Citation: Mahmood, A. & Sismeiro, C. S. (2017). Will they come and will they stay? Online Social Networks and news consumption on external websites. Journal of Interactive Marketing, 37, pp. 117-132. doi: 10.1016/j.intmar.2016.10.003

Will they come and will they stay? Online Social Networks and news consumption on external websites

Ammara Mahmood[1]

Cass Business School, London

Catarina Sismeiro

Imperial College Business School

Abstract

This study explores the role of online engagement, homophily and social influence in explaining traffic and news consumption by social network users at an external news website. The authors jointly model visits and page views for a panel of users who registered with the news site using their Facebook accounts. In their model, the authors account for homophily using a latent space approach, and account for endogeneity, heterogeneity, and unobservable correlates. The results show that measures of an individual's activity on Facebook are positively associated with that individual's actions at the news site. In addition, knowing what a user's Facebook *friends* do at the content website provides insights into a focal user's behavior at that website, as visitors with friends who visit external news sites are more likely to visit the news website studied. In addition, news consumption (not just visits) also depends on friend's actions but such an impact varies with the individual's underlying browsing mode. We highlight the importance of social influence in news consumption and further show that homophily bias in news consumption is similar to prior research in other categories. Our study also highlights that visitors' past browsing patterns are important predictors of future content consumption, although social network information significantly improves prediction beyond the effect of such more traditional behavioral metrics. Finally, we find that Managers can use readily available data for both prediction and targeting.

[1] Corresponding author ammara.mahmood.1@city.ac.uk,

Acknowledgements:The authors would like to thank Charles Hofacker, Ko De Ruyter, Caroline, Wiertz and Tom Van Laer for their insightful comments and suggestions.

Highlights

- Social network data helps predict and describe news consumption at third party news sites
- Personal preferences, Homophily and social influence explain news consumption amongst users connected via social networks
- Personal and social motivations for news consumption provide the underlying mechanism for the positive association between social network use and news consumption at external sites
- Visitors are 10% more likely to visit a news site if their Facebook friend also visited
- Failure to account for homophily can bias estimates of social influence by up to 40%

Keywords: Online Social Networks, News, Social Influence, Homophily, Latent Space Approach

Introduction

In the past few years, the market has witnessed the development of online social networks and their rise as one of the most influential online forces. With over a billion users engaging with sites like Facebook and Twitter, it is clear how important these media are for creating connections and communicating with customers. Social media can also provide benefits beyond engagement and advocacy, and businesses are fast recognizing the power of "social media analytics" that rely on ever more data tracking user actions. Sentiment analysis based on such data is now used to predict the outcome of elections (Golbeck and Hansen 2011; Tumasjan et al. 2010), success of movies at the box office (Rui et al. 2013; Asur and Huberman 2010), marketability of consumer goods (Shimshoni et al. 2009), and even stock performance (Bollen et al. 2011). Practitioners also use social media data to analyse customer satisfaction by gauging the impact of competitive marketing campaigns (Bradbury 2013).

The power of social networks lies in information users provide about themselves, their preferences, their "friends", and the influence friends exert on each other. Because network friends could be interested in the same products (e.g. homophily; McPherson Smith-Lovin and Cook 2001) social advertising, social targeting, and social customer scoring on platforms like Facebook have revealed great potential (Hill et al. 2006; Goel and Goldstein 2013). As online groups and communities that revolve around brands become more visible and easy to monitor, they attract greater interest from businesses because of the strategic opportunities they provide (Libai et al. 2010). As a result, today, online social networks (and notably Facebook) collect significant amounts of information and provide platforms that businesses can harness to implement strategies like social advertising and community development.

Despite the importance practitioners attach to social network analysis and social media use, there is currently a dearth of academic studies in marketing that illustrate the value of online social network data in targeting and predicting behavior of connected individuals *outside* the network (Goel and Goldstein 2013). More importantly, online content consumption at third-party external websites has largely been over-looked in social media studies, despite social networks being inherently platforms to share content, making them simultaneously complementary and competitive entities. The neglect by researchers is even more surprising given that the online content sector supports, directly and indirectly, a significant portion of economic activity online and the flourishing global online advertising with spending over $100 billion (E-marketer 2013).

Among companies whose business models revolve around content provision, news websites are facing some of the most significant challenges. As 46% of social network users discuss news stories (Anderson and Caumont 2014), online social networks have been progressively making inroads into the news delivery business. Snapchat's introduction of news and content distribution is a good example, as are Facebook's continuous improvements in news hosting. As a result, news websites and organizations are ever more focused on trying to understand how consumers interact with news, at news sites or while visiting social media platforms, to develop subscription plans and increase the visibility of their content (Mitchell, Jurkowitz and Olmstead 2014). Similarly, the ability to predict traffic and engagement at content sites using the browsing behavior of social network members is of financial relevance as it can lead to better placement of ads and content (Lerman and Hogg 2010).

To benefit from the prominent role of social networks, news websites encourage their readers to engage with news by registering using their social network accounts. In the case of Facebook pages, this stimulates interaction among group members, peer-to-peer sharing and greater visibility of news in users' Facebook newsfeed. It also gives businesses the access to a rich set of personal information including what users post and reveal on their personal pages, and the identity of their friends. Such information could help the selling of premium targeted ads and it may provide the opportunity to understand how connected individuals interact with external sites and whether peer-to-peer interactions (including news sharing) results in positive or negative effects on news reading. However, there is currently a dearth of studies trying to link social network activities and content consumption at external websites. Most of the existing studies, whether aggregate or disaggregate, are survey based (e.g., Bernoff and Li 2011) and often provide contradicting results depending on the methodology used and data employed (e.g., De Waal and Schoenbach 2010; Dimmick et al. 2004; Nguyen 2010; Tewksbury 2003).

As a result, it is not entirely clear to content providers how valuable the information they collect from Facebook is in predicting user behavior at their website, including traffic and page requests. This is an important issue for content websites, whose ad revenues depend directly on the traffic and on the page views they are able to generate. Finally, little research has focused on online content providers and their users, as most studies of online social media tend to be focusing on movies (Dellarocas Zhang, and Awad 2007), games (Zhu and Zhang 2010), micro lending (Stephen and Galak 2012), and books (Chevalier and Mayzlin 2006). With this work, we not only study the relationship between news reading and online social network activity (more specifically activity on Facebook), we further contend that the potential impact of Facebook on news consumption relies on a series of complex mechanisms, and that the net effect depends on how active friends are, and on how active the focal user is on social networks. Drawing on the communications research, we outline that personal and social goal fulfillment could explain the association between users' engagement with Facebook and news sites, as well as the influence of connected social network peers.

We focus our analysis on two browsing decisions: (1) a user's decision to visit the content website, and (2) the decision on how many pages to view. These variables are frequently monitored due to their direct bearing on the revenue generated by news sites, 80% of which relates to advertising placement on their pages (Clemons et al. 2002). Ours is the first study to combine Facebook data from a panel of Facebook users registered with a major news website with browsing activity at the third party news site to jointly model website visitation (traffic) and the number of pages viewed over time (engagement) at the news website.

We adopt a flexible modeling approach using a random-coefficients Poisson Hurdle model to find associations between behavior at news sites and at online social networks. We distinguish between users' *own* activity and the impact of users' *social* network peers, and in doing so we also carefully separate the role of social influence and homphily in driving news consumption of Facebook users at external sites. Our paper also contributes to research on the role of groups formed around a focal brand within social networks (e.g. Libai et al. 2010). In doing so we help managers understand the role of these groups in predicting and understanding the behavior of users while visiting brand sites.

Our results show that Facebook-related information can help predict site visits and number of page requests. We are able to improve the predictions from already extremely accurate models that use individual-level browsing data at the news site. Strikingly, browsing information from friends provides greater improvements in prediction than knowing the individual's own behavior at the online social network. This shows that information on the behavior of connected friends, the essence of social networks, is the information one can extract from online social networks that provides the greatest benefits.

In addition, news consumption and being active on Facebook appear to be complementary. Our results further suggest that engagement with online social networks could be associated with specific patterns of news consumption and site visits as individual-specific motivations and characteristics make both behaviors (i.e., the use of social networks and news reading at external sites) more likely. More importantly, our findings confirm that online news consumption is a shared experience, with new consumption activity of social network friends associated with similar behavior by other network members. Whereas most of these co-movements can be attributed to social influence dynamics, homophily also plays a role. We find a 12.4% increase in the likelihood a focal user will visit the news website when the user's Facebook friends also visit the same news website. However, 79% of this effect is due to social influence and 21% due to homophily. Because both phenomena (influence and homophily) suggest alternative strategies available to companies, knowing that both are of relevance is fundamental to make strategic decisions (even if the split between the two forces might depend on the approach employed, our findings highlight the importance of both forces).

Our results also suggest that engagement with online social networks could be associated with specific patterns of news consumption and site visits because individual-specific motivations and characteristics make both behaviors (i.e., the use of social networks and news reading at external sites) more likely. We unveil several complexities behind social network effects. We find that individuals visiting the news website on the same day as their friends seem to consume content differently once at the news site. When friends are active, focal users exhibit more shallow reading compared to when friends are inactive (visitors' underlying browsing mode moderates this effect). Hence, friends seem to attract traffic to the website (promotional effect), but that traffic is directed and comprises fewer pages viewed. These changes in content consumption might depend on the type and structure of the website, however, our results indicate that the type of information individuals are exposed to within their networks (directly or indirectly) may have an effect on the information they search outside the social network. This widens greatly the scope of social networks influence.

We note that aggregate-level statistics, readily available to website managers, mask the mechanisms in place and cannot provide the clear operational recommendations that the separation of effects from homophily, influence, and correlated unobservables can. We further highlight that the measures available to us in this study are the measures easily accessible to business whose users register using their Facebook accounts. Hence, our findings provide businesses and policy makers' with important insights into the value of widely accessible data. This makes our results of particular managerial significance as business can use this information to improve predictions of traffic and engagement at sites external to Facebook at little additional cost, and use these predictions to implement social advertising, social targeting, and social customer scoring (Hill, Provost and Volinsky 2006; Goel and Goldstein 2013).

