Competitive Versus Complementary Effects in Online Social Networks and News Consumption:

A Natural Experiment

Catarina Sismeiro

Associate Professor
Imperial College Business School
Imperial College London
South Kensington Campus
London, SW7 2EA
sismeiro@imperial.ac.uk

and

Ammara Mahmood

Assistant Professor
Ammara Mahmood
Lazaridis School of Business and Economics
Wilfrid Laurier University
Waterloo, Canada
ammahmood@wlu.ca

July 2017

Competitive Versus Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment

Abstract

Using hourly traffic and readership data from a major news website, and taking advantage of a global Facebook outage, we study the relationship between social networks and online news consumption. More specifically, we test if online social networks compete with content providers or instead play a complementary role by promoting and attracting traffic to external websites. During the outage, consistent with a promotional effect, we observe a significant decrease in traffic and unique visitors to the news website lasting beyond the outage hours. We further find that direct referrals from Facebook links, grossly underestimated the actual impact of Facebook in generating traffic. Instead, during the outage, we observe a more significant reduction in visitors arriving at the news website from search engines or directly typing the website URL or using bookmarks. Additionally, readership of articles and types of pages viewed also changed during the outage. Although we observe a drop in news consumption during the outage hours for all news categories, the subsequent news consumption differs across categories. Time sensitive categories like Sports and Local News see an increase in consumption whereas news on Women Issues or Health topics see a decrease. Analysis of individual-level visit and readership behavior during the outage also reveals that Facebook not only introduces selectivity bias by attracting shallower readers, Facebook also changes readership patterns (in the absence of Facebook, visitors engage in more in-depth reading). To test the generalizability of our results we study the impact of the outage on referrals from other social media outlets, on other news sites, and on other content and e-commerce sites. We find similar effects on other news providers whereas data from non-news sites, including e-commerce, show no major outage effects. Overall, our results have important managerial implications. We highlight how our results unearth the importance of search engine optimization and of strong branding for news websites, if providers want to harness fully the power of their social media presence.

Keywords: Online Social Networks, News, Natural Experiment, Social Media Metrics, Referrals, Content Providers

1. Introduction

"Nothing attracts news organizations like Facebook. And nothing makes them more nervous." (Somaiya, Isaac, and Goel 2015)

Content represents a significant portion of economic activity online, and supports the online advertising market that accounts for over $100 billion in global ad revenues (E-marketer 2013). Among companies whose business models revolve around content provision and ad-revenues, news websites are facing some of the most significant challenges in revenue generation. The difficulty of news sites in charging for content after years of giving it away for free, reduced readership of their offline arms, and a shift in consumers towards more informal sources of information, are all factors contributing to news websites' difficulties (Marshall 2015) and to their increased dependence on advertising. Today, advertising accounts for the majority of revenues for US news websites, with the remaining obtained via subscriptions or other value added services (Vinceguerra 2012).

In light of the significant reliance on advertising revenues, site traffic has become crucial for the survival of news websites. In a world dominated by information overload and content shock (Schaefer 2014), it can be difficult for news providers to attract visitors to their sites because individual consumers do not necessarily have access to, or awareness of, news and information that is relevant to them. In such an environment, content aggregators and online recommendation systems can provide a valuable aid to newsreaders through the curation and dissemination of information (Goldenberg et al. 2012) although a significant debate today surrounds their role in attracting traffic to news websites.

Some commentators and scholars worry about substitution effects, with users accessing news stories through aggregators and relying on the headline-like information while reducing website visits. Others contend that news aggregators can bring additional traffic to news providers due to exposure to new and diverse content (Chiou and Tucker 2017), in which case news providers benefit from the activity of aggregators. Previous research has demonstrated, for example, that collaborative filtering algorithms might be useful in recommending content to individual users by enhancing the diversity of content discovered (Zhou et al. 2010). As user satisfaction improves with content diversity, researchers propose that such positive effects can spillover to news providers. Content aggregators would then have a *complementary* instead of *competitive* role driving traffic to news websites (Susarla et al. 2012).

Recently, the increasing popularity of social networks has resulted in speculation that social networks are the new content aggregators. About 63% of US adults engage with news while using social networks (Barthel et al. 2015) and half of Facebook participants share news from external links while using online social networks (Baresch et al. 2011), making social networks a platform more of user-distributed content than of user-generated content (Oeldorf-Hirsch 2011). Hence, in a similar vein to the debate surrounding news aggregators, industry experts now fear that social networks might be potential competitors of news sites.

The launch of services that integrate news with online social networks such as Facebook's Instant Articles and Snapchat's Discover platform have further fueled news publishers' anxiety. For some time-constrained digital users the availability of headline-like information on social networks (e.g. on Facebook's news feed) would result in these users dedicating more of their time to social networks and less to news providers for information gathering. Evidence seems to suggest that such negative impact could be real. As online social networks have grown in popularity, so have the difficulties of news sites in maintaining adequate traffic and revenues and, while Facebook now absorbs about 12% of the global advertising spend (Seetharaman 2016), leading news sites report around a 10% drop in ad-revenues (Barthel 2016).

On the other hand, consumers might use social networks as a platform for the discovery of news and content, in which case positive effects similar to those previously discussed for news aggregators could occur. As a result, even though some contend that the amount of traffic originating directly from social media is insignificant, with only 9% referrals directed from social media (Mitchell et al. 2012a), other researchers argue that news and content providers are ever more dependent on traffic originating from online social networks (Ju et al. 2014, Olmstead et al. 2011). These positive effects might be greater in the context of social networks than in the context of content aggregators, as previous research has suggested that individuals benefit more if input from social media and from online social networks is used in content discovery (Goldenberg et al. 2012). More importantly, previous research on news aggregators is not able to provide a conclusive answer for the role of online social networks because of the significant differences between these platforms.

First, whereas aggregators rely on algorithms, online social network users can rely on friends to filter and recommend content, which is more likely to result in greater "referral" traffic to news sites, as users may be more willing to click on links aligned to their interests and tastes. Secondly, the scope of activities performed at online social networks is far richer; unlike news aggregators, the primary motivation to visit online social networks is not to read news; instead, individuals visit these sites to connect and interact with others and to develop a public image and, during these visits, individuals might incidentally discover news. About 78% of Facebook users see news on Facebook not due to an active search for news but for "other reasons," whereas only 34% of Facebook users like news organizations pages to get updates (Matsa and Mitchell 2014). These numerous social interactions could also prompt users to actively search for news at external sites creating traffic that is beyond the traffic associated with following direct links on online social networks. Hence, the impact of online social networks is likely more varied than that of news aggregators encompassing both direct social influence and mere exposure effects due to the multitude of activities people engage in.

Despite the discussions in the field of communications, a clear consensus on whether online social networks compete for users' time or instead complement news publishers' does not yet exist. To the best of our knowledge, no empirical work using online traffic data attempts to determine how online social networks are associated with content consumption at third-party websites. Do social networks perform a function similar to news aggregators? Does the exposure to content on social media generate spillover effects beyond that of referral traffic? Does Facebook attract different type of users to the news site, or does it change news readership? Answering these very important questions would help develop social media strategies for publishers.

We contribute to the literature on online content consumption, news aggregators and communication studies by analyzing news browsing data in the backdrop of a global Facebook outage. The Facebook outage constituted an exogenous shock and it was unprecedented and unexpected (users had no previous information about the disruption and were unable to adapt their behavior prior to the outage). The outage lasted approximately four hours and hit Western Europe in the early morning of October 21, 2013. Users were not able to add any updates on Facebook including posting links (including news stories) nor could they comment on existing posts. Because news stories are highly time dependent, and require frequent and timely updates to generate value for readers, we expect the lack of updates on Facebook to have a significant impact on news websites. This impact is also more likely during the early hours of the morning when users perform daily rituals such as starting their day by checking emails, social media pages, and their favorite news providers.

For our study, we collected hourly traffic and readership data for site visitors of a large news website before, during, and after the outage, enabling us to test the effect of the outage on traffic and readership at this focal site. Through this event study, our aim is not to assess long-term substitution patterns in news consumption. Instead, our objective is to use the exogenous variation in traffic and readership due to the outage to analyze the short-term dynamics in online news consumption (during and up to a day after the outage) due to Facebook unavailability. We present both an aggregate analysis for the whole site to measure effect sizes and perform a cohort analysis that relies on individual-level data to provide insights on the potential underlying mechanisms driving the results.

We also obtained data from other news websites from the same country and tested to see if our results are robust and generalizable across news outlets. To ensure that results do not depend on other social media events that might have kept users away from news reading, we further control for the referral traffic originating from other popular social networks. We also collected hourly traffic data from a network of four e-commerce websites and two technology-related blogs to serve as controls, as traffic to such websites allows us to compare the effect of Facebook on online news consumption to its effect on other types of online venues.

Our results suggest that online social networks are *positively associated* with news reading. During the outage, we observe an immediate and significant decrease of 38% in site visitors and 44% in content page views, which corresponds to a loss in advertising revenue of more than $21,000 in just over a day. The decrease in traffic and content page views lasts beyond the immediate hours of the outage but gradually declines over time. We observe a negative impact, albeit a reduced one, at least one day after the outage, which in turn is consistent with a potential carry-over effect. We find the same effects for the other two (less popular) news sites from the same country, but we find no impact of the outage on the e-commerce website and the technology blogs. The fact that other sites did not register changes in traffic provides further evidence that news and social media share a unique relationship and, despite fulfilling similar needs for entertainment and informational content, social media and news sites seem to share important synergies.

We also find that referrals from Facebook (typically used to measure websites' dependence on online social networks) grossly underestimate Facebook's impact. For our focal news website, and for the two minor news sites, traffic lost was far larger than the page requests typically associated with Facebook referrals. Instead, we find that the loss of traffic arriving via search engines and by directly typing the website URL on the browser was far greater in magnitude than the loss associated to lack of Facebook referrals. This indicates that Facebook has spillover effects via other online platforms due to the complex mechanism driven by exposure to social conversations and content curated by friends and family. Only an event study of a global outage like the one we investigate could allow us to uncover these effects.

We also report that social networks influence the browsing behavior and readership at the news site. For example, during the Facebook outage, we observed a significant increase in *articles read per user*, and a decline in *home page views per user*. Home pages include titles and very short descriptions of the content of each article, and links posted on Facebook do not typically direct users to the home page. This could mean that, during the outage period, once at the site, users immerse themselves in more in-depth perusal of content pages while exploring fewer headlines.[1] Our analysis also indicates a differing impact of Facebook across news categories. For example, although the outage affects negatively all news categories during the outage hours, we find a subsequent increase in consumption of Sports and Local News and a subsequent decrease in consumption of articles associated with Women Issues and Health. In this vein, our results are similar to the recent research suggesting that news aggregators influence not only traffic but also the type of content consumed at news sites (see Chiou and Tucker 2017, Athey, Mobius, and Pal 2017).

The aggregate changes in news reading we observe also indicate that Facebook could attract different users to the website when compared to those arriving at the news site independently from Facebook (a selectivity bias effect associated to Facebook). The findings from our cohort analysis that

[1] We note that this result is not simply due to a heightened interest in the news about the outage itself, as the article discussing the Facebook outage appeared hours later on the focal news site.

controls for individual differences show strong evidence of selectivity bias, but we also find evidence that users change their readership during the outage hours, reading more articles and visiting the home page fewer times. Facebook appears not only to attract different users to the site (those who are shallower readers) but it also seems to influence news readership (when available, the online social network seems to encourage shallower reading of news).

Our findings have important implications for content websites. We find that social networks have a traffic expansion effect similar to news aggregators. However, unlike previous research, we find that this expansion effect far exceeds the impact that we would infer by investigating only "referrals." Hence, we are able to not only refute speculations regarding competitive effects but we also open an important discussion on how to better measure the real impact of online social networks. In addition, search engine optimization (SEO) and brand building seem to be essential for news sites to harness the power of online social networks as changes in traffic arriving from search engines or by directly inputting the URL seem to be significantly affected during the outage. Finally, online social networks seem to attract different users to news sites but they also seem to have an impact on the type and depth of news readership. Our results confirm the importance of social media in informing citizens and disseminating news, issues that have been widely discussed in the aftermath of the 2016 US elections, as fake news stories transmitted via social media were noted as having had an impact on democratic participation.

2. Literature Review

Despite the lack of empirical studies using actual browsing data, a recurrent theme in the literature and popular press is whether social media pose a threat to an already burdened business model, or whether it represent mostly an opportunity for added revenue. Though not conclusive, previous research has provided some clues regarding the potential effects of online social networks on news providers.

Indeed, the study of the *complementarity* and/or *competition* of alternative media sources has a long history in the Information Systems and Communications literature. For example, previous work has investigated the impact of internet adoption on print newspapers (Cho et al. 2016), the complementarity between alternative news channels in the context of mobile technology (Xu et al. 2014), and the competing role of legitimate and illegitimate digital movie broadcasts on DVD sales

(Smith, Michael, and Telang 2009). Huang, Hong, and Burtch (2016), in a forthcoming study, analyze the role of social network integration across platforms. However, the focus is on user-generated content and the authors do not study the browsing behavior at third party websites.

Similarly, Mahmood and Sismeiro (2017), rely on browsing data to study the potential social influence in the context of online social networks and news reading. However, the work provides only correlational evidence of a potential complementarity and it is unable to distinguish between different traffic sources including referrals from Facebook. Perhaps our work shares the most similarities with the previous research on news aggregators (e.g., Palme et al. 2016, Prawesh and Balaji 2014) and on their impact on third party news sites (e.g., Jeon and Nasr 2016, Chiou and Tucker 2017, Athey, Mobius, and Pal 2017). However, such studies do not investigate online social networks and cannot provide a conclusive answer on their competitive or potential complementary role.

In light of extant research, we contend that substitution or promotional effects may very well exist in the interplay between social networks and news sites.

2.1 Substitution Effect

Some researchers propose that the main impact of online social networks is the audience stealing effect due to competition for consumers' limited resources (Kayany and Yelsma 2000). Because digital users have limited time available for media consumption, as time spent with social networks increases, individuals would be less likely to engage with other content websites including online news publishers.

