Story-focused Reading in Online News and its Potential for User Engagement

Janette Lehmann
Freie Universität Berlin[*]

Carlos Castillo
Sapienza University of Rome[†]

Mounia Lalmas
Yahoo Labs London

Ricardo Baeza-Yates
Yahoo Labs Sunnyvale

Abstract. We study the news reading behaviour of several hundred thousand users on 65 highly-visited news sites. We focus on a specific phenomenon: users reading several articles related to a particular news development, which we call *story-focused reading*. Our goal is to understand the effect of story-focused reading on user engagement and how news sites can support this phenomenon. We found that most users focus on stories that interest them and that even casual news readers engage in story-focused reading. During story-focused reading, users spend more time reading and a larger number of news sites are involved. In addition, readers employ different strategies to find articles related to a story.

We also analyse how news sites promote story-focused reading, by looking at how they link their articles to related content published by them, or by other sources. The results show that providing links to related content leads to a higher engagement of the users, and that this is the case even for links to external sites. We also show that the performance of links can be affected by their type, their position, and how many of them are present within an article.

1 Introduction

Online news reading is one of the most common activities of Internet users. A survey published by [33] reported that 39% of news readers get their news online. Users may have different motivations to visit a news site. Some users want to remain informed about a specific news story they are following, such as an important sport tournament or a contentious important political issue. Others visit news portals to read about breaking news and remain informed about current events in general.

---

[*]This work has been done while Janette Lehmann was a PhD student at Universitat Pompeu Fabra and it was carried out as part of her PhD internship at Yahoo Labs Barcelona.

[†]This work was carried out while Carlos Castillo was working at Qatar Computing Research Institute.

The Web has totally changed the news landscape, as users have the possibility to read news from diverse news sites as well as other sources. A recent study found that 57% of users routinely get their news from between two to five news sites [32]. Although users increasingly use social media sites to share news they read and find worth sharing, search engines continue to be an important tool for users to look for articles on news of interest to them; more than 33% of users use search engines regularly to find news [33]. Several search engines offer *news verticals* specifically designed for users to search for news published by online news sites e.g. [3, 10, 24].

While reading news, users sometimes become interested in a particular news item they just read, and want to find more about it. They may want to obtain various angles on the story, for example, to overcome media bias [36] or to confirm the veracity of what they are reading. A study from the New York Times reported that many users still visit established news outlets to confirm a story, no matter from which source the information initially came from [41].

When asked about news reading online, many users of news sites have said that links to related information on a news article page are important [32]. News sites recognise that users want to further inform themselves, and provide information on different aspects or components of a story they are covering. They also link to other articles published by them, and sometimes even to articles published by other news sites or sources. In this paper, we study this type of reading behaviour. Our goal is to understand its effect on user engagement in the context of news reading and provide insights into how online news sites can support this type of news reading behaviour.

We describe *story-focused news reading*, or simply *story-focused reading*, which occurs when users read multiple articles about a particular news development or event. In this paper, *article* refers to a single document, and *story* refers to a set of related articles. Our main contributions are:

1. To the best of our knowledge, this is the first study that investigates story-focused reading in this context and scale, and that also accounts for the interactions between news sites and other sources.
2. We study the characteristics of story-focused reading and show that users exhibit a different reading behaviour when focusing on a story.
3. We show that having links to related content within the text of a news article keeps users engaged with the news site. We also show that the type, the position and the number of links matter, and that even external links are beneficial.

Our study is based on a large sample of user interaction data on 65 popular news sites publishing articles in English. We analysed 4.9M news reading sessions covering a total of 2,536 stories comprising 25,703 news articles. Stories range from policy issues such as the threat of the US government shutdown (October 1 to 16, 2013), and the NSA spying scandal, to less important issues such as the Draconid meteor shower, and specific sport events.

2 Related Work and Motivation

Our work spans across several research areas. We discuss them, and position our work in their context.

2.1 Reading Behaviour and User Engagement

How users browse the Web has been studied in many contexts [5, 19] and some research has also focused on user reading behaviour on news sites [9, 44, 30]. Knowing how users interact with a site is used to evaluate their engagement with the site. For instance, [29] developed a process model of user engagement to determine the factors (*e.g.*, aesthetic appeal, interactivity, novelty) that make an online experience engaging. The model, which has been also applied in the context of online news reading [28], consists of four stages: (1) point of engagement (entering the news site), (2) period of engagement (browsing through the news site), (3) disengagement (leaving the news site), and (4) re-engagement (returning to the news site).

The *point of engagement* corresponds to the beginning of an engaging experience; in our context it relates with users' motivation to read news. Studies have shown, for instance, that many users read news to stay informed but also for relaxation [25, 14]. This was further confirmed in other studies [38, 18, 15], who also found that news reading is a *browsing task*, where users visit news site(s) as part of their daily routine to keep informed about current events.

The studies of [25] and [14] also showed that users turn to news sites and other sources to follow a breaking story and to search for updates or specific information about a story. Such users might want to obtain various angles on the story [36] or confirm the veracity of what they are reading [41]. These users become *focused* on a story, as they are "interested in the event details" [25]. Although such interest might be driven by curiosity [28], user personal preferences play an important role as well [2]. Indeed, user and story might share the same geographical location [23, 39], or the user is in general interested in the topic of the story [22], *e.g.*, users interested in football are likely to be interested in stories about FIFA World Cup. "These readers may gather different sources to fill in the gaps on a single story" [25], using search [26], social media sites [16] and Wikipedia [17], or even specialized tools for exploring stories [13, 43]. Referring to the work of [18], such behaviour can be considered as an *information gathering task*.

Also, [18] also showed that user browsing behaviour depends on the type of task users perform. For instance, users dwell longer and visit more pages during information gathering tasks compared to simple browsing tasks. In the context of news reading this might imply that user reading behaviour during the *period of engagement* differs depending on whether users perform their daily consumption of news (browsing) or consume news articles related to a specific story (information gathering). As stated by [25], such differences should be considered when designing news sites to increase the reading experience of users.

We investigate into this by characterising story-focused reading and comparing it with non-story-focused reading. We look at this in the context of news reading across news sites, because recent studies have demonstrated that news reading often involves not only one but many sites [32, 20].

2.2 Hyperlinks and Story-focused Reading

The work of [28] also investigated factors that influence engagement during news reading. Her results suggest that depending on the novelty and quality of the content, as well as the interactivity and aesthetic appeal of a news site, users either remain longer in the period of engagement or disengage earlier. Re-engagement was not analysed due to the laboratory setting of her study.

In our work, we look at one aspect of website interactivity and its effect on user engagement during story-focused reading: hyperlinks (or links for short) within news article pages that point to additional information about the corresponding story. Such links have been shown [28, 31] to influence engagement, as users have more control over the content they wish to consume, and how they consume it. During her study [28], some interviewees even stated that they followed such links, or that they would have appreciated the opportunity to do so. We hypothesise that users remain engaged for longer period of times (*i.e.*, read more articles) and re-engage sooner (*i.e.*, return earlier to the site to gather new information about the story or a different one), if a news site promotes story-focused reading by embedding links to related content within their article pages.