Literature Review

In this study, we focus on whether social networks are valuable for predicting behavior at external sites, and the underlying mechanisms that would make the association between behaviors in two very distinctive online platforms likely. Prior research in marketing has demonstrated that social networks influence their members' behavior (de Valck, van Bruggen, and Wierenga, 2009). Several recent studies also investigate the correlation between aggregate measures of social network activity and aggregate measures of performance in a variety of business contexts such as movies (Dellarocas et al. 2007), games (Zhu and Zhang 2010), microlending (Stephen and Galak 2012), and books (Chevalier and Mayzlin 2006).

Many of the existing studies on online content consumption and social networks focus on motivations for sharing content by using small-scale experiments (e.g., Wilcox and Stephen 2013) or social transmission using behavioral data from content sites including 'most read' and 'most emailed' stories or the inclusion in Google News (e.g. Berger and Milkman, 2012). Other studies focus on the impact of public endorsement on site navigation and reader's attitudes (see Hallahan 1999, and Johnson and Kaye 2004), which also seem to affect individual patterns of news consumption (e.g., Thorson 2008; Jeon and Esfahani 2012). Other studies using aggregate level data find limited association between Facebook and engagement on third-party news sites based on direct referrals (Sismeiro and Mahmood 2016, Mitchell et al. 2014a). However, none jointly analyze individual level behavioral data from online social networks and news sites.

Individual Preferences and Social Connections as Predictors of Online News Consumption

At the individual-level, what one does at online social networks may be associated with actions at external news sites because of (1) individual level factors (e.g., heterogeneity and changing individual preferences) and (2) the impact of the individual's social connections. Individual preferences and heterogeneity is a research topic in marketing with a long tradition and its relevance is well demonstrated in predicting online behavior. Previous studies have also established that tracking changing individual preferences (explicitly or implicitly through browsing histories) can help prediction of online behavior beyond what heterogeneity modeling can do (Sismeiro and Bucklin 2004).

Research on social connections also has a long history, with a focus on the role of social influence and homophily. In sociology, for example, research has indicated that social connections in their own right could help forecast behavior due to the existence of homophily (McPherson Smith-Lovin and Cook 2001) but the degree of improvement depends on the behavioral information already available on the targeted users (Liu and Tang 2011). Beyond homophily, social connections can be predictive of online behavior because individuals may influence each other, directly or indirectly. Social influence is a deeply elusive research topic and it is not always easy to identify and measure its effects (Aral 2011; Nair, Manchanda, and Bhatia 2010). However, it is now well documented that individuals influence each other offline and online (Watts and Dodds 2007) and in contexts like product adoption and diffusion and in the formation of product ratings (e.g., Van den Bulte and Stremersch 2004; Chevalier and Mayzlin 2006; Dellarocas et al. 2007; Godes and Silva 2012).

Though individual preferences, homophily, and social influence seem to be of importance in a variety of contexts, few studies exist in the context of online news and online social networks that attempt to demonstrate their role. Borrowing from existing literature on communication research, we employ the classic typology of Katz, Blumler and Gurevitch's

(1974) to posit that individuals' motivations for media consumption (*personal* and *social*) can provide the underlying mechanisms to explain why such factors might indeed support the link between what people do at online social networks and at third-party news sites.

Personal Motivations for Media Consumption

Personal motivations have long been recognized as very important in explaining media and news consumption (Tsfati and Cappella 2005). Individuals might consume news to develop a personal identity or for their referential function (i.e., to learn accurate information). News also fulfills surveillance functions (Wright 1960) when audience members wish to get the information necessary for their daily lives (e.g., learning about a strike, weather, traffic), or to gratify their cognitive needs (e.g. learning about a variety of issues and topics).

These *personal* motivations can be fulfilled by content consumption at news sites and online social networks. In addition, visits to social networks to fulfill these personal motivations could promote visits to news sites as content shared by network peers may trigger a personal desire to seek more news content. Actively participating in online activities could also reveal an individual's preference to spend available time in seeking online content. When users are actively engaged with social networks we can expect simultaneous or subsequent consumption of other online content (for instance by having another browser open, or by visiting news website and reading complete articles instead of news headlines on Facebook). We therefore envisage co-movements of activities in online social networks and at third party websites at the individual level, and knowing what a user does on social networks would reveal individuals' personal preferences and or their changing interest over time.

Social Motivations

Previous research has also shown that news reading can satisfy *social needs*. News reading can be instrumental to the development of a social identity and the feeling of being included. Individuals want to belong to reference groups and being aware of the news the group consumes, helps them develop and sustain social relationships, support a specific social identity, build social capital, and develop a sense of belonging (Tsfati and Cappella 2005).

We expect social motivations for news consumption to be significant for users who actively participate in online social networks. Existing literature on contagion caused by cohesion also support the basic view that people connected via social networks may induce their friends to behave in a similar way (Centola 2010). In addition, and in expectation of social interaction and discussion with others, individuals who see a specific topic on social media sites or their friends' pages are more likely to search for more information about it outside the online social network environment because they do not want to appear uninformed when conversing with others (Walther et al. 2008). Such "communicatory utility" is often used to explain how interpersonal motivations drive (mass) media information-seeking in order to fulfill interpersonal goals (Atkin 1972). Hence, social influence may also occur indirectly through conversations around content on social networks (Shirky 2008). Such interest would go beyond homophily explanations or that of clicking shared links, behaviors traditionally used as an example of social influence.

Previous research further suggests that image related motivations are strong and more significant than intrinsically utilitarian motives in the context of social networks such as Twitter (Toubia and Stephen 2012). We therefore predict that when friends of a focal user visit the news site they are more likely to share news from that site, and to comment or discuss online news stories they had read about (Baresch et al. 2011). As a result, the focal user exposed to these activities and be more likely to also visit the news site to know what her friends know and participate in the same discussions. Visits to the news site friends had visited could be the result of convenience in simply following the links posted by friends or, in the absence of a posted link, because of the common preference of focal user and her friends for that news outlet. In sum, individual preferences, the existence of homophily and social influence would suggest association between Facebook and news sites.

Empirical Approach

We model individual-level data obtained from a leading European newspaper with an online and an offline presence (the second largest newspaper in its country of origin in terms of readership and circulation). We collected the browsing activity for a panel of online users who registered with the news website using their Facebook accounts (joined Facebook fan page), indicating their preference for the news site to their social network.

The dataset comprises the daily site visits and pages viewed by news category for each user during the month of March 2012 and information from their Facebook accounts (when registering with the news site, users consent to share the details of their Facebook profile). Previous studies often use social network data that is not readily available to most businesses (e.g. Goel and Goldstein 2013), instead, we use of Facebook information that is readily available to any company with a Facebook page. Although the dataset does not include the entirety of users' daily posts and comments, we have access to their demographics (e.g., age and gender) and we observe all Facebook pages the users liked (these are not specific articles or friends' posts, but Facebook pages of celebrities, sports stars, news sites, businesses etc.). For example, if a person liked the Rolling Stones Facebook page, we see the date that person became a fan, the exact name of the page, and that it is a music related page. This information is extremely as it provides insights regarding the interests and preferences of the individuals as revealed by their behavior observed outside the content website under study. More importantly it allows us to know when individuals were actively engaged with Facebook.

We also observe the number of Facebook friends of each user (i.e., the degree of the network), the identity of each friend, and whether friends registered with the news website. By studying Facebook friends our focus is on networks formed due to shared interest in news as well as factors exogenous to the news site. We included in the panel those individuals with at least one Facebook friend registered with the news website, allowing us to monitor the activity of a visitor's Facebook friends while they visited the focal news website.[2]

We observe the page requests to the news site over time and by news category for our sample of Facebook users. Because the news website classifies its content into 35 categories, we grouped similar categories and use such information in our model estimation. For example, we grouped categories discussing celebrities, TV, and life-style under the category "Entertainment." We also grouped categories those with limited interest into one residual category called "Other." The final categories considered included Home Page, Entertainment, Local News, Sports, Politics and Other.

To allow enough time to initialize variables reflecting previous experiences (e.g., lagged variables, time since last visit, etc.), we reserve the first week of data as an initialization period and used the remaining three weeks (23 days, from March 9th to March 31st) for model estimation. We believe this is an adequate initialization period as users return to the website, on average, after 2.4 days and less than 2% of the users have an inter visit time of more than 15 days. To ensure we include a complete browsing history for our panel users, we kept only users who visited the website at least once during the first week of March, and returned to the website at least once during the estimation period.[3] Hence, we retained 1,562 users in our sample with 35,926 daily observations (23 observations per user) of which 15,864 correspond

[2] Our data does not allow us to observe direct influence as measured through direct referrals when users click and follow the links posted on Facebook pages by their friends, taking them to external websites. We believe this is not a limitation in our context. Previous studies find that traffic arriving at the news site through direct referrals is very limited with some authors reporting that only 9% of online news consumers follow direct recommendations from Facebook and other social media (Mitchell, Rosenstiel, and Christian 2012a). In addition, by not limiting our analysis to direct referrals we are able to include other forms of social influence such as when Facebook friends discuss their personal life or current affairs and engage in interpersonal communication that does not directly relate or include a link to the news site. Such interactions can create interest and curiosity in news stimulating visits to news sites for further information gathering. This phenomenon can have a potentially greater impact in driving visits by Facebook users than direct referrals as demonstrated by previous research (Sismeiro and Mahmood 2016).

[3] We exclude all users who do only one visit to the website during the entire month but not those who visited at least once in the first week and then only once during the estimation period. Browsers with only one visit in a month are casual browsers and correspond to a small percentage of users (17% of users registered using a Facebook account), and to an even smaller percentage of pages viewed (0.5% of page views by Facebook users). The amount of traffic lost in our analysis is hence very small.

to website visits (non-zero page views).

