprise for you!***

We are a group of researchers running music surveys online and we had a small number of left-over paid subscriptions that were incentives for participation in our surveys. We decided that it is better to give out these left-over subscriptions to the public instead of just letting them expire and you were lucky to be randomly selected to receive one as a gift. We do not ask for any action or any commitment on your side. We do not ask for anything in return. Just enjoy your gift!

We hope that our gift will let you enjoy your music even more!

Figure 2 **Message accompanying the gift**

Appendix B: Algorithms

Algorithm 1 Algorithm of resampling test

$[A] \leftarrow$ all users who were eligible to receive a gift but did not.

$[i] \leftarrow 1$

while $[i] \leq 700$ **do**

  $[NM_i] \leftarrow$ random 1,000 users from $[A]$

  Compute groups T_i and C_i based on M and NM_i

  $[C_i] \leftarrow$ number of adopters in C_i

  $i \leftarrow i + 1$

end while

Compute the histogram of $[C_i]$

Algorithm 2 Algorithm for establishing heterogeneity by FriendCount

$[p] \leftarrow 50\%$

while $[p] \leq 100\%$ **do**

  $[T_p] \leftarrow$ bottom $[p]$ percentile of group T by FriendCount

  $[C_p] \leftarrow$ bottom $[p]$ percentile of group C by FriendCount

  Estimate the effect size using the resampling approach for $[T_p]$ versus $[C_p]$

  $[p] \leftarrow [p] + 10\%$

end while

Plot the effect sizes for each subsample $[p]$

Algorithm 3 Algorithm for reshuffling test

$[i] \leftarrow 1$

while $[i] \leq 800$ **do**

  $[M_i] \leftarrow$ mix random 500 users from true group M + random 500 users from group NM

  $[NM_i] \leftarrow$ the leftover 500 users from true group M + leftover 500 users from group NM

  $[T_i] \leftarrow$ number of adopter friends of $[M_i]$

  $[C_i] \leftarrow$ number of adopter friends of $[NM_i]$

  $[D_i] \leftarrow [T_i] - [C_i]$

  $[i] \leftarrow [i] + 1$

end while

Appendix C: Additional Discussion of Robustness

C.1. Interference of Treatment with Natural Adoptions in the Network.

Per our experimental design, the users both in the treatment group and in the control group may potentially be subject to peer influence from other parts of the network, which may change the way the users respond to the treatment. Hypothetically, had a user been part of a "sterilized" network, where everyone else stopped adopting, a user may have responded differently to our treatment since there would not be any interference. However, the following considerations mitigate this concern:

- Both the treatment and control group are subject to statistically the same background influence. Therefore, there is no threat to internal validity of our experiment.

- We define our treatment as having a gifted friend while being a part of a real-life social network and being subject to the normal background influence. In a sense, we are interested in studying the net treatment effect of an extra friend's adoption on top of a background influence level typically experienced by users from various sources, be it TV ads, email promotions or endorsements by some other friends of the user. While it is potentially true that the effect of our treatment may have been different had the entire network been artificially sterilized and prohibited from adopting except for 1,000 users in group M, in order to remove any possible interference of influences, in this study we are asking the question about the impact of our treatment in a real-life network. The effect of our treatment is, therefore, by definition the effect induced by gifting a user's friend on top of whatever else is going on naturally in this network. In this study, we do not make any claims about what would happen to a user in an artificial scenario when the only new adopters in the entire network are group M and the rest of the network is forced into non-adoption.

C.2. Replication of Prior Results

For robustness, we also independently replicated several effects that are intuitive and confirmed by the literature. Specifically:

- We observe that even after our gift manipulation had expired in group M, some people in group M decided to renew the subscription on their own. The count of "renewers" in group M was statistically larger than the count of "new adopters" in group NM despite the fact that initially the groups were chosen at random, confirming the well-known effect of free promotions.

- The estimation results suggest that subscribers and adopters tend to be older and to have registered earlier than the general population. This is in sync with earlier findings by Oestreicher-Singer and Zalmanson (2013) 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 a premium subscription gives many more features to people outside of the USA, 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 the Last.fm social network is a domain that is subject to traditional economic laws. We hope that the insights gained from the Last.fm domain are a manifestation of more fundamental phenomenon applicable to other domains as well.

Appendix D: Predictive analysis: peer influence and homophily

Motivation

It is natural to explore whether the dominant force behind the observed clustering of adoptions is peer influence or homophily and other confounders. In this appendix, we explain how we can use our experimental data in an attempt to shed a light on this question. However, before we present our results, we would like to discuss how we can meaningfully ask about the comparison of peer influence with homophily and other confounders.