This *time displacement theory* resonates with many industry insiders and seems to find support in several studies. For example, based on a Pew Research Centre survey, in December 2011, Facebook users spent, on average, 423 minutes on Facebook and only 12 minutes on news websites (Mitchell et al. 2012a). Other survey studies find that social media websites now account for a quarter of time spent online (e.g., Nielsen 2011) and that this widespread use of social media has been "sucking up online time" from other media and online content sites (Bernoff and Li 2008). Hence, consumers with limited temporal budgets may reduce the consumption of one medium as their usage of another medium increases (James et al. 1995, Robinson, Barth, and Kohut 1997, Kayany and Yelsma 2000). The audience stealing effect of social networks would be similar to the "news scanning effect" (Chiou and Tucker 2017, Jeon and Nasr 2016) of news aggregators that reduces the time spent on news sites.

More importantly, Facebook could have an even greater displacement impact than that of news aggregators. Online social networks, like Facebook, satisfy social and personal functions some of which are similar to those satisfied by news providers and some of which go beyond the needs that news reading can satisfy. For example, personal motivations for news reading including surveillance, personal identity construction, and entertainment (Katz et al. 1974). The desire to satisfy these needs also motivate users to engage with online social networks (Kuss and Griffiths 2011). Facebook users develop their own personal narratives and identities through their digital wall, discover relevant information for their daily lives, and enjoy their leisure time. Similarly, individuals consume news to develop a personal identity and for their referential function, that is, to learn accurate information (Tsfati and Cappella 2005). News audiences also wish to get information necessary for their daily lives (e.g., learning about a strike, weather, traffic, or stock prices) and to understand and familiarize themselves with a variety of issues. News also satisfies surveillance and cognitive needs (Wright 1960) and, just as other entertaining content, news could simply help pass the time (Rubin 1993, 1984, Cutler and Danowski 1980) or aid mood management (Schramm 1949). Online social networks also satisfy these same personal needs.

Beyond *personal* motivations, news reading and online social networks also satisfy similar *social* needs. These include the need to connect with others and to achieve a sense of belonging. For these social gratification seekers, knowing the news the group reads or hears about, and consuming the content that the group consumes, will help achieve the social goals of inclusion and belonging (Wenner 1985). Again, users may choose to fulfill these social needs via engagement with content *within* their social networks. For instance, users may use their friends or family as curators of news and users may decide the information displayed on the newsfeed is enough, and reduce visits to the news websites. Hence, the significant degree of niche overlap (Dimmick et al. 2004) in terms of the benefits that news consumption and social network engagement provide, suggest the possibility of media substitution.

Beyond niche overlap, there could also be a degree of niche superiority (Dimmick et al. 2004), whereby, compared to traditional news outlets, social networks satisfy additional needs that may revolve around content consumption and that might displace traditional forms of engagement with news sites. For instance, the time users spend in updating their Facebook page, commenting on posts, sending messages, and reading the content posted by others could leave them with little or no time to visit news sites or other content providers. In some extreme cases, individuals get so enthralled with Facebook interactions that it can become an outright addiction leaving little time for other activities (Kuss and Griffiths 2011). Thus, the displacement effects theory would predict that the use of Facebook would result in less engagement with news sites and content consumption.

Hence, in the absence of Facebook (e.g., due to a disruption), we would expect more visits and greater time spent at news sites and perhaps other social media platforms. Furthermore, based on individual heterogeneity in terms of personal and social motivations, we may expect to observe differences in the impact of social networks on the readership and visit behavior amongst newsreaders.

2.2 Complementarity and Promotional Effect

Though the potential effects of time budgets are widely discussed in the literature, other media researchers suggest that, for multi-tasking digital users, news reading is not a "distinct and purposeful activity, but instead is placed among many other information seeking activities" (Yadamsuren and Erdelez 2011, p. 2), with social network engagement being one such activity. McLeod and Chaffee (1973) further note that the use of media, such as news, is interwoven in a user's social life and news reading could be part of a daily ritual that simultaneously includes other activities. An example of such a ritual would be waking up and browsing the Facebook news feed, reading the news, and checking emails, in this order. Once one of these activities is disrupted, instead of dedicating more time to the remaining activities of the ritualistic bundle, users could simply stop the entire ritual altogether.

Given the central role of social networks in our lives, it is clear that consuming news in association with social networks is now a natural "order of things" (Savolainen 1995, p.262). Online social network users could also browse news or search for content online with the sole purpose of posting content on their Facebook pages, that is, as part of their daily process of creating a Facebook persona. Hence, visits to news sites could be caused by the need to search for information in expectation of creating Facebook posts or sharing. In such cases, without the ability of updating their online profile/page (as during a Facebook outage), news reading loses its relevance.

Other media scholars point out that consumers prefer "a varied diet of media" (Ahlers 2006), users often supplement visits to news sites with social network engagement to gain different perspectives on news. Such synergies between social media and news media are known to increase the value of news (Tanriverdi and Venkatraman 2005). In a world with an abundance of content users may reduce the cognitive cost and effort of seeking content themselves by relying on links curated by social networks (e.g. posted by peers and shown on the Facebook news feed). In this respect, social networks can have a market expansion impact on news websites by providing a platform that facilitates content discovery like news aggregators (Chiou and Tucker 2017, Athey and Mobius 2017).

Beyond their role as a content aggregator, social networks can also spark interest in news through social interactions. Online friends often discuss their personal life, current affairs and engage in interpersonal communication that does not directly relate or include a link to the news site, but that could create interest and spark curiosity in news or content, leading to a visit to the news provider of their preference. Previous research has also shown that, in a context where users can have conversations with one another, as in online social networks, they will begin discussing the information they have shared (Shirky 2008). As a result, in expectation of such interactions, individuals who see a specific topic on social media sites 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). This concept of "communicatory utility" has often been used in the past to explain how interpersonal motivations drive (mass) media information-seeking in order to fulfill interpersonal goals (Atkin 1972) and can also explain a promotional effect of *online social networks* on news sites.

Whereas some users may consciously use social media as a vehicle for finding and propagating news, other online social network users discover news incidentally. Tewksbury and colleagues (2001) note that users often stumble upon news stories when they are online and that incidental news consumption is common. Indeed, more than 85% of U.S. adults turn to social media for connecting with friends or family (Asay 2014) but while engaged at the social network users are exposed to news stories in their personalized newsfeed. If a user discovers interesting articles on her newsfeed, she could follow the link to visit the news providers and read the content, even though she normally does not visit news, sites creating a promotional effect. This unintentional news discovery may have an effect, similar in vein to the well-established "mere exposure effect" in marketing communications (Obermiller 1985).

In sum, these alternative mechanisms comprising (1) direct ones, that rely on intentional curation of content to post on Facebook and clicked links due to reduced cognitive cost and effort, and (2) indirect ones, that rely on "mere exposure" and ritualistic behavior, would predict a complementarity effect of social networks on news websites. The unavailability of any one aspect of these entwined mechanisms due to an outage could lead to a conscious (or unconscious) change in behavior, albeit a short-term one.

3. Data and Analysis

We collected hourly traffic and readership data from the second largest online news website operating in a major Western European country.[2] The dataset corresponds to hourly traffic and hourly page views for all visitors of the news site. To test the competitive versus complementary effects of online social networks on online news reading, we take advantage of the exogenous variation in Facebook traffic created by a global Facebook outage that lasted approximately four hours in the early morning of Monday, October 21, 2013. During the outage, it was not possible to add new posts, comment on previous posts, and there were no newsfeed updates, although users could access the information previously loaded on their device (computer, tablet, or mobile).

The outage was one of the longest worldwide disruptions Facebook has faced. It struck Western Europe in the early morning, while users started their day by checking Facebook and reading the news. This temporary disruption was exogenous and unexpected for both Facebook and its users, allowing us to identify the relation between Facebook and third party websites.[3] News websites are indeed the ones likely to be the most affected by an outage of this type, because news is a time-sensitive product that expires quickly and is constantly updated.

[2] For confidentiality reasons we cannot reveal website's name nor the European country in which it operated. It is a major player in its country and it dominates the news market together with another major player. There also other multiple news websites operating, though with differing sizes.

[3] Considering that other short-lived errors frequently occur in the online domain, users were unlikely to be forward-looking and unlikely to adapt their behavior in expectation of a long outage.

The news website we study is one of the main news providers in the country, has a very large and stable audience, and has Facebook presence in the form of an official page users can like. During the period under analysis the site did not engage in any Facebook advertising campaigns, hence we can attribute any change in traffic and readership patterns during the outage to individuals' actions on Facebook or as part of their daily routine, not to a change in advertising effectiveness or spend. In order to validate our results we also study the effect of the outage on other (smaller) news sites and on other types of websites (including e-commerce and technology blogs). We provide more details on these additional analyses in the validation section.

We collected data over the period of October 13, 2013 to October 29, 2013 (17 days) for a variety of metrics for the major news website. We eliminated automatic refreshes,[4] we distinguished between home page views and the consumption of content by category (Local News, Sports, Women Issues, Health News, and so forth) and, using cookies, we determined how many unique users visited the site in each hour. Finally, we identified the page requests originating from Facebook, Twitter, and search engines[5], and those that are the result of users typing (or copying and pasting) a specific URL directly into the browser (we call these latter type of page views as coming from "undefined" referrals). This is an important feature of our dataset because we can determine if the number of Facebook referrals accurately capture the total impact Facebook has on news websites (referrals are the typical measure used by researchers and practitioners to assess the impact of social media on website traffic). We also checked for special political, social, entertainment or sports events and collected Google Trends metrics for the major news website, the major competitor, and specific major events. We do not find any specific event occurring on the day of the outage or during the outage hours, and we further note that any major event would have affected not only the focal news website, but also other websites we study in the validation section.

The overall traffic to the focal website during the period under analysis is stable at around an average 9 million impressions and 4 million unique visitors per day. Figure 1 shows the total hourly

---

[4] If left open on a browser, the home page is automatically updated at fixed time intervals. We removed these automatic refreshes to avoid confounding the effect of automatic updates with that of individuals' actions.
[5] Incoming URLs (referrals) enables us to distinguish between various sources of incoming traffic including those clicking on a link on Facebook, Twitter or search engines.

number of page requests and unique users visiting the site from October 13 until October 29. There are no trends in our data although there are significant day of the week effects (e.g., reduced traffic on weekends) and intra-day effects (e.g., traffic peeking in the early afternoon and reaching very low levels in the night). Using this data, we perform (1) an aggregate-level analysis, and (2) an individual-level cohort analysis

3.1 Aggregate-Level Analysis

We reduce the potential influence of external factors by limiting our aggregate analysis to the time interval that includes the day of the outage, the day before, and the day after the outage, that is, from Sunday, October 20 until Tuesday, October 22. This approach limits the impact of external events that may influence news consumption or any fluctuations in the time people have available to read news. Because of the significant intra- and inter-day variability in traffic observed, we take the average hourly traffic and content page views observed for the same weekdays (Sunday through Tuesday) and the same hour, one week before and one week after the outage, to build a baseline for each variable of interest. This baseline reflects the typical values for each variable in the absence of a Facebook disruption.

Figure 2 represents the timeline of our analysis and shows the day of the outage, the period of analysis we consider, and the days used to build the baseline. Figures 3 and 4 present the total page requests and total users per hour against the baseline across the three days of our analysis. We also build an alternative baseline considering only the data from the week before the outage and we find very similar results (results available from the authors upon request). Table 1 presents the summary statistics of the variables for the three days of our analysis period including the baselines used.

We denote y_t as the (24 ×1) vector of hourly values for variable y at day t (with t = 20, 21, 22), ${y_t}^*$ as the vector of deviations from the baseline, and $\bar{y}_t$ as the vector of hourly baselines. Hence:

$$ {y_t}^* = y_t - \bar{y}_t \text{ , and} \tag{1} $$

$$ \bar{y}_t = \frac{1}{2}(y_{t-7} + y_{t+7}). \tag{2} $$

For each variable of interest, we model 72 hourly deviation values that span the three days of analysis. Similar to the work of Goldfarb (2006), we estimate the immediate and subsequent effects of the outage by regressing the observed deviations from the baseline on two dummy variables. The first, D_{outage}, captures the immediate impact of the outage and takes the value 1 during the hours of the outage (between 7:00 AM and 9:00 AM) and 0 otherwise. The second dummy variable, D_{after_outage}, takes the value 1 during the hours *after* the outage, and 0 otherwise, and captures the impact of the disruption after the outage hours (this is also a short-term effect because we are modelling up to a day after the outage).[6] Hence, we estimate:

$$y_t^* = \alpha + \beta_1 D_{outage,t} + \beta_2 D_{after\_outage,t} + \varepsilon_t \,, \tag{3}$$

where α is the intercept, β_1 captures the impact during the two central outage hours and β_2 the impact after the outage (once Facebook is again available); ε_t is an identically and independently distributed vector of normal error terms (zero mean and σ^2 variance).

In the validation section, we present additional analyses that rely on expanding the period under study, and on alternative specifications for the baseline and for the dummies. We also test for potential wear-in effects of the outage by including two-hour interval dummies.

3.2 Individual Level Cohort Analysis

Some of the mechanisms underlying Facebook's potential impact are not separable without the analysis of individual-level data. For example, Facebook usage could be associated with a change in individual behavior or with Facebook attracting different people to the news websites. To better disentangle these alternative mechanisms of Facebook influence, we track what individual users do at the site before, during, and after the outage. We then perform a cohort analysis to isolate the effect of the outage while controlling for other factors that could explain site visitation and readership (e.g., whether users are "deep" or "shallow" readers).

We use cookies to identify visitors and find more than 22 million unique visitors during the period under analysis (17 days). We then separate users based on their browsing and visit behavior. We build two cohorts: the "7-9am Visitors" cohort includes users who visit the news site between 7:00 AM

[6] The outage was complete and affected all individuals worldwide. Facebook never revealed an exact timestamp for the start and the end of the disruption and the only information available is that the outage lasted for less than four hours and that all users felt its impact for the two complete hours between 7:00 AM and 9:00 AM. However, we do not know how many minutes before 7:00 AM and how many minutes after 9:00 AM the outage lasted. For this reason, we study the two complete outage hours (between 7:00 AM and 9:00 AM) as the hours of the disruption. We know that, during this two-hour interval, it was not possible to post or comment on Facebook and that no news updates were available. We did not consider the hours of partial outage (i.e., from 6:00 to 7:00 AM and from 9:00 to 10:00AM). We note further that only the content that had been loaded in the app or into the web page on individual PCs was visible and nothing else could be updated or uploaded.

and 9:00 AM, the two-hour outage interval; the "5-7am Visitors" cohort includes users who visit the news site between 5:00 AM and 7:00 AM, the hours prior to the outage. We note that we do not expect the outage to affect the actions of "5-7am Visitors" before the outage, although these visitors are subject to the day-specific dynamics associated with the outage day (e.g., specific events and news that are out that Monday) and can serve as a valuable control group.