Modeling Approach

Our goal is to model the behavior of individuals at the news site and determine if Facebook information can help prediction of individual's actions at an external news site. In addition, beyond determining the predictive power of Facebook information, we wish to study the importance of individual preferences, time dependent factors, homophily, and social influence in explaining the relationship between Facebook related activity and news consumption at an external news site. To do so we need to model individual actions at the news sites, more specifically the number of articles consumed by each individual from the website. Since not all visitors are active every day, the daily count of page views by individual visitors is typically characterized by excessive zeroes and over dispersion (variance greater than the mean). If not correctly modeled excessive zeroes and over dispersion can lead to incorrect inferences (Ridout, Demétrio, and Hinde 1998).

Although we adopt Poisson hurdle model, there are alternative specifications. For example, the zero inflated Poisson model (ZIP) can handle zero-inflation (Min and Agresti, 2005). A variety of Tobit specifications are also available relying, for example, on Log-Normal distributions for the non-zero positive variable. We believe though the approach we adopt is better suited for modeling our data. The hurdle model can handle not only zero-inflation but also zero-deflation and it handles the inflation/deflation process by modeling two separate but structurally connected decision processes, allowing a clear interpretation of results. In our context, this implies separation of the decision to visit the site (binary outcome) and, subsequently, the decision to read a certain number of articles from the website (the process accounting for the positive outcomes) making the hurdle model an obvious choice to answer our research question.

We further note that we could have adopted alternative specifications for the count part of our model. However, unlike the commonly used Binomial distribution, the Poisson does not place upper limits on the value of the observed count, and is more adequate than a Log-Normal specification, typically used in Tobit models and some selection models, which would model the counts as a continuous variable. Although the two decisions of the hurdle model could be estimated separately, we adopted a correlated random effects Poisson model to study simultaneously (1) the decision to visit the news website on a given day, and (2) the amount of daily content consumed once at the website (number of pages viewed). As we will see next, the correlated random effects specification allows us to capture unobserved heterogeneity across users, the possible non-independence in observations due to their clustering around individuals (Greene 2009), and non-independence of the two decisions.

We propose a generalized linear mixed model (GLMM) specification to estimate the correlated random effects Poisson Hurdle model (Breslow and Clayton 1993). The GLMM allows an efficient estimation of mixed non-linear models and an adequate handling of individual-level heterogeneity. In our model we let Y_{it} be the total number of pages viewed by individual i at day t (i = 1,…, N and t = 1,…, T). When a user does not visit the focal news website we observe zero page views (Y_{it} = 0), and when a visit occurs we observe a non-zero value (Y_{it} > 0). Following the GLMM framework (Min and Agresti 2005), we define p_{it} as the probability that individual i visits the website on day t, we adopt a logistic regression specification for this model component such that:

$$P(Y_{it} > 0) = p_{it}\ , \tag{1}$$

$$P(Y_{it} = 0) = 1 - p_{it}\ , \text{ and} \tag{2}$$

$$p_{it} = \frac{\exp(V_{it})}{1+\exp(V_{it})}\,, \tag{3}$$

where V_{it} represents the value associated with individual i visiting the news website at time t. Given a site visit, we model the probability that visitor i views y_{it} (non-zero) pages on day t

using a truncated Poisson process, such that:

$$P(Y_{it} = y_{it}) = p_{it} \cdot \frac{\exp(\mu_{it})\mu_{it}/y_{it}!}{1-\exp(-\mu_{it})} . \tag{4}$$

The term μ_{it} is the positive, individual, and time-dependent parameter of the Poisson distribution and corresponds to the mean of the (untruncated) Poisson distribution. To complete the model specification we model V_{it} and μ_{it} as a function of linear predictors assuming a logit and a log link-function, respectively[4]. Consider:

$$V_{it} = \log\left(\frac{p_{it}}{1-p_{it}}\right) = \beta Z_{it} + \theta_i Q_{it} + u_{it} \text{ , and} \tag{5}$$

$$\log(\mu_{it}) = \alpha X_{it} + \omega_i W_{it} + v_{it} \text{ ,} \tag{6}$$

where Z_{it} and X_{it} are vectors of covariates and the terms β and α are the associated parameters, common across individuals; similarly, Q_{it} and W_{it} are vectors of covariates (with Q and W including a column of ones for the intercept) and the terms θ_i and ω_i correspond to the associated individual-specific parameters. Note that the variables Z_{it} and X_{it} exert the same effect on all individuals (they might be different across individuals but are associated with a common parameter across individuals), whereas we allow the variables Q_{it} and W_{it} to elicit an individual-specific response (through individual-specific parameters). We further assume that the individual level parameters θ_i and ω_i are normally distributed random effects, such that:

$$\omega_i \sim N\ (\Delta_\omega, \Sigma_\omega), \tag{7}$$

$$\theta_i \sim N\ (\Delta_\theta, \Sigma_\theta) \tag{8}$$

Finally, we assume that u_{it} and v_{it} are jointly normally distributed random effects with zero mean and variance-covariance matrix Ω, such that:

$$\begin{bmatrix} u_{it} \\ v_{it} \end{bmatrix} \sim N\ (0, \Omega) \text{ , with} \tag{9}$$

[4] We note that, by coupling normally distributed random effects and the logit and log link-functions, this GLMM specification is part of the logistic/normal and the Poisson/normal family of models often used in the latent trait literature.

$$\Omega = \begin{pmatrix} \sigma_u^2 & \rho\sigma_v\sigma_u \\ \rho\sigma_v\sigma_u & \sigma_v^2 \end{pmatrix}, \tag{10}$$

where σ_u^2 and σ_v^2 are the variances of u_{it} and v_{it}, respectively; ρ is the correlation of u_{it} and v_{it} ($0 \leq \rho \leq 1$); σ_u and σ_v correspond to the standard deviations of the random effects.

There are several important features of this model to note. First, σ_u^2, the variance of the random factor associated with the value function of the website visit decision is not identified and we set this variance to one for identification purposes (Hadfield 2010). Second, we allow the two model components to correlate. A positive correlation would mean that an idiosyncratic positive shock that increases an individual's likelihood of site visitation simultaneously increases the expected number of page views requested on that day by that same individual (and vice-versa). We believe this specification is parsimonious yet flexible enough to accommodate the complex set of phenomena we observe. We can then define the log-likelihood as:

$$LL = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \int_{\vartheta} (1 - p_{it})^{1-d_{it}} \left[p_{it} g(Y_{it} = y_{it})\right]^{d_{it}} . d\vartheta , \tag{11}$$

where d_{it} takes the value 1 if visitor i visits the website on day t and zero otherwise, and $g(.)$ represents the zero-truncated Poisson distribution. Each individual's likelihood contribution is the product of the probability of crossing the hurdle and then selecting y_{it} pages to view (when a visit occurs) times the probability of no visit taking place (when no visit occurs).

Variable Definition

To study if information from Facebook can help prediction of news consumption at an external website, and to determine the impact of individual preferences, homophily, and social influence in determining a link between what people do on Facebook and at an third-party sites, we add a variety of observed and unobserved effects to the model. We make the individual's visitation and page decisions a function of (1) the individual's observed activity while on the social network site, (2) individual characteristics, (3) the individual's previous browsing activity on the news site, (4) the activity of online friends on the news site, (5) homophily corrections, and (6) exogenous temporal factors influencing interest in news.

We account for the individual's observed activity at the online social network using a Like Activity dummy that takes the value of one if the user was actively liking Facebook pages in a given day, and zero otherwise (see Figure 1 for an example of Facebook page likes). This variable is a good proxy for more general measures of user engagement, availability of resources (e.g., time), and interest in being online. Liking pages (which add to the personal narrative of the user), albeit imperfect, reveals that users had the time to visit their Facebook accounts and were not passively absorbing information but instead actively building their Facebook profile. We also tested for alternative specification including variables that measured the daily intensity of engagement such as the total number of pages liked in a given *day* (alternative specifications did not change the results and did not improve fit).

Because previous research has identified browsing history as an important predictor of website navigation (e.g., Moe and Fader 2004), we also account for individual browsing experiences. We find that the type of navigation and content consumed at the site differs significantly across visitors. For example, some site visitors visited the home page in search for content, others directly visited news articles without going through the home page (these visitors probably received links or viewed the recommendations of articles while online). We will consider visits with home page views as evidence of being in a more explorative mode, and visits without home page views as indicative of a directed browsing mode (this individual-specific measure changes over time). Our data confirms this pattern: visits without home page views are on average shallower than visits with home page views with a significant 12-page difference, confirming the directed vs. explorative patterns. In addition, to account for individual preferences in news consumption we created dummy variables for the content categories visited by users and include a lagged specification. We also create an inter-visit time variable that allows us to predict future visits based on the time since an individual's last visit (recency effects).

Because older readers may consume more pages, younger readers could visit news sites more frequently, and people with more friends might be exposed to more content on their newsfeed, we also tested for the inclusion of individual specific variables that do not change over time in addition to the previous dynamic measures. We tested for the inclusion of age, gender, number of Facebook friends and number of total page likes on Facebook. We further note that a random effects specification also accounts for unobserved individual heterogeneity.