News providers have different strategies to link their articles to additional information about a story. For instance, whereas blogs rely on hyperlinks to be reached by users and to direct users to other websites [34, 11], most traditional news sites do not link to each other, because of competition. Instead, they invest time and effort into connecting their own news articles with each other using hyperlinks [6]. This is to keep users on their site and increase user engagement [40, 42]. On the other hand, [7], [8] and [35] advocate that providing hyperlinks between news sites can actually increase profits in a costless way. In addition, it provides a more interactive, credible, transparent, and diverse news reading experience to users.

Also, [26] showed that news articles are not the only information source satisfying news-related queries in search engines. When searching for news, users like to see Wikipedia pages, blog posts, and tweets, which are sometimes sufficient to satisfy their information needs. Linking news articles to other information sources [27, 37, 6] allows users to learn about the story context, *e.g.* background and history of the story, as well as opinions and discussions around it.

However, there is still limited knowledge into how such links and thus story-focused reading affect user engagement. In this paper, we analyse different linking strategies of news sites and their impact on story-focused reading. Our goal is to provide insights into how news sites can support this type of news reading behaviour with the aim to increase the engagement with their site.

Table 1: News providers analysed, listed in alphabetical order.

abcnews.go.com	cnbc.com	huffingtonpost.com	nytimes.com
thehill.com	adweek.com	cnn.com	latimes.com
online.wsj.com	theonion.com	ajc.com	csmonitor.com
metro.co.uk	philly.com	thestar.com	azcentral.com
dailyfinance.com	miamiherald.com	rawstory.com	thesundaytimes.co.uk
bankrate.com	dallasnews.com	nationalpost.com	reuters.com
time.com	bbc.co.uk	denverpost.com	nationalreview.com
seattletimes.com	upi.com	bloomberg.com	digitalspy.co.uk
nbcnews.com	sfgate.com	usatoday.com	breitbart.com
economist.com	news.com.au	smh.com.au	usnews.com
businessweek.com	examiner.com	news.sky.com	theage.com.au
variety.com	cbc.ca	forbes.com	news.yahoo.com
theatlantic.com	voanews.com	cbsnews.com	foxnews.com
newsmax.com	theaustralian.com.au	washingtonpost.com	chicagotribune.com
heraldsun.com.au	nj.com	theglobeandmail.com	washingtontimes.com
chron.com	hollywoodreporter.com	nypost.com	theguardian.com
wnd.com			

3 Data Used and Data Processing

Our study is based on one month (October 2013) of anonymised user browsing data, consisting of tuples of the form ⟨timestamp, browser cookie, URL, referring URL⟩ from a random sample of users who gave their consent to provide data through the Yahoo Toolbar.[1] This sample consists of 800K users, and 325M page views.

3.1 News Providers

We selected the 100 most visited news sites according to the ranking provided by Alexa.[2] Alexa's ranking includes both traditional news outlets (*e.g.*, The New York Times or CNN) and news sites that mostly aggregate news from various sources (*e.g.*, Yahoo News). We then identified the news sites from the list that publish articles in English, to be able to compare articles from different news sites in a straightforward way. We also considered news providers that cover only specific genres (referred to as section); *e.g.*, bankrate.com and dailyfinance.com report mainly about investment and financial stories. The resulting news sites, which are based in the US, UK, Canada and Australia, are listed in Table 1.

To ensure that no strong bias in the browsing data affect our results and their applicability, we compared the Alexa ranking with our data on the basis of total user traffic, and found that the two correlate well (Kendall's $\tau = 0.62$, and Spearman's $\rho = 0.80$). Similar correlations were observed in a study comparing the reading behaviour of toolbar users and all users in Wikipedia [21]. Hence, the insights gained in this paper are not specific to Yahoo toolbar users.

---

[1] https://toolbar.yahoo.com/
[2] http://www.alexa.com/topsites/category/Top/News

3.2 News Stories

A *news article* or simply *article* is a single document on a news website (an HTML page). For each of the 65 sites, we used various rules, based on regular expressions, to distinguish between visits to an actual news article from visits to other parts of a news website, such as its homepage or section pages (*e.g.*, "politics" or "world news"). Articles visited by less than 5 users during a day, which correspond mostly to very old articles, were not considered for this study. This removed 8.9% of the browsing events, and left us with 98,241 news articles.

A *news story* or simply *story* is a collection of articles related to the same news event. We used the approach of [36] to identify the news stories in our dataset. For each day in our data period, we identified the articles that were visited on that day and extracted the words of the article texts. After stopword removal and lemmatization were applied, we computed the TD-IDF weight for each word t in each article d as follows:

$$tfidf(t, d, D) = \frac{f(t, d)}{max\{f(t, d) : t \in d\}} \cdot log(\frac{|D|}{|d \in D : t \in d|})$$

where D is the set of articles and f computes the frequency of a word t in an article d. We then modelled articles (nodes) and the similarity between them (edges) as a graph. The similarity between pairs of articles was determined by computing the cosine similarity between the corresponding word vectors. Finally, we partitioned the graph by removing all edges whose similarity value was below a given threshold, and each connected component of nodes (articles) that remained connected corresponds to a story. We experimented with several threshold values, as shown in Figure 1(a).

A too low threshold generates few stories containing many articles, as stories have often something in common (*e.g.*, they are located in a particular region). If the threshold is too high, many stories that are in fact related are not detected as being so (*e.g.*, they report about different aspects of the same story). This results in few stories made of only two or more articles.[3] For 0.4, we reach the maximum number of stories. We therefore used this value as our threshold,[4] which is comparable to what was used by [36].

We also removed *niche stories*, which are stories covered by very few news providers; these stories are likely to be region-specific (*e.g.*, "public chess games in San Francisco"). Such stories should be treated differently, as users exhibit a different reading behaviour on them. For instance, often solely one news provider (of our dataset) is reporting about a niche story and in this case users rely on regional news providers or even social media sites [36] when they want to gather more information about the story. Therefore, we only consider *top stories* in our analyses and leave for future work a detailed study about whether and how users focus on niche content. To identify niche stories in our dataset, we calculated the number of stories identified based on the minimum number of news providers we considered. We experimented with several numbers, and

---

[3]Note that a story must involve at least two articles.
[4]Using a threshold of 0.3 or 0.5 yields a similar number of stories.

(a) Number of stories generated at different similarity thresholds.