We consider two alternative cases for both cohorts. The first case considers users who visited the website on the Monday prior to the outage (October 14) during the two-hour intervals considered (7-9am and 5-7am), and return to the website on the day of the outage during those same time intervals. The second case considers users who are *loyal* visitors to the website and visit the news website *every day* prior to the outage during each of the two-hour intervals we consider.[7] We then assess how many of these users returned on the Monday of the outage during their cohort hours (7-9am or 5-7am), how many return the Monday after the outage, and how they compare in terms of readership. We distinguish these two loyal cohorts by denoting them as the "Loyal 7-9am Visitors" and the "Loyal 5-7am Visitors." Table 2 presents more details on the size of each cohort and return rates.

4. Results

We run the regression model outlined in Equation 3 for each variable separately including number of visitors, referrals, and page views (page views will include home page views and the requests for pages with specific content, unless otherwise indicated). We tested for the inclusion of the outage and the "after-outage" dummies, and the full model provided the best fit. Table 3 reports the full model results.

4.1 Facebook Outage Impacts Website Visitation and Page Views

The results show that Facebook has a significant impact on news site visitation. During the outage, we observe a significant decrease in the total number of visitors with a drop of 70,096 visitors per hour (*p-value* < 0.001). This corresponds to a 38% decrease in visitors considering the typical traffic during

---

[7] We also tested for alternative definitions of loyal readers. Because the behavior and traffic on weekdays seems different from that on weekends, with a significant decrease in news reading on weekends, in this second alternative we considered a loyal user of the news site between the predefined two-hour interval if a user visits the site every weekday (Monday till Friday) before the outage. Another definition required that users visited the site at least five of the eight days prior to the outage. We also tested the two hourly intervals separately. The analyses using alternative definitions produce the same substantive results (albeit with, naturally, a different number of visitors).

those morning hours. Even after the outage, we observe 9,100 fewer site visitors per hour (*p-value* < 0.001). Although statistically significant, this is a small reduction accounting for about 4% of visitors. We also observe a 44% reduction in the total number of page views (a drop of 201,814 pages per hour; *p-value* < 0.001). After the outage, total page views are also lower than expected (44,798 fewer pages each hour; *p-value* < 0.001), a drop of about 9% of the pages typically viewed during those hours.

The fact that we do not observe an increment in visitors or page views, once Facebook is restored, suggests that users did not hold back their website visits in anticipation of a resumption of Facebook. These results further suggest a (net) promotional Facebook effect on news reading: Facebook seems to help news websites to attract visitors and lead to more page requests. More importantly, we find that considering only the typical levels of Facebook referrals for this website we would have grossly underestimated Facebook's impact. For the website under study, only 3% of the traffic originates from links posted on Facebook and we find an hourly decrease of 3,956 page views originating directly from Facebook during the outage (*p-value* < 0.05). This decrease is substantially lower than the reduction in total page views during the outage (about 170,000 fewer pages) implying that Facebook has an effect that goes beyond the traffic originating from clicks on links to the news site posted on Facebook. The type of mechanisms we suggested, including mere exposure effects, searching for news or specific topics previously seen on Facebook, and a break in daily rituals, may all contribute to the lingering effects and might not allow a full recovery even after the outage hours.

4.2 Beyond Direct Facebook Referrals: Impact on Search and Direct Traffic

Following previous research, we suggested that mere-exposure effects, rituals and social motivations of social network users could explain the promotional effect of Facebook beyond the direct Facebook referrals. To test for these mechanisms, we look at the remaining sources of traffic (i.e., other referrals) during and after the outage. We find that, during the outage hours, referrals from search engines and undefined referrals (i.e., people directly typing the URL, using their own bookmarks, or copying and pasting URLs) decreased far more than Facebook referrals. We observe 29,470 fewer referrals from search and 142,020 fewer undefined referrals (a drop of 62% and 69% respectively). These substantial reductions are statistically significant (*p-value* < 0.001).

Hence, Facebook seems to exert an influence that transcends the dyadic relation between social networks and news sites, and this influence involves other key online players. For example, triggered by a topic of interest mentioned during online conversations or seen in online comments on social networks, users might search for more information using search engines and then follow a search result that takes them to the news site. Previous studies show that search engines (and Google in particular) are the top sites visited after a social network website visit (Simmons 2011). This would indicate that news publishers appearing higher in the list of results would get more traffic. To the best of our knowledge, no other study has thus far indicated the importance of search engine optimization as instrumental for news sites to harness fully the power of social media.

Users might also copy and paste the URL of the links seen on Facebook directly to a browser, and in that case users arrive to the news site as "undefined" traffic. Indeed, Chohan and Francis (2014) report that users prefer to copy and paste links to share instead of clicking on social media buttons. During the outage, we observe a drop in undefined referrals, with *home page views* with undefined referrals seeing the greatest drop (about fewer 117,821 "undefined" home page views). This suggests that users type the home page URL, or might use their own bookmarks, to access the website and read more content after seeing information or engaging in conversations while using Facebook, which would support the "mere-exposure" effect (such an effect would benefit mostly news outlets high on users' consideration).

Facebook could also be part of a daily ritual in which news reading might constitute a whole together with other daily activities. Once one of the activities is disrupted, the entire ritual is jeopardized (including news reading). This mechanism would also explain the significant drop in direct access to the website's home page during the outage. Both mechanisms are plausible (the website under study is very popular and likely to be top of mind) and supported by our results. Although further studies are necessary to understand how Facebook interacts with online search and the websites' brand strength, it is now clear that the possible paths of influence are far more complex than previously thought.

4.3 Differences in Readership by News Category

To uncover other possible dynamics of Facebook's influence, we further look at the performance of different news categories during the outage. The news categories we study include Local News, Sports, Women Issues, and Health (see Figure 5). As before, we report the hourly traffic to each news category and compare it against a baseline.[8] We further re-estimate the model in Equation 3 for each one of the categories and report the results in Table 4 (the variable being modelled is the difference to the baseline of the traffic to each news category).

Our results show a reduction in traffic of all news categories during the outage hours (the drop is significant at 5% for Local News and Health News and at 10% for Sports News and Women issues). In contrast, after the outage, traffic recovery varies by category. Sports and Local News see a significant increment after the outage (significantly above the baseline) whereas Women Issues and Health related sections remain below the baseline (all of these effects are statistically significant as it can be seen from Table 4). Sports and Local News will likely comprise of time-sensitive articles, and these could be more resistant to access difficulties associated to social media outages and people could still check back afterwards for those articles. Users normally catch up on the main sports events from the weekend on Monday mornings and sports is a popular "water cooler" topic with affective resonance allowing individuals to engage in socially and culturally meaningful social conversations (Lotz 2010).

Other articles could be less time-sensitive. For such articles, if readers did not get access at some point in time, it is less likely they would try to catch up in readership afterwards. This could explain the continued decline in Women Issues and Health related pages beyond the outage hours. The content in these categories tends to be less time-dependent and includes scientific and technical details, making casual conversations cumbersome and less popular in offline contacts (that is not the case for Sports and Local News).

The differences in readership we find across categories during the outage further reveal that Facebook might work as a (better) substitute for some time sensitive news categories but not for others. It also suggests that offline interactions might be a substitute of online ones, which in turn explains the lingering effects of the disruption, whose impact lasted beyond the outage hours.

[8] Analyzing single news articles before, during, and after the outage (instead of news categories) would not be informative, as the lifecycle of news stories is in general extremely short and the period under analysis is too broad.

4.4 Facebook Impacts News Readership: Selectivity Bias and Behavioral Changes

Our previous aggregate-level results suggest that user engagement as measured by news readership changed. During the outage, we observe a decrease in the number of home page views per user of 0.71, and an increase in the number of content page views per user of 0.52 (content pages exclude home page views and include, for example, health and sports news content). These correspond, to a reduction of 66% and an increase of 37% compared to their baselines, respectively. The net effect is a slight reduction of 0.2 total pages per user, which correspond to a marginally significant 8% decrease (see Figures 6 and 7).[9]

The drop in home pages per user and an increase in content pages per user during the outage seems to suggest that fewer shallow users (i.e., users who read mostly headlines from the home page and do not read many articles) arrive at the site during the outage. If this is the case, then Facebook seems to introduce a selectivity bias by attracting shallower users to the site. Alternatively, it could also mean that during the outage visitors behaved differently and read more in-depth content.

We test for these effects using the individual-level cohort analyses which considers how different users might be more or less likely to visit the website and consume differing amounts of content to avoid possible confounds. First, we study the return likelihood of users and find that it increases as the number of page views consumed increases. We considered the “7-9am Visitors” and the “5-7am Visitors” cohorts and grouped users in each cohort by the number of pages they requested on October 14 during their cohort hours. [10] We then compare their return likelihood on October 21 and October 28. Figures 8 and 9 show the return likelihood for the different cohorts and for the different levels of page requests. The results are as expected. Heavier newsreaders are also those who possibly

[9] This result cannot be explained simply by the presence of fewer updated links on Facebook and fewer referrals from the online social network (the vast majority of links posted on Facebook send users to specific articles, not the home page, and hence we should have expected the opposite result). The increase in articles read is also not due to an increase in interest for the article discussing and explaining the outage itself, as the article appears on the site several hours after the outage.

[10] We considered only up to 20 page requests on Monday, October 14, during the time intervals studied to build the groups. There were few cases with more than 20 page and, for both cohorts, considering up to 20 pages allows us to account for 99% of all users visiting the site during the intervals considered. For higher values of page views, results get very unstable because of the small base for computing the empirical return probabilities.

assign the most value to reading the news from the site and hence the more likely to return (irrespective of outages or any other event) which creates an *intrinsic selectivity bias*.

However, Figures 8 and 9 also reveal that the return probabilities for October 21 and October 28 are not different at each readership level for the cohort that visits the website between 5:00 AM and 7:00 AM (before the outage). Instead, for those who visit the website on October 14 during the outage hours (i.e., between 7:00 AM and 9:00 AM), the return likelihood at those same hours on October 21 seems to be lower than that of returning the following Monday (October 28).

To test if cohorts differed in return likelihood, we perform a two-way ANOVA analysis of the difference in return likelihood (we compute the difference between day 21 and day 28 for each group). We considered two main factors that could influence return likelihood: the cohort (i.e., whether users belong to the "7-9am" versus the "5-7am" cohorts) and whether the users are light versus heavy users (if users consumed more than 10 page views they were deemed to be heavy users, and if they consumed less than 10 pages they were classified as light users). Our results reveal that the two main effects are significant (F_{cohort} = 107.14 with *p-value* < 0.001; $F_{\text{light/heavy}}$ = 8.28 with *p-value* = 0.007), although the interaction is not significant ($F_{interaction}$ = 0.048 with p-value = 0.828) which means that indeed the "7-9am Visitors" are less likely to return on Monday 21 than on the 28 although differences also depend on their news consumption level.[11]

These results suggest the possibility of significant user-heterogeneity driving the changes in readership during the outage. Hence, we need to compare users with similar levels of return probabilities to disentangle user-heterogeneity from potential behavioral changes. To do so, we consider the "Loyal 5-7am Visitors" and the "Loyal 7-9am Visitors". These groups include those with higher return likelihoods and the heavier newsreaders.[12] For these loyal visitors we compute the page views per user

[11] We also run a regression model on the difference of return likelihood of each group using as independent variables a cohort dummy, a heavy users dummy(if users request more than 10 pages they are deemed heavy users), and the interaction between the two dummies (R^2 = 0.7623). The estimate of the cohort dummy is -0.0681 (p-value < 0.001), for the heavy users dummy is 0.0171 (*p-value* = 0.0482) and interaction estimate is 0.0028 (p-value 0.8283). Similar results were obtained using the actual level of page views instead of the heavy usage dummy.

[12] We note that even for the loyal cohorts the evidence is that the outage caused a drop in return likelihood. Of those 2,658 loyal users who visited during the outage hours before the day of the outage (Loyal 7-9am Visitors) we observe a reduction in the return likelihood on the outage Monday which is partially recovered on the following Monday (a return likelihood of 53.5% on October 21 and of 69.8% on October 28). For the 481 users from the Loyal 5-7am Visitors we see a very different pattern: 81.9% return on the outage Monday during the hours before

on October 14 (the Monday before the outage) and we test for differences between visitors and non-visitors.

Table 5 provides a summary of our results which suggest that *selectivity bias* is indeed present during the outage, even after controlling for differences in return likelihood. For the “Loyal 7-9am Visitors”, those who did not visit the website on October 21 during the outage hours requested fewer overall pages, fewer home pages, and fewer content pages the Monday before (note that all differences are statistically significant). For the “Loyal 5-7am Visitors” (i.e., our control group) we find no statistically significant differences. Hence, among loyal users to the website, we find that Facebook seems to attract more “shallow” readers, which is why we observe a drop in such readers during the outage, a change we do not observe that same day hours prior to the outage. This would of course explain in part the aggregate changes in readership we previously identified.

However, the aggregate readership differences could also be due to individuals changing their behavior during the outage, in addition to the selectivity bias we just identified. To test for this alternative mechanism, we consider those visitors from the “Loyal 7-9am Visitors” and from the “Loyal 5-7am Visitors” who visit the website during the Monday of the outage (October 21) and compare their readership with that of the prior Monday (October 14). We report these readership results in Table 6.

As we can see from Table 6, users from the “Loyal 7-9am Visitors” cohort who visited on October 21 *and* October 14 show a change in behavior. “Loyal 5-7am Visitors” (the control group) who also visited the news site both days do not show statistically significant differences in readership. This means that in the absence of Facebook those who still go to the website to read news visited the home page less often but read more news articles, an effect we do not find for the hours prior to the outage. The net result is a reduction of overall page views per user during the outage.