While it is relatively easy to define what one means by *causal effect of peer influence*, it is non-trivial to define *causal effect of homophily* since manipulating homophily is ill-defined. Specifically, in the case of peer influence, we can specify very precisely what we mean by manipulating peer influence while keeping everything else constant. For example, in our experiment, we define causal effect by comparing two potential outcomes: (1) adoption outcome of user A if A's friend was gifted as compared to (2) adoption outcome of user A if A's friend was not gifted.

It turns out that defining the counterfactual outcome for homophily is non-trivial, since it is not well-defined what it means to manipulate homophily while keeping everything else constant. Therefore, if one were to compare the causal effect of peer influence and homophily, the comparison is not straightforward since the term *causal effect of homophily is not well-defined*[1].

Given this difficulty of defining the causal effect of homophily, we will instead compare the predictive strength of peer influence and homophily since our experimental data is unique in that it allows us to observe both the natural adopters and the randomly gifted adopters at the same time as possible predictors of the future adoptions.

Predictive Strength of Peer Influence and Homophily

In order to define what we mean by *predictive strength*, consider the following example. Assume we are assigning "adoption scores" to users based on a predictive model with a good predictive performance. That is, a user with an adoption score of 0.0 is predicted by the model to be a very likely non-adopter, while a user with an adoption score of 1.0 is predicted by the model to be a very likely adopter.

Our scoring model has good predictive power (which we demonstrate below) and is based on the data available so far. As new events keep happening we revise our scores[2].

The following statements can then be made:

1. The predictive model should bump user A's score up by X if it suddenly sees that an extra friend of user A was randomly gifted by us.

2. The predictive model should bump user A's score up by Y if it suddenly sees that an extra friend of user A suddenly adopts on her own.

[1] Unless certain restrictive assumptions are made about the nature of homophily.

[2] The change in score only means that the predictive model's belief about the user changes. The user's actual propensity to adopt may remain constant. This is similar to, say, the FICO score assigned to individuals by credit agencies. The individual's true probability of default may actually remain constant, while the credit agency keeps revising the FICO score upward or downward as new information about the user is observed. In a sense, this section of our paper studies the impact of new peer influence information or homophily information on the predictions of the scoring model rather than on the true propensity to adopt.

Variable	Estimate	Std Err	t-value	Pr> $\|t\|$
Intercept:	−4.55183	3.92637	−1.16	0.2463
NumGiftedFriends	**1.57825**	0.69554	2.27	0.0233
NumGiftedFriends * log(FriendCnt)	**−0.25804**	0.14654	−1.76	0.0783
NumSelfAdopterFriends	**2.06548**	0.55210	3.74	0.0002
NumSelfAdopterFriends * log(FriendCnt)	**−0.35363**	0.10190	−3.47	0.0005
log(FriendCnt)	0.04331	0.18366	0.24	0.8136
log(SubscriberFriendCnt)	0.44144	0.17691	2.5	0.0126
Age	0.02392	0.01500	1.6	0.1108
Gender	−0.33316	0.21373	−1.56	0.1193
LastfmCountry	−0.40246	0.23819	−1.69	0.0914
RegDate	−0.00030	0.00020	−1.51	0.1306
log(SongsListened)	0.12724	0.08840	1.44	0.1501
log(Posts)	0.06899	0.06498	1.06	0.2884
log(Playlists)	0.37427	0.15997	2.34	0.0193
log(Shouts)	−0.01703	0.07766	−0.22	0.8265
log(LovedTracks)	0.24741	0.06400	3.87	0.0001

Table 1 The effect of NumGiftedFriends and NumSelfAdopterFriends on the adoption score

As a predictive model we employ a logistic regression that is estimated in Table 1. In order to interpret the impact of coefficients on a predicted score, assume we look at a median user A from group T that has 40 friends on Last.fm. According to Table 1, if the predictive model suddenly observes new information about user A and

- the new information shows that an extra friend of user A was gifted a premium subscription, then the predictive model will revise user A's score upward by $X = e^{1.5782-0.2580\log(40)} = 1.87$; that is, by 87%.
- the new information shows that an extra friend of user A adopted premium subscription on her own, then the predictive model will revise user A's score upward by $Y = e^{2.0655-0.3536\log(40)} = 2.14$; that is, by 114%.

The key surprising result of this section is that the difference between X and Y is small and is statistically insignificant ($p > 0.85$). In other words, our result can be summarized as follows. For a predictive model, there is no big predictive signal difference between the two kinds of information, when user A's friend suddenly adopted a subscription by herself as compared to when user A's friend is suddenly gifted with a subscription.