To account for the potential effects of social influence, we measure when a focal user's Facebook friends are active on the news website. We built a Friends' Activity Dummy variable that takes the value one if at least one friend is active on a given day at the news site, and the value zero otherwise. We tested for alternative specifications (e.g., total number of pages viewed by a user's friends on a given day at the news site) but the activity dummy provided the best fit. Finally, we include daily dummy variables to account for exogenous time varying factors that could influence news consumption.[5]

Table 1 provides a description of each one of the variables we considered. Table 2 provides the summary statistics of Facebook profile information for the 1,562 users in our sample. Finally, Table 3 provides the summary statistics of the browsing and like data for the users included in the estimation sample. The majority of visitors in our sample (78%) are male and the average age is 39 years old (mapping well with the readership audience of the newspaper). Visitors to the site had on average 424 friends, with a minimum of five friends and a maximum of 1,000, and liked an average 178 Facebook pages. Visitors viewed on average five pages per day but not all visitors visited the site daily. On average visitors returned after 3.7 days and made about 0.6 visits in a given day, though there were instances in which visitors

[5] We also tested for the inclusion of a variable that accounts for the average page views by registered users who used emails to register with the website We obtained average daily page views for the sample of users who registered using their email accounts. However, including daily dummies in our model provided a better fit.

visited the site 11 times (compared to the original sample we are missing a very small percentage of users who rarely visited the website). Users visited the home page about twice a day and requested 1.09 Entertainment pages. In addition, visitors were active liking Facebook pages 13% of the time (there was an average gap of 2.39 days in between like activities during the last three weeks of March) and in about 15% of the time at least one friend of each user visited the news site.

Homophily, Endogeneity, and Correlated Unobservables

In a model of influence it is important to separate homophily from social influence, and to consider and to model external factors influencing both social ties and behavior (e.g. Cohen-Cole and Fletcher 2008; Lyons 2011; Shalizi and Thomas 2011). External factors include endogeneity (simultaneity) and correlated unobservables (Hartmann et al. 2008). We now outline our empirical strategy to address each of these challenges.

To account for and assess the impact of homophily, we follow previous research and adopt a latent space approach that exploits the social network linkages and other user-specific information including demographics, interests, network degree, and pages liked (see Ansari, Koenigsberg and Stahl 2011). Considering the potentially missing network edges in our specific context, a latent space approach is more appropriate to avoid biases than other approaches such as dynamic propensity score matching (e.g. Aral, Muchnik, and Sundararajan 2009; Eckles 2015). In addition, as demonstrated by Davin, Gupta, and Piskorski (2014), the latent space coordinates are indicative of the unobservable traits that would drive individuals to be close together, hence are able to correct for homophily in models like the one we estimate.

The approach assumes that connections between individuals are the result of their relative distances in a K-dimensional latent space (the closer individuals are, the more likely they are to be connected). Latent space is transitive, that is, individuals who have common friends (even if not connected directly) will automatically be closer to each other in the network compared to those individuals who do not have any common linkages. To implement this approach we create an $N \times N$ matrix, $M = [m_{ij}]$, of dyadic ties m_{ij} takes the value one if individuals *i* and *j* are friends on Facebook, and zero otherwise. In our empirical application $N = 1{,}562$ (the number of individuals in our sample). We allowed for undirected ties so that if *i* and *j* were identified as friends, *i* could influence *j* and vice-versa. To determine the latent space we essentially estimate the probability that observable factors and unobservable latent traits explain a link between two individuals.

We formally model $\Pr(m_{ij}|\tau, C_i, k_i, k_j)$ (probability of a dyadic tie between individual *i* and *j*) as a logistic regression model in which the probability of a tie depends on the distance between individual *i* and *j* in a social latent space such that:

$$logit\left(m_{ij} = 1 \,\middle|\, \tau, C_i, l_i, l_j\right) = \tau c_{ij} - |k_i - k_j|, \tag{12}$$

where τ is the estimated coefficient associated with c_{ij}, and c_{ij} ($i = 1, \ldots N$; $j = 1, \ldots N$) a vector of covariates that help explain the proximity between *i* and *j*; k_i and k_j are K-dimensional vectors of individual positions in the K-dimensional latent space. To estimate the latent space we minimized the Euclidean distance $|k_i - k_j|$ between individuals using the MCMC algorithm in the R package "latentnet" (Krivitsky and Handcock 2008, 2009).

In our empirical application, c_{ij} corresponds to the distance between individuals *i* and *j*, computed using observable information. To determine how distant or close individuals are, we first create Facebook clusters using a k-means algorithm applied to data from 12,700 Facebook profiles. We used information on (1) user demographics (i.e., the age and gender stated by Facebook users), (2) social network characteristics for each user (i.e., the size of the network and the number of total Facebook likes), and (3) the detailed information on the categories of pages liked on Facebook by each user.[6] Then we assigned each of the 1,562 users in our sample

[6] We note that each visitor is associated to a set of Facebook pages they have liked (we have access to all the pages liked by each visitor, including those pages liked before our observation period) and each page is assigned to a specific category. We condensed the more than 200 original page categories from our Facebook data into a more

to their nearest cluster and set c_{ij} for these users to be the distance between individuals i and j when they are assigned to different clusters, and zero when they are assigned to the same cluster. We tested for a specification of up to eight clusters, and found that a six-cluster specification was adequate to account for the similarity and differences in Facebook users. We also tested for alternative specifications for c_{ij} in which we used directly the variables observed for each of the 1,562 Facebook users in our sample without first building clusters. However, the cluster specification provided the best results, perhaps because cluster formation relied on a larger sample of users and provided more stable similarity measures). We note that our conclusions regarding social influence do not change with these alternative specifications and we believe this provides an additional robustness check of our findings.

After estimating the latent space and obtaining the K-dimensional location vectors for each individual, we follow Davin et al. (2014) and directly add the latent space coordinates as variables to our random effects hurdle model. We test for the number of latent space dimensions to include by re-estimating the latent space assuming a different number of dimensions and adding the estimated vectors of coordinates to the hurdle Poison model. We tested for up to seven-dimensions and found that using a latent space of six-dimensions provided the best fit for the hurdle model (see Figure 2).

Beyond the latent space approach, the use of panel data mitigates some of the difficulties of separating homophily and social influence, as actions have a temporal structure. In addition, just as in Nair et al. (2010), the network is external to the news site and its definition does not depend on the use and visitation of the news site (instead, the social network we consider is defined by Facebook connections). However, simultaneity might still create challenges in adequately identifying social influence. To disentangle whether the focal agent's behavior influenced other members of the group, or whether the activity of other

manageable set, comprising categories like Entertainment, Sports, Politics, Local Interests, Economy, and Travel. These categories are not the same as the ones used to categorize the sections of the news website.

members influenced the focal agent's actions, we follow previous research and adopt a lagged specification: we estimate our model using a lagged Friend Activity variable.[7]

Finally, as we stated previously, we have included daily dummy variables (with and without individual specific effects) and have tested for the inclusion of a variable that accounts for the average page views by registered users who are not part of our Facebook sample. This allows us to account for fluctuations of news interest over time that are common across individuals (and to control for possible correlated unobservables). Our model specification also follows the recommendations of previous research to avoid confounding social influence with other factors that may be driving networks members to exhibit similar consumption patterns (Nair et al. 2010; Shalizi and Thomas 2011). For example, we include a random effects specification and add to the model the total number of friends, total number of page likes, age, and gender, to account for observed similarity of users.[8]

Model Estimation

We use a hierarchical Bayesian approach to estimate the proposed model. We adopted conjugated priors for all parameters whenever possible including an inverse Wishart as prior for variance-covariance matrix of the joint normal distribution of all random effects, and we allow the covariance of the error terms to be unstructured (for a discussion of Bayesian estimation of such correlated random effect Poisson Hurdle models using a GLMM specification see Draper 2008). We used the first 50,000 iterations for burn-in, checked for chain convergence and ran an additional 25,000 iterations to compute posterior distributions..

[7] We also tested an exclusion restrictions approach (Angrist 2001), similar to the typical instrumental variable methodology applied to continuous variables. Because the best performing model included the model with the lagged Friend Activity variable, we report the results using this approach. Again, the results obtained on social influence did not change substantially across alternative specifications (results available upon request).

[8] The variables age and total number of likes, for instance, can proxy the total time spent online that can affect engagement with social networks as well as news sites. We also tested for the inclusion of a "Lag Friend Category Match" variable we created. This variable takes the value of one if an individual sees the same category on day t as at least one of his/her Facebook friends saw on day $t-1$, and zero otherwise (see Table 1). It indicates whether, in any given day, a user is seeing the same categories previously seen by his/her Facebook friend, and accounts for similarity of tastes and interests between a focal user and his/her friends without suffering from endogeneity problems. The variable did not improve fit and hence not included.

To determine the final model specification, we tested for the inclusion of variables sequentially and compared model performance in-sample using the deviance information criterion (DIC).[9]

Results

Tables 4 and 5 provide the fit for the final model specifications, in and out-of sample. The models include a “No Facebook Model” (baseline model with browsing variables, random effects and additional daily controls), an “Own Facebook Activity Model” (that adds the users’ Facebook like activity and Facebook static variables as additional controls to the baseline model), and a “Own and Friend Activity Model” (that further adds the activity of Facebook friends while at the news site; “Own Facebook Activity Model” and the “Own and Friend Activity Model” are compared *with* a latent space corrections). Tables 6 and 7 present the detailed posterior means and the 95% probability intervals for the hierarchical parameters for all models and all variables we retained after extensive tests.

We use the DIC to compare models in-sample. To compare the models out-of-sample, we re-estimated the four alternative model after removing the last two days of data from the original sample (the two last days of March 2012). We then built a holdout sample that included the page views and visit information made by 1,559 site visitors during the excluded days and for each model we predicted the number of page views and visit likelihood.[10] The “Own and Friend Activity Model” with a six-dimensional latent space specification is the best fitting model in- and out-of-sample. The in-sample DIC of 109,927.0 is the best across all models and in holdout it is the model the lowest mean squared error (MSE) and lowest

[9] We tested the two parts of the model separately as well as simultaneously and kept only those variables that improved model fit. For all variables that might have a non-linear effect we also considered logarithmic and quadratic functional forms. We present as our final results the best fitting specification for each variable (e.g., the quadratic specification seemed to best describe the effect of “time since last visit” on visit behavior). Though there are alternative ways of selecting the final model specification (O'Hara and Sillanpää 2009) the approach we adopt has been widely used in marketing.