These changes in behavior do not seem to last. Table 7 reports the behavior of visitors in the loyal cohorts who visited the website the three Mondays (14, 21 and 28 of October) during their cohort hours. We take October 14 as baseline and we find no difference in behavior for the Monday after the

---

the outage and 70.7% return for the following Monday. Note that the two loyal cohorts do not show a significant difference in return likelihood for October 28 and instead a significant difference for the outage Monday. This result corroborates our previous findings, as even for the loyal cohort there is evidence that the outage caused a reduction in site visitation (see Table2).

outage (October 28) for both cohorts. The only statistically significant differences are those related to the behavior on October 21 for the "Loyal 7-9am Visitors" who visited during the outage hours (corroborating our previous finding). No other differences are significant for that group and we find no differences in readership for the control group (Table 7).

In sum, we find evidence in favor of both (1) selectivity bias, with more in-depth readers visiting during the outage[13], and (2) change in visitors' readership patterns during the outage hours (visitors request fewer home pages and read more content page).

4.5 Validation

To validate our results we perform several additional analysis. Our objectives are to (1) determine if our results are generalizable to other sites, (2) test if our results could emanate from other phenomena, (3) determine if alternative specifications for the model would provide differing answers, and (4) test the effects of the outage over a longer period.

4.5.1 Analysis of Data from Other Websites

To test the generalizability of our results and to rule out alternative explanations, we collected traffic data at other websites including two smaller news sites, niche content (technology-related) blogs, and four e-commerce sites that operate in the same country as the focal website. Whereas the outage should have had a similar impact on the smaller news sites (not only major ones), it is unlikely that e-commerce websites or other types of content websites would have been disrupted in the same way, as these serve different individual needs and follow a different commercial logic. Similar to the focal website, all websites maintained a Facebook presence with their official page but ran no campaigns during the period of analysis, and there is no overall growth trend in readership and traffic.

We collected hourly data for the day of the outage (October 21), which occurred on a Monday, and for the Monday before (October 14) and the Monday after (October 28) to build a baseline. We also

[13] News heterogeneity could be an alternative explanation. However, we find that during the outage Monday users did not behave differently in terms of readership before the outage hours. We compared the pages requested by the "Loyal 5-7am Visitors" during the early morning hours on the Monday of the outage and the same hours on the Monday a week before the outage. We find no significant differences for total pages, home pages, and other pages (details available from the authors upon request).

built the baseline only using data from the Monday of the previous week[14] and we also tested the models using logs, but our conclusions are robust to such alternative specifications (results available from the authors upon request). In addition, we focus our analysis on the total number of page views, unique users, and on referrals from Facebook, search and undefined. Figures 10 and 11 show the total number of page views and unique users (and their baselines) for the two lesser-known news websites.

Using the 24 data points for each website (news, e-commerce, and technology blogs), we regressed the deviations of these variables from their baselines (Equations 1 and 2) against the dummies presented previously (Equation 3). In line with the effect on the focal news site, we find that the smaller news sites also register a significant reduction in page views and unique visitors during and after the outage.[15] We observe a decrease of 5,013 page views and of 2,644 website visitors per hour during the outage hours. These correspond to a drop of 29% and 32%, respectively, and are both statistically significant (*p-value* < 0.05). After the outage, the decrease in page views and users is still significant: 4,074 fewer page views per hour (which corresponds to 11% fewer page views considering the normal readership for that time of the day; *p-value* < 0.01), and 2,324 fewer visitors per hour (which corresponds to 16% less of the normal traffic of those hours; *p-value* < 0.01).

We again find direct Facebook referrals to be an inadequate measure of the impact of Facebook. During the morning hours of the outage, these two websites only receive about 680 referrals per hour from Facebook, which is again far less than the loss in page views during the outage. Just like in our main analysis, we see a decrease in search referrals during the outage (about 770 page views per hour against the baseline) and in undefined referrals (about 400 page views per hour against the baseline). We note that because these two websites are less well-known and do not have the brand power or the instant recall for most visitors, they cannot benefit as much from the traffic of their established audience.

These additional results demonstrate that the findings from our focal website hold across other, albeit smaller, news websites. However, one potential alternative explanation would suggest that the

[14] Because the data provider is not the same as the one supplying the data for our previous analysis, we could not obtain traffic data for the day before and the day after the outage and we could not distinguish home pages and general content. Hence, we limit our analysis to the hourly data for the outage day using the data from the corresponding day (Monday) of the previous and subsequent weeks as a baseline.
[15] The adjusted R^2 for the page views regression is 0.971 and for the regression of unique visitors it is 0.972.

decrease in visits had little to do with Facebook's influence on news sites. Instead, if users became preoccupied in solving their "Facebook problem" (i.e., the outage) they would also lack the time and attention to perform other activities online (including browsing for news). If this is the case, we would expect e-commerce websites to decrease in visits to other types of websites including e-commerce *and* an increase in engagement with technology-related websites. Users might feel that tech sites would provide an explanation or a solution for the Facebook disruption.

In order to test for such alternative explanations, we collected and analyzed the traffic data of four e-commerce websites and two technology-related blogs. We find no significant differences in page views and in unique website visitors for the e-commerce and the technology-related blogs (see Table 8 for the full results). Hence, the decrease in traffic we observe at news sites highlights the specific relation shared by news and online social networks that is not present for e-commerce and other technology-related blogs.

4.5.2 Analysis of the Focal Website for an Extended Time Period

To test whether the model specification could influence results and if the effects of the outage are likely to linger for a longer period, we performed an alternative validation analysis. We reanalyzed our hourly data considering the entire period from October 20 (the day before the outage) until October 29 (the last day in our dataset). Hence, we re-do the analysis to include 10 days of hourly observations (240 observations; see Figure 12) including and following the day of the Facebook outage. As before, we account for the variability that is present within a day and across weekdays. To do this, we use the hourly data from the first week of the sample (from October 13 until October 19) as our baseline, and model the deviations from this baseline for key variables.

We tested for multiple alternative specifications for each variable (both linear and in logs) after testing for the presence of unit roots (all deviation series were stationary). Among the independent variables we tested, we included a dummy variable for the outage hours (just as before), a dummy variable for the hours after the outage (but still during the same day of the outage), and eight additional dummies that indicate post-outage observations (one for each day after the outage day, i.e., from the 22nd until the 29th). This allows for a flexible and non-parametric specification of time-specific effects.[16] We further tested for the inclusion of lagged variables of the dependent variable, whose significance varied depending on the variable under analysis.

Finally, one major challenge we face is that across the 10 days of data (instead of three) changes in the appeal of news stories and the interest for the website could vary significantly from the baseline because of major newsworthy events in the country (unrelated to the outage). Though no significant events had occurred during the outage, two events of relevance affected the country several days after the outage: one related to a major natural phenomenon that attracted significant attention of news providers and web users, and the second a specific event connected to the refugee crisis that has assailed Europe in the past years. We then searched for the daily Google Trends scores for these newsworthy events, and for the country under analysis, using a combination of relevant keywords.

Though we found little evidence of searches related to the refugee event (flat or inexistent Google Trend scores across the period), we did find a significant peak of interest for the major natural event (during the weekend of October 26 and 27). To account for this increased search activity we added the Google trend scores to our model. We also collected daily Google Trends scores for the focal news website and for its major competitor in the country to account for the effect of general interest for the website under study, and news in general, and tested for the inclusion of these scores in our models. Figure 13 presents the evolution of the Google Trend Scores over the entire sample.

Tables 9, 10 and 11 provide the results of the best models for each variable of interest.[17] Few important notes on the estimation results are in order. First, the estimated effects for the hours of the outage, and the estimated effects for the hours after the outage (still during the outage day) were extremely robust and were always significant irrespective of the specification used for the other variables. The values of the estimates were also extremely similar across model specifications and to the results in our main three-day analysis, strengthening our confidence in our results.

---

[16] We also tested for dummies to account for early morning, afternoon and evening hours, though these were not significant in explaining differences from the baseline.

[17] We report the models with the full dummy specification and omit the specifications with the set of Google Trend scores. The inclusion of these scores as we will explain in this section was problematic in most cases and the models the models reported were always the best fitting across all estimated. However, we note that, if we are to exclude the eight daily dummies (and avoid the multicollinearity problem) we find typically that the Google Trend score for the website has a positive effect on the variables under analysis and that the score for the competitor website has a negative effect (lending also face validity to our model and results).

Second, specifications including daily dummies were always the best fitting models across all variables of interest. Including Google Trend scores introduced multicollinearity with such daily dummies. Plotting the scores and the estimates for the daily dummies for some of our variables we find correlated patterns as the daily dummies already reflect the change in traffic and news reading due to events other than the Facebook outage (e.g. the major natural event).

Finally, the estimated effect of the day after the outage was always significant and robust to the specification chosen. For the remaining daily dummies (corresponding to the second until the eighth day after the outage) the estimated values and their significance change depending on the specification used. This result reflects that, as we get further away from the outage, the differences from the baseline might be the result of changing interest in specific events and news (measured in our case through the Google Trend scores) and not associated to the outage per se, lending credence to our original three-day approach which excluded distant noisy observations. These results provide further support that any impact of the outage is likely short-lived, not leading to long-term changes in news reading behavior.

4.5.3 Analysis of Referrals from Other Social Networks

Another possible explanation for our findings could be external factors that may change interest across social media outlets. Because Facebook and Twitter referrals are highly correlated (correlation greater than 0.7), it is possible to use the evolution of one variable to control for underlying trends affecting social media in general. Therefore, we conducted an additional differences-in-differences analysis on Facebook related traffic by using our longer estimation sample of 10 days (240 observations) and modeling the difference of Facebook referrals from Twitter referrals (both variables taken as differences from their baselines). We then model these differences as a function of the outage dummy and the hourly and daily dummies. Our findings reveal that the drop of Facebook referrals during the outage, and the drop observed in the hours after the outage, are robust to this alternative specification. We also used Twitter traffic as a control variable for the remaining regression analyses we report, and we find that adding Twitter traffic does not change any of our results, making our findings robust to such alterative explanation (details available from the authors upon request).

5. Conclusion

A key debate amongst news providers today is whether online social networks drive individuals to news sites or, instead, the excessive time spent engaging with news and other content while browsing social media platforms creates scanning effects that divert traffic. The few studies that attempt to answer such a question are mostly survey based and often provide mixed results depending on the methodology and data employed (e.g., Bernoff and Li 2008).

With this work, we contribute to this debate by studying how a global Facebook outage affected content consumption at a major news website. By taking advantage of an exogenous Facebook disruption, we investigate the impact of Facebook on traffic at a major news website and at other websites. To the best of our knowledge, this is the first study that uses actual browsing data to determine the relationship between social networks and third party websites. Our work relates, conceptually, to previous studies on the short-term impact of brand unavailability (e.g. Campo, Gijsbrechts, and Nisol 2000) and, methodologically, to studies that rely on major platform outages to investigating the impact of word of mouth, willingness to generate content, and brand unavailability on future brand choices (Seiler et al. 2017, Zhang and Zhu 2011, Goldfarb 2006).

Our findings suggest that online social networks can have a complementary role in news consumption. In the absence of Facebook, we find a substantial loss of unique visitors and total page views that persisted even after the end of the disruption, although the subsequent impact was less prominent. The total revenue loss associated with the outage was significant, the focal website lost about $2,000 per hour during the outage, and $448 per hour after the outage, for a total of more than $21,000 lost in less than two days.[18] We note that this estimate is conservative as it only includes estimates for revenue generated by banner ads, whereas videos and home pages would fetch higher revenue per impression. If the relationship between news sites and Facebook were competitive, we would have observed instead an increase in traffic during the outage and a gain in revenue.

Although greater in magnitude, the promotional effect we observe is similar in direction to the effect previously attributed to news aggregators. In the case of online social networks, the positive

[18]News websites produce an average of $10 of revenue per thousand impressions (Johnston 2003).

relationship seems to result of a direct impact (via referrals or clicking on links) but also of an indirect spillover effect via complex mechanisms of social influence. Hence, unlike news aggregators that rely on users actively seeking news, online social networks can drive users to news websites even if users engaged in other social network activities rather than actively seeking news. In this regard, our research also relates to previous work exploring how consumers search for online content (Goldenberg, Oestreicher-Singer, and Reichman 2012) emphasizing the need for a better measurement of influence.[19]

Our results also reveal very different patterns in news engagement during the outage as on average users viewed fewer *home pages* and more *article pages,* indicating more in-depth reading. We find evidence of selectivity bias and behavioral changes during the outage that explain the identified aggregate differences in readership. The two results would imply very different recommendations for websites and policy makers: in one case, it is a matter of the type of population attracted to the website; in the other case, online social networks have a potential impact on the type and amount of content consumed. These results have important implications on the existing discussions regarding the impact of social networks on the formation of active and informed citizens, deemed by many a challenge in a connected world in which individuals might lack exposure to differing views (Iyengar and Hahn 2009).

News publishers today seem to recognize the need to harness the significant potential of online social networks (Somaiya, Isaac, and Goel 2015). Many news websites incentivize readers to register using social network accounts, and allow users the easy embedding of news article recommendations. These are indeed steps in the right direction. However, not all news outlets will be able to take advantage of social media. The outlets with most value added, specialized and original content are probably the ones that have the most to gain. Finally, content and news sites need to carefully consider the site's SEO performance and their brand strength. Further studies are required to address how these results might generalize to other online social media platforms might depend on the specific platform dynamics

[19] The weakness of direct referrals in predicting Facebook's true impact reminds the debate that dominated the area of internet advertising with many authors and practitioners insisting on clicks as the sole measure of advertising effectiveness. This perspective has now changed in light of studies demonstrating that banner ads can have an impact on consumers' attitudes independently of click-through (Briggs and Hollis 1997), and that these forms of advertising still have an impact despite people's conscious ad avoidance (Drèze and Hussherr 2003).

References

Ahlers, Douglas (2006), "News Consumption and the New Electronic Media," *Harvard International Journal of Press/Politics*, 11(1), 29–52.

Athey, Susan and Markus Mobius (2017), "The Impact of News Aggregators on Internet News Consumption: The Case of Localization," *Working Paper No. 3353*

Athey, Susan, Markus Mobius, and Jeno Pal (2017), "The Impact of News Aggregators on Internet News Consumption", *Stanford University Graduate School of Business Research Paper No. 17-8*.

Atkin, Charles K. (1972), "Anticipated Communication and Mass Media Information-Seeking," *Public Opinion Quarterly*, *36*(2), 188–199.

