



# DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

## Essays on the Social Consumer: Peer Influence in the Adoption and Engagement of Digital Goods

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters.

|                     |                                                                                                                                                                                                                                                                                                                                                  |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Citation</b>     | Davin, Joseph. 2015. Essays on the Social Consumer: Peer Influence in the Adoption and Engagement of Digital Goods. Doctoral dissertation, Harvard Business School.                                                                                                                                                                              |
| <b>Accessed</b>     | July 21, 2018 5:34:51 PM EDT                                                                                                                                                                                                                                                                                                                     |
| <b>Citable Link</b> | <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:16881890">http://nrs.harvard.edu/urn-3:HUL.InstRepos:16881890</a>                                                                                                                                                                                                                            |
| <b>Terms of Use</b> | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a> |

*(Article begins on next page)*

## Essays on the Social Consumer: Peer influence in the adoption and engagement of digital goods

### Abstract

In this dissertation, I study how consumers influence each other in the adoption and engagement of digital goods.

In the first essay, I study peer influence in mobile game adoption. Although peer effects are expected to influence consumer decisions, they are difficult to identify in observational studies due to selection bias: Friends share common characteristics and behave in similar ways even without peer effects. I use a novel approach to estimate unobserved characteristics which endogenously drive tie formation and use the estimates to control for selection, without need for instruments. This is the first paper to use latent space to reduce bias in peer influence estimates. I find that peers account for 27% of mobile game adoptions, and that ignoring latent homophily would bias the estimates by 40%, in line with previous studies. In some samples, ignoring latent homophily can result in overestimation of social effects by over 100%.

In the second essay, I examine the effect of zero rating on consumer behavior in a social network. I use Facebook data on millions of users to quantify direct, peer, and long-term effects of zero rating, a campaign where consumers can access digital media over mobile networks for free, on social network activities. I find that zero rating does not have the same effect on all social network activities. While the direct impact of zero rating is positive on all activities, users with more friends on zero rating create less, consume more, and give more feedback on content. In addition, zero rating does not have a uniform effect across consumers. Some consumers benefit more from zero rating than others, and I show that network characteristics can help identify those consumers whose network benefits the most from zero rating.

# Contents

|     |                                                        |    |
|-----|--------------------------------------------------------|----|
| 1   | Introduction                                           | 1  |
| 2   | Peer Influence in the Adoption of Social Games         | 3  |
| 2.1 | Homophily and social influence . . . . .               | 5  |
| 2.2 | Data . . . . .                                         | 12 |
| 2.3 | Model and estimation . . . . .                         | 13 |
| 2.4 | Results . . . . .                                      | 20 |
| 2.5 | Discussion . . . . .                                   | 23 |
| 3   | The Effect of Zero Rating on Social Network Activities | 26 |
| 3.1 | Pricing and social network activities . . . . .        | 29 |
| 3.2 | A model of social network activities . . . . .         | 33 |
| 3.3 | Results . . . . .                                      | 38 |
| 3.4 | Conclusion . . . . .                                   | 44 |
| 3.5 | Further extensions . . . . .                           | 45 |
| 4   | Conclusion                                             | 49 |
|     | References                                             | 57 |

# Listing of figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                           |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Log marginal likelihood (LML) by dimensions of latent space . . . . .                                                                                                                                                                                                                                                                                                                                     | 19 |
| 2.2 | Distribution of peer effects as a driver of mobile app adoption . . . . .                                                                                                                                                                                                                                                                                                                                 | 22 |
| 2.3 | Drivers of mobile app adoption (% of total adoptions) . . . . .                                                                                                                                                                                                                                                                                                                                           | 23 |
| 2.4 | Bias to effect size ratio in each sample . . . . .                                                                                                                                                                                                                                                                                                                                                        | 24 |
| 3.1 | Data timeline of zero rating campaign and pricing . . . . .                                                                                                                                                                                                                                                                                                                                               | 31 |
| 3.2 | Average daily activity per person: Opt-in curve includes anyone who opted into zero rating at any point during the campaign; normalized so non-opt in is 1.0 at the beginning of the data period . . . . .                                                                                                                                                                                                | 34 |
| 3.3 | Covariate balance (average degree and pre-campaign activity level) before and after matching (FB stories example). Before matching, the circle and triangle represents average degree and pre-campaign activity for zero rated and non-zero rated individuals respectively. After matching, there are nine subclasses, so there are nine pairs of statistics. Scale redacted for confidentiality. . . . . | 39 |
| 3.4 | Average daily stories viewed per person by subclass: Opt-in curve includes anyone who opted into zero rating at any point during the campaign; normalized so non-opt in is 1.0 at the beginning of the data period . . . . .                                                                                                                                                                              | 40 |
| 3.5 | Significant heterogeneity in post-stratified estimates by subclass (FB stories coefficients example); Scale redacted for confidentiality. . . . .                                                                                                                                                                                                                                                         | 43 |

This dissertation is dedicated to my family, both natural and chosen.

# Acknowledgments

I would like to thank the Marketing Unit at HBS, the doctoral office, staff and faculty at HBS, research collaborators, and colleagues at Harvard University, Facebook, and beyond, especially Elie Ofek, Thales Teixeira, Doug Chung, Pete S. Fader, Bruce G.S. Hardie, Mikolaj Jan Piskorski, Eytan Bakshy, Asim Ansari, Kamel Jedidi, Lingling Zhang, Clarence Lee, Pavel Kireyev, Taylan Yalcin, and Tuan Phan.

I would like to extend special thanks to my dissertation committee members Eduardo M. Airoidi and Thomas J. Steenburgh for their professional guidance and generosity.

I am greatly indebted to Sunil Gupta for his incomparable mentorship. There are few words that describe the extent of my appreciation for the chance to work with him. I want to express my deepest gratitude for his critical yet kind spirit, openness to some of my crazy ideas, patience with my not-so-crazy ideas, wisdom through all our projects together, and perhaps most of all, time that he could be spending on more important people. Without his help, my voice would not be what it is today.

Last but not least, I thank my family, especially my parents Paiboolya and Supranee Gavinlertvatana and my partner Alan, for standing by me through everything.

*People influence people. Nothing influences people more than a recommendation from a trusted friend.*

Mark Zuckerberg, CEO Facebook

# 1

## Introduction

Advertising guru David Ogilvy (2013) once said that word of mouth was “manna from heaven, but nobody knows how to do it on purpose.” In this dissertation, I build on existing work to introduce two new techniques and empirical settings that help advance our understanding about “the social consumer” and how peers influence consumers’ adoption and engagement of digital goods.

Call this “manna” what you will – word of mouth, social influence, peer effects, peer spillovers,

network effects – these ideas all revolve around the idea that one individual’s decision depends not just on their own personal circumstance but also on those around them. In this dissertation, I use the above terms synonymously to refer to the general phenomenon of individuals making decisions not just based on their own characteristics but also based on the actions of those around them.

This dissertation addresses how peers influence adoption and engagement in two new and exciting markets. According to Nielsen, the average US consumer spends 37% of time on smartphones using games and social networking. This statistic is even higher abroad: 40% and 47% in Japan and the UK respectively (Nielsen 2014). From my research, there is a big opportunity to gain more customers and to keep them more engaged using peer influence within gaming and social networking.

In the first essay, I explore how peers influence the adoption process of social games. In order to separate social influence from shared traits between friends (“homophily”), I introduce latent space models as an approach to impute those shared traits between friends, and use these as covariates in a hierarchical Bayesian model. I find that controlling for latent traits dramatically decrease the estimated size of social influence, although social influence still accounts for a quarter of all adoptions. The effect varies significantly from sample to sample, underscoring the need to use multiple samples and large datasets.

In the second essay, I study how peers influence engagement in social networks. I use data from an Internet.org campaign where Facebook is zero rated, i.e., users do not pay for data charges when accessing Facebook, on millions of users and billions of connections. Using matching and fixed effects regression to isolate the causal effect of social influence, I find that changing mobile data prices has both direct and indirect effects, and have differential effects across activities and consumers.



*Right now we spend three billion hours a week playing online games. Some of you might be thinking, "That's a lot of time to spend playing games. Maybe too much time, considering how many urgent problems we have to solve in the real world." But actually, according to my research at The Institute For The Future, it's actually the opposite is true. Three billion hours a week is not nearly enough game play to solve the world's most urgent problems.*

Jane McGonigal, game designer

# 2

## Peer Influence in the Adoption of Social Games

Mobile gaming is a large industry and is relevant to consumers as well as companies. Revenues are expected to grow to \$35 billion by 2017 (Takahashi 2014). Companies are increasingly looking to connect with consumers via mobile games. Coca-Cola, in partnership with gaming company Ubisoft, launched a mobile dance game to connect with teens and healthy lifestyles (Maytom

2014). Red Bull produces its own mobile games, including one that was voted by Google as one of the best games in 2014 (RedBull 2014).

Intuitively, mobile games spread via social networks, but there is no study to quantify the role of peers in the adoption of mobile games.

The study of peer influence is important to marketing because social influence has the potential to affect not just mobile games but many areas of consumer decision making. McKinsey & Company estimates that a third of all consumer spending, c. \$940 billion in annual consumption in the US and Europe alone, will be influenced by social interactions (Chui 2013). The ubiquitousness of social influence in consumption decisions makes it an attractive marketing opportunity for firms. As a consequence, researchers want to quantify contagion effects in a myriad of consumer settings (Godes et al. 2005, Hartmann et al. 2008).

To be effective in social networks, marketers have to take actions in line with consumer decision making processes, not just for the individual but for the individual's peers. The first possible driver is social influence, where one individual influences the other. If purchase decisions are mainly driven by social influence, then marketers need to augment their marketing strategy to target the most influential individuals in the social network or build product features that promote peer influence (Tucker 2008, Trusov et al. 2010, Hartmann 2010, Aral & Walker 2011a). Another possible explanation is homophily, which is the concept that people who share things in common are more likely to be friends. Because of this, friends act in similar ways because they possess similar traits. If homophily is the dominant force, then there are latent traits responsible for purchase. This means marketers should uncover these traits for segmentation and targeted marketing actions (Hill et al. 2006). In order to be effective, it is important to take the right action in response to the right force governing consumer behavior.

However, identifying peer effects is not straightforward because social influence can be confounded with homophily. When one friend uses a game and another installs a new game, is it

their latent similarity that drives adoption or is it because of peer influence?

I use *latent space models* to untangle homophily from social influence. First, I use the social network structure to extract latent traits that govern friendship formation. Then, I use the estimated latent space coordinates to control for homophily when estimating the effect of peer influence on adoption. I find that homophily can inflate the effect of peer influence by 40%, and even after controlling for homophily, peers still drive more than a quarter of all mobile game installations.

The rest of the paper is laid out as follows: I discuss homophily and social influence in section 2.1, illustrating how ignoring latent traits could introduce bias in peer influence estimates, and introduce latent space models. Section 2.2 covers data of games adoption in a social network. Section 2.3 introduces the model and estimation approach. Results are in section 2.4, and I conclude in section 2.5.

## 2.1 Homophily and social influence

Social influence has profound impact on many aspects of customer decision making and marketing, including the adoption of new technologies (Tucker 2008, Aral et al. 2009, Aral & Walker 2011a, 2012), social network usage and adoption (Trusov et al. 2010, Aral & Walker 2011b, Ghose & Han 2011, Katona et al. 2011), eCommerce (Stephen & Toubia 2010), new prescriptions of pharmaceutical drugs (Manchanda et al. 2008, Nair et al. 2010, Iyengar et al. 2011), ad effectiveness (Bakshy et al. 2012), group decision making (Hartmann 2010), and customer retention (Nitzan & Libai 2011). Effect size varies by the empirical setting. For example, Hartmann (2010) finds that 35% of customer value is attributable to peer effects while Trusov et al. (2010) found that only 1 in 5 customers actually influence their peers.

One challenge for social social influence studies is separating latent homophily and social influence (Shalizi & Thomas 2011, Angrist 2014). Homophily, the phenomenon that “birds of a

feather flock together,” has been widely documented in sociology and organizational behavior (Kandel 1978, Ibarra 1992, McPherson et al. 2001, Kossinets & Watts 2009). Latent homophily arise from correlated unobservable traits among friends (sometimes called endogenous tie formation) and can inflate estimates of social influence if those latent traits also cause the outcome of interest. This is analogous to omitted variable bias or confounding which induces correlation between friendship and the outcome of interest. The critique is not new; Manski (1993) discusses identification problems in endogenous social effects for linear-in-means models.

Marketing literature dealing with social influence has shed some light on how to identify social influence effects. Van den Bulte & Lilien (2001) show that marketing efforts, which is similar among doctors who know each other in a network, can explain the pattern of new drug adoption. Once they control for these marketing actions, they find no evidence of social influence among doctors in adoption of a new pharmaceutical drug. Tucker (2008) suggest that not properly controlling for endogenous group ties could bias peer effects by 50% in technology adoption among employees in a firm.

In order to make use of observational data, one common attempt to separate the confounding of homophily and social influence is to include covariates in regressions and matching (e.g., Aral et al. 2009, Nitzan & Libai 2011, Iyengar et al. 2011). Covariates could contribute to social ties as well as the outcome, and so controlling for them should remove bias arising from observed homophily. In the adoption of a new instant messaging system, Aral et al. (2009) show that naive models that do not control for covariates could bias peer effect estimates by 300-700%.

Our use of latent space to control for homophily is motivated by Shalizi & Thomas (2011) who suggest that in order to get identification of social influence, unobserved traits that influence tie formation must be made observable. I address this by using latent space models to infer unobserved traits that underlie social tie formations. Latent space provides a model of tie formation in a network by co-locating friends together and separating apart strangers in a  $k$ -dimensional space

(Hoff et al. 2002, Hoff 2005, Krivitsky et al. 2009).

Although the approach itself is no stranger to marketing, I use latent space for a different purpose than research in the past. Previous researchers focus on the *distance* between individuals in latent space for community detection. Ansari et al. (2011) use latent space to model distances between managers and musicians. Braun & Bonfrer (2011) build a latent space using cellular phone call network, to predict future calls within the network based on how far apart people are in latent space. Instead of focusing on the distance between individuals, I focus on the *coordinates* of the estimated latent space. This paper is the first to use latent space models to control for homophily in peer influence studies.

In the next section, I will sketch a latent space model and how its coordinates control for homophily.

### 2.1.1 Social network model based on latent space

I begin by defining a social network by its sociomatrix  $\mathcal{A}$ , where  $\mathcal{A}_{ij} = 1$  if persons  $i$  and  $j$  are friends and 0 otherwise. Suppose that each person has a set of observed characteristics  $X_i$  and unobserved characteristics  $\xi_i$ ; assume they are both dimension 1 for illustrative purposes. Following Hoff et al. (2002), I assume that the log odds of two persons being friends is linear in the distance between them according to observed and unobserved characteristics:

$$P(\mathcal{A}_{ij} = 1) = \frac{\exp(\gamma_0 + \gamma_1|X_i - X_j|^2 + \gamma_2|\xi_i - \xi_j|^2)}{1 + \exp(\gamma_0 + \gamma_1|X_i - X_j|^2 + \gamma_2|\xi_i - \xi_j|^2)} \quad (2.1)$$

The latent space model assumes dyadic independence, that is, the probability of a tie between two individuals depends only on the distance between them in the latent space. The latent characteristic  $\xi$  is identified by the network structure and the absence of shared observed characteristics.

For example, if two people are friends, and they look completely different according to their observed characteristics, it is highly likely that they share some common latent trait that induced their friendship.  $\xi$  could capture a range of unobserved characteristics, from unmeasured traits to beliefs and preferences.

To summarize, a latent space model takes data the sociomatrix and observed characteristics, then infers the unobserved characteristics (coordinates) which gives rise to the distances between individuals which fit the data. I will show how leaving out unobserved characteristics will introduce bias, and how the coordinates help reduce this bias.

### 2.1.2 Latent homophily introduces bias

If I leave out latent traits related to friendship, it could introduce bias when estimating peer effects. As an example, suppose that individuals become friends based on age (observed) and intelligence (unobserved): the more similar in age and/or intelligence, the more likely two individuals are to become friends. Now, suppose that higher age and intelligence increases the likelihood to download a new smartphone game that improves brain activity, and that there is no social influence. Because of homophily, there will be clusters of friends who are similar in age and/or intelligence. Since friends have similar likelihood to download the game, I would see clustering of game downloads within the social network. If I did not control for either age or intelligence, I might conclude that the clustering of game adoption within the social network is due to social influence. If I was able to control for both age and intelligence, I would be able to correct for the likelihood of app download for each individual, which would explain the clustering of downloads among friends with similar profiles.

More formally, I assume that the outcome variable depends on its lag, peer effects, and observed and latent characteristics,  $X_i$  and  $\xi_i$  (Shalizi & Thomas 2011).

$$Y_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 \frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}} + \beta_3 X_i + \beta_4 \xi_i + \epsilon_{it}, i = 1, \dots, N \quad (2.2)$$

where  $\beta_1$  is the effect of lagged outcome,  $\beta_2$  captures the effect of average lagged outcome of connected peers,  $\beta_3, \beta_4$  are the effect of observed and unobserved characteristics on the outcome variable  $Y_{it}$ , and  $\epsilon_{it}$  is an i.i.d. error term with mean 0. Assume I estimate the coefficients using linear regression.

If  $X_i$  and  $\xi_i$  are observed, there is no problem since every variable on the right hand side is observed. However,  $\xi_i$  is a latent characteristic and is unknown to the analyst. Suppose I wrongly assume that there is no latent characteristic and estimate the following:

$$\hat{Y}_{it} = \beta'_0 + \beta'_1 Y_{i,t-1} + \beta'_2 \frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}} + \beta'_3 X_i \quad (2.3)$$

Comparing the above two equations, it should be apparent that this is an omitted variable bias problem. The omission of  $\xi_i$  induces bias in the coefficients through its correlation with other variables.

I formalize this by regressing  $\xi_i$  on observed right hand variables:

$$\xi_i = \tau_0 + \tau_1 Y_{i,t-1} + \tau_2 \frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}} + \tau_3 X_i + \epsilon_i^1 \quad (2.4)$$

Then, from omitted variable equations, I know that  $\beta'_2 = \beta_2 + \tau_2 \beta_4$ .

To summarize, bias is introduced to the social influence coefficient  $\beta_2$  based on the correlation between  $\frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}}$  and  $\xi_i$  above and beyond its correlation with  $X_i$ . Shalizi & Thomas (2011)

point out that current methods that use  $X_i$  (such as matching methods in Aral et al. 2009) to control for observed homophily also controls for any latent homophily to the extent that  $X_i$  and  $\xi_i$  are correlated.

Omitted variable bias depends on two necessary conditions. The first condition is that  $\xi_i$  is correlated to the outcome variable, given all the other variables, i.e.,  $\beta_4 \neq 0$ . The second condition is that  $\xi_i$  is correlated with peer activity  $\frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}}$  after controlling for  $X_i$ , i.e.,  $\tau_2 \neq 0$ . If either  $\beta_4 = 0$  or  $\tau_2 = 0$ , then we can just estimate using observed variables without introducing bias.

To address omitted variable bias, Goldsmith-Pinkham & Imbens (2013) cluster the social network to estimate  $\hat{\xi}_i$ , and then using them to control for  $\xi$  when estimating the outcome variable, then use the estimated latent characteristics to predict the outcome variable.

$$\tilde{Y}_{it} = \beta_0'' + \beta_1'' Y_{i,t-1} + \beta_2'' \frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}} + \beta_3'' X_i + \beta_4'' \hat{\xi}_i \quad (2.5)$$

The new coefficients now depends on a new regression:

$$\xi_i = \tau_0' + \tau_1' Y_{i,t-1} + \tau_2' \frac{\sum_{j=1}^N A_{ij} Y_{j,t-1}}{\sum_{j=1}^N A_{ij}} + \tau_3' X_i + \tau_4' \hat{\xi}_i + \epsilon_i^2 \quad (2.6)$$

So  $\beta_2'' = \beta_2 + \tau_2' \beta_4$  and bias is reduced if  $|\tau_2'| < |\tau_2|$ , that is, if  $\xi$  and  $\hat{\xi}$  are positively correlated.

### 2.1.3 Continuous latent space

One of the problems with Goldsmith-Pinkham & Imbens (2013) is that their latent variable  $\xi_i$  is binary (Jackson 2013). A more realistic formulation is to let the latent characteristics take on continuous values. Many variables that determine network ties naturally lend themselves to continuous space, such as geographic location or age. A model with continuous variables will be more



empirically plausible as well as contain more information. Continuous variables can be reduced to binary variables, but not vice versa, and so using continuous latent space is more flexible.

#### 2.1.4 Estimation of latent space

There are several considerations for estimating a latent space model.

First, an important consideration is that  $\xi_i$  are not identifiable by rotation or reflection (called procrustation). This is because friendships are determined by distances in  $\xi_i$ . If everyone in the latent space is rotated around an axis or reflected across a plane, then their distances remain the same, and the likelihood of friendship formation remains the same despite the rotation and/or reflection. Also,  $\gamma_2$  and  $\xi_i$  are no longer separately identifiable. To fix this, I procruste estimated coordinates to some fixed set of points. The linear transformation does not affect the estimated social influence effect  $\beta_2''$  because the bias is determined by the correlation between the actual and estimated latent characteristics. Since correlations are invariant to linear transformations, then the actual fixed point I procruste to is irrelevant. This is the same reason regressing some  $Y$  on some  $X$  gives the same fit as regressing on linear transformations of  $X$ , as long as the span of  $X$  and its transform are the same.

Second, latent space models requires a connected graph, i.e., you can draw a path from an individual to any other individual. Suppose there are disconnected subgraphs in the social network. Latent space models fails because it cannot identify how far those two disconnected subgraphs should be from each other. In fact, it would try to place them as far apart as possible. With MLE, the distance between the two subgraphs would approach infinity; under Bayesian estimation, their distance would be highly dependent on the prior. Empirical applications of latent space constrain the data to come from a connected graph.

Third, it is computationally intensive and used mainly on small datasets. Most use cases examine social networks with up to a few hundred actors. Braun & Bonfrer (2011) use latent dirichlet

as a way to cluster nodes and estimate latent space on the clustered social network with about a hundred thousand actors. Many cases of social networks contain millions of actors. To be fair, modeling big data is a challenge to many other statistical approaches as well and I leave extending latent space models to large datasets for future research. In this paper, I take standard approaches. First, I use snowball samples, using actors at the edge of the snowball sample to estimate latent space but leaving them out in the regression estimation since I have incomplete network information on them. I also use multiple samples within the social network. This allows us to quantify when controlling for homophily has the most impact on the estimated social influence effect. I combine estimates across samples using Bayesian regression.

Finally, if one had panel data, fixed effects can be used to absorb individual latent traits that induce bias. This approach can be an easier and relatively effective way to control for homophily. It is also relatively easy computationally. However, fixed effects requires its own set of assumptions such as the exclusion of lagged variables in the data generating process.

## 2.2 Data

I use data from mixi, a leading social network in Japan. The data was shared by the company using a password protected medium. User IDs were scrambled to prevent identification of users on the platform. Other personally identifiable information were also scrambled. The data consists of the complete social graph of all its members, a total of 600 million connections among its 22 million users in October 2010. The social network provides a platform for apps, mostly games, like a mobile analog of Zynga for Facebook.

The dataset records the time when an individual uses and installs mobile games over seven days. Although this is a short timeframe, if we are able to detect influence within this short time period, then longer panel data which companies have access to would provide even more predic-

tive power. I use game usage and installation information from the first six days to predict mobile game installation in the seventh day, avoiding simultaneity issues. Because adoption for individual games is low in the observation window of the data, I could limit the analysis only to games with a large number of installations. However, this would ignore a majority of games. In order to include information of all games, especially those in the long tail, and to focus on the installation of games on a social network not on individual apps, I aggregate usage and installation information across all games.\* This is plausible because watching a friend play a mobile game may drive consumers to the app market and encourage game adoption in general.

I use snowball samples from the network because current latent space approaches have sample size limitations and I am not able to use the entire network (22 million users). Because mobile app adoption is sparse on a single day, I sample from a seed who adopts a mobile app in day seven to ensure that at least one actor in the network adopts, then include all actors two degrees from the seed in the sample.† I keep only samples that contain 300 and 600 actors to simplify estimation of latent space, for a total of 28,007 actors across 62 snowball samples. Data is summarized in Table 2.1 below.

### 2.3 Model and estimation

I estimate latent space coordinates from the social network, then use them as controls in a model of social games adoption. I choose a two-step estimator over joint estimation to sidestep computational complexity. Two-step estimators are known to have standard error and coverage concerns; I leave this for future research as our focus is on using the approach to reduce bias.

---

\*If I had app meta data, I would group by genre. However, apps have been anonymized.

†This resembles case-control study design where sampling is based on outcome variable(Prentice & Pyke 1979).

**Table 2.1:** Data summary statistics (62 network samples consisting of 28,007 people)

|                                                         | mean  | SD    | min   | max   |
|---------------------------------------------------------|-------|-------|-------|-------|
| % install app in day 7                                  | 0.068 | 0.04  | 0.028 | 0.033 |
| Peer usage (% friends who used mobile apps in days 1-6) | 0.141 | 0.043 | 0.084 | 0.094 |
| Own usage (% who used mobile apps in days 1-6)          | 0.148 | 0.613 | 0.068 | 0.528 |
| Degrees (number of friends), log                        | 4.025 | 0.313 | 3.046 | 4.819 |
| Female                                                  | 0.541 | 0.119 | 0.257 | 0.82  |
| Age                                                     | 27.98 | 4.187 | 19.9  | 37.45 |
| Number of photos, log                                   | 2.638 | 0.449 | 1.463 | 3.759 |
| Number of comments, log                                 | 3.012 | 0.442 | 0.934 | 3.648 |

### 2.3.1 Model of latent space

I assume conditional independence of dyads so that the probability of forming a link between two actors in the network depend only on their observed and latent traits:

$$P(A|X, \xi, \theta) = \prod_{i \neq j} P(A_{ij} = 1 | X_i, X_j, \xi_i, \xi_j, \theta) \quad (2.7)$$

where  $X = \{X_1, \dots, X_N\}$  are observed covariates and  $\xi = \{\xi_1, \dots, \xi_N\}$  are latent traits for each individual  $i = 1, \dots, N$ ,  $\theta = \{\gamma_0, \gamma_1\}$  is the set of parameters linking traits with friendship status.

I assume presence of dyadic ties as a function of distance in this latent space:

$$\text{logit} [P(A_{ij} = 1 | X_i, X_j, \xi_i, \xi_j, \theta)] = \gamma_0 + \gamma_1 \|X_i - X_j\| + \|\xi_i - \xi_j\| \quad (2.8)$$

where  $\gamma_1 > 0$  so that actors who are far apart have smaller chance of being friends. The model can accommodate multiple covariates: each  $X_i, \xi_i$  can be vectors capturing  $k$  and  $l$  traits of an individual, which means that  $\gamma_1$  is a vector of length  $k$ . There is no coefficient on distances of  $\xi$  because I cannot separately identify the coefficient from the scale of  $\xi$ .

I use Euclidean distance although any other distance metric satisfying the triangle inequality can also be used (e.g, cosine distance). The number of dimensions for the latent space can be determined by BIC or other likelihood-based criteria (Krivitsky et al. 2009).

An appealing feature of latent space models is its ability to capture more complex network relationships, for example, relationships that are reciprocal and transitive. Even though the likelihood depends on dyadic ties, the location of an individual depends not only on its direct ties but also on how everyone else in the network is located.

I use diffuse prior distribution  $\theta \sim MVN(\eta, \Psi)$ .  $\pi(\xi_i)$  is  $MVN(0, I_k)$  where  $\eta, \Psi$  are hyperparameters to be specified by the analyst. I set  $\eta = 0, \Psi = 9 \cdot I$  to allow for a wide range of parameter values. The R package *latentnet* estimates latent space models (Krivitsky & Handcock 2008, Krivitsky et al. 2009).

### 2.3.2 Model of social games adoption

The dependent variable of interest in this paper is whether an individual installs a mobile game or not. This latent utility of installing a game is allowed to vary by peer usage, own usage, and personal traits, with heterogeneous coefficients across samples:

$$u_{sit} = \beta_{s0} + \beta_{s1}Peer_{i,t-1} + \beta_{s2}M_{i,t-1} + \beta_{s3}Deg_i + \beta_{s4}X_i + \beta_{s5}\hat{\xi}_{si} + \epsilon_{sit} \quad (2.9)$$

where  $s$  denotes the snowball sample,  $i$  denotes the individual, and  $t$  denotes the time period.  $Peer_{i,t-1} = \frac{\sum_j A_{ij}M_{j,t-1}}{\sum_j A_{ij}}$  is the proportion of friends who use mobile games in days 1-6, and

$M_{j,t-1} = 1$  if person  $j$  used mobile games in days 1-6, and 0 otherwise,  $Deg_i = \log \sum_j A_{ij}$  is log number of degrees, or the number of friends person  $i$  has,  $X_i$  are time-invariant covariates for person  $i$ , namely, gender, age, log number of photos and log number of comments. Photos and comments control for how active the person is in the social network. Although these two variables may vary over time, this is unlikely to change significantly over the observation period of a week.  $\hat{\xi}_{si}$  are  $k$ -dimensional latent space coordinates for individual  $i$  in sample  $s$  estimated from the latent space model. Latent space coordinates are aimed at capturing unobserved covariates which influence friendship formation (via homophily) and are associated with game downloads. Examples of unobserved covariates might include being an early technology adopter or being interested in gaming.  $\epsilon_{sit}$  are independent type I extreme value random variables. Parameters  $\beta_s$  are the coefficients to be estimated in the model.

I formulate this as a logistic regression problem where individual  $i$  from sample  $s$  installs ( $Y_{sit} = 1$ ) if the latent utility of installing is greater than 0 at time  $t$ :

$$Y_{sit} = 1 \text{ if } u_{sit} > 0 \tag{2.10}$$

$$= 0 \text{ otherwise} \tag{2.11}$$

I want to estimate the model across all snowball samples to increase power and to make population as well as sample specific statements. There are two main issues. First, I expect heterogeneous treatment effects across samples. Peer effects may be large for some people and small for others. Second, estimated latent space coordinates  $\hat{\xi}$  differ across snowball samples. For example, the first dimension of latent space in snowball sample 1 might correspond to education level, while the first dimension of latent space in snowball sample 2 might correspond to tech savviness. Thus, I cannot simply include the latent space coordinates into a regression with the expectation that the latent space dimensions align across samples.

A parsimonious way to address these two issues is to impose a hierarchy when estimating the coefficients in the model:

$$\beta_s = \delta + \omega_s \quad (2.12)$$

$$\omega_s \sim N(0, V_\beta) \quad (2.13)$$

The hierarchy allows us to make population level statements about social influence while allowing sample-specific heterogeneous treatment effects, shrunk towards a set of population hyperparameters, as well as allowing for sample-specific latent space coordinate effects on probability of adoption. This is important because while I might expect that homophily plays a big role in mobile app adoption and therefore generate a large bias in social influence if ignored, the phenomenon may vary by consumers in the population, so the bias may be large or small. The hierarchy will help uncover which groups of consumers are most affected by homophily bias and which are not.

I place prior distributions:

$$\delta \sim N(\bar{\delta}, A_\delta^{-1}) \quad (2.14)$$

$$(V_\beta)^{-1} \sim \mathcal{W}(\nu, V) \quad (2.15)$$

where  $\mathcal{W}(\nu, V)$  is the Wishart distribution with  $\nu$  degrees of freedom and  $V$  location parameter.

I estimate the model with MCMC using the Metropolis algorithm. At iteration  $t$ , I accept a

candidate draw for each sample with probability proportional to the ratio of posterior density:

$$P(\text{accept}\beta_s^{new}) = \min \left[ 1, \frac{\pi(\beta_s^{new}|\delta^{(t-1)}, V_\beta^{(t-1)}) p(Y|\beta_s^{new}, X, \hat{\xi})}{\pi(\beta_s^{(t-1)}|\delta^{(t-1)}, V_\beta^{(t-1)}) p(Y|\beta_s^{(t-1)}, X, \hat{\xi})} \right] \quad (2.16)$$

otherwise, keep  $\beta_s^{(t)} = \beta_s^{(t-1)}$

Then, given the draws for all samples  $\{\beta_s^{(t)}\}$ , I draw parameters from distributions centered at the weighted average of the priors and the mean and variance of these samples:

$$\delta^{(t)} \sim N[\bar{d}, V_{beta}^{(t-1)} \otimes (S + A_\delta)^{-1}] \quad (2.17)$$

$$(V_\beta^{(t)})^{-1} \sim W(\nu + S, V + O) \quad (2.18)$$

where  $(2.19)$

$$\bar{d} = (S + A_\delta)^{-1} (\sum_s \beta_s + A_\delta \bar{d}) \quad (2.20)$$

$$O = \sum_s (\beta_s - \delta^{(t)}) (\beta_s - \delta^{(t)})' \quad (2.21)$$

$$S = \text{Number of snowball samples} \quad (2.22)$$

$$(2.23)$$

### 2.3.3 Alternate models

I estimate four models for mobile app adoption. The first model is a naive model, including only peer usage, own usage, and degrees. The second model adds covariates (gender, age, photos, comments) but no latent space, similar to existing methods that control for observed homophily but not latent homophily. The third model adds latent space coordinates but no covariates, reflecting a scenario where latent space is estimated but no variables are available to control for observed homophily. The last model controls for both observed and latent homophily using covariates and latent space coordinates.



I standardize variables and estimate the model with MCMC. The MCMC chain is run for 400,000 iterations, keeping every 5 draws. Convergence is checked visually with plots of coefficients after a burn in of 200,000 draws. I check model fit and choose the number of latent space dimensions with log marginal likelihood (Chib 1995). I set  $\bar{\delta} = 0, \mathcal{A}_\delta = 0.1, \nu =$  number of X variables + 3,  $V = \nu I$ .

### 2.3.4 Model selection

I report the log marginal likelihood of different model specifications in a plot in Figure 2.1. The full model (Model 4) with both covariates and latent space of 8 dimensions (for each sample) fits the data best. It improves over either the covariates only model or the latent space only model (Models 2 and 3). This suggests latent space captures an aspect of the data that is separate and additional to observed homophily.

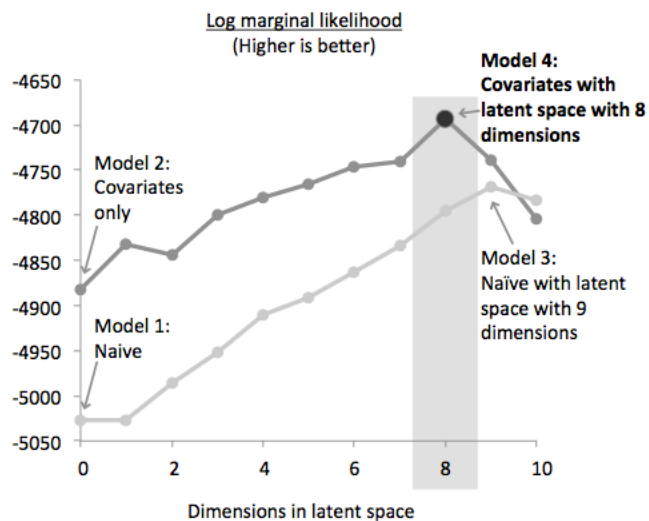


Figure 2.1: Log marginal likelihood (LML) by dimensions of latent space

## 2.4 Results

### 2.4.1 Peer influence estimates

Table 2.2 shows the estimated posterior mean and 95% posterior interval for population parameters across different specifications of the model. In all models, own mobile app activity from the past six days is positively correlated with new app installation. Female consumers are more likely to install apps than their male counterparts.

The peer activity coefficient, a measure of peer influence, is positive and significant. The parameter estimate decreases when I control for latent homophily with latent space (from Model 2 to Model 4).

As expected, the coefficients for latent space only make sense at the sample level and therefore are not significantly different from zero at the population level. This is because latent space captures different aspects of latent homophily across different samples from the population. To check the face validity of the latent space model in fitting the social network, I compare the estimated latent distances between friends and non-friends in the sample. On average, non-friends are twice the distance from each other as friends are in latent space. This supports the idea that there is significant latent homophily in the data.

To quantify the impact of social influence, I simulate two worlds: one with social influence and another without social influence. In the current dataset, individuals can be influenced by peers who use mobile apps. For the world without social influence, I predict the level of mobile app adoption when peer influence is removed.

When I remove peer influence, the average mobile app adoption likelihood falls from 6.80% to 4.98% (under model 4). The decline of 1.82% is significantly different from zero. From these numbers, I can attribute  $1.82\% / 6.80\% = 27\%$  of all mobile app installations to social influence (95% posterior interval: [25%, 28%]).

**Table 2.2:** Regression coefficients estimates (Logistic regression population parameters)

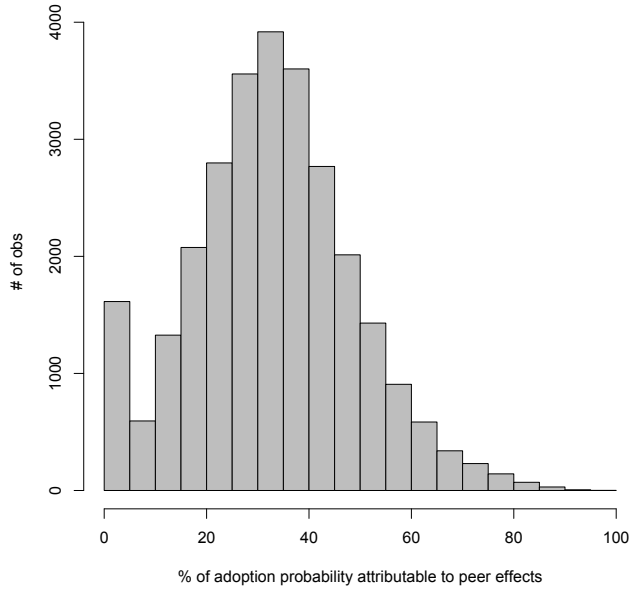
| Model         | <b>1: Naive</b>             | <b>2: Covariates</b>        | <b>3: Naive &amp; latent</b> | <b>4: Covariates &amp; latent</b> |
|---------------|-----------------------------|-----------------------------|------------------------------|-----------------------------------|
| Intercept     | <b>-4.48</b> (-4.82, -4.15) | <b>-4.59</b> (-5.05, -4.15) | <b>-4.77</b> (-5.07, -4.45)  | <b>-5.28</b> (-5.67, -4.85)       |
| Peer activity | <b>3.21</b> (2.66, 3.79)    | <b>3.48</b> (2.89, 4.15)    | <b>2.74</b> (2.15, 3.23)     | <b>2.53</b> (2.17, 2.88)          |
| Own activity  | <b>2.73</b> (2.53, 2.93)    | <b>2.76</b> (2.54, 2.98)    | <b>3.08</b> (2.83, 3.34)     | <b>3.13</b> (2.87, 3.40)          |
| Degree        | 0.07 (-0.04, 0.18)          | 0.01 (-0.13, 0.16)          | 0.01 (-0.14, 0.15)           | 0.04 (-0.13, 0.21)                |
| Female        | -                           | <b>0.21</b> (0.02, 0.41)    | -                            | <b>0.28</b> (0.05, 0.50)          |
| Age           | -                           | -0.01 (-0.011, 0.10)        | -                            | -0.01 (-0.14, 0.13)               |
| Photos        | -                           | 0.06 (-0.05, 0.18)          | -                            | 0.08 (-0.07, 0.23)                |
| Comments      | -                           | 0.02 (-0.11, 0.15)          | -                            | 0.03 (-0.14, 0.19)                |
| $\hat{\xi}_1$ | -                           | -                           | 0.01 (-0.12, 0.14)           | -0.01 (-0.15, 0.13)               |
| $\hat{\xi}_2$ | -                           | -                           | 0.01 (-0.14, 0.12)           | 0.01 (-0.13, 0.15)                |
| $\hat{\xi}_3$ | -                           | -                           | 0.00 (-0.12, 0.13)           | 0.00 (-0.14, 0.14)                |
| $\hat{\xi}_4$ | -                           | -                           | 0.02 (-0.11, 0.15)           | -0.01 (-0.15, 0.13)               |
| $\hat{\xi}_5$ | -                           | -                           | -0.01 (-0.13, 0.12)          | 0.00 (-0.14, 0.14)                |
| $\hat{\xi}_6$ | -                           | -                           | 0.00 (-0.13, 0.13)           | 0.00 (-0.13, 0.14)                |
| $\hat{\xi}_7$ | -                           | -                           | 0.01 (-0.12, 0.13)           | -0.01 (-0.14, 0.14)               |
| $\hat{\xi}_8$ | -                           | -                           | 0.00 (-0.13, 0.12)           | 0.00 (-0.14, 0.14)                |
| $\hat{\xi}_9$ | -                           | -                           | 0.00 (-0.13, 0.12)           | -                                 |

Mean of posterior distribution,  $\hat{\xi}$  contain 95% posterior credible interval; bold = 95%

posterior credible interval does not cover 0

Next, I plot the change in probability of adoption with and without social influence for individuals (Figure 2.2). For 5% of individuals, peers play no role at all because none of their peers

use mobile apps. For the middle half of individuals, peer influence accounts for 23% to 43% of the motivation for app adoptions.



**Figure 2.2:** Distribution of peer effects as a driver of mobile app adoption

#### 2.4.2 Magnitude of bias from homophily

To assess the importance of accounting for latent homophily, I repeat the simulation in the previous section but use model 2 which does not control for latent homophily. With this model, I find that social influence accounts for 38% of all installations (see figure 2.3). Thus, ignoring latent homophily results in a  $\frac{38\% - 27\%}{27\%} = 40\%$  inflation of social influence effects ( $p < 0.001$ , based on posterior distributions of difference in peer effects). The size of the bias due to latent homophily is in line with prior research which find that latent homophily can inflate social influence by 10-50% (Tucker 2008, Hartmann 2010).

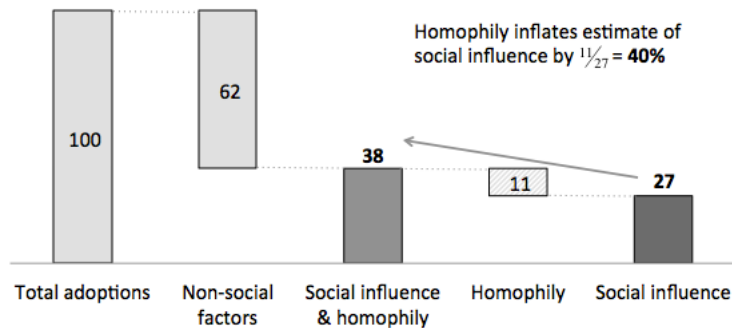


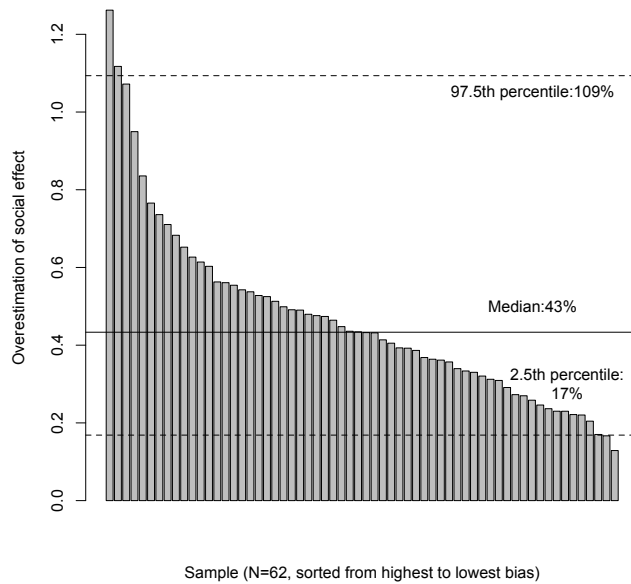
Figure 2.3: Drivers of mobile app adoption (% of total adoptions)

Our modeling approach allows us to estimate the bias distribution across different samples from the network. I plot the histogram of biases as a percent of effect size in Figure 2.4. This is how much social influence is overestimated if I do not account for latent homophily. I find positive bias in every sample. On the low end, ignoring latent homophily would bias social influence estimates by 15-20%. On the high end, the bias could inflate the effect by more than 100%. The variability across samples could be due to sampling error or because homophily and social influence have heterogeneous effects in the network. At this point, I am not able to tease these two effects apart.

## 2.5 Discussion

I study peer influence in social games. To reduce bias in our estimates, I propose using latent space to model latent homophily and using them as controls in regressions. I apply the method to investigate whether mobile game adoption is influenced by peers and find that peers drive more than a quarter of all mobile game adoptions, even after controlling for homophily.

Ignoring homophily could result in extremely biased estimates. In our setting, I find that latent homophily could inflate the proportion of adoption attributed to social influence by 40%. Alarm-



**Figure 2.4:** Bias to effect size ratio in each sample

ingly, there is large variation across samples; the bias in some samples was as large as the effect size itself. Based on our simulations and empirical findings, I conclude that it is necessary to control for latent homophily when estimating social effects. Latent space presents an appealing and intuitive way to do so.

There are some methodological gaps for further research. The approach I use involves two step estimation which may have implications for standard errors. I do not explore this issue because the primary focus of this paper is to address a more fundamental social science question: whether I can detect and reduce the bias induced by latent homophily on the coefficient of social influence using latent space.

Separating social influence from homophily continues to be a challenging area for research. Providing bounds for the effect of social influence is of great interest to the social influence research community but require additional work (Aral et al. 2009, Shalizi & Thomas 2011). If cer-

tain assumptions are satisfied, bounds for the regression coefficients in the presence of proxy variables can be stated (Klepper & Leamer 1984, Bollinger 2003, Bollinger & Minier 2015). Since latent space coordinates act as proxy variables for latent traits that drive homophily, there may be potential to provide bounds for social influence effects. I leave this for future research.

*In locations like Sub-Saharan Africa where 69% of people live on less than \$2 per day, only 53% of the population can afford the internet with a cap of 20MB, an amount that provides just 1-2 hours of web browsing a month.*

State of connectivity: 2014, Internet.org

# 3

## The Effect of Zero Rating on Social Network Activities

Digital media and social networks are increasingly significant parts of consumers' lives. In a single day at the end of 2014, 2 billion photos are shared over Facebook and Instagram, over 7 billion likes are made on Facebook, and over 30 billion messages are sent via Whatsapp (Zuckerberg 2015). The top five largest online social networks collectively boast over 3 billion monthly active



users (Ballve 2013). Among iOS and Android users, access of digital content accounts for 50-60% of all time spent on a smartphone, with social networks accounting for more than a quarter of all time spent on mobile phones (Khalaf 2014).

Yet, not everyone is online. The percentage of people who access the internet at least once a year is around 30% in developing countries, compared to 80% in developed countries (Internet.org 2015). This disparity has led to organizations such as internet.org to work on connectivity around the world, because connecting people to the internet has significant impacts on national economic outcomes and consumer livelihood (Qiang et al. 2009, Katz 2012, Nottebohm et al. 2012, Chhachhar et al. 2014, Meltzer 2013, Manyika et al. 2013, Trucano et al. 2012).

One of the barriers to connectivity is affordability. Only 46% of those in developing countries can afford the internet (as measured by the cost of 250MB per month), compared to 99% in developed countries (Internet.org 2015). In response, many mobile operators, especially in developing countries, have partnered with digital media companies, including Facebook, Google, and Wikipedia, to “zero rate” mobile data access of digital content (Bergen 2014). Under zero rating, consumers on partner mobile networks can access zero rated digital content for free. These efforts have resulted in massive increases in mobile data adoption. For example, Globe, a large mobile carrier in the Philippines, doubled the number of mobile data users since the introduction of zero rated Facebook (Globe 2014). Indeed, partnering with digital media companies is a standard practice with over 45% of mobile operators engaging in zero rating globally (Morris 2014).

Another barrier to connectivity is relevance (Internet.org 2015). Globe in the Philippines identifies that “customer awareness of the value of the internet” as a barrier to adoption (Globe 2015). On communication platforms such as Facebook, there needs to be a critical mass before the platform becomes relevant for consumers. By making connectivity affordable for masses of people, consumers may find being online more relevant since they can communicate with more peers.

Existing studies of zero rating focus on connecting unconnected consumers. A case study

of zero rating by Telenor in Pakistan showed that not only did more people come online during the zero rating campaign, but the campaign accelerated connectivity in the country as more and more people came online (MEF 2015). A study by Globe in the Philippines found that zero rating fueled growth in new subscriber acquisitions (Globe 2015). This chapter extends existing knowledge about zero rating beyond unconnected users, for whom benefits to zero rating are obvious and fairly well documented, by studying how zero rating affects *connected* users – those users who are already online.

Zero rating should have multiple effects by addressing affordability and relevance of connectivity for consumers who are already connected. At the basic level, there is a direct impact of a price change. By lowering the price of connectivity to zero, zero rating directly benefits consumers economically and may influence their social network activities. Since peers also benefit from the price change, then having more peers on zero rating should also influence activity levels. Finally, since connectivity is an experience good, we expect a long term effect of zero rating, even after the pricing regime reverts back to pre-campaign levels.

In this study, I model consumer behavior during a Facebook zero rating program to answer three questions around zero rating partnerships. First, what is the direct effect of zero rating on customer activities in the social network? Second, do zero rated users affect activity levels of their peers? Finally, what are the persistent effects of zero rating after the pricing change has ended?

I contribute to the literature by quantifying the direct and spillover effects of marketing actions on social network activities across people and time. I find that zero rating does not have the same effect on different activities. While the direct impact of zero rating is positive on all activities, users with more friends on zero rating create less, consume more, and give more feedback on content. In addition, zero rating does not have a uniform effect across consumers. Some consumers benefit more from zero rating than others, and I show that network characteristics can help identify those consumers whose network benefits the most from zero rating.

The paper is structured as follows: I describe the study context and relate it to extant literature in section 3.1. I build the model in section 3.2. Results and managerial implications are in section 3.3. I conclude in section 3.4, and end with a few extensions to the current study in 3.5.

### 3.1 Pricing and social network activities

There is an active research agenda on how consumers influence each other. Many studies find that peers who adopt new technologies increase the probability that their connected ties will adopt as well (Tucker 2008, Aral et al. 2009, Aral & Walker 2011a, 2012). Studies on activities in social network find similar patterns: higher activity individuals influence their friends to engage in more activities as well (Trusov et al. 2010, Aral & Walker 2011b, Ghose & Han 2011, Katona et al. 2011).

To make these research findings actionable, attention is given to how social influence interacts with firm actions. Some studies examine peer influence changes the role of communication channels such as detailing (Manchanda et al. 2008, Nair et al. 2010), the value of product and platform design to influence consumer chatter (Godes & Mayzlin 2009, Aral & Walker 2011a, Manchanda et al. 2014), social advertising (Bakshy et al. 2012), or pricing in group settings (Hartmann 2010). Understanding how managers can influence social network engagement is important because social networking is an increasingly important aspect of how consumers spend their time. There is also evidence that increased engagement is associated with higher consumer well-being (Burke et al. 2010). Understanding activity levels can also shed light on the evolution of social capital within a network (Shriver et al. 2013). Finally, different social network activities vary in their impact on business outcomes like adoption and product virality (Aral & Walker 2011a). Therefore, understanding how to influence one activity over another is critical to marketers in general.

Marketers are familiar with the relationship between prices changes and sales in traditional

settings (e.g., Gupta 1988, Bell et al. 1999, Jedidi et al. 1999). In the past two decades, extensive research has documented not just how price changes affect the product in the offer but spillover to related products (Van Heerde et al. 2003, Leeflang & Parreño-Selva 2012) as well as long run effects (Mela et al. 1997, Nijs et al. 2001, Pauwels et al. 2002). Although higher peer activities increase activity level (Trusov et al. 2010), certain activities may benefit more from a price change than others.

To my knowledge, there is the first study on how pricing and peer effects affect social network activities, and on long-term peer effects.

### 3.1.1 Study context

Internet.org, a global partnership with the goal of making internet access available to the next 5 billion people, introduced a partnership between Facebook, the largest social network in the world, and a mobile operator, second largest operator in a duopoly in a large country, to zero rate access to Facebook<sup>\*</sup>. A nationwide marketing campaign ensured high awareness of zero rated pricing; post-campaign offline surveys confirmed this finding. When mobile subscribers of this partner opted into zero rating, all data access charges incurred by using Facebook are waived. Everyone is eligible to opt into zero rating by logging in to Facebook on the partner mobile network and clicking on a box agreeing to terms of service. As with any marketing campaign where some consumers find the offer attractive or not, some users did not opt into zero rating. Most consumers are on pre-paid mobile plans and few use mobile data during the time of the campaign.

Our dataset begins several months before the campaign launch date. Before the campaign launched, users pay for access to Facebook over the mobile carrier's network. The campaign lasted for several months, during which time users who opt-in to the zero rating campaign receive free unlimited mobile bandwidth when accessing Facebook. For these users, mobile data

---

<sup>\*</sup>Some details of the campaign have been redacted for confidentiality

to use Facebook, such as creating content such as posting status updates and uploading photos, consuming content by reading stories, and giving feedback via comments or likes, are free on the partner mobile carrier network. After the campaign, data charges for accessing Facebook revert to the pre-campaign price (see Figure 3.1). Users who do not opt in pay for data associated with all their activities throughout the entire observed time period.

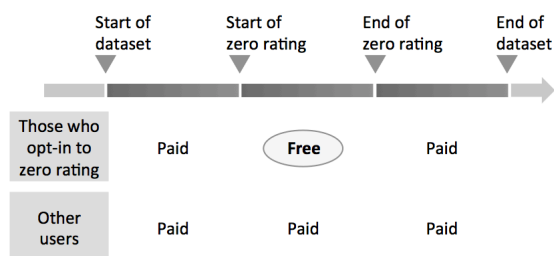


Figure 3.1: Data timeline of zero rating campaign and pricing

The analysis includes only existing Facebook users because although zero rating resulted in significant new user growth, new users are an obvious win for all parties involved and is fairly straightforward to count. Complexity arises from those who were already Facebook users and may change their social networking activities in response to price changes, or their friends' access to free data. I keep users who had at least one activity in the first month of the data period before zero rating, and use the remainder of the data before zero rating to establish a baseline for what would have happened if there were no campaign. Therefore, this analysis is conservative in estimating the benefits to consumers because new to Facebook consumers are excluded from analysis.

### 3.1.2 Social network activities

Not all social networking activities are created equal; they play different roles and have different impact on consumer and company well-being. Since they play different roles in relationships in a social network, I expect zero rating to have different effects on each activity type.

On social networks, people can create, consume, or give feedback on content. *Lurkers*, users who consume but do not create content, have been a major concern for user-generated content platforms (Hughes et al. 2005). Content consumption without creation induces three issues. First, if not much content is being created, then the platform becomes less compelling for other users. Content engages other consumers to be active on the social network platform (Zhang & Sarvary 2011). To combat this freeriding issue, Burke et al. (2009) find that feedback encourages users to create more content. Second, research into user well-being finds that content consumption is correlated with decreased social capital and increased loneliness (Burke et al. 2010). Finally, although one might assume that online activities are positively correlated, these social networking activities actually compete for the user's time. After controlling for individual characteristics, Ghose & Han (2011) find that content creation is negatively correlated with consumption.

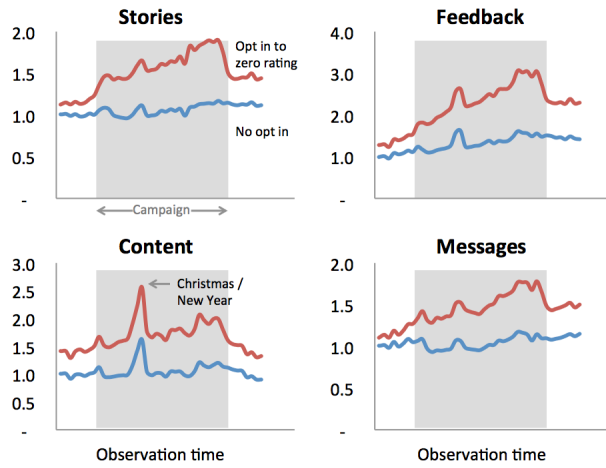
Zero rating may have a direct effect of increasing content creation, consumption, and feedback, since they are all free with the mobile operator. Having peers with zero rating could impact these activities differently. Since peers create more content under zero rating, then an individual's newsfeed potentially has more interesting content, so peer zero rating should increase consumption and feedback. Peer zero rating could increase or decrease individual content creation. Although increased peer feedback and peer consumption of content might also motivate higher content creation, consumers may also decrease their activity if they spend more time consuming content or if they feel that their peers have generated sufficiently interesting content (and therefore they do not have to bear the burden of creating content themselves). On the other hand, seeing their peers create content also set norms of content creation and spur higher activity levels as well. In essence, individuals may view more content but it is unclear whether content creation will increase in the network or not.

After zero rating, those who opt in start paying for mobile data if they continue using it. Habituation might suggest that individuals who had higher levels of activity might continue having using Facebook more. Zero rating may also improve brand positioning in consumers' mindset. However, if they also adjust their reference point during the campaign, then suffering the loss of zero rating, or the loss of peer zero rating, could result in lower activity levels than before the campaign period (Kahneman & Tversky 1979). Therefore, we might see either higher or lower levels after the campaign.

### 3.2 A model of social network activities

I model each measure of activity level (stories, feedback, and content) as a function of whether the individual has or had zero rating and whether their peers have or had zero rating.

I start with an ego-centric network where  $N(i)$  is individual  $i$ 's neighborhood, the set of users



**Figure 3.2:** Average daily activity per person: Opt-in curve includes anyone who opted into zero rating at any point during the campaign; normalized so non-opt in is 1.0 at the beginning of the data period

who are directly connected to individual  $i$ , i.e., her friends. Although a consumer's Facebook network is dynamic, I take a static view because tie-formation is not the focus of the study. Let zero rating start at time  $T_0$  and end at time  $T_1$ . Before the campaign ( $t = 1, \dots, T_0 - 1$ ), I model digital media activity  $Y$  by individual  $i$  at time  $t$  as a function of individual-level ( $\mu_i$ ) and time ( $\gamma_t$ ) fixed



effects:

$$Y_{it} = \mu_i + \gamma_t + \epsilon_{it}, t < T_0 \quad (3.1)$$

with idiosyncratic error terms  $\epsilon_{it}$  (Angrist & Pischke 2008). To control for individual differences across consumers, I take the average activity level pre-campaign for each individual:

$$\bar{Y}_i^{pre} = \frac{1}{T_0-1} \sum_{t=1}^{T_0-1} Y_{it}, t < T_0 \quad (3.2)$$

$$= \mu_i + \frac{1}{T_0-1} \sum_{t=1}^{T_0-1} (\gamma_t + \epsilon_{it}) \quad (3.3)$$

$$= \mu_i + \bar{\gamma}^{pre} + \bar{\epsilon}_i^{pre} \quad (3.4)$$

During the campaign (from  $T_0$  to  $T_1$ ), individuals can opt into zero rating. I denote the time that an individual  $i$  opts in by  $\tau_i$  and end of zero rating for that individual as  $\eta_i$ . If the individual does not opt into zero rating, then  $\tau_i = \eta_i = \infty$ . Define  $Z_{it} \equiv I(t \geq \tau_i, t \leq \eta_i)$ , which equals 1 if individual  $i$  has zero rating at time  $t$ . I address endogeneity and propose fixed effects regression with matching to combat self-selection issues in the next section.

During the campaign, i.e.,  $T_0 \leq t \leq T_1$ , I model activities as a function of fixed effects as before and add individual and peer zero rating status:

$$Y_{it} = \mu_i + \gamma_t + \alpha_1 Z_{it} + \beta_1 \sum_{j \in N(i)} Z_{jt} + \delta_1 Z_{it} \cdot \sum_{j \in N(i)} Z_{jt} + \epsilon_{it} \quad (3.5)$$

where  $\alpha_1$  is the *direct* average treatment effect of an individual having zero rating,  $\beta_1$  is the incremental indirect or *peer* effect of each peer that opts into zero rating, and  $\delta_1$  is the effect of interaction between peer and individual zero rating statuses. Even though an individual's activity level could be affected by peers' zero rating status via peers' activity level, adding peer activity levels would introduce issues of simultaneity. I abstract away from this issue by letting the treatment be

individual and peer zero rating status, thereby getting an intention-to-treat (ITT) estimate of zero rating.

After the campaign ends, define  $Z_{it}^{had} \equiv I(\tau_i \leq t, \eta_i < t) = I(\eta_i < t)$  which equals 1 for those who had but lost zero rating. For  $t > T_1$ , I get:

$$Y_{it} = \mu_i + \gamma_t + \alpha_2 Z_{it}^{had} + \beta_2 \sum_{j \in N(i)} Z_{jt}^{had} + \epsilon_{it} \quad (3.6)$$

$\alpha_2, \beta_2$  are the long run effects of having had zero rating or having peers who had zero rating.

I account for individual fixed effects by subtracting the pre-campaign average activity level from observed activity levels during and after the campaign. During the campaign, I get:

$$\Delta Y_{it} = \eta_t + \alpha_1 Z_{it} + \beta_1 \sum_{j \in N(i)} Z_{jt} + \delta_1 Z_{it} \cdot \sum_{j \in N(i)} Z_{jt} + \epsilon'_{it} \quad (3.7)$$

where  $\Delta Y_{it} = Y_{it} - \bar{Y}_i^{pre}$ ,  $\eta_t = \gamma_t - \bar{\gamma}^{pre}$ , and  $\epsilon'_{it} = \epsilon_{it} - \bar{\epsilon}_i^{pre}$ . Post-campaign, I get:

$$\Delta Y_{it} = \eta_t + \alpha_2 Z_{it}^{had} + \beta_2 \sum_{j \in N(i)} Z_{jt}^{had} + \epsilon'_{it} \quad (3.8)$$

We estimate three models in this paper. Equations 3.7 and 3.8 form our basic fixed effects regression model (Model 1). I take daily average activity for each week. Since opt-in need not occur at the beginning of the week, to be conservative I consider that an individual had zero rating that week if they had at least one day of zero rating. I regress average daily activity each week against dummy week variables and covariates using linear regression. Because I observe many weeks per person, clustered standard errors for necessary for inference. However, analytic calculation is prohibitive due to data size (in the hundreds of millions person-week observations). Instead, I use bootstrap to calculate robust standard errors for clustered data (Efron 1981, Efron & Tibshirani

1986). I winsorize the data at the 95th percentile to avoid outliers.

For model 2, I control for the number of friends (degree) and pre-campaign activity level as covariates in addition to using pre-campaign activity in the fixed effects regression (Equations 3.9 and 3.10). Model 2 during the campaign:

$$\Delta Y_{it} = \eta_t + \alpha_1 Z_{it} + \beta_1 \sum_{j \in N(i)} Z_{jt} + \delta_1 Z_{it} \cdot \sum_{j \in N(i)} Z_{jt} + \nu_1 X_i + \epsilon'_{it} \quad (3.9)$$

where  $X_i$  are degree and pre-campaign activity level.

Model 2 after the campaign:

$$\Delta Y_{it} = \eta_t + \alpha_2 Z_{it}^{had} + \beta_2 \sum_{j \in N(i)} Z_{jt}^{had} + \nu_2 X_i + \epsilon'_{it} \quad (3.10)$$

The inclusion of these covariates allows for a baseline change in difference in  $Y$  to interact with covariate terms (and thus it does not get absorbed by the individual fixed effect).

For model 3, I allow for heterogeneity in the dataset by dividing the data into nine subclasses: three subclasses by degree crossed with three subclasses by pre-campaign activity level, at the 33rd and 67th percentile. I then estimate the fixed effects regression with degree and pre-campaign activity as covariates for each subclass. This allows for all coefficients, including fixed effects, to vary by subclass.

### 3.2.1 Endogeneity

One issue is that opt-in is endogenously chosen by the consumer. I expect that opting into zero rating is mainly driven with two factors. First, those who use Facebook more benefit more from the reduced price, and therefore are more likely to sign up for zero rating. Second, those with more friends also benefit more from Facebook, so are more likely to sign up for zero rating.

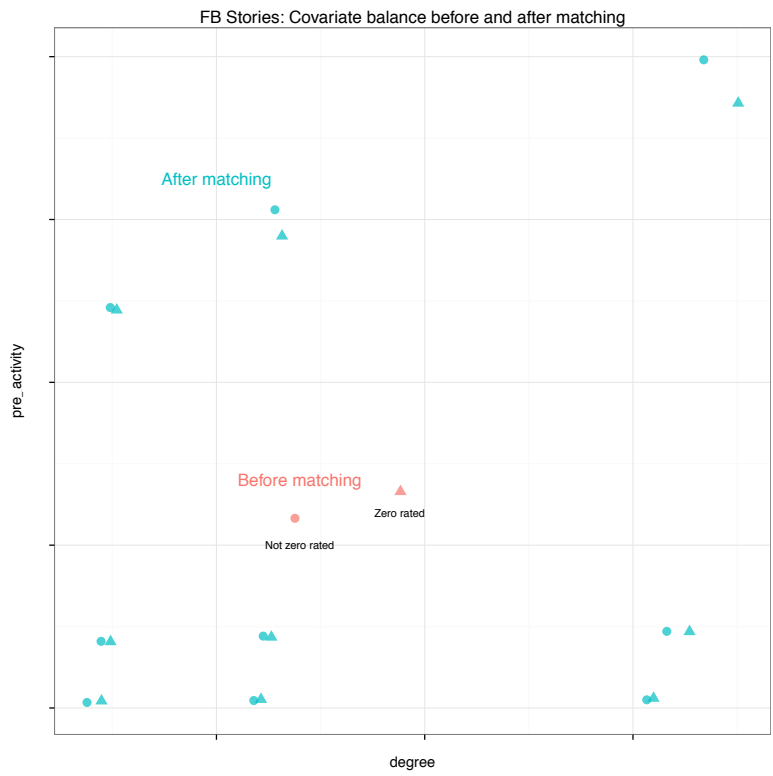
Another issue is that omitted variables can induce spurious correlations through homophily. For example, young people may be more likely to be connected to each other, more likely to opt into zero rating, and also more likely to increase activity over time. Shalizi & Thomas (2011) demonstrate that even simple models of homophily, where ties are more likely between similar individuals, can bias social influence estimates when tie-forming traits are unobserved and uncontrolled for in the model. Researchers have several strategies to control for homophily bias in observational studies (for a review, see Angrist 2014), including post-stratification and matching (Aral et al. 2009, Eckles & Bakshy 2015) and fixed effects regression (Bramoullé et al. 2009, Nair et al. 2010).

I employ both matching and fixed effects regression to address these issues. Individual and weekly fixed effects address unobserved individual traits and time trends, but there may be other unobserved variables which vary by both consumer and time, for example, local marketing activities. Post-stratification based on degree and pre-campaign activity level takes care of unobserved variables that vary by subclass, for example marketing activities that could affect active users differently than less active ones. The balance improves significantly after matching (Figure 3.3).

Figure 3.4 shows the average daily activity for stories viewed per person, for those who opt in and those who did not. By matching on pre-campaign activity, using those who did not opt in as a control group seems more plausible. We also see distinctly different treatment effects by subclasses.

### 3.3 Results

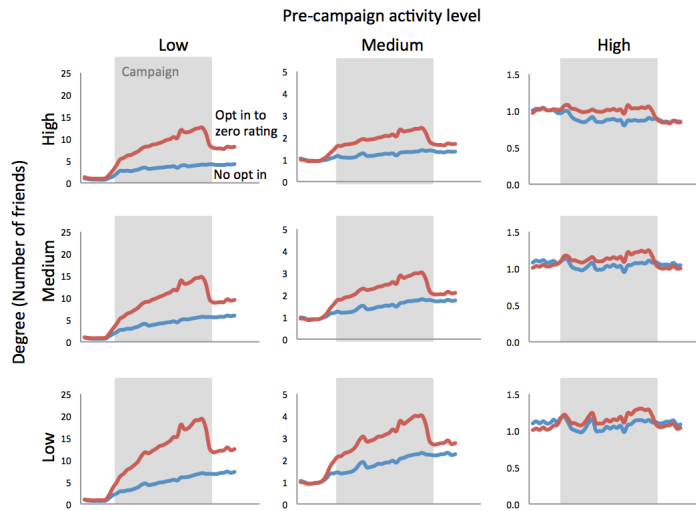
Based on estimated model coefficients of model 3 (full coefficients are redacted for confidentiality), I estimate the average treatment effect on the treated (ATT) by taking the difference in  $\hat{Y}$  when zero rated individuals get zero rating and when zero rated individuals do not get zero rating.



**Figure 3.3:** Covariate balance (average degree and pre-campaign activity level) before and after matching (FB stories example). Before matching, the circle and triangle represents average degree and pre-campaign activity for zero rated and non-zero rated individuals respectively. After matching, there are nine subclasses, so there are nine pairs of statistics. Scale redacted for confidentiality.

Estimated percentage changes averaged across subclasses are found in Table 3.1.

I find that zero rating increases content consumption. The direct impact of zero rating is a 14% increase in stories viewed per day on average. In addition to the direct impact of zero rating, zero rated peers induce an increase in content consumption, resulting in a total lift of 37% in



**Figure 3.4:** Average daily stories viewed per person by subclass: Opt-in curve includes anyone who opted into zero rating at any point during the campaign; normalized so non-opt in is 1.0 at the beginning of the data period

content consumption from pre-campaign levels (even after controlling for time effects).

After the campaign, having had zero rating still results in higher content consumption, but peers losing zero rating reduces the number of stories viewed. As a consequence, total lift after the campaign ends is lower, we find significant positive lift of 5.6% above pre-campaign levels.

Zero rating also increased feedback but peer effects are even more pronounced, and remain

**Table 3.1:** Effect of zero rating on zero rated and non-zero rated users, as percent of pre-campaign activity level (Bold: 95% confidence interval in parentheses does not cover 0)

| <b>Per zero rated user</b>     | <b>Stories</b>           | <b>Feedback</b>          | <b>Content</b>              |
|--------------------------------|--------------------------|--------------------------|-----------------------------|
| <i>During campaign</i>         |                          |                          |                             |
| Zero rating                    | <b>14.0</b> (13.7, 14.2) | <b>1.0</b> (0.7, 1.3)    | <b>7.9</b> (7.6, 8.2)       |
| Peer zero rating               | <b>22.5</b> (21.7, 23.2) | <b>78.8</b> (78.1, 79.3) | <b>-7.0</b> (-7.5, -6.4)    |
| Total effect                   | <b>36.5</b> (35.4, 37.4) | <b>79.8</b> (78.8, 80.6) | <b>0.93</b> (0.13, 1.81)    |
| <i>After campaign</i>          |                          |                          |                             |
| Zero rating                    | <b>12.0</b> (11.8, 12.1) | <b>17.1</b> (16.6, 17.4) | <b>9.6</b> (9.2, 9.9)       |
| Peer zero rating               | <b>-6.4</b> (-7.3, -5.7) | <b>33.7</b> (32.6, 34.5) | <b>-27.3</b> (-27.9, -26.6) |
| Total effect                   | <b>5.6</b> (4.5, 6.4)    | <b>50.8</b> (49.2, 51.9) | <b>-17.6</b> (-18.6, -16.7) |
| <b>Per non-zero rated user</b> | <b>Stories</b>           | <b>Feedback</b>          | <b>Content</b>              |
| <i>During campaign</i>         |                          |                          |                             |
| Peer zero rating               | <b>4.1</b> (3.5, 4.6)    | <b>26.2</b> (25.8, 26.5) | <b>-12.5</b> (-12.9, -11.9) |
| <i>After campaign</i>          |                          |                          |                             |
| Peer zero rating               | <b>-5.6</b> (-6.3, -5.1) | <b>19.8</b> (19.1, 20.3) | <b>-23.2</b> (-23.6, -22.8) |

positive even after the end of the campaign. Feedback increases by 80% and 51% during and after the campaign respectively.

I find that zero rating does not increase content creation uniformly. While the direct impact of zero rating on content creation is positive, both during and after the campaign, peers having zero rating reduce the amount of content being created. This may be due to a number of possible reasons, such as satiation to increased content generated by peers. During the campaign, the net effect is still positive (0.9% increase in content on average), but after the campaign, when peers lose zero rating, the total change is negative (-18% decrease from pre-campaign levels).

Zero rating has direct impact on those who opt in, but also affect their friends due to peer spillover effects. Non-zero rated users increase their content consumption during the campaign, and decrease after their zero rated peers lose zero rating. Feedback increases dramatically both during and after the campaign (26.2% and 19.8% respectively). Content creation goes down

among those who do not have zero rating.

The analysis suggests a few takeaways for managers. First, to summarize, zero rating does not have a uniform impact on consumers. Clearly, consumers benefit economically. They also benefit socially with increased interactions with content shared by friends. However, this does not mean they increase all activities on social network. While content consumption and feedback go up for all consumers on average, content creation goes up for zero rated individuals, but not for non-zero rated individuals.

Second, a significant proportion of the effect of zero rating is manifest through peer interactions. To conduct an analysis only on those who received zero rating while keeping those who did not receive zero rating separate would be erroneous, because the latter group is indirectly affected by the campaign. Further, to consider benefits to consumers who opt into zero rating only is to paint a partial picture of the benefits of zero rating to existing customers. Those who do not opt into zero rating also benefit, based on their peers increased activities online.

Third, ending a zero rating campaign should be done with care, since activities decline after the end of a campaign. Thus, this suggests that zero rating, once started, should be continued to avoid activity decline.

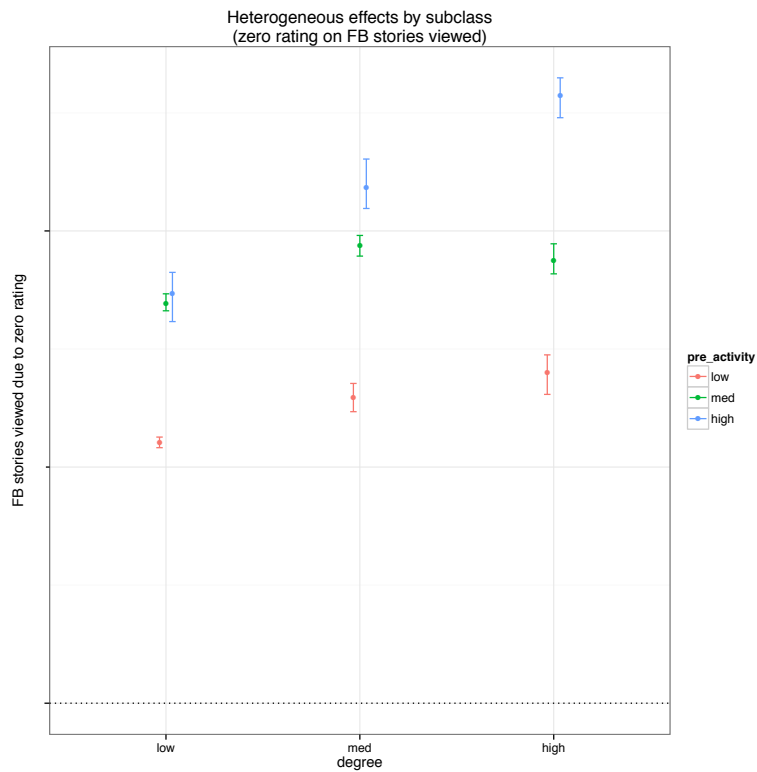
Fourth, zero rating has different value to different consumers. Consumers have heterogeneous response to zero rating, and are affected by their friends in different ways. The subclass-level coefficients, coupled with network structure and the characteristics of peers, allows deeper understanding of how the value of zero rating varies across the population.

### 3.3.1 Heterogeneous consumer response to zero rating

In our model, I estimate separate regressions and get different sets of coefficients for each subclass. Comparing the effects of zero rating and peer zero rating across subclasses, I find significant heterogeneity among the user base (Figure 3.5). For example, the effect of zero rating on the



daily stories appear to be higher for those with higher degree and pre-campaign activity level.



**Figure 3.5:** Significant heterogeneity in post-stratified estimates by subclass (FB stories coefficients example); Scale redacted for confidentiality.

This supports our modeling choice to post-stratify by degree and pre-campaign activity. We dig into this further by comparing the value to consumers for different groups in the population. An individual might generate more spillover if they have more friends and if their friends are more responsive to peer effects. I simulate the impact of giving zero rating to each individual, taking

into account their response to zero rating, number of friends, and their friends’ response to peer zero rating. The simulation uses individual-level data and subclass-level coefficients to derive an estimate of the causal change in activity level by zero rating each individual.

I compare the population average to top decile based on degree and based on network characteristics (degree and what bucket of consumers the friends fall into) (Table 3.2). The network of consumers that benefit the most from zero rating increase their activity significantly more than the population on average. They view twice as many stories, and give three times as many feedback compared to the population average due to zero rating. They also create 130% more content as a consequence of zero rating. A consumer’s network characteristics identify consumer needs better than just using degree as measured across all social network activities.

**Table 3.2:** Heterogeneity in network sensitivity to zero rating: Percent lift in activity over population average

| Activity | Top decile by degree | Top decile by network |
|----------|----------------------|-----------------------|
| Stories  | 93%                  | 104%                  |
| Feedback | 183%                 | 195%                  |
| Content  | 126%                 | 133%                  |

### 3.4 Conclusion

In this paper, I examine direct and spillover effects of zero rating. I find that zero rating increases content consumption, feedback, and creation among those who signed up for zero rating. I find significant spillover to peers and to the period after the end of the campaign. Peers make up a significant proportion of the total campaign effect, in part due to the large number of people who did not get zero rating pricing but were affected indirectly through their social circle.

My work provides an insight that zero rating programs do not have a uniform effect across all zero rated activities, can affect consumers not on zero rating (if their peers are on zero rating), and have implications on consumer behavior after the program ends.

My approach uncovers variability in consumer response to zero rating. Differences in response can be attributed to consumers' characteristics, past behavior, and their friendship network. Zero rating, as one of many products targeting connectivity for consumers in developing countries, help solve problems for some consumers better than for others. We find network characteristics to be a better indicator of consumer needs than degree alone.

One limitation is that my approach is that I exclude new consumers who come online because of zero rating itself. Thus, the results are conservative and conclusions apply only to existing users.

This analysis on activities on Facebook covered by zero rating. To understand the financial implications of zero rating for mobile operators, related data of mobile phone and app use would provide a fuller picture of how zero rating affects consumer communication activities. My analysis benefits from including the Facebook network, but supplementing this with phone calling and SMS network could help capture a more complete social influence process.

### 3.5 Further extensions

In this section, I discuss four potential extensions to the analysis in this chapter.

#### 3.5.1 Improved matching

In the analysis, I matched using degree and pre-campaign activity level. Matching by additional variables could improve control for self-selection and homophily. Other matching variables could include demographics, geographic location, and time since joining Facebook. Another potential variable is birthday, which could cause someone to self-select into zero rating and be more active.

The current matching scheme splits the dataset into subclasses. This approach becomes prohibitive with a large set of variables. In response, we could use a dynamic propensity score match-

ing approach (Aral et al. 2009). One limitation to propensity score matching is that it only considers a single binary treatment. Here, we have two dimensions of treatment: ego zero rating status, which is binary, and number of peers with zero rating, which is continuous. Instead of using propensity score matching, we can use propensity function which accommodates continuous as well as multivariate treatments (Imai & Van Dyk 2004).

To illustrate the modeling approach, consider two treatments:  $Z_1 \in \{0, 1\}$ ,  $Z_2 \geq 0$ . Let  $M$  be matching variables and  $Y$  be the outcome of interest. First, predict  $Z_1, Z_2$  based on  $M$  to get predicted values  $\hat{Z}_1, \hat{Z}_2$ . Next, subclassify similar values of  $\hat{Z}_1, \hat{Z}_2$ . Finally, estimate  $E(Y|Z_1, Z_2)$  within each subclass.

Since we observe many observations per individual, we conduct the analysis at the week level. For each week, a consumer will have a (1) subclass based on propensity score, (2) treatment status, and (3) outcome variable (change in activity level). Standard errors for the whole process (including uncertainty around propensity score matching) can be estimated using bootstrap.

### 3.5.2 Deeper content analysis

There is an opportunity to gain greater consumer insight beyond those contained in this paper by examining what type of content is affected by zero rating. In my analysis, I used the number of activities as a measure of each activity type. For example, content consumption is measured by the number of stories viewed. I can break this down further to understand which sub-types of activities are affected.

For example, content creation can be broken down into status updates, link sharing, or photo uploads (among others). Because these activities differ in data cost, zero rating should have different effects across the activities. Understanding the effect of zero rating on different content creation activities for consumers can give better insight into consumer needs.

### 3.5.3 Combined measure of network activities

Instead of measures corresponding to each social network activity, we can aggregate them into a single metric. For example, we could index a network's content consumption and creation.

Let  $S_{it}$ ,  $C_{it}$  be the number of stories consumed and content created by person  $i$  at time  $t$ . Then, an index  $R_{it} = \frac{C_{it}}{S_{it}}$  measures the number of content created for every story consumed. A high  $R$  means the person creates more content than consumed (relative to average), and a low  $R$  means the person is more of a lurker.

We can take this measure to the network level by considering the ratio  $R_{N(i),t} = \frac{\sum_{j \in N(i)} C_{jt}}{\sum_{j \in N(i)} S_{jt}}$  where  $N(i)$  is a network of person  $i$ , and  $j \in N(i)$  means person  $j$  is a friend of  $i$ . Thus, we can compare two different networks  $N_1, N_2$  and see the ratio of content created to stories consumed in those networks.

We can also measure how zero rating affects this index for individuals over time. Understanding which consumers are affected by zero rating to become more content active or content inactive relative to content consumption can help give product ideas to customize to fit consumer needs.

### 3.5.4 Extended model of peer influence

In my model, activity level is modeled as a function of a consumer's zero rating status and peer zero rating. Realistically, peer activity level could enter into consumers' decision making. I excluded this in the current analysis due to simultaneity and reverse causality issues but discuss it here for exploratory purposes.

I sketch a vector autoregressive (VAR) model at the individual level, by modifying equation 3.9:

$$\begin{aligned}
 Z_{it} &= \max(Z_{i,t-1}, f(\Delta Y_{it}, \Delta Y_{jt})) \\
 \Delta Y_{it} &= \eta_t + \alpha_1 Z_{it} + \beta_1 \sum_{j \in N(i)} Z_{jt} + \delta_1 Z_{it} \cdot \sum_{j \in N(i)} Z_{jt} + \kappa \sum_{j \in N(i)} Y_{jt} + \nu_1 X_i + \epsilon'_{it}
 \end{aligned} \tag{3.11}$$

where  $f$  maps continuous variables to  $C(0, 1)$  (e.g, probit function). In the first equation, we directly model the choice of the consumer to opt into zero rating or not. We let this choice depend on their change in activity level and peer activity level. In the second equation of 3.11, we let  $\Delta Y_{it}$  be a function of  $Y_{jt}$ . This reflects that peer activity may have additional effects on consumer behavior above and beyond peer zero rating.  $Y$  can be a multidimensional vector to reflect the interdependence of activities, for example, own stories viewed depends on how much content is created by peers. We pool across all individuals and use OLS to estimate coefficients, then get standard errors using bootstrap. Impulse response function can be used to derive the impact of zero rating on activities.

# 4

## Conclusion

In this dissertation, I study how consumers influence each other in two different digital settings with implications for businesses and regulators. There are a few takeaways from the dissertation.

First, peer influence plays a large role in consumer decision making. Peer influence accounts for a large proportion of adoption of social games. Yet, consumers may respond to peers by being more or less active in a social network. The implications for managers is that it is important to understand how peers influence each other because it is a big component of marketing and

consumer decision making.

Second, it is critical for causal inference methods to be able to accommodate Big Data and distributed computing. In the second essay, every analytic step from data processing to subclass matching, regression and simulation was performed in Hive and Presto. Many statistical approaches, such as latent space models in the first essay, and marketing models are implemented in statistical software that was created to accommodate sampled data. As companies acquire larger datasets and rely more on A-B testing to inform tactical direction, what is the role of causal inference and observational studies on marketing strategy? The ability to take on Big Data and utilize new technologies of distributed computing is a must for research to stay relevant.

Third, observational social influence studies still need further validation with gold standard methods such as randomized experiments. There are many critiques, from Shalizi & Thomas (2011) and Angrist (2014), which highlight difficulties in identifying social effects. Understanding peer influence especially in digital social networks where randomized experiments may be possible would be a positive first step.

Fourth, for managers who have already bought into the idea of social influence, the next question is how the effect varies from individual to individual. While there are studies that examine the question of heterogeneous treatment effects (Trusov et al. 2010), there is opportunity for more studies to quantify how social influence varies across consumers.

Finally, understanding drivers of consumer decision making and the effect of pricing in a social network are just first steps. Managers are interested in linking marketing levers they control with financial outcomes. For example, how do free and freemium app experience affect social activities? How do distribution channels change the impact of peer influence in adoption?

To conclude, there is a rich research agenda to develop for how we understand consumers and influence each other, with important practical and research implications for businesses and regulators alike.



# References

- Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30, 98–108.
- Angrist, J. D. & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Ansari, A., Koenigsberg, O., & Stahl, F. (2011). Modeling multiple relationships in social networks. *Journal of Marketing Research*, 48(4), 713–728.
- Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51), 21544–21549.
- Aral, S. & Walker, D. (2011a). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9), 1623–1639.
- Aral, S. & Walker, D. (2011b). Identifying social influence in networks using randomized experiments. *Intelligent Systems, IEEE*, 26(5), 91–96.
- Aral, S. & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, 337(6092), 337–341.
- Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social influence in social advertising: evidence from field experiments. In *Proceedings of the 13th ACM Conference on Electronic Commerce* (pp. 146–161).: ACM.
- Ballve, M. (2013). Our list of the world's largest social networks show how video, messages, and china are taking over the social web.
- Bell, D. R., Chiang, J., & Padmanabhan, V. (1999). The decomposition of promotional response: An empirical generalization. *Marketing Science*, 18(4), 504–526.
- Bergen, M. (2014). The secret sauce for twitter's global growth strategy: Subsidized data. *Advertising Age*.
- Bollinger, C. R. (2003). Measurement error in human capital and the black-white wage gap. *Review of Economics and Statistics*, 85(3), 578–585.

- Bollinger, C. R. & Minier, J. (2015). On the robustness of coefficient estimates to the inclusion of proxy variables. *Journal of Econometric Methods*, 4(1), 101–122.
- Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1), 41–55.
- Braun, M. & Bonfrer, A. (2011). Scalable inference of customer similarities from interactions data using dirichlet processes. *Marketing Science*, 30(3), 513–531.
- Burke, M., Marlow, C., & Lento, T. (2009). Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 945–954): ACM.
- Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1909–1912): ACM.
- Chhachhar, A. R., Qureshi, B., Khushk, G. M., & Ahmed, S. (2014). Impact of information and communication technologies in agriculture development.
- Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432), 1313–1321.
- Chui, M. (2013). *The social economy: Unlocking value and productivity through social technologies*. McKinsey.
- Eckles, D. & Bakshy, E. (2015). Bias and high-dimensional adjustment in observational studies of peer effects.
- Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, 68(3), 589–599.
- Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, (pp. 54–75).
- Ghose, A. & Han, S. P. (2011). An empirical analysis of user content generation and usage behavior on the mobile internet. *Management Science*, 57(9), 1671–1691.
- Globe (2014). Globe free facebook offer lifts subscriber uptake; more than doubles mobile browsing adoption. Press Release.
- Globe (2015). Driving internet adoption in developing markets and the role of mobile carriers: A philippine case study. <http://www.globe.com.ph/documents/50301/15939122/2015-03-05+FB+Globe+Whitepaper>.
- Godes, D. & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, 28(4), 721–739.

- Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M., & Verlegh, P. (2005). The firm's management of social interactions. *Marketing Letters*, 16(3-4), 415–428.
- Goldsmith-Pinkham, P. & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3), 253–264.
- Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing research*, (pp. 342–355).
- Hartmann, W. R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science*, 29(4), 585–601.
- Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K., & Tucker, C. (2008). Modeling social interactions: Identification, empirical methods and policy implications. *Marketing letters*, 19(3-4), 287–304.
- Hill, S., Provost, F., & Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, (pp. 256–276).
- Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469), 286–295.
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098.
- Hughes, D., Coulson, G., & Walkerdine, J. (2005). Free riding on gnutella revisited: the bell tolls? *Distributed Systems Online, IEEE*, 6(6).
- Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative science quarterly*, (pp. 422–447).
- Imai, K. & Van Dyk, D. A. (2004). Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467).
- Internet.org (2015). State of connectivity: 2014 a report on global internet access.
- Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(2), 195–212.
- Jackson, M. O. (2013). Unraveling peers and peer effects: Comments on goldsmith-pinkham and imbens“social networks and the identification of peer effects”.
- Jedidi, K., Mela, C. F., & Gupta, S. (1999). Managing advertising and promotion for long-run profitability. *Marketing science*, 18(1), 1–22.

- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, (pp. 263–291).
- Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, (pp. 427–436).
- Katona, Z., Zubcsek, P. P., & Sarvary, M. (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, 48(3), 425–443.
- Katz, R. (2012). The impact of broadband on the economy: Research to date and policy issues. *International Telecommunications Union, Geneva*.
- Khalaf, S. (2014). Apps solidify leadership six years into the mobile revolution. <http://blogs.plos.org/mfenner/2011/07/21/how-to-formally-cite-a-blog-post/>.
- Klepper, S. & Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica: Journal of the Econometric Society*, (pp. 163–183).
- Kossinets, G. & Watts, D. J. (2009). Origins of homophily in an evolving social network1. *American Journal of Sociology*, 115(2), 405–450.
- Krivitsky, P. N. & Handcock, M. (2008). Fitting position latent cluster models for social networks with latentnet.
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3), 204–213.
- Leeflang, P. S. & Parreño-Selva, J. (2012). Cross-category demand effects of price promotions. *Journal of the Academy of Marketing Science*, 40(4), 572–586.
- Manchanda, P., Packard, G. M., & Pattabhiramaiah, A. (2014). Social dollars: The economic impact of customer participation in a firm-sponsored online customer community. *Forthcoming, Marketing Science*.
- Manchanda, P., Xie, Y., & Youn, N. (2008). The role of targeted communication and contagion in product adoption. *Marketing Science*, 27(6), 961–976.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3), 531–542.
- Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. (2013). *Disruptive technologies: Advances that will transform life, business, and the global economy*, volume 180. McKinsey Global Institute San Francisco, CA.

- Maytom, T. (2014). Coca-cola launches campaign around mobile game and wearables. <http://mobilemarketingmagazine.com/coca-cola-campaign-mobile-game-wearables>.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, (pp. 415–444).
- MEF (2015). How telenor and facebook partnered to drive internet adoption in pakistan. <http://mefminute.com/2015/03/17/how-telenor-and-facebook-partnered-to-drive-internet-adoption-in-pakistan/>.
- Mela, C. F., Gupta, S., & Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing research*, (pp. 248–261).
- Meltzer, J. (2013). The internet, cross-border data flows and international trade. *Issues in Technology Innovation*, (22).
- Morris, A. (2014). Report: 45% of operators now offer at least one zero-rated app.
- Nair, H. S., Manchanda, P., & Bhatia, T. (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research*, 47(5), 883–895.
- Nielsen (2014). How smartphones are changing consumers' daily routines around the globe. <http://www.nielsen.com/us/en/insights/news/2014/how-smartphones-are-changing-consumers-daily-routines-around-the-globe.html>.
- Nijs, V. R., Dekimpe, M. G., Steenkamps, J.-B. E., & Hanssens, D. M. (2001). The category-demand effects of price promotions. *Marketing science*, 20(1), 1–22.
- Nitzan, I. & Libai, B. (2011). Social effects on customer retention. *Journal of Marketing*, 75(6), 24–38.
- Nottebohm, O., Manyika, J., Bughin, J., Chui, M., & Syed, A. (2012). Online and upcoming: The internet's impact on aspiring countries. *McKinsey & Company Publishing*.
- Ogilvy, D. (2013). *Ogilvy on advertising*. Vintage.
- Pauwels, K., Hanssens, D. M., & Siddarth, S. (2002). The long-term effects of price promotions on category incidence, brand choice, and purchase quantity. *Journal of marketing research*, 39(4), 421–439.
- Prentice, R. L. & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66(3), 403–411.
- Qiang, C. Z.-W., Rossotto, C. M., & Kimura, K. (2009). Economic impacts of broadband. *Information and Communications for Development 2009: Extending Reach and Increasing Impact*, (pp. 35–50).

- RedBull (2014). Games. <http://www.redbullmediahouse.com/products-brands/games.html>.
- Shalizi, C. R. & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, 40(2), 211–239.
- Shriver, S. K., Nair, H. S., & Hofstetter, R. (2013). Social ties and user-generated content: evidence from an online social network. *Management Science*, 59(6), 1425–1443.
- Stephen, A. T. & Toubia, O. (2010). Deriving value from social commerce networks. *Journal of marketing research*, 47(2), 215–228.
- Takahashi, D. (2014). Mobile games revenue will double by 2017 as growth spreads worldwide (exclusive). <http://venturebeat.com/2014/07/23/mobile-games-revenue-will-double-by-2017-as-growth-spreads-worldwide-exclusive/>.
- Trucano, M., Hawkins, R., & Iglesias, C. (2012). Ten trends in technology use in education in developing countries that you may not have heard about. *EduTech: A World Bank Blog on ICT use in Education Retrieved Jan, 12, 2013*.
- Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining influential users in internet social networks. *Journal of Marketing Research*, 47(4), 643–658.
- Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. *Management Science*, 54(12), 2024–2038.
- Van den Bulte, C. & Lilien, G. L. (2001). Medical innovation revisited: Social contagion versus marketing effort1. *American Journal of Sociology*, 106(5), 1409–1435.
- Van Heerde, H. J., Gupta, S., & Wittink, D. R. (2003). Is 75% of the sales promotion bump due to brand switching? no, only 33% is. *Journal of Marketing Research*, 40(4), 481–491.
- Zhang, K. & Sarvary, M. (2011). *Social media competition: Differentiation with user generated content*. Technical report, Working Paper, INSEAD.
- Zuckerberg, M. (2015). Our facebook community. <https://www.facebook.com/photo.php?fbid=10101878930744521&set=a.529237706231.2034669.4&type=1&theater>.