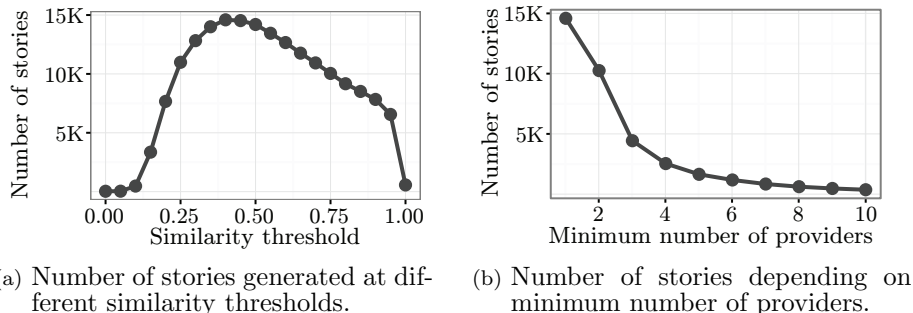(b) Number of stories depending on minimum number of providers.

Figure 1: Selecting the threshold to decide when two articles relate to the same story, and the threshold to decide when a story is niche (only stories with more than one article are included).

plotted the outcome in Figure 1(b). We see that many stories are published by only few providers. The number of stories decreases fast as the minimum number of considered providers increases, and then slows down at around 3-4 providers. As a conservative setting, we define a story to be *niche* if it is covered by 3 providers or less and *top* if it is covered by 4 providers or more.

This process results in 2,536 top stories, about 82 per day, and 25,703 articles. On average, each of these stories has 14 articles (median 8), and is covered by 7 providers (median 5). The distribution of visits to stories is very skewed, with an average of 2,482 users per story (median 758). The number of visits correlates moderately with the number of articles about the story (Spearman's $\rho = 0.67$) and with the number of providers that cover it ($\rho = 0.54$).

3.3 Inline Article Links

In this paper, we also analyse how news providers promote story-focused reading by offering links to related content within their articles pages.

We downloaded the HTML content of the article pages visited by the users in our one-month dataset. Articles from the Wall Street Journal were not considered, as those articles could not be freely downloaded. We assume that any link within the body of an article, an *inline link*, connects that article to a page that is related to it. A manual inspection of several of the news websites under study showed that inline links point in most cases to articles belonging to the same story. This is a common strategy to provide additional information about a news item [6].

Although news sites may provide links to related articles in a specific panel (*e.g.*, "Related articles"), identifying such cases proved to be complex and introduced some level of ambiguity across providers. Therefore, we did not attempt to distinguish classes of non-inline links. However, we also calculated the total number of links in each article page, considering inline and non-inline links.

In our dataset, 75.45% of the article pages have inline links, but on average,

only 6.4% of the links in an article page are inline links.

3.4 News Reading Sessions

The browsing activities of users were split into sessions, where a session ends if more than 30 minutes elapse between two successive pageviews [5]. A *news reading session* is a session in which *at least one* news article of the selected stories is accessed. A news reading session is *story-focused*, if the user visits two or more articles related to a story. Otherwise the news reading session is labelled as *non-story-focused*. In our dataset, we extracted a total of 4.9M news reading sessions, whereas 245K of them are story-focused.

The sites users navigate from when viewing a news article and the sites they are navigating to afterwards were categorised using the following schema. We distinguished whether the user came from (navigated to) a page of the same provider (*internal traffic*), or from (to) somewhere else in the Web (*external traffic*). For both cases, we also differentiated the traffic from (to) articles ([Internal/External Article]) or other pages ([Internal/External Non-Article]) of the provider sites in our list. We annotated the remaining (all [External]) sites using Alexa's schema and the one described by [20] as follows:

- 1022 further news sites and blogs [News Non-Top]
- 42 news aggregators and online RSS feeds (*e.g.*, Google news, FriendFeed) [News Aggregator]
- 39 social media sites (*e.g.*, Twitter) [Social Media]
- 5 mail sites [External Mail]
- 25 multimedia sites (*e.g.*, YouTube) [Multimedia]
- 52 reference sites (*e.g.*, Wikipedia) [Reference]
- 10 search engines (*e.g.*, Google, Bing) [Search]
- 812 organisation sites (*e.g.*, nasa.gov) [Organisation]
- 7 front pages (*e.g.*, AOL) [Front page]
- 17K uncategorised sites [Other]

Depending on whether we analyse the upstream or downstream traffic, we made the following simplifications. In the 5.3 Section, the multimedia, reference, and organisation sites are annotated as [Ext. Other]. In the 7 Section, news aggregator sites are merged to [Ext. News Non-Top], and mail and search sites, and front pages are part of [Ext. Other].

4 Does Story-focused News Reading Exist?

We show that story-focused reading exists and that it is not a trivial phenomenon, *i.e.*, not merely a consequence of how articles are distributed among stories.

4.1 Shuffle Test

We first determine whether story-focused reading occurs because many articles belong to a story, or because users are interested in reading articles related to a specific story. To answer this question we perform a *shuffle test* (similarly to [1]). We create an alternative dataset of news reading sessions that has the same distribution of session length, but with random articles in them. We call this alternative dataset the *shuffled dataset*.

We next calculate the *probability of story-focused reading* for a given story s:

$$\frac{\#\text{story-focused sessions of s}}{\#\text{sessions including s}}$$

where the story-focused sessions for a story s are those in which a user visits two or more articles related to the story s. The distribution of this probability across all stories is shown in Figure 2(a).[5] We also calculate the *probability of multi-provider reading* for a given story s:

$$\frac{\#\text{story-focused sessions of s in two or more providers}}{\#\text{story-focused sessions of s}}$$

which is shown in Figure 2(b).

We observe a clear difference between the actual dataset and the shuffled one. The probability of story-focused reading is about 4 times larger with the actual data (0.019 vs. 0.005), and the probability of multi-provider reading is about two times larger in the actual dataset (0.48 vs. 0.25). A Kolmogorov-Smirnov (K-S test) test confirms that the difference between the distributions is statistically significant (p-value < 0.01). This indicates that story-focused reading is observed due to users deciding to read multiple articles associated with a story.

4.2 Popularity and Providers

We analyse whether story-focused reading depends on the popularity of a story, on its number of articles, or on the number of news providers that cover it. The *popularity* of a news story is defined as the number of sessions where users have read articles related to that story.[6] We again compare our dataset with the shuffled dataset.

In Figure 3(a) we plot the *probability of story-focused reading* based on the story popularity. We observe that story-focused reading is not necessarily related to popularity. Even stories that are not popular engage users in story-focused reading. The probability of story-focused reading given its popularity is lower than what is observed with the shuffled dataset (Spearman's $\rho = 0.30$ vs. $\rho = 0.57$ in the shuffled dataset). Overall, the probability of story-focused reading is comparable across all levels of popularity; this indicates that personal interests trigger users into story-focused reading.

---

[5]In this and other log-scale plots, we added to each value a small constant (0.0001) to represent zeros in the log scale.

[6]The traffic volume has been scaled with an arbitrary but constant factor for confidentiality.

(a) Distribution of the probability of story-focused reading.