[10] We removed three visitors with no activity prior to the last two days in March. After their visits to the website during the first week of March (the initialization period) these three visitors returned to the website only on the two last days used for holdout. We can predict page views and site visitation for these users using the population means of the parameters, though the overall performance results do not present any significant variation from the ones presented. The proposed final model predicts an average 2.8 page views per visit in holdout (the actual average number of page views is about 3.0).

normalized mean square deviation for page views, and with the lowest mean squared error, the highest hit-rates, and the greatest improvement in the lift chart for the visit decision (See Figure 3).[11] We also find that homophily is an important force driving the co-movements of friend's behaviors as latent space corrections improve fit (the DIC of the final specification without latent space controls is 110,238.4), and that the two dependent variables are positively correlated (estimated covariance is 0.83 and $p < 0.001$), which means that the our joint estimation is preferred.

We note that we are comparing our full model with very strong alternative specifications (e.g., the "No Facebook Model" includes random effects, temporal effects, and previous browsing history as predictors) and that the news site has currently over 1.2 million Facebook fans making the conservative 3% improvement in hit rate a very significant result in terms of revenues. Next, we will study the results from this model to understand the role of the different factors we outlined previously.

Role of Individual Preferences in Driving Visits and Page views

The detailed results we present in Tables 6 and 7 confirm the importance of previous browsing behavior as predictors of visit and page view decisions. Individuals who allow a long time period to pass in between their visits are less likely to visit the site, indicating the importance of remaining salient in users' minds and be part of their daily choice for news consumption. In addition, knowing that users are reading specialized content, and not simply browsing for headlines, helps predict future visits to the site (visitors who read categories like Sports, Local News, and check TV related content are more likely to return). This further supports existing theories of involvement and selective exposure (see Dutta and Bergman 2004). Similar to the

[11] To create the lift charts we sorted the holdout observations by predicted visit probabilities. We then took 10% of all (holdout) observations with the highest predicted probability and computed the percentage of actual visits associated to these observations. We repeat this procedure for 20% of the observations, 30%, and so on. We then plotted the fraction of visits that each model would have been able to capture at different targeting percentages.

findings of previous research, demographic variables did not improve fit (the random effects specification seems to account for the individual level heterogeneity).

User Actions at the Social Network and their News Consumptions at External News Sites

The effect of Like Activity is positive and significant for site visits and page views. The size of the effects is also substantial: when the focal user is actively engaged on Facebook, she is 26% more likely to visit the news site and, given a site visit, she reads 13% more articles than when she is not active on Facebook (Table 8 reports the effect sizes). The net effect on pages requested is an average increase of 41% in pages requested from the external news site on those days users also engage with Facebook.

These results seem to support that Like Activity is a proxy for interest in content consumption, and for time availability by the focal user, which in turn will be associated with higher probabilities of visits to the news site. Personal motives for news consumption such as mood management, entertainment and passing time (Rubin 1993) could not only drive users to visit both social networks *and* news sites, but the availability of time and the desire to engage with content, would likely result in deeper browsing at the news site alongside engagement with Facebook. Our results provide evidence that support this contention. Furthermore, Wilcox and Stephen (2012) show that Facebook can lower people’s self-control. It is therefore plausible that users actively engaged in browsing content on Facebook would be prone to greater content seeking once at an external website, in which case we would also expect to observe changes in browsing behavior at the news site with deeper browsing when users are engaged with the online social network.

Friends’ Actions at the News Site and their Effect on Site Visits and News Consumption

The Friend Activity dummy is significant and positive for the visit decision. We further simulated the impact of friends’ actions and we estimated that when friends are active at the news site, focal users are about 12.4% more likely to visit that same news site, an effect that is also managerially meaningful given its size (see Table 9 for more details). In addition, we compare the results with and without the correction for homophily (i.e., with and without latent space corrections). By comparing the impact of Facebook friends' news consumption on the focal user we find that 79% of the 12.4% increase in visit likelihood (i.e., a 9.82% increase) is due to social influence and 21% of that increase (i.e., 2.64%) is due to homophily.

We also consider the impact of friends' actions on page views (see Table 7 for the estimated parameters). We find that friends' actions have a negative and significant main effect on page views. This *negative* effect of friends actions on page views seems to be due to social influence (Table 8) which further supports the notion that users directed browsing as a result of friends' actions is likely due to the focal users relying on their friends for content filtering. As Levinson (2009, p. 122) states "Facebook and Twitter connections may behave as a real-time knowledge base resource." Had homophily effects been stronger than social influence, we might have found the opposite net effect of Friend's Activity (see Table 9).

The fact that the behavior of social network peers can also influence the amount of content consumed once at the site (we observed focused browsing) is an important finding because it reveals that social integrative motives might indeed be at the base of news consumption online. More importantly, we find this result holds despite the control for number of Facebook friends, as users with more friends are likely to be exposed to more links.

User's personal preferences as a Moderator of the role of Facebook Friends

Previous research suggests that users' browsing behavior could reveal social and intrinsically personal motivations. Browsing mode (explorative vs. directed) has been previously studied in the literature in the context of e-commerce (e.g., Janiszewski 1998; De Nie 2012) and previous work has noted that consumers in an explorative mode spend more time before making a purchase than those in a directed mode. Similarly, in our context, we contend that visitors in an explorative mode are higher in need for cognition, are expected to spend more time visiting the news website, try new links, and browse more extensively for various topics. This leads them to visit the home page of the website (instead of only reading the articles shared by friends and discussed by friends) and use it as a navigating tool for further content discovery.

As a result, it is likely that the browsing mode or need for cognition will moderate peer influence. We tested for these effects by including Home Page dummy as a main effect and as an interaction with Friend Activity. First, we note that the estimated main effect of home page views on the number of articles read is positive and significant, providing further evidence that the Home Page Dummy is a good metric, albeit imperfect, for when users are in an explorative browsing mode. Second, we simulated the impact on page views of Friends' Activity depending on whether users visited the home page or not. We find that Friends' Actions have a mixed effect depending on whether users are in a directed versus explorative browsing mode. From Table 10 we can see that the net effect of Friends' Actions is positive when the focal user is in explorative browsing mode, and negative when in directed browsing mode (i.e., when friends are active and a focal user is in directed mode, that focal user will read fewer articles).

It follows that, for users high in social motivations for news reading, the impact of Friends' Actions is likely to attract them to the website but reduce the amount of content consumed as individuals would rely on friends for information filtering. In contrast, if the motivations for media use are intrinsically personal (e.g., surveillance, personal entertainment, and identity construction; Katz et al 1974), users seeking content for personal needs are more likely to rely on friends as a tool for discovery, not filtering. In this case, news consumption activity by friends could foster the discovery of content. As one of the respondents to a Pew Research survey stated "I believe Facebook is a good way to find out news without actually looking for it," which exemplifies the role of friends as a discovery tool (Mitchell et al 2012b).

Conclusion

Currently, the news publishing industry faces important challenges such as the difficulty in charging for content and declining advertising revenues. Simultaneously there has been a significant increase in the importance of online news within social media outlets, with 63% of Facebook users actively engaging with news on social networks, and an even higher percentage engaging with news passively, that is, reading headlines without clicking on links (Mitchell et al 2014). In recent years, news websites have also incentivized users to engage with their content within social networks and encouraged users to join their Facebook brand pages. In light of these developments, understanding the interplay between social networks and online news providers becomes fundamental for digital marketers.

For struggling news sites, access to users' Facebook profiles provides obvious benefits including the selling of premium-targeted ads. However, there has been limited research investigating other potential benefits. With this study, we aim at addressing this gap in the marketing literature and investigate whether information obtained from social networks can be valuable in describing and predicting traffic and engagement of online news readers, while disentangling the importance of social influence and homophily in the link between the behaviors at both platforms.

For the purpose of this study we rely on data that is readily available to businesses once users register to their websites using Facebook accounts. Our results demonstrate the value of social network data in not only describing behavior of social network users at external news sites but also predicting their behavior both in and out of sample. In addition, our individual level model uncovers that the personal preferences for news browsing along with their own Facebook activity, and the activity of their Facebook friends influence the news consumption behavior at *external* third party news sites. Users seem to return to their preferred content categories and long gaps in visits signal lack of interest in news. The Facebook data further reveals that being active on Facebook correlates positively with visits and page views. Instead, users arriving at the content website in association with friends' actions engage in more directed browsing and view fewer pages (although the net effect of friends' actions on content consumption is positive). More importantly, this result confirms that social networks provide a platform to facilitate connections and content seeking behavior.

Our contribution also lies in borrowing from theories of information seeking from the media literature to assert that *personal* and *social* motivations of social network users are the underlying mechanisms through which social networks (like Facebook) may be of value to news sites. This is a novel perspective thus far in the marketing literature. Considering the controls in our model, we can also conclude that *both* homophily and social influence seem to be playing an important role in news reading. We also note that our findings are not simply the result of newsfeed effects (e.g., some users receiving updates and others not receiving them) as *all* users in our sample receive newsfeeds updates from the news site. The news site did not engage in Facebook advertising during this period and hence targeting of specific users cannot explain our results either.

The patterns of social influence we identify suggest that in a world dominated by information overload friends might serve a major role in directing content consumption. It is not surprising then that one of the important features of several social networking systems is to make salient what one's friends are doing, not just what a diffuse group of anonymous peers have to say (Lohner 2012). News sites can attract larger audiences by better targeting the most influential users (taking advantage of the social influence effect) or targeting users with the characteristics that make them potential avid readers (taking advantage of the homophily effect). In our context, being influential is associated with being active at the news site (not only at the online social network). Web site managers need to realize that direct referrals may not reveal the complete mechanisms. In this dynamic environment managers need to encourage more links with social media and make news items salient so that users may engage with news stories. Also, with the growing popularity of ad-blockers, the role of sponsored and native advertising at social networks is likely to generate more revenues as direct ad revenues are declining (Marshall, 2015).