Asay, Matt (2014), "Facebook And Twitter Want To Be Each Other—But Shouldn't," *RW Labs*, [accessed June 15th, 2016, at http://readwrite.com/2014/02/12/facebook-twitter-news-friends-freaky-friday/].

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.

Barthel, Michael (2016), "5 key takeaways about the State of the News Media in 2016," Technical Report, Pew Research Centre.

Barthel, Michael, Elisa Shearer, Jeffrey Gottfried, and Amy Mitchell (2015) "The evolving role of News on Twitter and Facebook," Technical Report, Pew Research Centre.

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

Briggs, Rex and Nigel Hollis (1997), "Advertising on the Web: Is there Response Before Click-Through?" *Journal of Advertising research*, *37*(2), 33–45.

Campo, Katia, Els Gijsbrechts, and Patricia Nisol (2000), "Towards Understanding Consumer Response to Stock-outs," *Journal of Retailing* 76(2), 219–242.

Chiou, Lesley and Catherine Tucker (2017), "Content Aggregation by Platforms: The Case of the News Media," *Journal of Economics & Management Strategy*, March.

Chohan, Ashraf and Graham Francis (2014), "GOV.UK social sharing buttons: the first 10 weeks," [accessed 26th of January 2017, https://insidegovuk.blog.gov.uk/2014/02/20/gov-uk-social-sharing-buttons-the-first-10-weeks]

Cutler, Neal E. and James A. Danowski (1980), "Process Gratification in Aging Cohorts," *Journalism Quarterly*, 57(2), 269–76.

Cho, Daegon, Michael D. Smith, and Alejandro Zentner (2016), "Internet Adoption and the Survival of Print Newspapers: A Country-Level Examination," *Information Economics and Policy*, 37, 13–19

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.

Drèze, Xavier and François-Xavier Hussherr (2003), "Internet Advertising: Is Anybody Watching?" *Journal of Interactive Marketing*, 17(4), 8–23.

E-marketer (2013), "Digital to Account for One in Five Ad Dollars," *e-Marketer*, [accessed July 27, 2013, at http://www.emarketer.com/Article/Digital-Account-One-Five-Ad-Dollars/1009592].

Simmons, A. (2011), "The 2011 Social Media Consumer Trend and Benchmark Report," *Market Research Bureau*, New York, NY.

Goldenberg, Jacob., Gal Oestreicher-Singer and Schashar Reichman (2012), "The Quest for Content: How User-Generated links can Facilitate Online Exploration," *Journal Marketing Research*, 49(4), 452–468.

Goldfarb, Avi (2006), "The Medium-Term Effects of Unavailability*," Quantitative Marketing and Economics*, 4(2), 143–171.

Huang, Nina, Yili Hong, and Gordon Burtch (2016), "Social Network Integration and User Content Generation: Evidence from Natural Experiments," forthcoming *MIS Quarterly*

Iyengar, Shanto, and Kyu S. Hahn (2009), "Red media, Blue Media: Evidence of Ideological Selectivity in Media use," *Journal of Communication*, 59(1), 19–39.

James, Michael L., C. Edward Wotring, and Edward J. Forrest (1995), "An Exploratory Study of the Perceived Benefits of Electronic Bulletin Board Use and Their Impact on Other Communication Activities," *Journal of Broadcasting & Electronic Media*, 39(1), 30–50.

Jeon, Doh-Shin, and Nikrooz Nasr (2016), "News Aggregators and Competition Among Newspapers in the Internet," *American Economic Journal: Microeconomics,* 8(4), 91–114.

Ju, Alice, Sun Ho Jeong, and Hsiang Iris Chyi (2014), "Will social media save newspapers? Examining the effectiveness of Facebook and Twitter as news platforms," *Journalism Practice*, 8(1), 1–17.

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

Kayany, Joseph M. and Paul Yelsma (2000), "Displacement Effects of Online Media in the Socio-Technical Contexts of Households," *Journal of Broadcasting & Electronic Media*, 44(2), 215–229.

Kuss, Daria J. and Mark D. Griffiths (2011), "Online Social Networking and Addiction—A Review of the Psychological Literature," *International Journal of Environmental Research and Public Health*, 8, 3528–3552.

Lotz, Amanda. (Ed.) (2010), *Beyond prime time: Television programming in the post-network era*. Routledge.

Mahmood, Ammara and Sismeiro, Catarina (2017), "Will They Come and Will they Stay? Online Social Networks and News Consumption on External Websites," *Journal of Interactive Marketing*, 37, 117–132.

Matsa, Katerina Eva and Amy Mitchell (2014), "8 Key Takeaways about Social Media and News," *Pew Research Journalism Project*, [accessed 13 March 2015 at: http://www.journalism.org/2014/03/26/8-key-takeaways- about-social-media-and-news/

McLeod, Jack M., and Steven H. Chaffee (1973), "Interpersonal Approaches to Communication Research," *American behavioral scientist*, *16*(4), 469–499.

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

Nielsen (2011), "State of the Media: The Social Media Report," Technical Report, Q3.

Obermiller, Carl (1985), "Varieties of Mere Exposure: Effects of Processing Style on Affect Formation," *Journal of Consumer Research*, 12(1), 17–30.

Oeldorf-Hirsch, Anne (2011), "Engagement with News Content in Online Networks," doctoral dissertation, The Pennsylvania State University.

Olmstead, Kenneth, Amy Mitchell and Tom Rosenstiel (2011), "Facebook Is Becoming Increasingly Important," Pew Research Center Report.

Palme, Elia, Chrysanthos Dellarocas, Mihai Calin, and Juliana Sutanto (2016), “Attention Allocation in Information-Rich Environments: The Case of News Aggregators,” *Management Science*, 2543–2562.

Prawesh, Shankar and Padmanabhan Balaji (2014), “The ‘Most Popular News’ Recommender: Count Amplification and Manipulation Resistance,” *Information Systems Research*, 25(3), 569–589.

Robinson, John P., Kevin Barth, and Andrew Kohut (1997), “Social Impact Research Personal Computers, Mass Media, and Use of Time,” *Social Science Computer Review*, *15*(1), 65–82.

Rubin, Alan M. (1984), “Ritualized and instrumental television viewing,” *Journal of Communication*, 34, 67–77.

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

Savolainen, Reijo (1995), “Everyday life Information Seeking: Approaching Information Seeking in the Context of “Way of Life”, *Library & Information Science Research*, 17(3), 259-294.

Schaefer (2014), “Content Shock: Why Content Marketing is not a Sustainable Strategy,” *Marketing Social Media Humanity*, January 6th 2015, [accessed December 1, 2015, http://www.businessesgrow.com/2014/01/06/content-shock/#sthash.UCQjk67M.dpuf].

Schramm, Wilbur (1949), “The Nature of News,” *Journalism Quarterly*, 26(3), 259–269.

Seiler, Stephan, Song Yao, and Wenbo Wang (2017), “Does Online Word-of-Mouth Increase Demand? (and How?) Evidence from a Natural Experiment,” Stanford University Graduate School of Business Research Paper No. 15-62.

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

Smith, Michael D. and Rahul Telang (2009), “Competing with Free:The impact of Movie VBroadcasts on DVD Sales and Internet Piracy,” *Management Information Systems Quarterly*, 33(2), 312–338.

Somaiya, Ravi, Mike Isaac, and Vindu Goel (2015), “Facebook may Host News Sites’ Content,” New York Times, March 23, 2015, New York Times, [accessed April 1, 2015 at www.nytimes.com/2015/03/24/business/media/facebook-may-host-news-sites-content.html].

Susarla, Anjana, Jeong-Ha Oh, Yong Tan (2012) “Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube,” *Information Systems Research*, 23(1), 23–41.

Tewksbury, David, Andrew J. Weaver, and Brett D. Maddex (2001), “Accidentally Informed: Incidental News Exposure on the World Wide Web,” *Journalism and Mass Communication Quarterly*, 78(3), 533–554

Tanriverdi, H., and N. Venkatraman (2005), “Knowledge Relatedness and the Performance of Multibusiness Firms,” *Strategic Management Journal,* 26(2), 97–119.

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 Scepticism and Exposure,” *Media Psychology*, 7(3), 251–271.

Walther, J. B., B. Van Der Heide, D.S.Y. Kim, Westerman and S.T. Tong (2008), “The role of friends’ behavior on evaluations of individuals’ Facebook profiles: Are we known by the company we keep?” *Human Communication Research*, 34, 28–49.

Wenner, Lawrence A. (1985), “The Nature of News Gratifications,” *Media Gratifications Research: Current perspectives*, 171–193.

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

Seetharaman, Deepa (2016), "Facebook Revenue Soars on Ad Growth," *The Wall Street Journal,* [accessed 15th June, 2016, http://www.wsj.com/articles/facebook-revenue-soars-on-ad-growth-1461787856].

Marshall, Jack (2015), "Is Facebook Driving Less Traffic to Publisher's Site*?" The Wall Street Journal,* [accessed 5th December, 2015, http://www.wsj.com/articles/is-facebook-driving-less-traffic-to-publishers-sites-1449041402].

Vinciguerra, Mark (2017), "The increasing importance of revenue derived from your audience," New Your Press Service Report, [accessed 24th May, 2017, http://nynewspapers.com/wp-content/uploads/2017/04/NYPA-v16.pdf].

Xu, Jiao, Chris Forman, Jun B. Kim, and Koert Van Ittersum (2014), "News Media Channels: Complements or Substitutes? Evidence from Mobile Phone Usage," *Journal of Marketing* 78(4): 97–112.

Yadamsuren, Borchuluun and Sanda Erdelez, (2011) "Online News Reading Behavior: From Habitual Reading to Stumbling upon News," *Proceedings of the American Society for Information Science and Technology* 48(1), 1–10.

Zhang, Xiaoquan and Feng Zhu (2011), "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia," *The American Economic Review*, 101(4), 1601–1615.

Zhou, Renjie, Samamon Khemmarat, and Lixin Gao (2010), "The Impact of YouTube Recommendation System on Video Views," *IMC'10*, Melbourne, Australia.

Table 1: Summary Statistics (day of the outage, one day before, and one day after)

	Day 20 Sunday	Day 21 Monday	Day 22 Tuesday	Baseline Sunday	Baseline Monday	Baseline Tuesday
Unique Users	3,281,197	4,438,833	4,369,408	3,338,717	4,790,076	4,593,135
Page Views (All Pages) (no automatic refreshes)						
Arriving from Facebook	256,456	213,594	193,695	241,766	320,792	290,495
Arriving from Search	729,728	992,022	971,256	746,295	1,173,685	1,051,829
Arriving from Undefined	3,142,628	4,623,999	4,725,284	3,151,933	5,217,045	4,927,489
Arriving from Other Sources	3,190,702	4,256,703	3,710,218	3,032,023	4,674,326	3,936,373
Total	7,319,514	10,086,318	9,600,453	7,172,017	11,385,848	10,206,186
Home Page Views (no automatic refreshes)						
Arriving from Facebook	233	113	4,474	113	894	306
Arriving from Search	379,779	592,577	612,691	371,530	702,245	667,921
Arriving from Undefined	2,220,163	3,128,257	3,209,123	2,210,241	3,588,113	3,322,492
Arriving from Other Sources	384,252	415,592	414,099	408,471	564,307	477,828
Total	2,984,427	4,136,539	4,240,387	2,990,355	4,855,559	4,468,547
Per User Metrics						
Page Views per User (All Pages)	2.23	2.27	2.20	2.15	2.38	2.22
Home Page Views per User	0.91	0.93	0.97	0.90	1.01	0.97

Table 2: Number of Visitors in Each Cohort and their Return Likelihoods

	7-9am Visitors		5-7am Visitors		Loyal 7-9am Visitors		Loyal 5-7am Visitors	
	Number	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**
Total Number of Users	318,092	100%	79,132	100%	2,658	100%	481	100%
Users who returned on Monday 21 during the same hours	43,072	13.5%	9,413	11.9%	1,421	53.5%	394	81.9%
Users who returned on Monday 28 during the same hours	64,331	20.2%	9,236	11.6%	1,855	69.8%	340	70.7%
Users who returned on Monday 21 and 28 (visited the site the three Mondays) during the same hours	22,687	7.1%	4,864	6.2%	1,059	39.8%	294	61.1%

Table 3: Regression Results for the Main News Website

Focal Variable Modeled as Differences From Baseline Per Hour	Change During the Outage: $\widehat{\beta_1}$		Change After the Outage: $\widehat{\beta_2}$		Adjusted R2
	Estimate (Std. Errors)	Percentage Change	Estimate (Std. Errors)	Percentage Change	
Page Views (All Pages)	$-201{,}813.70^{***}$ (22,458.47)	-44.20%	$-44{,}797.72^{***}$ (7,407.21)	-8.71%	0.58
Unique Users	$-70{,}095.99^{***}$ (6,658.42)	-38.49%	$-9{,}100.06^{***}$ (2,196.07)	-4.09%	0.62
Referrals					
From Facebook	$-3{,}955.52^{*}$ (1,970.60)	-43.58%	$-5{,}215.23^{***}$ (668.40)	-35.91%	0.45
From Search	$-29{,}470.47^{***}$ (3,113.15)	-61.83%	$-4{,}654.33^{***}$ (1,026.77)	-8.72%	0.57
Undefined	$-142{,}020.45^{***}$ (13,444.68)	-69.43%	$-13{,}065.84^{**}$ (4,434.30)	-5.41%	0.61
Home Page Traffic					
Home Page Views	$-161{,}375.99^{***}$ (14,442.90)	-80.75%	$-15{,}892.04^{**}$ (4,763.53)	-7.12%	0.64
Home Page Views from Search	$-26{,}365.81^{***}$ (2,400.62)	-82.17%	$-3{,}160.79^{***}$ (791.77)	-9.62%	0.63
Home Page Views from Undefined	$-117{,}821.00^{***}$ (10,590.87)	-82.62%	$-9{,}278.47^{***}$ (3,493.06)	-5.59%	0.63
Per User Metrics					
Pages per User (All Pages)	-0.191^{**} (0.057)	-7.74%	-0.128^{***} (0.019)	-5.80%	0.41
Content Pages per User (Excludes Home Pages)	0.518^{***} (0.042)	37.24%	-0.108^{***} (0.014)	-8.48%	0.78
Home Pages per User	-0.707^{***} (0.040)	-65.78%	-0.018 (0.013)	-1.93%	0.81

*significant at 0.05; ** significant at 0.01; *** p-value < 0.001.
Note: We report the absolute change of the hourly values for each variable, the robust standard errors (within brackets), the percentage change considering the baseline values for that variable during the corresponding period, that is, during and after the outage depending on the estimated effect.

Table 4: Results for the Four Category Specific Regression Models (Dependent Variable = Total Page Views to each news category)

	Local News (Model R^2 = 0.65)	**Sports News** (Model R^2 = 0.34)	**Women Issues** (Model R^2 = 0.57)	**Health News** (Model R^2 = 0.54)
Change During the Outage: $\widehat{\beta_1}$	**$-50{,}427.36^{*}$**	$-46{,}547.00$	$-10{,}252.14$	**$-3{,}257.86^{***}$**
Change After the Outage: $\widehat{\beta_2}$	**$64{,}616.58^{***}$**	**$35{,}671.97^{*}$**	**$-18{,}453.51^{***}$**	**$-5{,}050.50^{*}$**

* significant at 5%; **significant at 1%; *** significant with p-value<0.001 (the remaining estimates significant at 10%)

Table 5: Comparing Returning and Non-Returning Visitors within Loyal Cohorts

	Difference of Readership for the Loyal 7-9am Visitors§	Difference of Readership for the Loyal 5-7am Visitors§
Page Views (All Pages)	-1.34^{***}	-0.05
Home Page Views	-0.28^{**}	-0.29
Content Page Views	-1.06^{***}	0.24

* significant at 5%; **significant at 1%; *** significant with p-value<0.001;
§ we compare those who returned on October 21 with those who did not (a negative value means that users who *did not return* on October 21 viewed fewer pages on October 14, than those who returned on October 21)

Table 6: Comparing the Behavior of Loyal Visitors over Time
Comparison Using the Cohort Corresponding Time-Intervals§

	Behavior of the 1,421 visitors of the Loyal 7-9am Visitors who visit the website on Day 14 and Day 21			Behavior of the 394 visitors of the Loyal 5-7am Visitors who visit the website on Day 14 and Day 21		
	October 14	October 21	Difference	October 14	October 21	Difference
Page Views (All Pages)	5.12	4.30	-0.82^{***}	4.70	4.94	0.25
Home Page Views	2.42	0.90	-1.52^{***}	2.33	2.54	0.21
Content Page Views	2.70	3.40	0.70^{***}	2.37	2.41	0.04

* significant at 5%; **significant at 1%; *** significant with p-value<0.001.
§ the corresponding time-intervals are from 7:00 AM until 9:00 AM for the 7-9am Visitors and from 5:00 AM until 7:00 AM for the 5-7am Visitors)

Table 7: Comparing the Behavior of the Loyal 7-9am Visitors and the Loyal 5-7am Visitors who Visited the Website All of the Three Mondays§

	Loyal 7-9am Visitors (N = 1,059)				Loyal 5-7am Visitors (N = 294)			
	Day 21		Day 28		Day 21		Day 28	
	Average	Difference from Baseline	Average	Difference from Baseline	Average	Difference from Baseline	Difference from Baseline	Average
Page Views (All Pages)	4.42	-0.78^{***}	5.41	0.21	5.15	0.43	4.87	0.15
Home Page Views	0.91	-1.57^{***}	2.51	0.03	2.76	0.32	2.55	0.11
Content Page Views	3.51	0.79^{***}	2.90	0.18	2.40	0.11	2.33	0.04

* significant at 5%; **significant at 1%; *** significant with p-value<0.001.
§ baseline = Monday, October 14

Table 8: Regression Results Using Data from Other Websites for the Validation Analysis (we report the estimates for each model and indicate their statistical significance)

	Smaller News Sites	E-commerce Websites	Technology Blogs
Page Views (All Pages)			
Change During the Outage: $\widehat{\beta_1}$	**-5012.93***	-7,485.93	-28.04
Change After the Outage: $\widehat{\beta_2}$	**-4073.86****	2,523.11	-171.25
Unique Users			
Change During the Outage: $\widehat{\beta_1}$	**-2644.29*****	-1,296.61	-40.82
Change After the Outage: $\widehat{\beta_2}$	**-2323.89*****	916.14	-138.14
Referrals			
From Facebook			
Change During the Outage: $\widehat{\beta_1}$	**-559.43***	2.29	3.58
Change After the Outage: $\widehat{\beta_2}$	**-1,846.83*****	**64.52****	**-41.73***
From Search			
Change During the Outage: $\widehat{\beta_1}$	**-769.61***	73.32	-21.54
Change After the Outage: $\widehat{\beta_2}$	**-536.79***	397.10	-30.32
Undefined			
Change During the Outage: $\widehat{\beta_1}$	**-399.43****	20.61	-8.50
Change After the Outage: $\widehat{\beta_2}$	**-256.40****	635.62	-70.63
Pages Per User (All Pages)			
Change During the Outage: $\widehat{\beta_1}$	-0.08	0.04	0.07
Change After the Outage: $\widehat{\beta_2}$	-0.05	0.07	**-0.14***

* significant at 5%; ** significant at 1%; *** significant with p-value<0.001.

Table 9: Regression Results Using the Estimation Sample of 10 Days for the Validation Analysis (we report the estimates for each model and indicate their statistical significance)

	Number of Pages (All Pages) (Model $R^2 = 0.42$)	**Number of Users** (Model $R^2 = 0.49$)	**Facebook Referrals** (Model $R^2 = 0.36$)	**Undefined Referrals** (Model $R^2 = 0.55$)	**Search Referrals** (Model $R^2 = 0.47$)
Intercept	29,825.00***	7,079.00***	1,474.20*	7,485.00**	1,437.70
Change During the Outage $\widehat{\beta_1}$	-197,500.00***	-70,740.00***	-5,981.70*	-146,183.00***	-27,173.70***
Change Between 9 and 11 AM	-121,970.00***	-30,370.00***	-7,676.20***	-71,030.00***	-12,615.40***
Change Between 12 and 2 PM	-80,207.00**	-21,850.00***	-12,223.50***	-23,831.00**	-6,028.70*
Change Between 3 and 5 PM	-65,818.00**	-10,740.00	-8,630.50***	-15,303.00	-6,313.70*
Change Between 6 and 8 PM	-74,324.00**	-12,760.00*	-6,540.80***	-21,034.00*	-8,591.70***
Change Between 9 and 11 PM	-71,575.00**	-13,930.00*	-9,973.50***	-17,806.00*	-5,349.70*
Change 1 Day After Outage	-67,782.00***	-15,770.00***	-6,451.30***	-17,835.00***	-5,989.40***
Change 2 Days After Outage	-31,496.00**	-8,296.00**	-4,004.80***	-10,421.00**	-4,984.30***
Change 3 Days After Outage	-23,849.00*	-8,028.00**	-6,115.90***	-5,868.00	-4,466.20***
Change 4 Days After Outage	-42,195.00***	23.40	-2,692.70**	-1,728.00	-2,277.50*
Change 5 Days After Outage	-50,544.00***	-9,924.00***	-2,501.00*	-11,697.00**	-5,061.60***
Change 6 Days After Outage	32,194.00**	12,400.00***	2,053.40*	11,642.00**	3,561.40**
Change 7 Days After Outage	-12,932.00	-5,060.00	-3,654.20***	-8,091.00*	1,579.30
Change 8 Days After Outage	-55,262.00***	-9,747.00***	-3,361.70**	-11,334.00**	-3,826.60***
Lagged Dependent Variable	--	-0.02**	--	--	--

*significant at 0.05; ** significant at 0.01; *** p-value < 0.001.

Note: We report the absolute change of the hourly values for each variable and the p-values; when the inclusion of lagged variables does not improve the AIC of the model we omit these from the estimation; we also tested for the inclusion of other variables including Google trend Scores for the website under analysis, its main competitor and major newsworthy events occurring in the country; we also tested for trends and we run the models also in logs and we obtained substantively the same results (results available from the authors upon request).

Table 10: Regression Results Using the Estimation Sample of 10 Days for the Validation Analysis – Home Page Metrics (we report the estimates for each model and indicate their statistical significance)

	Home Pages (Model R^2 = 0.55)	**Home Pages from Search** (Model R^2 = 0.55)	**Home Pages from Undefined** (Model R^2 = 0.55)
Intercept	10,8027.00**	1,750.00***	7,891.00**
Change During the Outage $\widehat{\beta_1}$	-164,600.00***	-26,150.00***	-119,300.00***
Change Between 9 and 11 AM	-81,196.00***	-10,570.00***	-59,990.00***
Change Between 12 and 2 PM	-27,990.00*	-2,714.00	-14,890.00
Change Between 3 and 5 PM	-20,580.00	-3,088.00	-11,250.00
Change Between 6 and 8 PM	-27,200.00*	-4,995.00**	-17,920.00*
Change Between 9 and 11 PM	-22,490.00*	-3,820.00*	-14,480.00
Change 1 Day After Outage	-22,750.00***	-4,492.00***	-15,790.00***
Change 2 Days After Outage	-7,727.00	-2,597.00***	-5,722.00
Change 3 Days After Outage	-7,087.00	-1,445.00	-5,888.00
Change 4 Days After Outage	-10,860.00*	-2,185.00**	-8,564.00*
Change 5 Days After Outage	-17,800.00***	-3,581.00***	-13,080.00***
Change 6 Days After Outage	11,960.00*	1,325.00	8,111.00*
Change 7 Days After Outage	-12,440.00*	-1,504.00	-10,510.00**
Change 8 Days After Outage	-21,620.00***	-3,548.00***	-17,820.00***
Lagged DV (one period)	-0.03*	-0.04**	-0.02

*significant at 0.05; ** significant at 0.01; *** p-value < 0.001.
Note: We report the absolute change of the hourly values for each variable and the p-values; when the inclusion of lagged variables does not improve the AIC of the model we omit these from the estimation; we also tested for the inclusion of other variables including Google trend Scores for the website under analysis, its main competitor and major newsworthy events occurring in the country; we also tested for trends and we run the models also in logs and we obtained substantively the same results (results available from the authors upon request).

Table 11: Regression Results Using the Estimation Sample of 10 Days for the Validation Analysis – Per User Metrics (we report the estimates for each model and indicate their statistical significance)

	Pages Per User (All pages) (Model $R^2 = 0.58$)	**Home Pages Per User** (Model $R^2 = 0.59$)	**Content Pages Per User** (Model $R^2 = 0.63$)
Intercept	0.0658	0.0131	-0.2019***
Change During the Outage $\widehat{\beta_1}$	-0.2178**	-0.7232***	0.4312***
Change Between 9 and 11 AM	-0.2459***	-0.1713***	-0.1699***
Change Between 12 and 2 PM	-0.1693**	-0.0230	-0.1520**
Change Between 3 and 5 PM	-0.2128***	-0.0455	-0.1642***
Change Between 6 and 8 PM	-0.2554***	-0.0750*	-0.1619***
Change Between 9 and 11 PM	-0.2607***	-0.0565	-0.1791***
Change 1 Day After Outage	-0.2156***	-0.0427**	-0.1500***
Change 2 Days After Outage	-0.1256***	0.0167	-0.1348***
Change 3 Days After Outage	-0.0704*	0.0071	-0.0667**
Change 4 Days After Outage	-0.2413***	-0.0502**	-0.1731***
Change 5 Days After Outage	-0.2018***	-0.0447**	-0.1608***
Change 6 Days After Outage	0.0484	0.0204	0.0228
Change 7 Days After Outage	-0.0719*	-0.0337*	-0.0527*
Change 8 Days After Outage	-0.2137***	-0.0624***	-0.1362***
Lagged DV (one period)	0.0482*	--	0.2737***

*significant at 0.05; ** significant at 0.01; *** p-value < 0.001.
Note: We report the absolute change of the hourly values for each variable and the p-values; when the inclusion of lagged variables does not improve the AIC of the model we omit these from the estimation; we also tested for the inclusion of other variables including Google trend Scores for the website under analysis, its main competitor and major newsworthy events occurring in the country; we also tested for trends and we run the models also in logs and we obtained substantively the same results (results available from the authors upon request).

Figure 1: Total Hourly Page Views (Excluding Refreshes) and Total Unique Users from October 13 until October 29 (408 hourly observations across the 17 days of our dataset)

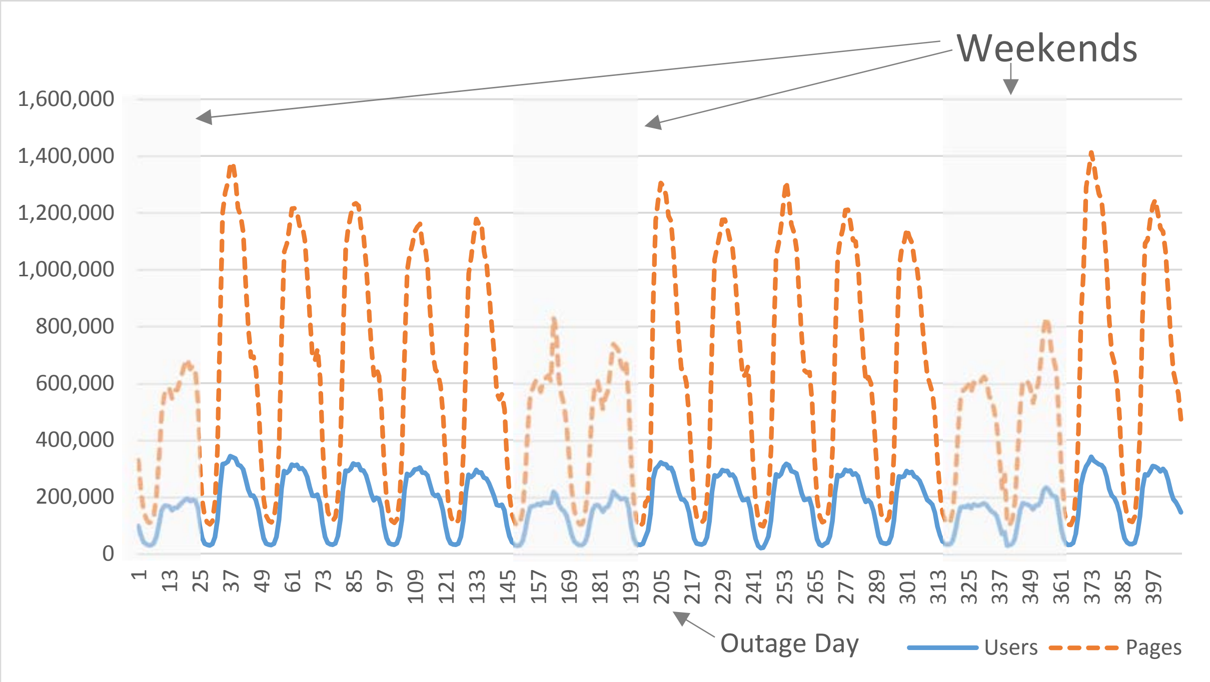


Figure 2: Timeline of the Analysis

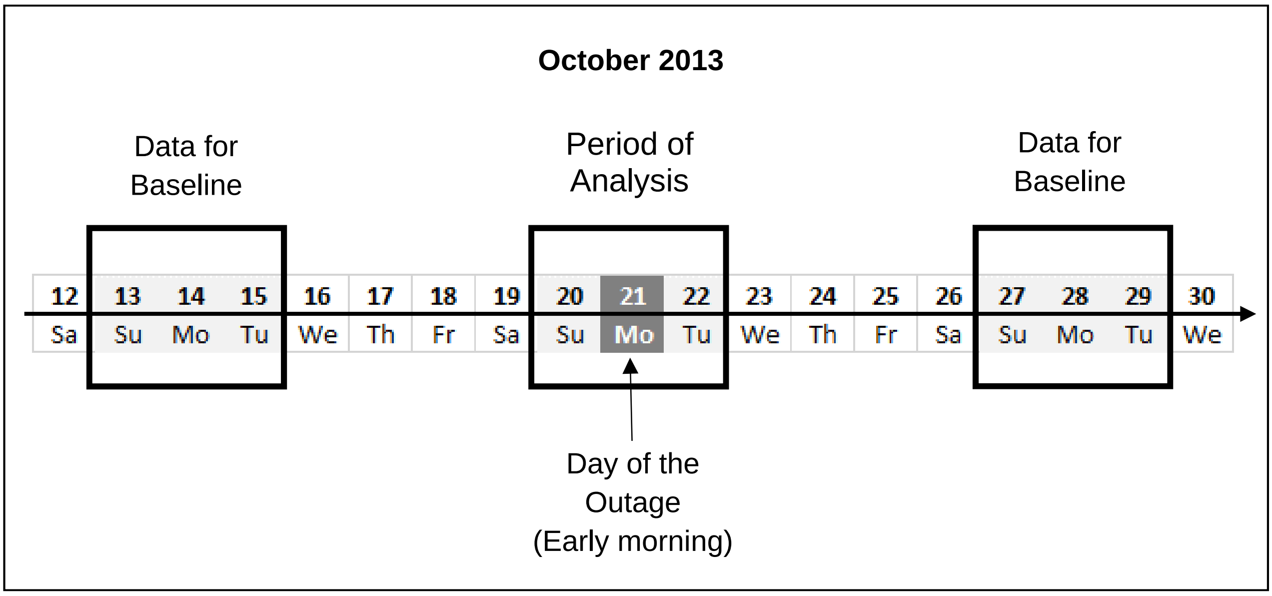


Figure 3: Unique Users Against Baseline
(shaded area corresponds to outage hours)

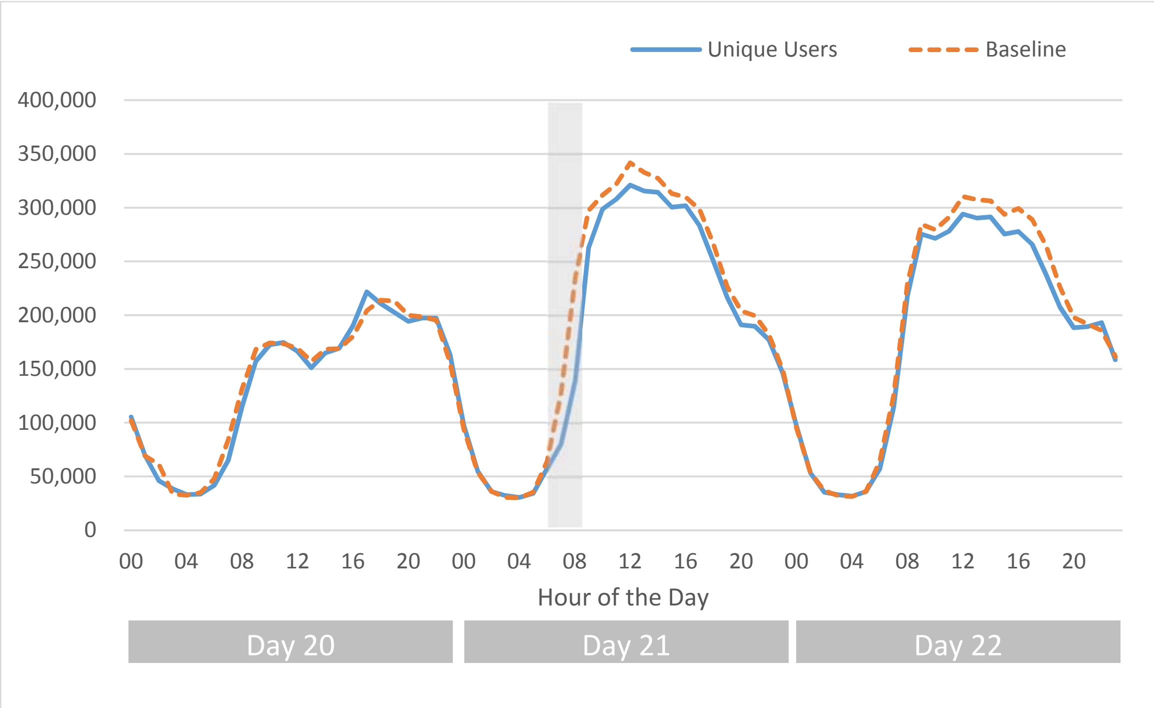


Figure 4: Total Page Views Against Baseline
(shaded area corresponds to outage hours)

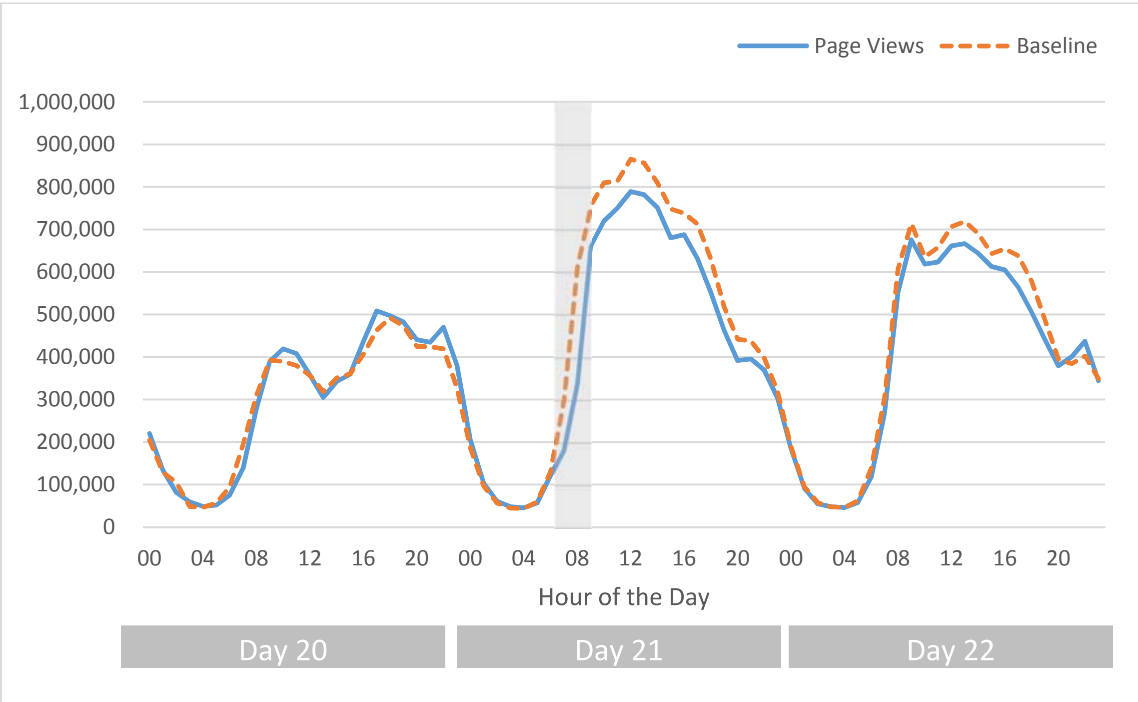


Figure 5: Total Page Views by Category to the Main News Site in our Analysis

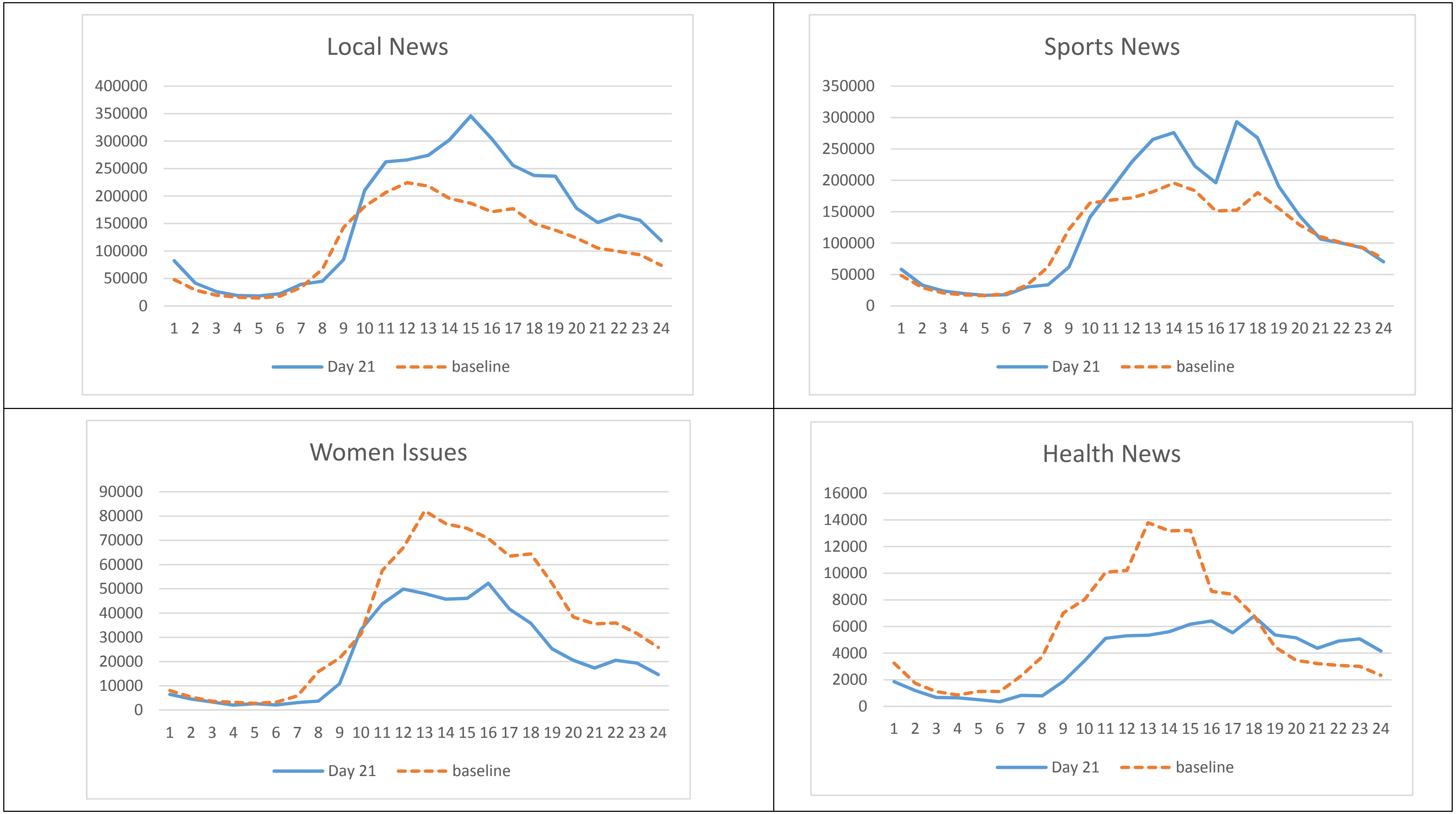


Figure 6: Home Page Views per User Against Baseline (shaded area corresponds to outage hours)

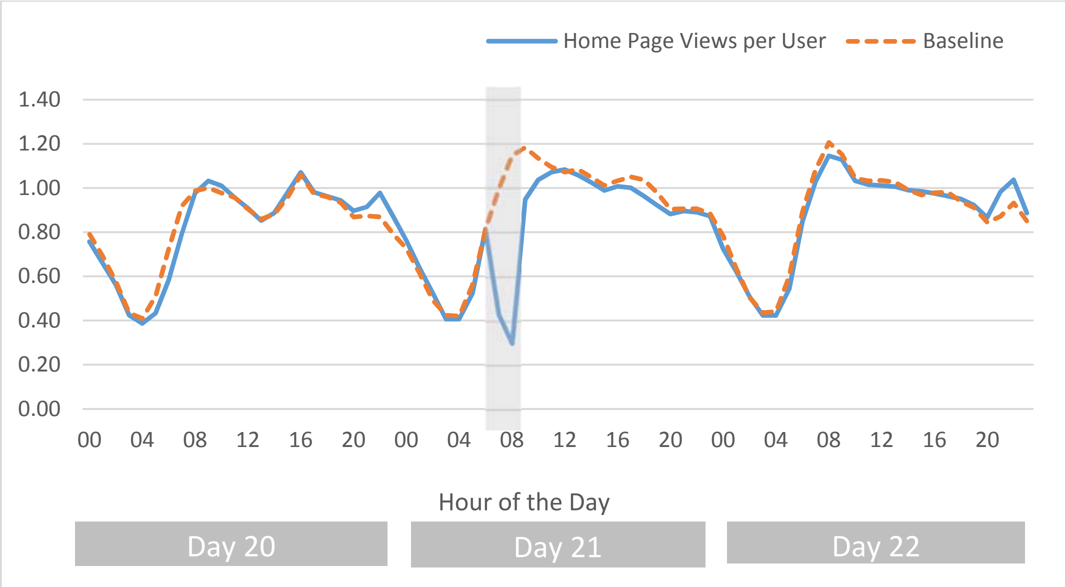


Figure 7: Content Page Views per User Against Baseline (shaded area corresponds to outage hours)

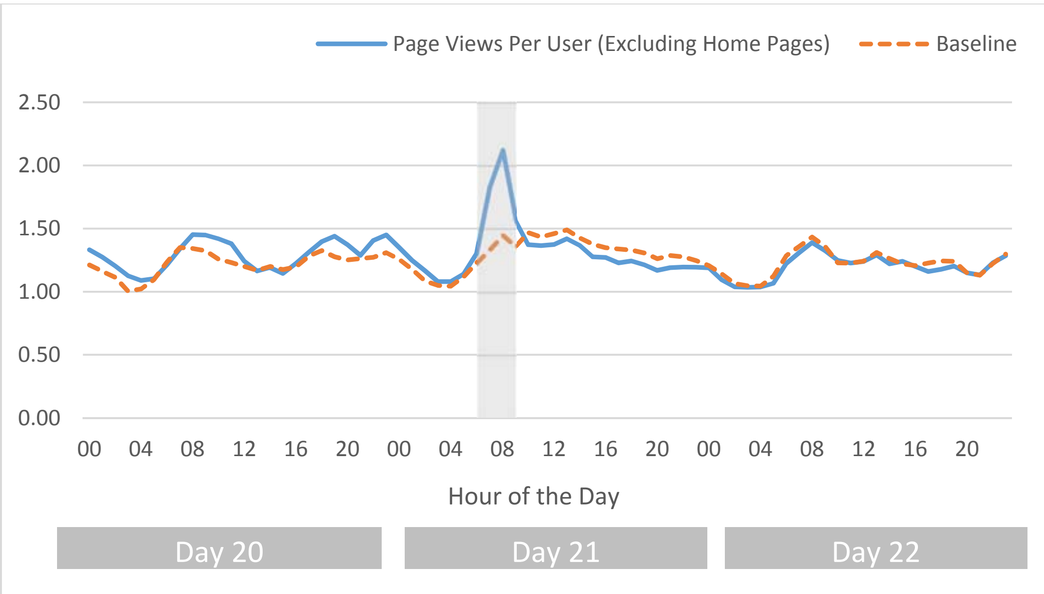


Figure 8: Return likelihood for the 7-9am Visitors during the outage hours for day 21 and day 28

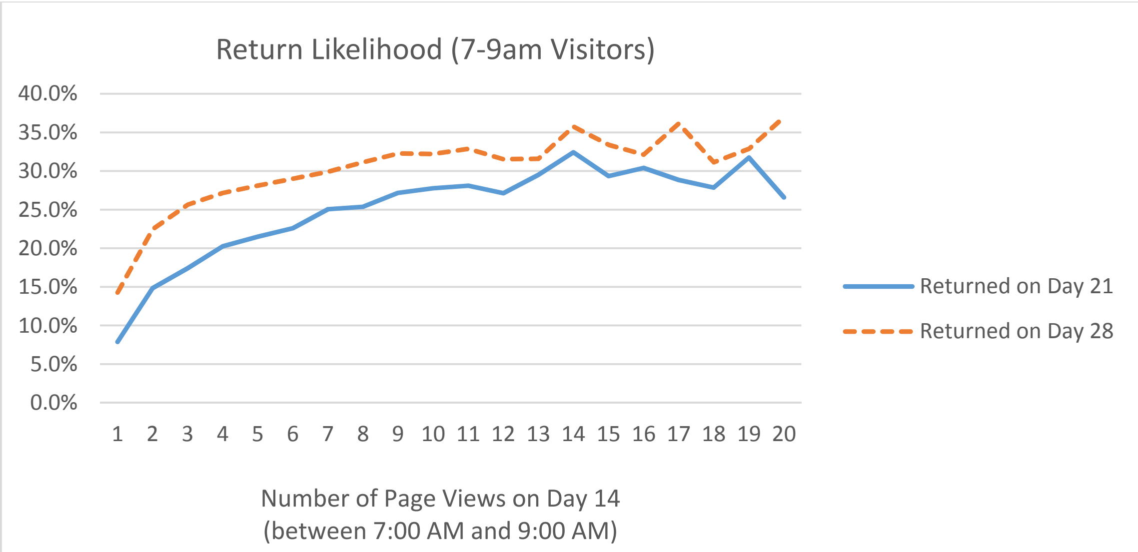


Figure 9: Return likelihood for the 5-7am visitors during the outage hours for day 21 and day 28

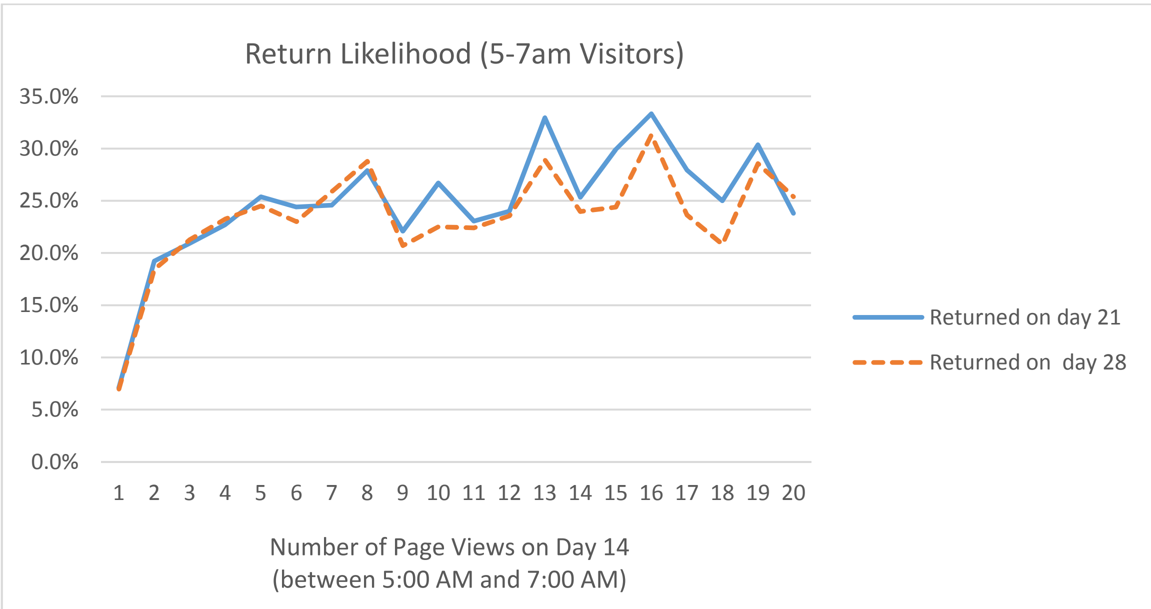


Figure 10: Total Page Views Against Baseline for Two Additional News Websites (shaded area corresponds to outage hours)

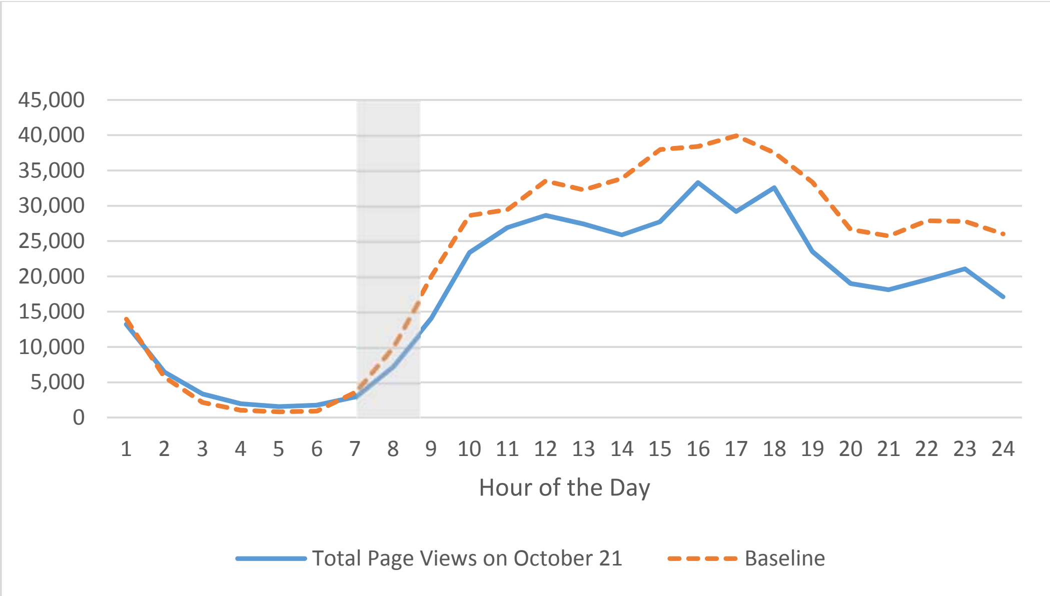


Figure 11: Total Unique Users Against Baseline for Two Additional News Websites (shaded area corresponds to outage hours)

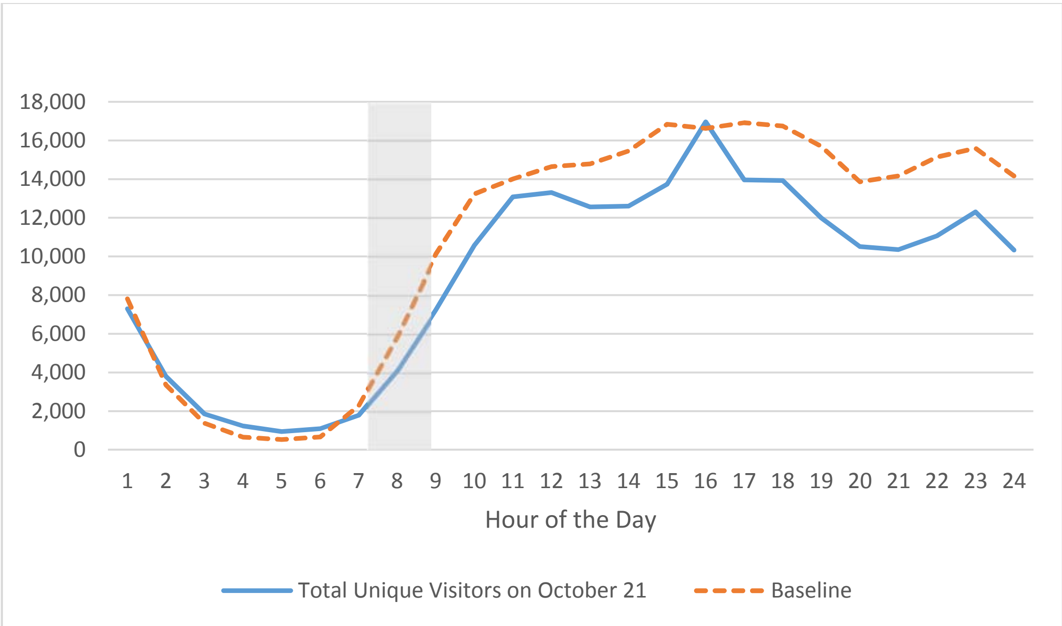
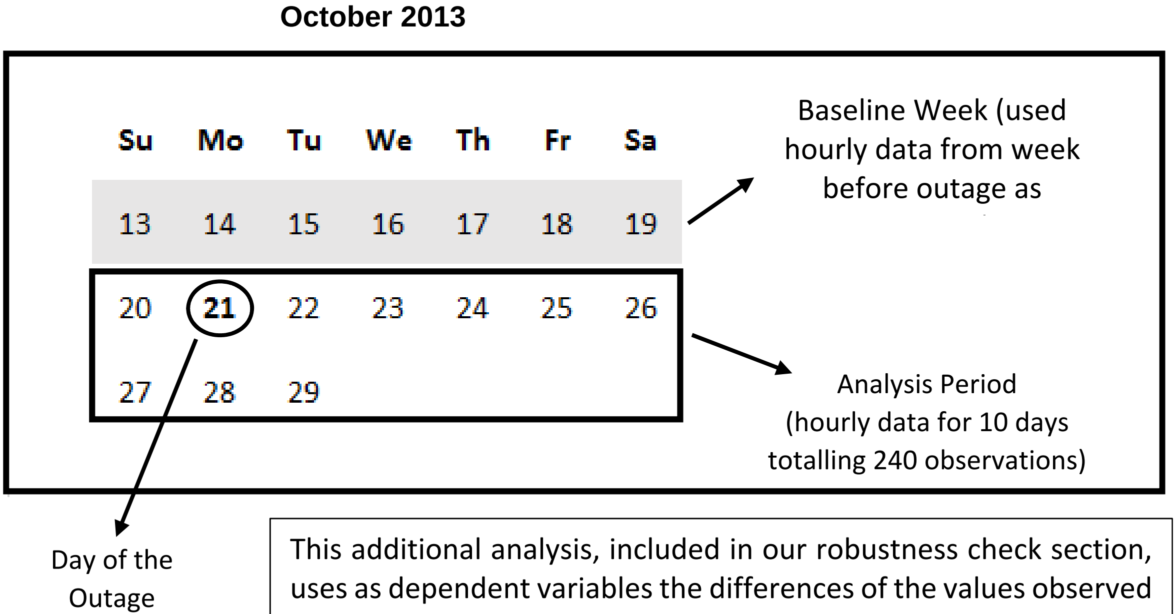


Figure 12: Baseline and Estimation Periods for the Additional Validation Analyses Using the Data from Focal Website



This additional analysis, included in our robustness check section, uses as dependent variables the differences of the values observed for the Analysis Period and the baseline values observed the corresponding hour of the corresponding day from the baseline week. For example, the variable Visitors at 8 AM of the 27th of October is computed as the total number of visitors at 8 AM of day 27 minus the total number of visitors observed at 8 AM on day 13. The 27th was a Sunday, and hence we use the data from the Sunday of our baseline, which is day 13th, to compute the differences. For each variable to be studied we built a dataset of 240 differences from baseline which we use to estimate the models.

Figure 13: Google Trend Scores