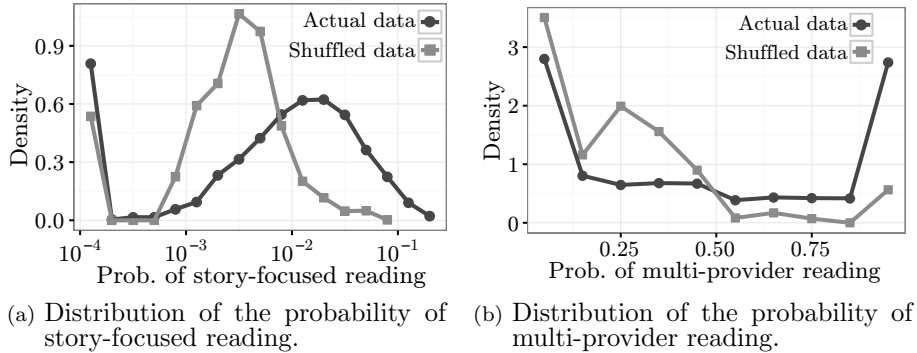(b) Distribution of the probability of multi-provider reading.

Figure 2: Shuffle test showing the difference in story-focused reading and multi-provider reading between the actual dataset and the shuffled dataset.

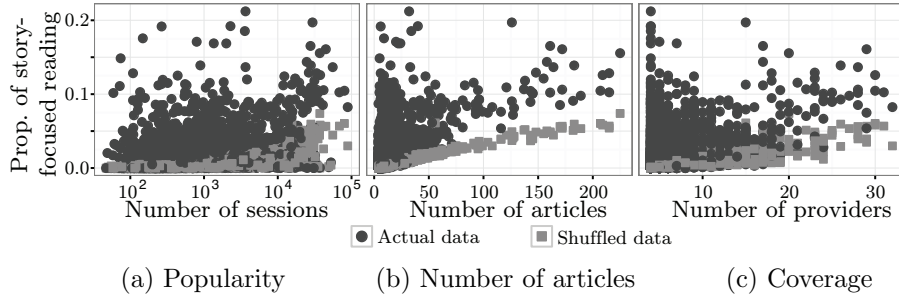(a) Popularity   (b) Number of articles   (c) Coverage

Figure 3: Story-focused reading and popularity, number of articles, and coverage.

Story-focused reading is also not merely a consequence of having a story reported through many articles. Figure 3(b) shows the probability of story-focused reading as a function of the number of articles published about that story. We observe that even stories having few articles written about them engage users in story-focused reading. Compared with the shuffled dataset, the correlation between this probability and the number of articles is lower in the actual dataset (Spearman's $\rho = 0.61$, vs 0.80 in the shuffled dataset). For instance, in our dataset we obtain a probability of story-focused reading equal to 0.1 for the two stories "Royal Christening of Prince George" (October 23, 2013) and "Draconid Meteor Shower" (October 7, 2013). The former story has 95 articles associated with it, whereas the latter story has only 8 articles.

We reach similar conclusions when relating the probability of story-focused reading to how many news providers are reporting about a story, as shown in Figure 3(c). The fact that several news sites report the same news story is not what promotes story-focused reading. The correlation is lower than in the shuffled dataset (Spearman's $\rho = 0.36$ vs 0.62 in the shuffled data).

Table 2: Comparison of story-focused sessions with non-story-focused sessions at different session lengths, including the probability that a session is story-focused, the total duration of the session, and the number of distinct news providers visited. Maximum values are highlighted in boldface.

% Sessions	Num. Articles	% Focused	Duration [minutes]		Num. providers	
			Non-focused	Focused	Non-focused	Focused
74.93	1	—	3.09	—	1.00	—
17.31	2	15.16	5.77	6.66	1.20	1.26
4.95	3	29.00	8.57	9.83	1.37	1.48
1.66	4	41.33	11.24	12.92	1.52	1.67
0.63	5	51.02	13.60	15.96	1.65	1.83
0.27	6	60.85	15.92	18.35	1.81	1.98
0.12	7	67.13	18.61	21.59	1.91	2.14
0.13	>7	**78.15**	**21.29**	**27.91**	**2.12**	**2.55**

4.3 Users

The percentage of users that engage, at least once, in story-focused reading is 16% in our one-month dataset. As expected, avid news readers are more likely to engage in story-focused reading: 64% of the users with at least 15 reading sessions in our one-month dataset have at least one story-focused session. However, this does not imply that the more articles a user is reading, the more often s/he is engaging in story-focused reading. In fact, the correlation between these two variables is only $\rho = 0.45$.

Overall, we can conclude that story-focused reading is not simply a consequence of some stories being more popular or some users being heavy news consumers. Users focus on a story because they are interested in gathering more information about it.

5 How are Story-focused Reading Sessions Characterised?

We characterise *story-focused reading sessions*, or simply *story-focused sessions*, in particular how they differ from non-story-focused sessions, their depth and how users start engaging in such sessions.

5.1 Story-focused vs. Non-story-focused Sessions

Story-focused sessions have several characteristics that distinguish them from non-story-focused sessions (sessions where no story-focused reading is observed). Table 2 compares the two, grouping them by session length (number of articles visited in a session). We note that 75% of the sessions contain only one news article, and by definition cannot be story-focused. Using a K-S test, we can confirm that the described differences with respect to the duration and number of providers are statistically significant (p-value $\ll 0.01$).

Table 3: Percentage of story-focused sessions of different depth (in-story articles), and their average number of in-story and out-story articles, duration, and number of distinct providers.

| % | Number of articles | | Duration [minutes] | | Num. |
Sess.	In-story	Out-story	Total	Per-article	providers
85.03	2	0.89	6.67	3.34	1.27
11.48	3	1.09	10.48	3.49	1.53
2.43	4	1.45	14.29	3.57	1.79
0.69	5	1.67	18.23	**3.65**	2.05
0.23	6	1.80	20.09	3.35	2.31
0.08	7	2.67	23.09	3.30	2.36
0.06	**>7**	**3.05**	**25.03**	2.79	**3.19**

We observe that when the session length increases (more articles are read), the probability that a session is story-focused increases (column "%Focused"). For instance, 41% of the sessions with 4 articles are story-focused sessions. This shows that story-focused reading becomes more predominant the longer the user spends time reading news.

We also see that users spend more time in their news reading activity when focusing on a specific story, compared to when they access articles about different stories (column "Duration"). Story-focused sessions are at least 15% longer, and the difference increases with the session length. For instance, for sessions with 4 articles, the session duration is on average 11.24 minutes (in non-focused sessions) and 12.92 minutes (in focused sessions).

Finally, we observe similar patterns when looking at the number of news providers (column "Num. providers"). For sessions with 4 articles, the average number of news providers from which users are reading articles is 1.52 (non-focused sessions) and 1.67 (focused sessions).

5.2 Depth of Story-focused Reading

Session length only takes into account the number of articles visited during the session. These articles may not necessarily relate to the same story. We study now the browsing behaviour depending on how many articles of the same story are accessed, called the number of *in-story articles* or the *story depth*. We also report the average number of *out-story articles* (articles that do not belong to the story the user is focusing on), and other averages (see Table 3). In cases where the user is focusing on several stories within a session, we calculated the browsing behaviour with respect to each story. We note, however, that only 2.36% of the story-focused sessions have the user focusing on more than one story.

Deeper story-focused sessions are naturally longer. They also involve a larger number of news providers. The number of out-story articles is higher as the session depth increases; however, the in-story articles always constitute the majority of articles read in story-focused sessions (columns "Number of articles").

In sessions with 5 or less in-story articles (99.6% of the sessions), we see an

increase in the *per-article dwell time*: users spend time reading the individual articles they are accessing. For 0.4% of sessions with 6 in-story articles or more, the dwell time decreases. This is in accordance with results reported by [20] about general user online behaviour, and suggests that users are skimming the articles, probably because the articles contain increasingly more redundant information. To verify this, we measure for each article read in a session the number of words (information) that do not occur in any other article of that session. In sessions with 2 in-story articles, each article contains on average 49% unique words, whereas in sessions with 7 or more in-story articles, the percentage of unique words per article is on average 33%. This suggests that with each article read in a session, users skip increasingly more parts of it (*e.g.*, information known from previous articles), and skim the article for specific or new facts.

5.3 Upstream Traffic

We want to understand how users reach the article when they actually engage in story-focused reading. Using the HTTP referrers available in our browsing dataset and the categorisation schema described in the 3 Section, we study which sites users navigate from when engaging in story-focused reading. As story-focused sessions involve at least two articles by definition, we consider *upstream traffic* in relation to the second, third, *etc.*, article being accessed.

Table 4 shows the percentage of upstream traffic for the considered site categories, grouped by story depth. Most of the traffic to an article comes from other pages of the same news provider. In sessions with 2 in-story articles, 78.8% of the traffic is coming from another page of the same provider (*internal traffic*), and only 21.2% of the traffic originates from somewhere else in the Web (*external traffic*). However, the dominance of internal traffic decreases as more articles are read. For example, if users read more than 7 articles about the same story, only 55.1% of the traffic comes from another page of the same provider.

Interesting is that the internal traffic is mainly driven from non-article pages of the news provider. Looking again at story-focused sessions with depth 2, in only 17.8% of the article views the users navigated from another article of the news provider; in 61.0% of the article views the users employ other means on the provider site to access related articles. With respect to this, we can report that for on average 57% (median is 67%) of the article views per news provider, the user clicked on a link on the front page of that provider. This suggests that the linking strategy of many providers does not support story-focused reading *at* the article level, as users are more likely to return to the front page to search for another article related to the story.

We look now at the external upstream traffic, and discuss in particular the values obtained for sessions with more than 7 in-story articles. If the story-focused session is very deep, many articles are accessed from webmail sites and other sources (3.8%, and 11.6%, respectively). The same applies for less popular news sites and social media sites. We see that 11.2% of the upstream traffic comes from less popular news sites ("News Non-Top"). This shows that in the

Table 4: Upstream traffic sources as a function of the depth of a story-focused session. The traffic is divided into two main categories (internal and external), and for the latter, in seven sub-categories.

| | Number of in-story articles | | | | | | |
	2	3	4	5	6	7	>7
Internal	**78.8**	77.2	75.5	73.2	73.3	69.5	55.1
Article	17.8	22.9	26.4	28.4	**30.2**	28.3	16.8
Non-Art.	**61.0**	54.3	49.1	44.8	43.1	41.2	38.3
External	21.2	22.8	24.5	26.8	26.7	30.5	**44.9**
Other	5.5	5.4	6.7	7.3	8.0	9.4	**11.6**
News Non-Top	1.5	1.9	2.1	2.4	2.2	2.5	**11.2**
Social media	1.2	1.0	1.3	1.7	2.0	3.1	**8.4**
Front page	5.3	5.0	4.0	4.2	3.7	3.8	**5.7**
Mail	1.3	1.8	2.1	2.5	2.7	3.6	**3.8**
Search	3.7	4.4	5.1	**5.3**	5.0	5.2	2.8
News Article	0.6	0.8	0.9	**1.0**	0.9	0.8	0.6
News Aggregator	1.3	**1.4**	1.4	1.4	1.2	1.1	0.3
News Non-Article	0.8	**1.1**	0.9	1.0	1.0	1.0	0.5

context of story-focused reading, inter-provider linking can increase the traffic to the most popular news providers. In addition, the usage of social media sites increases (8.4%). This showcases the increasing importance of social media sites as a source of traffic for people interested in having in-depth information about a story. For instance, Twitter allows users to click on a hashtag or search for it (*e.g.*, *#Obamacare*), and then access multiple related articles. We also saw that front pages are frequently used to access related articles. Although the traffic coming from the front pages decreases as the story depth increases, it increases again for sessions with more than 7 in-story articles (5.7%).

For all other types of sites, we observe that the traffic increases first (*e.g.*, search increases until 5 in-story articles), and then the traffic decreases as the story depth increases. In the previous section, we observed the same behaviour for the dwell time per article. This suggests that when users are skimming many articles (because they are redundant, or to search for a specific piece of information), the mentioned upstream traffic categories are less frequent. For instance, users often use search sites to find articles related to a story. However, search sites are less frequently used, if users access many articles and they are skimming the articles for specific facts.

In this section, we showed that story-focused sessions differ from non-story-focused sessions. We also showed that the reading behaviour and the strategies employed to find the articles depend on how many articles of the same story are accessed.

6 How does Story-focused Reading affect the Engagement with Online News?

We study how story-focused reading is actually supported by news sites. We investigate whether inline links (*i.e.*, links embedded in the article text, see Section: 3), which can be considered to promote story-focused reading, have an effect on user reading behaviour on a news site and hence the engagement with that site. We distinguish between *internal links*, which point to a page on the same news provider, and *external links*, which point to a page on a different site.

6.1 Methodology

For each news provider site, we extract the page views of the news reading sessions that belong to the provider site under consideration (views on article and non-article pages). We call these the *provider sessions* of the news site. A provider session is *internal-story-focused*, if the user clicked on at least one inline link (views at least two pages related to the same story). If the inline link brings the user to an external page (outside the provider), we refer to that provider session as *external-story-focused*. Otherwise, the provider session is labelled as *non-story-focused*.

We study all news sites having users with at least one story-focused and one non-story-focused session, resulting in a sample of 50 news sites, 57K users and 1M provider sessions. Restricting to users with both types of sessions ensure that any observed difference is not an artefact of user browsing behaviours (*e.g.*, users performing story-focused reading are always more engaged than users that do not focus on stories).

We analyse user engagement, more precisely, the period of engagement and the re-engagement [28], with three metrics. For each news provider site, we calculate the average number of *page views* and the *dwell time* per provider session of a user, which describes user reading behaviour during the period of engagement. We also calculate the loyalty of a user to that site, using *absence time*, which is the time elapsed between two provider sessions of a user. This metric was introduced by [12], where it was assumed that engaged users return sooner to a site, and hence their absence times are shorter. Here, we study whether story-focused reading has an effect on this metric (*e.g.*, leads to shorter absence time). We calculate the percentage of provider sessions with an absence time below 12 hours, which represents users who re-engage to the same news site within that time. The same results were observed using 6, 24 and 36 hours.

6.2 Effects of Hyperlinks

We first look at the relationships between inline links and provider sessions. The percentage of inline links to pages of the same news provider correlates moderately with the percentage of internal-story-focused sessions ($\rho = 0.62$, p-value < 0.01). In addition, the percentage of external inline links correlates
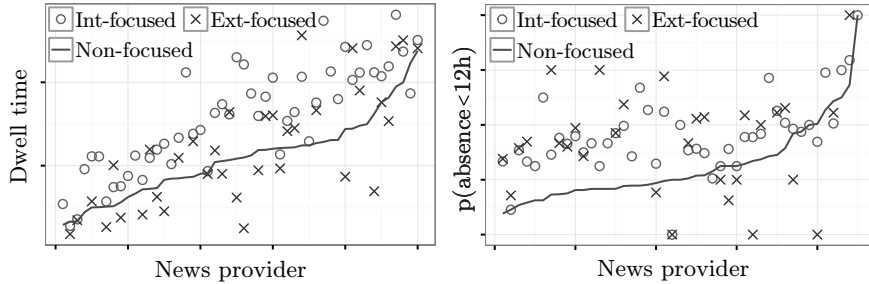
(a) Dwell time per provider session. (b) Absence time after provider session.

Figure 4: Session activity (dwell time) and loyalty (absence time) of users depending on the type of provider session.

moderately with the percentage of external-story-focused sessions ($\rho = 0.56$, p-value < 0.01). Therefore, providing inline links can lead to more story-focused reading within a news site. We now investigate whether this leads to higher user engagement with the news site as well.

Figure 4 depicts the average dwell time, and the probability that the absence time is below 12 hours per provider, depending on the type of provider session.[7] The x-axis represents the providers, ordered by increasing dwell time for non-story-focused sessions (represented by the line). The two types of dots represent internal-story-focused sessions (circle) and external-story-focused sessions (cross). A similar plot is obtained for the average page views metric (omitted as it does not bring additional insights). Now, we discuss the results for these two types of sessions.

Internal-story-focused. The dwell time is higher for internal-story-focused sessions, for almost all considered news providers. Only 3 (out of 50) providers have their corresponding average dwell time lower for the internal-story-focused sessions. The average increase in dwell time from non-story-focused to internal-story-focused sessions is 50%. The same can be observed with respect to the loyalty metric. For 78% of the providers, we find that there are more users that return earlier after they have an internal-story-focused session. The probability that users come back to the same news provider within the following 12 hours increases by 68%. The K-S test confirms that the differences are statistically significant (p-value $\ll 0.01$).

External-story-focused. Some providers do offer only few inline links to external content,[8] and we were not able to identify external-story-focused sessions for them (for these providers there are no values in the two plots in Figure 4). We focus on the remaining 35 news provider sites, consisting of 31K users and

---

[7]The axis values are removed for confidentiality.

[8]These providers have on average only 3.5% external inline links, compared to the remaining 35 providers that have on average 6.6% external inline links.

37K sessions.

We do not observe an effect on the dwell time (neither positive nor negative). The average increase is only 5.5%, and based on a K-S test we cannot confirm that the distributions are different (p-value = 0.36). Interestingly, for 70% of these news sites, the probability that users return within the following 12 hours increases (the average increase is 76%). The difference is statistically significant (p-value $\ll$ 0.01). This suggests that offering links to external sites does not necessarily hurt user engagement (with respect to the site providing such links).

In many cases, users navigating to external sites when engaged in story-focused reading are more likely to return sooner to the news site. Their experience was positive, and such users are inclined to return to consume more content, whether hosted by them or other sites they link to.

Overall, providing links (inline) – thus to promote story-focused reading – has a positive effect on user engagement, in terms of time spent on the site, the number of articles read on the site, and loyalty to the site. Although only few news sites provide links to content on other sites, we could show that offering such links have a positive effect on user engagement in terms of the user loyalty to the site.

7 How Do Hyperlinks Promote Story-focused Reading?

In the previous section, we observed that users increase their engagement with a news provider when following inline links to related content. However, we assume that the performance of such links (do they get clicked?) depends on the linking strategy of the news provider. We therefore analyse in this section how inline links promote story-focused reading, depending on their type (*e.g.*, linking to "Internal News Article" versus "External Multimedia"), position in the text (*e.g.*, top or bottom of the text), and total number of inline links provided.

7.1 Methodology

In the first two subsections, we group inline links based on their type and position in the text. We use two metrics, *popularity* and *performance*, to compare the various inline link groups. Popularity is concerned with how many inline links belong to the group, whereas performance relates to how often the inline links in the group are clicked. We measure the *popularity* of an inline link group as the percentage of inline links that belong to that group. To measure the *performance* of an inline link group, we calculate the probability that a user clicks on a link of that group using its frequency of occurrence (propLink = #inline links of that group / #inline links), and compare it with the real click probability (propClick = #clicks on inline links in that group / #clicks on inline

Table 5: Popularity and performance of inline links depending on their type.

Link type	%Articles	Popularity	Performance
Internal	87.73%	72.20%	+13.48%
Article	59.76%	28.97%	**+80.39%**
Non-Art.	**62.94%**	**43.22%**	-11.70%
External	51.14%	27.80%	-14.58%
News Article	14.74%	3.68%	-29.66%
News Non-Art.	2.24%	0.46%	-15.19%
News Non-Top	**27.33%**	**11.11%**	-2.60%
Other	17.33%	4.11%	-9.34%
Organisation	11.61%	3.09%	-21.88%
Social Media	7.50%	4.15%	-90.89%
Multimedia	3.58%	0.75%	**+60.75%**
Reference	1.59%	0.45%	-46.53%

links). The difference indicates the performance of the links:

$$\text{LinkPerf} = \frac{\text{propClick - propLink}}{\text{propLink}} \quad .$$

The last subsection is concerned with how the total number of inline links on an article page influences a link performance. The performance is measured by the total number of clicks on inline links and the average number of clicks per link.

7.2 Types of Inline Links

Table 5 shows the popularity and performance depending on the types of inline links, using the predefined site categories (see Section: 3). We also report the percentage of articles containing a link of that type (column "%Articles").

Internal links appear in 87.73% of the articles that have inline links. These links include internal links to articles, and internal links to non-article pages (the latter includes links to topic pages, profiles of politicians or celebrities, *etc.*). Both categories of internal links occur in about 60% of the articles, but the popularity of internal links to non-article pages is higher. On the other hand, internal links to article pages have a higher performance than those that point to non-article pages (+80.39%, and −11.70%, respectively). However, we know from our previous analysis that these links are not as frequently used as links on the front page of the news provider (see Section: 5.3). We hypothesise that the provided links are not well presented or they do not cover the full information need of the users; in other words, their potential in driving users to consume more content is not fully exploited.

External links appear in 51.14% of the articles that have inline links. The most common type of external links are links to news sites outside our sample of top English news sites ("External News Non-Top"), but also the popularity of links to popular news sites is high. However, the link performance is −2.60% for links to less popular news sites, and −29.66% for links to popular news sites. This suggests that users are more attracted to less known sources, as they provide new information related to the story.
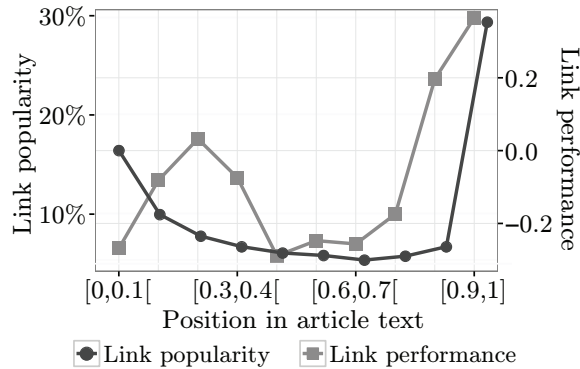
Figure 5: Popularity and performance of inline links depending on their position in the article text.

Links of types "External Other" and "External Organisation" also appear frequently in articles. For instance, the website of the Royal Astronomical Society of Canada[9] was linked from articles related to the story about the "Draconid meteor shower". Other articles link to valuable background information about a story; research studies related to the story "Volcanoes on Mars"[10] or insurance information for "Obamacare"[11]. The performance of inline links to external other content is the third highest with -9.34%.

Links to reference websites (such as Wikipedia) and social media sites are less frequent and less likely to be used, particularly in the case of social media links. For both categories of links, the link performance is the lowest (-90.89% and -46.53%, respectively). Only for links to multimedia content we observe that the link performance is above 0. This suggests that users are interested in gaining more information from multimedia channels (such as YouTube). None of the classes of external links have a performance that compare to the internal links to articles pages.

7.3 Position of Inline Links

We examine the effect of the link position in relation to story-focused reading. We define the position of an inline link in the article text by counting the number of words occurring before the link. We then normalise the position between 0 (beginning of the text) and 1 (end of the text). Figure 5 depicts the popularity (left y-axis), and the performance (right y-axis) of inline links depending on their position.

We see that 30% of the links occur at the end whereas 16% of them appear at

---

[9]http://www.rasc.ca

[10]http://redplanet.asu.edu/?p=2389

[11]http://kff.org/health-reform/perspective/how-buying-insurance-will-change-under-obamacare/

(a) Total number of clicks on inline links.
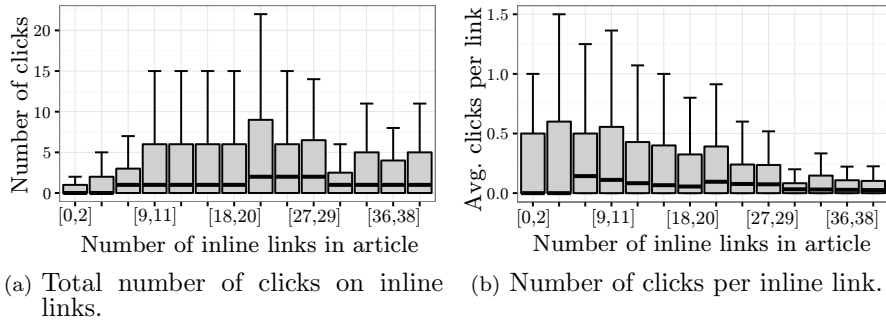
(b) Number of clicks per inline link.

Figure 6: Distribution of the performance of inline links depending on the total number of inline links. Outliers are removed for clarity.

the beginning of the text. The remaining 54% are distributed within the article text. However, the performance of links located at the beginning of the text is very low (-28%), whereas the best performance is achieved with links at the end of the article text ($+35\%$). We hypothesise that users interested in a story (thus focusing on it), like to read the whole article first, before eventually deciding to read more articles on the same story. To support story-focused reading, a good strategy seems to have a paragraph with inline links to related articles at the end of the article text. We also observed the same when restricted to inline links of types "Internal News Article", "External News Article", or "External News Non-Top".

Interestingly, the inline links located between the upper 20% and 40% of the article text perform also well. A manual inspection of the data shows that these are links to multimedia content. Many news providers have articles embedding a picture with a link to a gallery in their upper part. However, we could also find examples of solely text-based links that refer to multimedia content related to the article story.

7.4 Number of Inline Links

Figure 6 shows the distribution of the number of clicks on inline links and the average clicks per inline link with respect to the number of inline links in the article. There seems to be a "sweet spot" around 10 inline links per article. The number of clicks increases until reaching about 9 to 11 inline links in the article, and then stagnates. The average number of clicks per link also starts to drop around that same number, and articles having more than 29 inline links tend to elicit fewer clicks on inline links than articles with less inline links.

This suggests that (i) having less than 10 inline links per article may be wasting an opportunity, as users may be enticed to click to access related content by offering more links to them; (ii) having between 10 and 29 inline links per article does not result in more clicks, but simply spreads the clicks more;

and finally (iii) having more than 29 inline links may actually harm the user experience and make users less likely to click.

Overall, we showed that the performance of inline links, which allow users to engage in story-focused reading (as they bring users to content related to the story of an article), can be affected by their type, their position, and how many of them are present in an article.

8   Discussion

We performed a large-scale data analysis that focused on a specific aspect of online news consumption: when users focus on a story while reading news, *i.e.*, they read more than one article related to a specific story. We referred to this as *story-focused reading.*

We studied this type of news reading behaviour with several hundred thousand users during a one-month period, on 65 highly-visited news sites. Our study encompasses a large variety of news sites, which millions of users access on a daily basis. In addition, we showed that story-focused reading exists, and that it is not a trivial phenomenon. This type of news reading differs from a user daily consumption of news (*i.e.*, non-story-focused reading) and these differences should be noticed by news providers, as promoting story-focused reading can lead to increased user engagement with the news site.

8.1   Characteristics of Story-focused Reading

Most studies investigating online news consumption have been concerned with how users read news on a specific news site [9, 25]. We add to this body of work a new dimension: we analysed how users engage in news reading across news sites when focusing on a particular story.

We observed that story-focused reading is not simply a consequence of the fact that some stories are more read, have more articles written about them, or covered by more news providers. As in the case of general reading behaviour of users [23, 2], story-focused reading is driven by the interest of the users. In addition, even users that can be considered as casual news readers (*i.e.*, they only read few articles) engage in story-focused reading. Hence, news providers can potentially engage such readers by supporting story-focused reading.

We analysed story-focused sessions and compared them with non-story-focused sessions, and observed that users spend more time reading and visit more news providers when focusing on a specific story. Only when users read many articles about a story, the reading time decreases. Our analysis suggests that this could be due to news articles containing mostly the same information. [18] reported that news reading is a browsing task where users are visiting their preferred news site(s) to keep up to date with the latest news. This kind of behaviour can be found during non-story-focused sessions. However, story-focused reading reflects user browsing behaviour during an information gathering task which "involves the collection of information, often from multiple sources" [18]

around a story of interest. The user might do so to obtain various angles or background information on the story. Returning to the work of [28], our results demonstrate that user motivation – *why* they want to read news – influences the reading behaviour during the period of engagement.

The strategies that readers employ to find articles related to a story depend on how deep they want to delve into the story. If users are only reading a few articles about a story, they tend to gather all information from a single news site. In the case of deeper story-focused reading, where users are interested in the story details or specific information [25], they often use search and social media sites; two tools that are frequently employed to follow news events [26, 16]. Furthermore, many users are coming from less popular news sites and blogs, which makes sense, because blogs frequently link their posts to mainstream news sites when discussing an event [34] and users are following these links to likely gather further information or confirm the veracity of what they are reading [41]. Since news sites do not link to each other [40, 42], because of competition, they exchange less traffic. All these were validated with our analysis.

[25] already highlighted that different types of reading exists and recommended that this should be considered when designing news sites. News providers could adapt their sites when they identify a user engaging in story-focused reading. For instance, such information could be integrated in the personalised news recommender of the news site [23]. Story-related articles in the news feed could be highlighted or content frames containing information and links related to the story could be presented on the front page. It might be also beneficial to provide and link to topic pages containing latest updates, background information, blog entries, eye witness reports, *etc.* related to the story.[12] Story-focused reading also brings new opportunities for news providers to drive traffic to their sites by collecting the most interesting articles and statements around a story, *i.e.*, becoming a news story curator [4], and publishing them via social media channels or email newsletters.

8.2 Promoting Story-focused Reading

A number of previous works have aimed at identifying strategies that keep users engaged with a news site, such as recommending news articles to users [23, 30] or integrating interactive features (*e.g.*, multimedia content, social features, hyperlinks) into news articles [31]. We showed that news providers can promote story-focused reading and increase engagement by linking their articles to other related content.

Most related to our work is the study of [6] that developed an approach for automatically embedding links to related content into news articles, and the study of [28] who highlighted that hyperlinks (and interactivity in general) are an important factor that influences the stages of engagement (these are point of engagement, period of engagement, disengagement, and re-engagement). We contributed to this research by analysing the different linking strategies of news

---

[12]http://readwrite.com/2012/08/20/why-topic-pages-are-the-next-big-thing

providers and their impact on user engagement. Moreover, we extended existing research by showing that the findings apply to many news sites (previous research focused on a single site) and by investigating into the effects of hyperlinks on users' returning behaviour to the site (previous research did not study re-engagement).

We showed that having internal links within the article text promotes story-focused reading and as a result keeps users engaged. It leads to a longer period of engagement (reading sessions are longer) and ealier re-engagement (shorter absence time). We also showed that providing links to external content does not have a negative effect on user engagement; the period of engagement remains the same (reading sessions are the same), and the re-engagement begins even sooner (shorter absence time).

It should, however, be emphasised that this does not mean that news providers should just provide inline links, but the right ones in terms of quantity and quality. As demonstrated in this paper, the type, the position, and the number of links play an important role. Users tend to click on links that bring them to other news articles within the same news site, or to articles published by less known sources, probably because they provide new or less mainstream information. However, it is not a good strategy to offer too many such links, as this is likely to confuse or annoy users. This aligns with the user study reported by [6] showing that too many inline links can have detrimental effect on users' reading experience. Finally, users tend to click on links that are close to the end of the article text.

Overall, we can conclude that the linking strategies of news providers affect the way users engage with their news sites, which is in alignment with the findings of [28] and [6]. However, our results are in contradiction with the linking strategy of many news providers, which often comes down to keeping users as long as possible on their sites by linking to other content on their own site [40, 42]. Instead, our results confirm the assumption of [7] and [8] that it may be beneficial (long-term) to entice users to leave the site (*e.g.*, by offering them interesting content on other sites) in a way that users will want to return to it.

Some news sites (*e.g.*, CNN, Time) already offer links to content around the Web[13] using third-party news recommendation engines. In this context, a cost-per-click (CPC) pricing model is employed, in which the content owner is paying the news provider each time the link is clicked by a user. Such approaches could be also applied to promote story-focused reading. Both, the provider (because of increased engagement and the CPC model) and the user (because s/he is offered interesting content) can benefit from such links.

9   Future Work

Our definition of stories as hard clusters of articles can be relaxed. There are many stories that are related to each other (*e.g.*, *Obamacare* and the threat

---

[13]http://www.outbrain.com/

of government shutdown), and this could be taken into account when studying story-focused reading. It would be also interesting to analyse how users focus on niche stories, especially which other sites and information sources are involved. Furthermore, stories can have follow-up stories, and it would be interesting to study the temporal aspects of story-focused reading; more precisely, how stories evolve over time and how this influences the story-focused reading behaviour of users and user engagement.

Second, it will be important to improve the identification of the related links of an article. We did not consider, for example, links in "related content" boxes. The presentation (*e.g.*, position, colour, size) of the links is also of great importance. This investigation will lead to a better understanding of how to provide related links from an interaction design perspective.

Finally, we did not take into account how the novelty and quality of the related content influences user engagement with respect to the news provider; which has been shown to influence user engagement [28]. We expect that linking to low quality content will have a negative effect on users, who will not likely to click on related content in future interaction with the news site. A study of the post-click satisfaction would enhance the understanding of these interdependencies.

10 Concluding Remarks

This paper provides new insights about how users read news online. More precisely, it studies how users consume content related to a story. Understanding how a story is read can help news providers promoting this type of news reading. Indeed, supporting story-focused reading can promote a successful reading experience, and as a consequence, increase user engagement with the news site, both in terms of time spent and loyalty.

11 Acknowledgements

References

[1] ANAGNOSTOPOULOS, A., KUMAR, R., AND MAHDIAN, M. Influence and correlation in social networks. In *Proc. Conference on Knowledge Discovery and Data Mining* (2008), SIGKDD, ACM, pp. 7–15.

[2] Arapakis, I., Lalmas, M., Cambazoglu, B. B., Marcos, M., and Jose, J. M. User engagement in online news: Under the scope of sentiment, interest, affect, and gaze. *JASIST 65*, 10 (2014), 1988–2005.

[3] Arguello, J., Diaz, F., and Callan, J. Learning to aggregate vertical results into web search results. In *Proc. