Analysis of referral channels on online news consumption

Husna Sarirah Husin
Universiti Kuala Lumpur
Malaysia Institute of Information Technology (MIIT)
Malaysia
sarirah@unikl.edu.my

Nadiah Ruza
Victoria University,
Melbourne Australia
nadiahruza@gmail.com

***Abstract*—** When users visit a page, the browser records the referrer that sent the user to the page. If no referrer is recorded, it means that the user typed in the URL of the page. Referrer can comes from various sites such as search engine site, social media sites or other sites. In this paper, we analyze how referrals have any impact on news consumption. We use a dataset of web server logs from an online newspaper for 14 days from 13-27 January 2017. We found that half of the referrals are from Facebook. In terms of contents, there is a significant strong relationship between referrer and article content. The Facebook referrals use Android devices to access the pages and Sunday has the highest number of page accessed. Our findings can be an input for the news organization in their strategy to position themselves as the leader in online news.

Keywords— online news; referral; user behavior; web server logs

I. INTRODUCTION

Typically, users access a website by typing the URL. Besides that, the users may be referred by other channels such as search engine, social media, online advertisements or other websites, known as referrals. Referrals are one of the effective marketing strategies. Using referrals, companies can build relationship with their customers, increase their brand presence and expand their business with minimal cost.

In the case of web traffic, a referral can be defined as recommendation from one website to another. Users who visit the website based on referral, particularly through social sharing are interested with the website and potentially can be a returning user. Apart from testimonials, referrals can be powerful factor especially if the services offered are highly intangible and subjective [1].

The study of referrals from different channels is important for web manager to identify which referral channel bring the most traffic. Thus, the web manager can plan the marketing strategies to cater to the referred traffic. In addition, he can monitor the rate of online engagement and improve the brand presence of the website. Previous literature has shown visitors that are from different referral channels have different characteristics and behavior in terms of page views [2] or duration of the time spent on the website [3]

Not much attention has been paid to the how referrals have any impact on news website. Thus, this study aims to investigate the user behavior accessing the online newspaper based on the different referral channels, the types of content that they accessed and the different devices they use to access the news web site.

The next section on related work will discuss literature on referral channels and social referral. Then, we explain the data and methodologies used in this study. The results and findings are described in section IV and finally we present the discussion of the results and propose future work.

II. RELATED WORK

A. Referral channels

Visitors to a web site may come from different sources. These sources can be categorized into three major source or channels called, direct, referral and search [2], [3]. Direct visitors are the ones that type in the web site address or the URL. Referral visitors are from other web sites that provide a link to the destination web site. Search visitors use the search engine to get to the web site using keyword.

Direct visits to a website may come from repeat visitors. Those who visit an e-commerce site directly are incline to be existing customers, because consumers rarely type the URL on their first visit to a website [4]. Therefore, a direct visit could mean high probability of return visit; which is a key measurement of websites' performance [5] and a hint of a positive attitude toward that website [6]. Moreover, as customers make more visits to a website, they will acquire more information, in turn will increase the prospect of making future visits. Subsequently, when a user visits a website directly, he/she have high tendency of making a purchase.

On the other hand, consumers who visit an online store through search advertising are expected to exhibit more goal-directed search behavior [7]. Generally, these types of consumers enter the website to compare different brands or prices. Previous work has suggested that search advertising has a positive impact on sales [8], [9], however, the effect to be lower than for direct visits since these consumers are still at the information-gathering stage. As these shoppers gather more information

during each store visit, purchasing may occur after a series of store visits [7] [10].

Customers who enter the websites via referral are expected to be the least likely to purchase during the current visit. These shoppers may include hedonic browsers who enter a retailer without a particular product or even product category in mind, or they may be knowledge-building visitors, who have no intention of buying and are simply in the store to gather information about the products available [7].

A study found that the average returning visitor ratio from referral traffic source has the highest percentage of 25% compared to 17% from search traffic source and 19% from direct traffic source. The results showed that different traffic sources have significantly different returning visitor ratio especially between referral traffic source and the direct referral and search referral sources [11] . There are studies examining the effectiveness of these referral channels. Generally, direct visits are the most effective methods to cultivate return visits to a website [4]. On the other hand, referral or display ads has shown to be useful to increase search clicks and subsequent purchases in an e-commerce site [12].

B. Referrals from social media

Social referrals in an online context is distinctive because it involves direct communication between individuals with social connections to each other [13]. Thus, social referrals are more personal as compared to seller feedback or online review that are posted publicly [13]. The propagation of mobile devices and personal communication tools make it feasible to share information with wider audience, crossing geographical location and simplify sharing information process. These advantages attract the attention of companies interested in leveraging existing consumers' social networks to acquire potential consumers [13], [14].

Social media interactions and traffic to website can complement each other. More often than not, people who click on the shared link will subsequently visit the news website. Nowadays, people do not depend on homepages and search engines, instead news discovery is through social media and direct messaging on mobile applications [15]. Studies showed that consumers who are referred through social referrals outpace consumers referred through search engines or online advertisements [16].

Users in social media such as Facebook and Twitter share their posts using social referrals. As the number of social media is increasing rampantly, the referrals are becoming more important source of incoming traffic for news site. The news organization are now realizing the significance of social media because it can attract more visitors to the news website compared to other external referrer [17]. In America, newspapers that have weekday circulation of more than 100,000 run a Facebook page and spread links to their articles through the network to get more users to read their contents [18]. The amplified traffic from Facebook referral carries many potential advantages to the news site. From these increased traffic, news publisher can increase its advertising rates or convert the readers into subscribers [19].

Social media can influence media consumption by providing an avenue for news organizations to display their contents in a single location. Users can select variety of news from an array of news provider. This signifies a change from conventional ways of getting news of people getting news from a trusted source to users selecting news from a wide range of sources reckoned by friends to be interesting or important [20]. In the Facebook news feed, the article comprises of headline, couple of sentences about the news story, a picture and social endorsement cues such as comments or "likes" [21] . Social media has the potential to help the news consumer deal with information overload through socially mediated information selection and organization [22]

Facebook being amongst the most popular social networks is the front runner in news consumption because Facebook users spend more time on daily basis; therefore consuming more news stories while browsing their news feeds [23]. According to a research by Pew Research Center, about 66 percent of adults surveyed in America get their news on Facebook [24]. Another study claimed that the performance of a news site is highly dependent on how it performs in Facebook's news feed [19]. Moreover, with the widespread use of mobile devices, such as smartphones and tablets, users can access to online services anywhere and anytime [25].

There are a number of research deliberating the consequences of news exposure in Facebook and how the Facebook referrals affect users' engagement to the site [16], [26], [27]. The news content that user found on Facebook could lead to further news seeking and exposure. The users are likely to either use Facebook as a starting point to find more information by clicking the link or seek the information elsewhere [27]. When user scroll through Facebook news feed, it is suggested that this behavior encourage flow [28] which will eventually could lead users to recognize the existing news content presented, even though their initial content was not about reading news when they logged in to their Facebook account [29]

Social media should be expected to increase users' exposure to a variety of news and politically diverse information [20]. A study reveals that the social media users typically access news from entertainment, sports and local news compared to political, business, international news [25], [30]. Another study also suggest that most common news people see is entertainment news with 73% of Facebook users regularly see this kind of content on the site [31].

Although Facebook is an important source of website referrals for many news outlets, but the users who arrive via Facebook spend far less time and consume far fewer pages than those who arrive directly [31]. Direct visitors spend, on average, 4 minutes and 36 seconds per visit. That is roughly three times as long as those who wind up on a news media website through a search

engine (1 minute 42 seconds) or from Facebook (1 minute 41 seconds). Direct visitors also view roughly five times as many pages per month (24.8 on average) as those coming via Facebook referrals (4.2 pages) or through search engines (4.9 pages). And they visit a site three times as often (10.9) as Facebook and search visitors [32]. The data also suggest that converting social media or search eyeballs to dedicated readers is difficult to do. Most people that arrived at one of these popular news sites used only one of the three modes – suggesting that, at least on desktop/laptops, individuals tend to come to these news sites using a single method [32].

Facebook referrals generate only marginal traffic on news organization website [18]. Those who follow links from Facebook spend less time on news websites than direct visitors' exposure to Facebook news posts, thus, does not seem to significantly boost the use of other news sources.

III. DATA AND METHODOLOGY

The dataset consists of Web server logs from a Malaysian daily online newspaper called *Berita Harian* for 14 days period (13 – 27 January 2017). The Web server log contains information of date and time of access, URL that user requests, the HTTP status code representing that the request is successful. We only chose HTTP status code with 200 as it denotes a successful landing to the web page. The following is the list of the HTTP status codes

- 1xx Informational responses.
- 2xx Success.
- 3xx Redirection.
- 4xx Client errors.
- 5xx Server error.
- Unofficial codes.

The size of the page requested can also be found in the web server logs, as well as the referrer path. From the referrer path, we can identify the origins of the path; as in the previous page before the user make request to the page.

A user agent string determines the platform and device of the user and finally the IP address of the user can be identified in the last part of the web server logs. Table 1 shows the sample of the *Berita Harian* log.

The Web mining technique is adopted as the methodology in this study. Web usage mining consists of three main stages are preprocessing, user identification, and session identification [33]. In preprocessing, the logs are cleaned and filtered to ensure that only real requests to the news articles captured. Among the items that removed from the data set are unsuccessful requests, image files, automated requests, RSS requests, and bots.

TABLE I Sample of *Berita Harian* log entry

Date and Time	[13/Jan/2017:03:06:05]
URL requested	GET /node/220218?m=1 HTTP/1.1
HTTP Status code	200
Response content size	35129
Referrer path	http://m.facebook.com/
User agent	Mozilla/5.0 (Linux; Android 5.1.1; C6833 Build/14.6.A.1.236; wv) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Chrome/52.0.2743.98 Mobile Safari/537.36 [FB_IAB/FB4A;FBAV/104.0.0.17.71;]
IP Address	113.210.66.166, 203.114.28.26

As the Web server logs that we use do not have any data about the users and the users are not a subscriber and not logged-in, makes it almost impossible and very difficult to create a reliable user profile [30]. Therefore, we identify users based on a combination of the IP address and the user agent string that is type of browser and operating system. If a user has the same IP address, but different type of user agent, it is assumed that it denotes a different user [33].

In Table II, we can identify the source of referrals from the referral column. Generally, there are five types of referrals that users use to access this news site. Typically, users type in the URL to the news site, known as direct traffic. Social media referral is from social media sites such as Facebook, Twitter, Instagram or Pinterest. Facebook can either be accessed through mobile device or a desktop. Users also access the page from another page in the news site. Other than that, users may access the news site from search engines like Google.

TABLE II. Sample of referrals and the page requested

URL requested	Referral	Remark
GET /node/261505 HTTP/1.1	http://m.facebook.com/	Mobile Facebook
GET /node/238938 HTTP/1.1	https://www.facebook.com/	Desktop Facebook
GET /node/261507?m=1 HTTP/1.1	http://www.bharian.com.my/node/261507	Another page in the website
GET /hiburan?m=1 HTTP/1.1	http://www.google.com/	Google
GET /node/238545?m=1 HTTP/1.1	-	User types in the mainpage URL
GET /node/238428	https://t.co/SbFV2hIO6e	Twitter

IV. RESULTS AND FINDINGS

There are four components of results; which are findings on (i) different types of referrals, (ii) the content requested, (iii) the device used to access the news website, and (iv) analysis on day of access.

A. Analysis on different referral channels

Figure I show the overall page referrals for the duration of the study. We combine both Facebook mobile and desktop to one category. Overall, referral from Facebook takes half of the pie, followed by referral from other pages in the site. Only about 11 percent of the users type in the URL to get to the news site. Google and Twitter only have very thin slice of referrals in the pie.

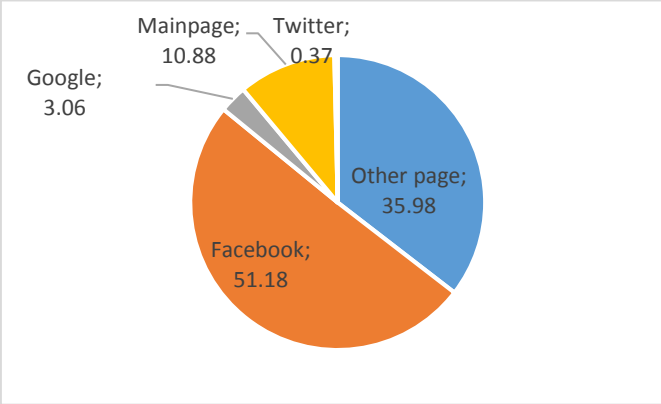

Fig. I Different types of referral to *Berita Harian Online*

B. Analysis on contents accessed using different referral channels

We divided the web page to three main categories, which are the main page, the section page and the article page. The main page is the landing page or the index page with the URL *www.bharian.com.my*.

TABLE III Correlation using Spearman rho of referrals and contents accessed

			Total Content
Article	referrer	Correlation Coefficient	-0.643**
		Sig. (2-tailed)	0.000
		N	90
Mainpage	referrer	Correlation Coefficient	-0.180
		Sig. (2-tailed)	0.342
		N	90
Section	referrer	Correlation Coefficient	-0.346
		Sig. (2-tailed)	0.061
		N	90

***. Correlation is significant at the 0.01 level (2-tailed).*

In terms of the contents accessed by different types of referral, we found a significant strong relationship between referrer and article content (r=0.643, p=0.000) as shown in Table III. However, for mainpage and section pages, there are no association shown between referrer and content.

C. Analysis on devices to access the news site

Since Facebook represent half of the page referral, we focus on the analysis of different devices to access the website just on Facebook referrals. In this analysis, we report the different types of devices that are used by users that are from Facebook to access the website. We divide the devices that are used to desktop, Android, iPhone and iPad. Desktop consists of Windows, Macintosh and Linux Epiphany. Figure II reveals that Facebook users mainly accessed the website using Android, followed by iPhone. Desktop users are the least number of users that access the news website from Facebook.

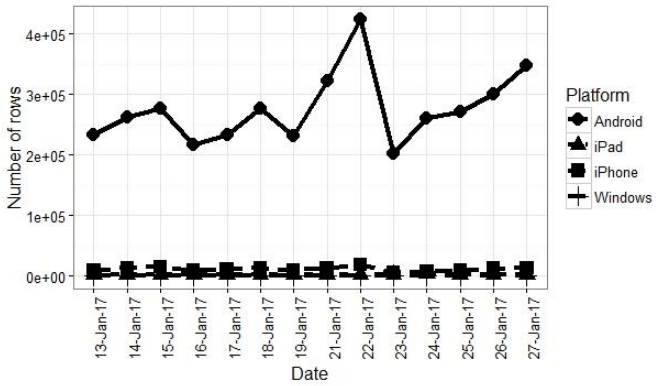

Figure II. Access of different devices

Spearman rho correlation is used to study the dependency between different days, access using either mobile or desktop, type of device and pages requested. Referring to Table IV, there is a significant correlation exists between the users accessing site using mobile or desktop device (rs=0.721) and type of device (rs=0.581) with the online newspaper user. Thus, whether the users using mobile or desktop have a significant impact on the user to access the newspaper site. The type of device also has a significant impact on the user to access the site Thus, the type of device has a significant impact on whether users browse the site using mobile or desktop.

Since there is a high referral from Facebook, we are interested to know what kinds of contents that the users are reading on. We divided the web page to three main categories, which are the main page, the section page and the article page. The main page is the landing page or the index page with the URL *www.bharian.com.my*.

The section page consists of list of headlines and abstract of articles in that particular sections. There are 15 main sections in the *Berita Harian Online* website. Some of these sections are Current, Sports, Entertainment, Business, Crime, Politics, Features, Property, Automotive and Information Technology. There are also one-time sections that are created specially to

cover an event such as elections, awards ceremony or sporting events like F1 or EPL. The last type of content is articles. Articles are individual news story about some certain events that happen.

TABLE IV Correlation between type of device and total user of online newspaper according to mobile or desktop

Access using device			Type of device	Page request
Mobile	*Type of device*	Correlation coefficient Sig. (2-tailed)	1.000	0.581** 0.000
	Page request	Correlation coefficient Sig. (2-tailed)	0.581** 0.000	1.000
Desktop	*Type of device*	Correlation coefficient Sig. (2-tailed)	1.000	-0.721** 0.000
	Page request	Correlation coefficient Sig. (2-tailed)	0.000	1.000

***. Correlation is significant at the 0.01 level (2-tailed)*

Figure III shows the total page requested from 13-27 January 2017 according to different content types. Most of the online newspaper users browsed article, followed by section and mainpage.

To see the how the different referral channels association with the contents, we perform a Phi & Kramer V. Association refers to coefficients which gauge the strength of a relationship. Phi and Cramer's V are based on adjusting chi-square significance to factor out sample size. In Figure IV, the contents are divided to the different referral channel.

Phi and Cramer's V are both tests of the strength of association. We can see that the strength of association between the variables is strong with 0.504 and 0.357.

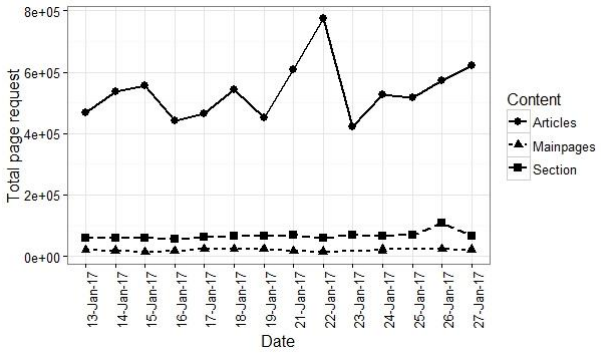

Figure III. Access to different types of contents by Facebook referral

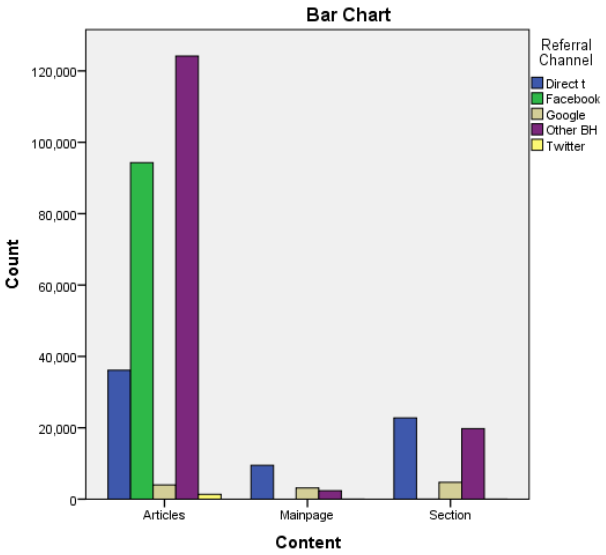

Figure IV Contents accessed by different referral channel

D. Analysis on different days of access

Figure V shows the behavior of online newspaper users who access the site according to specific days based on different content type. It is clearly shows that the online newspaper users mostly access specific article when browsing to the newspaper site. By comparing according to days, Monday is the least users who access the article in the site while Sunday has the most users who access the site based on specific article. Spearman rho correlation is used to study the dependency between day and total users of online newspaper.

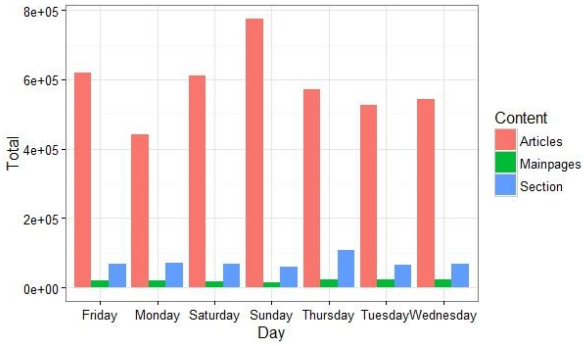

Figure V. Access to different types of contents by day

Referring to Table IV, there is no correlation exists between the total number of page accessed (rs=-0.16, p=0.908) with the different days of the week. The readers are always reading the newspaper, be it on weekends or weekdays. This trend was different from previous research that discussed the 'weekend effect'; where the access to online news site will decrease during weekends [34]. This is due to the fact that users only access the online news during weekdays, before they start their work. This study also suggested that users prefer to read the hardcopy newspaper on the weekends for their leisure [35].

TABLE IV Correlation using Spearman rho for content pages and days

		Total content pages	**Day**
Day	Correlation Coefficient	1.00	-0.16
	Sig. (2-tailed)		0.908
	N	90	90
Total content pages	Correlation Coefficient	-0.16	1.000
	Sig. (2-tailed)	.908	1.00
	N	90	90

V. DISCUSSIONS

Based on the findings, we found that almost half of the users that are accessing the *Berita Harian Online* website are from Facebook referral. Other referrals are from Google or users straight typed in the URL address. Based on the Facebook referrals users, we discovered that majority of users are accessing the news website using Android, followed by iPhone. On the other hand, desktop users are the least group of users that use Facebook to access the news site. Articles is the main types of content accessed using Facebook. In terms of the different days of week to access the news site, there is not much difference between weekends and weekdays. This trend of navigation is definitely different from previous work that suggested a decrease of traffic during weekends [34]. Perhaps the advancement in mobile devices make it easier for users to access the news site at any time or any day [25].

Facebook referral definitely play an important role for incoming traffic to this news website. With the proliferation of mobile devices and tablets, accessing the news can be done anywhere and anytime. Facebook users who are browsing their news feed may stumble upon news [27]. Therefore, the news organization should intensify their social media strategy to appeal to more Facebook users. In addition to that, the mobile version of the online news site should be more user friendly, together with the apps to engage more users in more contents.

For future work, we plan to investigate the dwell time of page accessed by Facebook or not Facebook. We also wish to examine if there are any significant difference on time spent on the page accessed based on the device used by users.

ACKNOWLEDGMENT

We thank Mr Aswandi Yasin from Media Prima for assistance with collection of the web server logs.

REFERENCES

[1] K. L. Keller, *Strategic Brand Management, Building Measuring & Managing Brand Equity*. Pearson Prentice Hall, 2008.

[2] M. A. Omidvar, V. R. Mirabi, and N. Shokry, "Analyzing the Impact of Visitors on Page Views with Google Analytics," *Int. J. Web Semant. Technol.*, vol. 2, no. 1, 2011.

[3] A. Prasetio, D. T. Alamanda, A. Partono, and O. O. Sharif, "The Impact of Traffic Source on Page Views," in *The Proceedings of The 7th ICTS*, 2013.

[4] B. Plaza, "Google analytics for measuring website performance.," *Google Anal. Meas. website performance.*, vol. 32, no. 3, pp. 477–481, 2011.

[5] D. E. Rosen and E. Purinton, "Website design: Viewing the web as a cognitive landscape," *J. Bus. Res.*, vol. 57, no. 7, pp. 787–794, 2004.

[6] E. J. Karson and R. J. Fisher, "Predicting intentions to return to the web site: Extending the dual mediation hypothesis.," *J. Interact. Mark.*, vol. 19, no. 3, pp. 2–14, 2005.

[7] W. W. Moe and P. S. Fader, "Dynamic conversion behavior at e-Commerce sites," *Manage. Sci.*, vol. 50, no. 3, pp. 326–335, 2004.

[8] O. J. Rutz and M. Trusov, "Zooming in on paid search Ads—A consumer-level model calibrated on aggregated data," *Mark. Sci.*, vol. 30, no. 5, pp. 789–800, 2011.

[9] S. Yang and A. Ghose, "Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence," *Mark. Sci.*, vol. 29, no. 4, pp. 602–623, 2010.

[10] A. Dobilinskas and V. Zakaitė, "Online Branding for SMEs: Media types, Channels and their Effectiveness," Lund University, 2014.

[11] A. Prasetio, P. K. Sari, O. O. Sharif, and E. Sofyan, "Analyzing traffic source impact on returning visitors ratio in information provider website," *Int. Conf. Innov. Eng. Vocat. Educ.*, 2016.

[12] P. Kireyev, K. Pauwels, and S. Gupta, "Do display ads influence search? Attribution and dynamics in online advertising," *Int. J. Res. Mark.*, vol. 33, no. 3, pp. 475–490, 2016.

[13] N. Shi, Y. Hong, K. Wang, and P. A. Pavlou, "Social commerce beyond word of mouth: Role of social distance and social norms in online referral incentive systems," in *Proceedings of the 34th International Conference on Information Systems (ICIS)*, 2014.

[14] Michael Stelzner, "2014 Social Media Marketing Industry Report," 2014.

[15] D. Wong, "In Q4, Social Media Drove 31.24% of Overall Traffic to Sites," 2015.

[16] A. Köster, C. Matt, and T. Hess, "Does the Source Matter ? How Referral Channels and Personal Communication Tools Affect Consumers ' Referral Propensity," pp. 3915–3924, 2017.

[17] C. Castillo, M. El-haddad, and M. Stempeck, "Characterizing the Life Cycle of Online News Stories Using Social Media Reactions," 2012.

[18] A. Ju, S. H. Jeong, and siang I. Chyi, "Will social media save newspapers? Examining the effectiveness of Facebook and Twitter as news platforms," *Journalism*, vol. 8, no. 1, pp. 1–17, 2017.

[19] R. Somaiya, "How Facebook Is Changing the Way Its Users Consume Journalism," *New York Times*, 2014. .

[20] S. Messing and S. J. Westwood, "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online," *Communic. Res.*, vol. 41, no. 8, pp. 1042–1063, 2012.

[21] I. Costera Meijer and T. Groot Kormelink, "Checking, Sharing, Clicking and Linking: Changing patterns of news use between 2004 and 2014," *Digit. Journal.*, pp. 1–16, 2014.

[22] I. Pentina and M. Tarafdar, "From 'information' to 'knowing': Exploring the role of social media in contemporary news consumption," *Comput. Human Behav.*, vol. 35, no. June 2014, pp. 211–223, 2014.

[23] T. Litsa, "How news consumption changed with social media," *ClikZ*, 2017. .

[24] J. Gottfried and E. Shearer, "News Use Across Social Media Platforms 2016," *Pew Research Center*, 2016. [Online]. Available: http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/. [Accessed: 17-Apr-2017].

[25] C. Liu, "Facebook Journalism: The Influences of Social Media on Journalistic Work in Taiwan," in *e-Proceedings. School of Communication Arts and Management Innovation, National Institute of Development Administration*, 2015, pp. 107–118.

[26] R. Meshulam, W. Street, N. York, U. States, W. Street, and N. York, "For Your Eyes Only : Consuming vs . Sharing Content," no.

1, pp. 725–729, 2016.

[27] P. Müller, P. Schneiders, and S. Schäfer, "Appetizer or main dish? Explaining the use of Facebook news posts as a substitute for other news sources," *Comput. Human Behav.*, vol. 65, pp. 431–441, 2016.

[28] M. Mauri, P. Cipresso, A. Balgera, M. Villamira, and G. Riva, "Why Is Facebook So Successful? Psychophysiological Measures Describe a Core Flow State While Using Facebook," *Cyberpsychology, Behav. Soc. Netw.*, vol. 14, no. 12, pp. 723–731, 2011.

[29] A. Valeriani and C. Vaccari, "Accidental exposure to politics on social media as online participation equalizer in Germany, Italy, and the United Kingdom," *New Media Soc.*, vol. 18, no. 9, pp. 1857–1874, 2016.

[30] M. Trevisiol and L. M. Aiello, "Cold-start News Recommendation with Domain-dependent Browse Graph Categories and Subject Descriptors," 2012.

[31] M. Anderson and A. Caumont, "How social media is reshaping news," *Pew Research Center*, 2014. [Online]. Available: http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/. [Accessed: 18-Apr-2017].

[32] A. Mitchell, M. Jurkowitz, and K. Olmstead, "Social, Search and Direct Pathways to Digital News," 2014.

[33] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowl. Inf. Syst.*, vol. 1, no. 1, pp. 5–32, 1999.

[34] E. Dans, W. Plaza, and D. Suite, "Internet Newspapers : Are Some More Equal than Others ?"

[35] S. Steensen, "The shaping of an online feature journalist," vol. 10, no. 5, pp. 702–718, 2009.