This result is non-trivial. If homophily and other confounders were indeed the dominant predictive "force" behind the observed new adoptions in this network, then a predictive model would naturally learn that Y is much greater than X, since in that scenario, having a self-selected adopter friend would bump one's score a lot higher as compared to just having a gifted friend, which carries no homophily signal. Surprisingly, this is not what we observe in the data.

We should emphasize that the results demonstrated in this section are predictive and not causal. In other words, we only explore the changes in scores that are assigned by a particular predictive model to users as the model applies new information, but we do not make any claims in this appendix that the actual probability of the user's adoption changes with the new information. This is an important distinction in our case. For instance, if NumSelfAdopterFriends were suddenly observed to go from 0 to 1 for some user A, we do not claim that user A's true probability of adoption increases (that would be a causal claim); we only say that it would be rational for the predictive model (as a third party) to revise its belief (the adoption score) upward because of this new information. The true probability may as well be unchanged by the increase of NumSelfAdopterFriends, but the predictive model clearly should update the score in the light of this new signal learned about the focal user.

Appendix E: Quasi-Experiment

E.1. Description

As Tables 1 and 2 in the paper demonstrate, paid subscriptions are rare events[3] 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 in order to estimate the appropriate sample size as well as to compare the bias of quasi-experiment as compared to a randomized experiment.

Based on this quasi-experiment, we discovered that the sample size of 1,000 would be an appropriate size for an experiment, while the sample size of 500 or less cannot yield statistical significance even in a potentially confounding setting of a quasi-experiment. As is evident from Tables 12 and 11 from the paper, the estimate of the treatment effect based on the quasi-experiment are indeed very close to the estimate of the treatment effect based on the randomized experiment.

We construct a quasi-experiment using observational data in a way that would simulate the randomized experiment by matching existing observed adopters based on observed characteristics. In order to introduce the quasi-experiment, consider three consecutive times in the evolution of our data: t_1, t_2 and t_3, each separated by one month. If we look into our data across the one month period $[t_1, t_2]$, we will see that more than a thousand users became subscribers in that time period $[t_1, t_2]$ as demonstrated by Table 2. We will refer to these users as "$0 \mapsto 1$" users and randomly select 1,000 of them into a group "M". Also, as demonstrated by Table 2, in the same time period $[t_1, t_2]$, we will see more than 1 million users who remained non-subscribers. We will refer to them as "$0 \mapsto 0$" users.

For every user u in group "M", we match her to a user who has the observed properties identical to user u but happened to remain a "$0 \mapsto 0$" user in the time frame $[t_1, t_2]$. In other words, we match every $0 \mapsto 1$ user from group "M" with a random $0 \mapsto 0$ "twin" based on the observed variables and thus form a group "NM" of 1,000 "twins".

Similarly to our experimental setup, we define quasi-treatment group "T" as all immediate friends of "M" who are not themselves in "M" and who are not friends of someone in "NM". Symmetrically, we define quasi-control group "C" as all immediate friends of "NM" who are not themselves in "NM" and are not friends of someone in "M". Clearly, because of the matching, groups "M" and "NM" are statistically identical in terms of the observed characteristics that were used for matching at time t_1. By comparing the subscription changes in groups "T" and "C" during the subsequent time period $[t_2, t_3]$ and controlling for all of the observed characteristics of each user, we are able to tell whether being a friend of "M" has any effect on the subscription behavior as compared to being a friend of "NM". Our logistic regression analysis of adoption in groups "T" and "C" is presented in Table 2. This analysis demonstrates that the *QuasiTreatment* dummy variable (which tells whether a given person is in "T" or in "C") is statistically

[3] For example, only 3% of active users are currently subscribers and 0.2% of users are recent adopters in the one month period. However, 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 the size of Facebook this 0.2% would correspond to more than 1.5 million unique users (not even counting the social multiplier effect). In addition, the rareness depends on the chosen time scale.

Variable	Estimate	Std Err	z-value	Pr> $\|z\|$
Intercept: adopter=0	−10.2465446	2.5243634	−4.059	<0.0001
QuasiTreatment	**0.5958513**	0.1666775	3.575	0.000350
log(FriendCnt)	−0.3739855	0.1218169	−3.070	0.002140
log(SubscriberFriendCnt)	0.8169019	0.1288109	6.342	<0.0001
Age	0.0007438	0.0111094	0.067	0.946622
Gender (Male=1)	−0.0604786	0.1626430	−0.372	0.710005
LastfmCountry	−0.5287179	0.1767493	−2.991	0.002778
RegDate	0.0001270	0.0001519	0.836	0.403234
log(SongsListened)	0.2612372	0.0766349	3.409	0.000652
log(Posts)	−0.0317030	0.0542661	−0.584	0.559077
log(Playlists)	0.1030531	0.1268074	0.813	0.416405
log(Shouts)	−0.0578087	0.0621945	−0.929	0.352639
log(LovedTracks)	0.2467524	0.0503097	4.905	<0.0001

Table 2 Quasi-experimental results: logistic regression

significant in explaining the decision to subscribe after controlling for a variety of demographic, network, and social activity variables.

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 (2013). 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."

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 in order to establish the effect of peer influence from observational data. The following variable is used as a manipulation variable in this particular analysis:

- $QuasiTreatment_i$. This quasi-manipulation variable represents the dummy variable that indicates whether user i is a friend of group "M" or group "NM"[4].

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

E.2. Results

As evident from Table 2, *QuasiTreatment* variable is statistically significant even after controlling for observed individual user characteristics. Moreover, since *QuasiTreatment* is assigned independently of characteristics of user i, this coefficient has a causal interpretation: *QuasiTreatment causes* the adoption of subscription, thus providing additional evidence for Hypothesis 1.

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

Since *QuasiTreatment* is a dummy variable, it is easy to estimate the average marginal effect of *QuasiTreatment* on odds of adopting the subscription: if *QuasiTreatment* changes from 0 to 1, the odds of adoption increase by $e^{0.59585}$; that is, by a factor of 1.81.

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.

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 1,000 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.

Appendix F: Logistic regression: Model and diagnostics

In Section 4 in the paper, we refer to a logistic regression to estimate the percentage increase caused by our treatment. The main theoretical rational for using the logistic regression is that it is a Maximum Entropy Classifier (Mount 2011), therefore, motivating our use of logistic regression is, in a certain sense, necessarily easier than motivating the use of any other model that would predict a binary variable.

The estimation of the logistic regression with random effects is shown in Table 3 below. In order to evaluate the fit of logistic regression, we present the following tests:

- Significance of the model (likelihood ratio test). Likelihood Ratio of our model = 75.1996 ($p < .0001$), demonstrating the significance of the model as a whole.
- Predictive accuracy of our logistic model is demonstrated by ROC curve in Figure 3. As demonstrated in the Figure 3, the area under the curve represents a respectable predictive performance of 0.765.
- Hosmer-Lemeshow test of goodness of fit of the logistic regression fails to reject the logistic regression hypothesis with $\chi^2 = 9.33$ ($p = 0.315$), indicating that the data is not inconsistent with the logistic regression assumptions.[5]
- Analysis of Pearson and Deviance residuals[6] obtained in our logistic regression is presented in Figure 4. Since we are using individual outcome data and not frequency data (that is, $n_i = 1$ for each observation), the interpretation of Pearson and Deviance residuals is not straightforward. Therefore, for comparison, in Figure 5 we present the results of a simulation showing what Pearson and Deviance residuals look like for the data that is truly generated by the following (arbitrarily-picked) logistic regression:

$$\text{logit}\left(Pr\{\text{Adopter} = 1\}\right) = -7.9 + 0.5 \cdot \textit{Treatment}$$

[5] In Hosmer-Lemeshow test (Hosmer et al. 2013) the null hypothesis is that the true model is a logistic regression. We acknowledge that a failure to reject the null hypothesis does not mean that the null hypothesis is true.

[6] We thank the anonymous reviewer for this suggestion.

Variable	Estimate	Std Err	z-value	Pr> $\|z\|$
(Intercept)	-4.2692	4.1617	-1.0260	0.3050
Treatment	**0.5203**	**0.2549**	**2.0410**	**0.0413**
Age	0.0266	0.0181	1.4700	0.1415
Gender (Male=1)	-0.4674	0.2566	-1.8220	0.0685
log(FriendCnt)	-0.3969	0.1885	-2.1050	0.0353
log(SubscriberFriendCnt)	0.5975	0.2045	2.9210	0.0035
LastfmCountry	-0.7398	0.3062	-2.4160	0.0157
RegDate	-0.0004	0.0003	-1.5120	0.1305
log(songsListened)	0.2818	0.1278	2.2050	0.0274
log(Posts)	-0.0029	0.0855	-0.0340	0.9728
log(Playlists)	0.4699	0.2214	2.1230	0.0338
log(Shouts)	0.0129	0.0978	0.1320	0.8953
log(LovedTracks)	0.2265	0.0793	2.8550	0.0043

Table 3 **Experimental results: logistic regression with random effects**

As is evident from comparing Figures 4 and 5, residuals obtained by fitting a logistic model to our data resemble residuals obtained by fitting a logistic model to the data that was generated precisely according to a logistic regression specification.

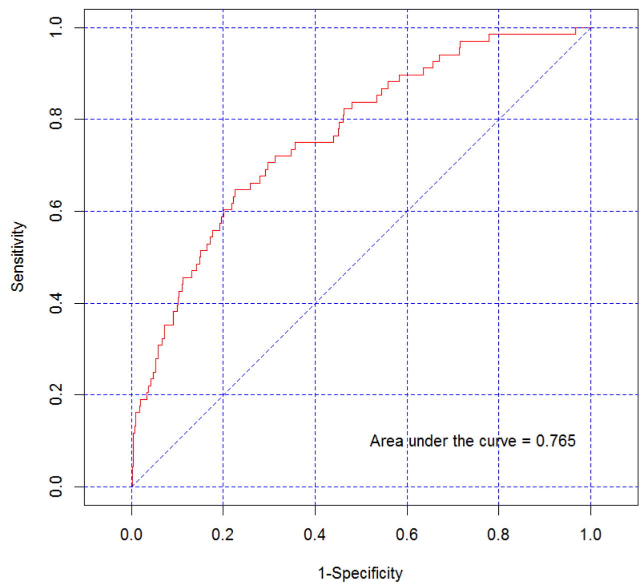

Figure 3 **ROC curve for predictive performance of logistic regression**

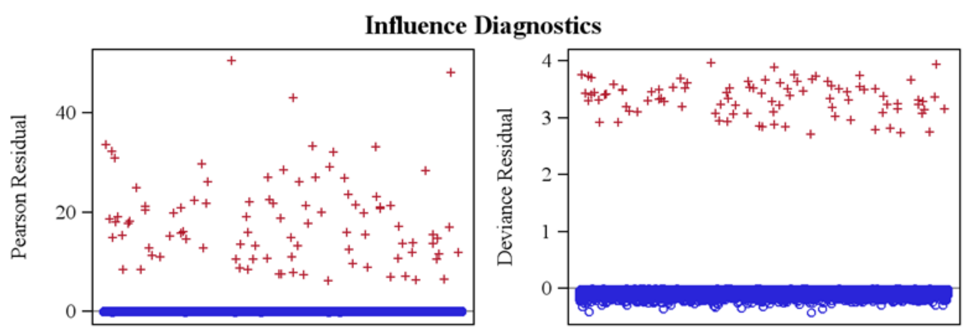

Figure 4 **Pearson residuals and Deviance residuals for our logistic model**

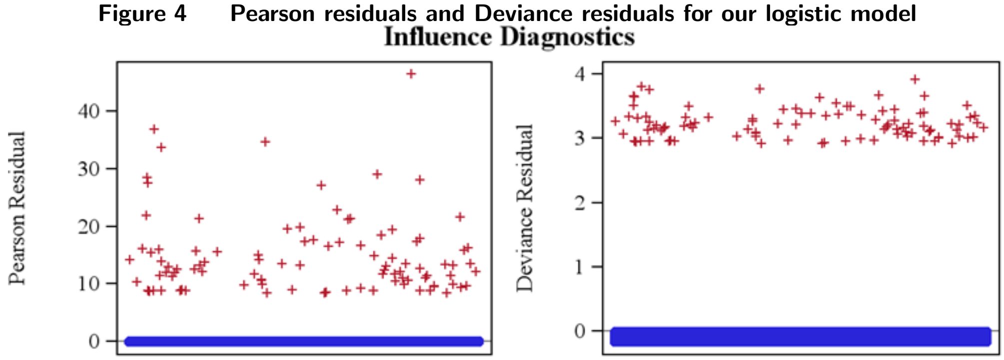

Figure 5 **Pearson residuals and Deviance residuals for a true logistic model**

References

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.

Hosmer, D., S. Lemeshow, R. Sturdivant. 2013. *Applied logistic regression, 3rd ed.*. Wiley. com.

Mount, J. 2011. The equivalence of logistic regression and maximum entropy models. Also available as `http://www.win-vector.com/dfiles/LogisticRegressionMaxEnt.pdf`.

Oestreicher-Singer, Gal, Lior Zalmanson. 2013. Content or community? a digital business strategy for content providers in the social age. *Management Information Systems Quarterly* **37**(2) 591–616.

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* .