Hence, a key take away from our research is the benefit to online content providers of developing groups within large social networks (e.g., through specialized pages on Facebook). Group membership not only allows the gathering of useful data fosters connections that can result in joint consumption and engagement with the news sites. How these results generalize to other online social media platforms might depend on their dynamics, as these can be different. For example, one is more likely to receive news information from family and friends on Facebook than on Twitter, as news sources on Twitter are more likely to include news organizations and experts (Mitchell et al. 2012a). In addition, there is a preference for personal and “authentic” sharing instead of content automatically produced by newsreader apps on Facebook (Herrman 2012). Future research could explore other platforms.

Despite the relevance of our findings, our research suffers from limitations. The social media data we considered does not include all possible behaviors of users while using the network. By finding a significant effect with limited measures of Facebook engagement is however promising; we believe that access to more detailed actions (e.g., frequency and content of comments) could provide even stronger results. Finally, it is encouraging to note that we were able to uncover the association between Facebook and the mechanisms of social influence amongst such a small sub-sample of members of the news site’s Facebook community that today has over 1.2 million members. We are hopeful our research will give impetus to future research in this area.

References

Anderson, Monica and Andrea Caumont (2014), "How Social Media is Reshaping News," Pew Research Centre, Retrieved from http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/, Accessed on 16th June, 2016.

Angrist, Joshua D. (2001), "Estimation of Limited Dependent Variable Models withDummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics*, 19(1), 1–16.

Ansari, Asim, Oded Koenigsberg, and Florian Stahl (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48(4), 713–728.

Aral, Sinan (2011), "Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30(2), 217–223.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009), "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences*, 106(51), 21544–21549.

Asur, Sitaram, and Bernardo A. Huberman, (*2010)* "Predicting the Future with Social Media." *Web Intelligence and Intelligent Agent Technology (WI-IAT),IEEE/WIC/ACM International Conference on*. Vol. 1. IEEE, 2010.

Atkin, C.K. (1972), "Anticipated communication and mass media information-seeking," *The Public Opinion Quarterly*, 36, 188–199.

Baresch, Brian, Lewis Knight, Dustin Harp, and Carolyn Yaschur (2011), "Friends Who Choose Your News: An Analysis of Content Links on Facebook," *International Symposium on Online Journalism*, 1(2), 1–24.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49(2), 192-205.

Bernoff, Josh and Charlene Li (2008), "Groundswell: Winning in a World Transformed by Social Technologies," Boston: *Harvard Business School Publishing*.

Bollen, Johan, Huina Mao, and Alberto Pepe (2011), "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-economic Phenomena," *ICWSM* 11: 450-453.

Bradbury, Danny (2013), "Effective Social Media Analytics," The Guardian, published 10th June 2013; Accessed: March 15 2016, [https://www.theguardian.com/technology/2013/jun/10/effective-social-media-analytics]

Breslow, Norman E. and David G. Clayton (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88(421), 9–25.

Centola, Damon (2010), “The Spread of Behavior in an Online Social Network Experiment," *Science* 329(5996), 1194-1197

Chevalier, Judith A. and Dina Mayzlin (2006), “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research*, 43(August), 345–54.

Clemons, Eric K., Bin Gu, and Karl Reiner Lang (2002), “Newly Vulnerable Markets in an Age of Pure Information Products: An Analysis of Online Music and Online News,” in *Proceedings of the 35th Annual Hawaii International Conference*, 2949–2958, IEEE.

Cohen-Cole, Ethan and Jason M. Fletcher (2008), “Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic,” *Journal of Health Economics*, 27(5), 1382–1387.

Davin, Joseph P., Sunil Gupta, and Mikolaj Jan Piskorski (2014), “Separating Homophily and Peer Influence with Latent Space,” Working Paper, Harvard Business School.

De Nie, Hubert W. (2012), “What Makes’ em Tick: The Effects of Shopping Mode and Culture on Time and Money Spent Shopping,” Working Paper, Erasmus University.

Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F. Awad (2007), “Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures,” *Journal of Interactive Marketing*, 21(4), 23–45.

De Waal, Ester, and Klaus Schoenbach (2010), “News Sites’ Position in the Mediascape: Uses, Evaluations and Media Displacement Effects Over Time,” *New Media and Society*, 12(3), 477–496.

De Valck, Kristine, Gerrit H. Van Bruggen, and Berend Wierenga (2009), "Virtual Communities: A Marketing Perspective." Decision Support Systems 47(3), 185-203.

Dimmick, John, Yan Chen, and Zhan Li (2004), “Competition Between the Internet and Traditional News Media: The Gratification-Opportunities Niche Dimension,” *Journal of Media Economics*, 17(1), 19–33.

Draper, David (2008), “Bayesian multilevel analysis and MCMC,” In *Handbook of Multilevel Analysis*, Springer New York, 77–139.

Dutta-Bergman, Mohan J. (2004), “Complementarity in consumption of news types across traditional and new media,” *Journal of Broadcasting & Electronic Media*, 48(1), 41–60.

Eckles, Dean (2015), “Identifying Peer Influence Effects in Observational Social Network Data: An Evaluation of Propensity Score Methods,” Working Paper.

E-marketer (2013), “Digital to Account for one in Five Ad Dollars,” http://www.emarketer.com/Article/Digital-Account-One-Five-Ad-Dollars/1009592 [accessed July 27, 2013]

Godes, David and Jose Silva (2012), “Sequential and Temporal Dynamics of Online Opinion,” *Marketing Science*, 31(3), 448-473.

Goel, Sharad, and Daniel G. Goldstein. (2013) "Predicting Individual Behavior with Social Networks." *Marketing Science* 33(1), 82-93.

Golbeck, Jennifer, and Derek Hansen (2011), “Computing Political Preference Among Twitter Followers,” *in Proceeding of Human Factors in Comp. Sys.*

Greene, William (2009), “Models for Count Data with Endogenous Participation,” *Empirical Economics*, 36, 133–173.

Hadfield, Jarrod D. (2010), “MCMC Methods for Multi–Response Generalised Linear Mixed Models: The MCMCglmm R package,” *Journal of Statistical Software*, 33(2), 1–22.

Hallahan, Kirk (1999), “No, Virginia, it’s Not True What they Say about Publicity’s “Implied Third-Party Endorsement” Effect,” *Public Relations Review*, 25(3), 331–350.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar, and Catherine Tucker (2008), “Modeling social interactions: Identification, Empirical Methods and Policy Implications,” *Market. Letters*, 19(3-4) 287-30.

Herrman, John (2012), “Facebook Social Readers Are All Collapsing,” available at http://www.buzzfeed.com/jwherrman/facebook-social-readers-are-all-collapsing.

Hill, Shawndra, Foster Provost, Chris Volinsky (2006), “Network-based Marketing: Identifying Likely Adopters via Consumer Networks,” *Statistical Science* 21 (2) 256- 276.

Janiszewski, Chris (1998), “The influence of Display Characteristics on Visual Exploratory Search Behavior,” *Journal of Consumer Research*, 25(3), 290–301.

Jeon, Doh-Shin and Nikrooz Nasr Esfahani (2012), “News Aggregators and Competition among Newspapers in The Internet,” Technical Report, NET Institute.

Johnson, Thomas J. and Barbara K. Kaye (2004), “Wag the Blog: How Reliance on Traditional Media and the Internet Influence Credibility Perceptions of Weblogs Among Blog Users,” *Journalism & Mass Communication Quarterly*, 81(3), 622–642.

Katz, Elihu, Jay Blumler, and Michael Gurevitch (1974), “The Use of Mass Communication,” Beverly Hills.

Krivitsky, Pavel N. and Mark Handcook (2008), “Fitting Position Latent Cluster Models for Social Networks with Latentnet,” *Journal of Statistical Software*, 25(5).

——— (2008), “Fitting Position Latent Cluster Models for Social Networks with Latentnet,” *Journal of Statistical Software*, 25(5).

Lerman Kristina and Tadd Hogg (2012), "Using Stochastic Models to Describe and Predict Social Dynamics of Web Users," *ACM Transactions on Intelligent Systems and Technology,* 3.4(62).

Levinson, Paul (2009), New new media. Allyn and Bacon.

Libai, Barak, Ruth Bolton, Marnix S. Bügel, Ko De Ruyter, Oliver Götz, Hans Risselada, and Andrew T. Stephen (2010), "Customer-to-Customer Interactions: Broadening the Scope of Word of Mouth Research," *Journal of Service Research*, 13(3), 267-282.

Liu, Tang, J., Gao, H., Hu, X., & Liu, H. (2013), "Exploiting Homophily Effect for Trust Prediction," *In Proceedings of the sixth ACM International Conference on Web Search and Data Mining,* 53-62.

Lohner, Jessica A. (2012), "Facebook, Twitter, and Interactivity: Changes in the Way Society Discusses News," *Doctoral dissertation, Ohio University*.

Lyons, R. (2011), "The spread of evidence-poor medicine via flawed social-network analysis," *Statistics, Politics, and Policy*, 2(1).

Nguyen, An. (2010). Harnessing the potential of online news: Suggestions from a study on the relationship between online news advantages and its post-adoption consequences. *Journalism*, *11*(2), 223-241.

Marshall, Jack (2015), "Growth of Ad-blocking adds to Publishers' Worries", *Wall Street Journal Blogs*, Retrieved from: [http://blogs.wsj.com/cmo/2015/04/09/growth-of-ad-blocking-adds-to-publishers-worries/].

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Min, Yongyi and Alan Agresti (2005), "Random Effect Models for Repeated Measures of Zero-inflated Count Data," *Statistical Modelling,* 5(1), 1–19.

Mitchell, Amy, Mark Jurkowitz and Kenneth Olmstead (2014), "Social, Search and Direct Pathways to News," Pew Research Centre, Retrieved from: [http://www.journalism.org/2014/03/13/social-search-direct/], Accessed [1st July 2015].

Mitchell, Amy, Tom Rosenstiel, and Leah Christian (2012a), "What Facebook and Twitter Mean for News," Technical Report, Pew Research Center.

———, Tom Rosenstiel, and Leah Christian (2012b), "State of the media: Mobile Devices and News Consumption: Some Good Signs for Journalism," Technical Report, Pew Research Center.

Moe, Wendy W. and Peter S. Fader (2004), "Capturing Evolving Visit Behavior in Clickstream Data," *Journal of Interactive Marketing*, 18(1), 5-19.

Nair, Harikesh S., Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47(5), 883–895.

O'Hara, R.B. and M.J. Sillanpää (2009), "A Review of Bayesian Variable Selection Methods: What, How and Which," *Bayesian Analysis*, 4(1), 85–118.

Ridout, Martin, Clarice GB Demétrio, and John Hinde (1998), "Models for Count Data with Many Zeros," in proceedings of *the XIXth International Biometric Conference*, 19, 179–192.

Rubin, Alan M. (1993), "Audience Activity and Media Use," *Communication Monographs*, 60, 98–105.

Rui, Huaxia, Yizao Liu, and Andrew Whinston (2013), "Whose and What Chatter Matters? The Effect of Tweets on Movie Sales." *Decision Support Systems* 55(4), 863-870.

Shalizi, Cosma Rohilla, and Andrew C. Thomas (2011), "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods Research*, 40(2), 211–239.

Shimshoni, Yair, Niv Efron, and Yossi Matias (2009), " On the Predictability of Search Trends," *Google Research Blog*

Shirky, Clay (2008). "Here Comes Everybody: The Power of Organizing without Organizations," Penguin.

Sismeiro, Catarina and Ammara Mahmood (2016), "Displacement vs. Promotional Effects: Role of Social Networks in News Consumption," *Working Paper*.

Sismeiro, Catarina and Randolph E. Bucklin (2004), "Modeling Purchase Behavior at an E-Commerce Website: A Task Completion Approach," *Journal of Marketing Research*, 41(August), 306–323.

Stephen, Andrew T., and Jeff Galak (2012) "The Effects of Traditional and Social Earned Media on Sales: A Study of a Microlending Marketplace," *Journal of Marketing Research*, 49(5), 624–639.

Tewksbury, David (2003), "What do Americans Really Want to Know? Tracking the Behavior of News Readers on the Internet," *Journal of Communication*, *53*(4), 694–710.

Thorson, Emily (2008), "Changing Patterns of News Consumption and Participation," *Information, Communication & Society*, 11(4), 473–489.

Toubia, Olivier and Andrew T. Stephen (2012), “Intrinsic versus Image-Related Utility in Social Media: Why do People Contribute Content to Twitter?” Working Paper.

Tumasjan, Andranik, Timm O Sprenger, Philipp G. Sandner and Isabell Welpe (2010), "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment,” *Proceedings of the Fourth International Conference on Weblogs and Social Media*, ICWSM 2010, Washington, DC, USA.

Tsfati, Yariv and Joseph N. Cappella (2005), “Why Do People Watch News They Do Not Trust? The Need for Cognition as a Moderator in the Association Between news Media Skepticism and Exposure,” *Media Psychology*, 7(3), 251–271.

Van den Bulte, C. and S. Stremersch. (2004), “Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test,” *Marketing Science*, *23*(4), 530–544.

Walther, Joseph B., Brandon Van Der Heide, Sang-Yeon Kim, David Westerman, and Stephanie Tom Tong (2008), “The role of friends’ appearance and behavior on evaluations of individuals on Facebook: Are we known by the company we keep?” *Human Communication Research*, 34(1), 28–49.

Watts, Duncan J. and Peter Sheridan Dodds (2007), “Influentials, Networks, and Public Opinion Formation,” *Journal of Consumer Research*, *34*(4), 441–458.

Wilcox, Keith and Andrew T. Stephen (2013), “Are Close Friends the Enemy? Online Social Networks, Self-esteem, and Self-control,” *Journal of Consumer Research*, *40*(1), 90–103.

Wright, Charles R. (1960), “Functional Analysis and Mass Communication,” *Public Opinion Quarterly*, 24(4), 605–620.

Zhu, Feng, and Xiaoquan Zhang (2010), “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics." *Journal of Marketing* 74(2), 133-148.

Table 1: Variable Description
(Variables used, or tested for, in the main model specification for Study 1)

Own Browsing Variables	
Page Views	Total number of pages viewed at the news website in given day by each user (it takes the value zero if a user does not visit the website on a specific day)
Site Visits	Total number of site visits a user makes to the news website on a given day (following previous research, a page view is assumed to start a new site visit after the user is idle for at least 30 minutes)
Lag Page Views (by Category)	Number of pages viewed (by category) during a user's last visit to the site; we will have five variables, one for each of the main categories (Local News, Sports, Entertainment, Politics, and Other News)
Inter Visit Time	Number of days since a user's last visit to the focal website (this variable measures how many days have passed since a user's last visit to the focal website)
Home Page Dummy	Dummy variable that takes the value one if a user visits the home page, and zero otherwise
Friend's Browsing Variables	
Friend Browsing	Dummy variable that takes the value one if a Facebook friend visited the news website on a specific day, and zero otherwise (i.e., for each user and for each day, if at least one Facebook friend is active on the focal news website, this variable takes the value one; if no friend is active, it takes the value zero)
Own Facebook Activity Variables	
Like Activity	Dummy variable that takes the value one if a user liked a page on Facebook on a given day, and zero otherwise (i.e., if a user is active liking pages on Facebook in a given day, this indicator variable takes the value of one and zero if the user is not active on Facebook that day)
Own Facebook Profile Variables	
Age	Age (in years) of the user as the user self-reports on his/her Facebook page
Gender	Dummy variable that takes the value of one if the user reports to be a female on her Facebook page
Number of Likes	Total number of Facebook pages a user liked; these are page likes that are visible on a user's profile page
Number of Friends	Total number of Facebook friends for each user
Additional Control Variables	
Average Page Views by Registered Users	Average number of pages requested from the news website by users who registered with the website using their email account (this variable does not include the activity of the users in our sample, who registered using their Facebook accounts)
Daily Dummies	Daily dummy variables that take the value one for a given day and zero otherwise (we created 22 daily dummies to account for daily specific effects; the estimation period comprises 23 days)

Table 2: Summary Statistics of Facebook Profile Information
(measured at the end of the estimation period for the 1,562 Facebook users of our sample)

	Mean	Standard Deviation	Minimum	Maximum
Gender	0.22	0.41	0	1
Age	38.61	13.17	16	80
Total Number of Likes	178.29	137.93	0	551
Total Number of Friends	424.36	293.45	5	1,000

Table 3: Summary Statistics of Browsing and Like Data

	Mean	Standard Deviation	Minimum	Maximum
Site Visits	0.59	0.83	0	11
Inter Visit Time (in days)	3.72	4.05	1	28
Home Page Dummy	0.40	0.49	0	1
Page Views	4.98	13.20	0	295
Page Views by Category				
Home Page	1.97	7.49	0	295
Entertainment	1.09	4.34	0	106
Local News	0.38	1.84	0	146
Sports	0.37	3.21	0	162
Politics	0.12	0.63	0	29
Other News	0.51	3.17	0	181
Friends' Browsing at the News Site	0.15	0.35	0	1
Facebook Like Activity	0.13	0.34	0	1

Note: All variables are user-specific and daily variables and the summary statistics are computed across the 35,926 daily observations generated by 1,562 users (23 observations per user); the mean page views, given a site visit, i.e., considering only non-zero values, is 11.29 with a standard deviation of 17.98 and is computed over 15,864 observations.

Table 4: In Sample Model Fit

Model Description	DIC
Browsing Only (browsing and controls variables only)	110,527.4
Own Facebook Activity (adds Facebook static variables and Facebook Like variable)	110,291.9
Own and Friend Activity (adds friends' actions)	110,238.4
Own and Friend Activity with Latent Space Correction (adds latent space correction to friends' actions)	**109,927.0**
Own and Friend Activity with Latent Space Correction without Home Page Interaction (removed the home page interaction from previous model)	110,304.7

Table 5: Out-of-Sample Predictive Performance

Model	Mean Squared Error	NRMSD	Page views	Hit Rates*
Browsing Only	48.8	3.6%	2.8	78.9%
Own-Facebook	49.3	3.6%	2.7	80.1%
Full-Facebook	48.3	3.5%	2.9	81.6%

** The cut-off probability for the hit rates was set at 0.5

Table 6: Correlated Random Effects Poisson Hurdle Model - Modeling Visit Decision

Variable	Browsing Only	Own FB Activity	Own and Friend Activity †
Random Effects Variables			
Intercept	-0.571***	-0.609***	-0.597
	[-0.792, -0.312]	[-1.088, -0.189]	[-1.386, 0.420]
Lag Friend Activity			0.167**
			[0.165, 0.315]
Like Activity		0.418***	0.399***
		[0.313, 0.532]	[0.294,0.516]
Inter Visit Time	-0.714***	-0.727***	-0.669***
	[-0.766, -0.667]	[-0.767, -0.681]	[-0.703, -0.630]
Inter Visit Time Squared	0.041***	0.042***	0.038***
	[0.036,0.045]	[0.038, 0.045]	[0.035, 0.041]
Lag Page Views for Local News	0.144***	0.129***	0.141***
	[0.086, 0.199]	[0.086, 0.171]	[0.097, 0.175]
Lag Page Views for Sports	0.085***	0.085***	0.106***
	[0.044, 0.123]	[0.052, 0.114]	[0.067, 0.158]
Lag Page Views for Entertainment	0.038***	0.038***	0.037***
	[0.023, 0.051]	[0.028, 0.051]	[0.027, 0.045]
Lag Page Views for Other News	0.034***	0.058***	0.032***
	[0.011, 0.063]	[0.034, 0.087]	[0.015, 0.050]
Lag Page Views for Politics	0.255***	0.250***	0.189***
	[0.119,0.365]	[0.139, 0.359]	[0.067, 0.295]
Fixed Effect Variables			
Number of Friends††		-0.011	-0.185
		[-0.047, 0.026]	[-0.895, 0.409]
Number of Likes††		0.007	0.069
		[-0.060, 1.000]	[-0.992, 1.050]
Age		0.002	0.001
		[-0.006, 0.009]	[-0.016, 0.015]
Gender		-0.212	-0.223*
		[-0.462, 0.029]	[-0.467, -0.009]

Note: We report posterior means and the 95% probability intervals; *** means Bayesian p-values of less than 1%, ** less than 5% and * less than 10%. For all those parameters that are individual specific (i.e., the random effects group), we only report the estimated population averages. All models include Daily Dummy Variables with associated random effects (parameters available from the authors upon request).

† Own and Friend Activity model includes correction for latent space coordinates (parameters available from the authors upon request).

†† Parameter results scaled by 1,000 for better readability.

Table 7: Correlated Random Effects Poisson Hurdle Model - Modeling Page Views

	Browsing Only	Own FB Activity	Own and Friend Activity†
Random Effects Variables			
Intercept	0.667*** [0.575, 0.764]	0.585*** [0.427, 0.770]	0.769*** [0.558, 0.962]
Lag Friend Activity			-0.254*** [-0.358, -0.168]
Lag Friend Activity * Home Page			0.241*** [0.138, 0.343]
Like Activity		0.118*** [0.069, 0.162]	0.117*** [0.071, 0.156]
Home Page	1.216*** [1.136, 1.309]	1.262*** [1.179, 1.334]	1.133*** [1.067, 1.278]
Inter visit time	0.012 [-0.009, 0.033]	0.017 [-0.001, 0.043]	-0.013 [-0.009, 0.030]
Inter visit time square	-0.004*** [-0.006, -0.002]	-0.004*** [0.006, -0.002]	-0.004** [-0.005, -0.002]
Lag Page Views for Local News	0.019*** [0.010, 0.029]	0.018*** [0.009, 0.028]	0.018*** [0.008, 0.027]
Lag Page Views for Sports	0.011*** [0.005, 0.019]	0.011*** [0.005, 0.018]	0.011** [0.005, 0.018]
Lag Page Views for Entertainment	0.0003 [-0.003, 0.003]	-0.0001 [-0.004, 0.003]	0.0004 [-0.003, 0.004]
Lag Page Views for Politics	0.025** [0.005, 0.048]	0.023** [0.002, 0.005]	0.026** [0.006, 0.047]
Lag Page Views for Other News	0.003 [-0.004, 0.009]	0.003 [-0.003, 0.009]	0.003 [-0.005, 0.009]
Fixed Effect Variables			
Number of Friends††		0.002 [-0.016, 0.016]	0.016 [-0.216, 0.143]
Number of Likes††		-0.020 [-0.054, 0.015]	-0.283 [-0.627, 0.055]
Age		0.002 [-0.001, 0.005]	0.001 [-0.003, 0.004]
Gender		-0.019 [-0.131, 0.069]	-0.023 [-0.137, 0.072]

Note: We report posterior means and the 95% probability intervals; *** means Bayesian p-values of less than 1%, and ** less than 5% and * less than 10%. For all those parameters that are individual specific (i.e., the random effects), we only report the estimated population averages. All models include Daily Dummy Variables with associated random effects (parameters available from the authors upon request).

† Own and Friend Activity model includes correction for latent space coordinates (parameters available from the authors upon request).

†† Parameter results scaled by 1,000 for better readability.

Table 8: Effect of Facebook Engagement (Like Activity) on Page Views and Site Visits*

	Overall Change due to Like Activity
Visit Probability	25.58%
Conditional Page Views (given a site visit)	12.46%
Unconditional Page Views	41.22%

*We simulated the impact of Like Activity using the six-dimensional latent space Own and Friend Activity Model.

Table 9: Effect Size of Friends' Activity*

	Total Impact	**Social Influence**	**Homophily**	**Homophily Bias***
Visit Probability	12.40%	9.82%	2.64%	22.61%
Conditional Page Views (given a site visit)	–2.50%	–3.22%	0.63%	–26.66%
Unconditional Page Views (net effect)	-3.60%	–5.85%	2.22%	40.66%

* We simulated the percentage changes of visit likelihood or page views using the six-dimensional latent space Own and Friend Activity model and then computed the average percentage changes across all observations. We computed the bias by comparing the overall impact of friends' activity estimated with two alternative specifications, one with and the other without latent space corrections. For example, a value of 22.61% for visit probability means that if we do not use latent space corrections we would estimate the overall social influence effect in visit probability as being 22.6% bigger than when we correct for homophily.

Table 10: Measures of Social Influence: Moderating Role of Browsing Mode

	Effect under a Directed Mode (Home Page Dummy = 0)	**Effect under an Explorative Mode** (Home Page Dummy = 1)	**Overall Effect**
Visit Probability	11.53%	7.19%	9.82%
Conditional Page Views (given a site visit)	–22.48%	–1.28%	–3.22%
Unconditional Page Views (net effect)	–13.49%	5.83%	–5.85%

Figure 1: Examples of Facebook Page Likes (Like Activity)

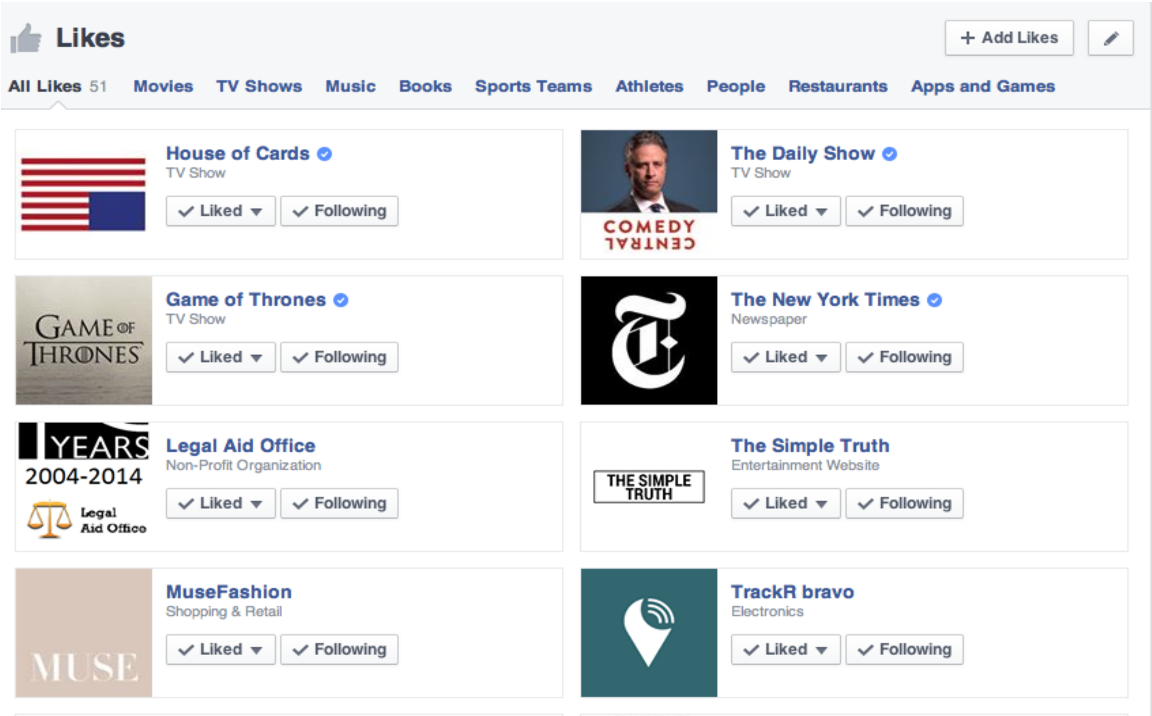

Figure 2: Model Fit for the Own and Friend Activity Model Considering Different Latent Space Dimensions

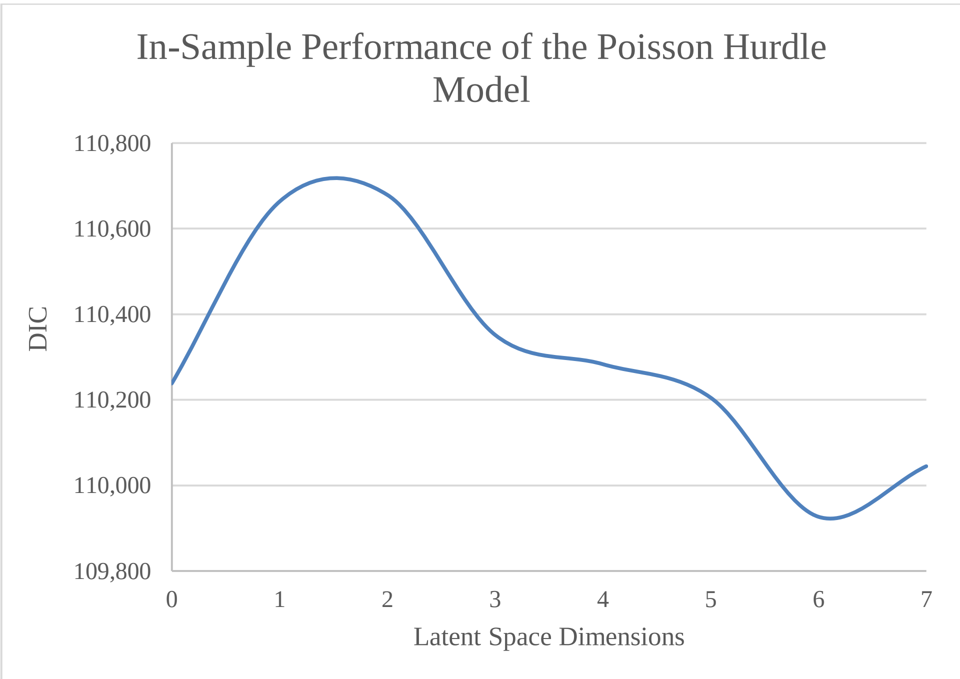

Figure 3: Out of Sample Performance: Hit Rate Visit Probability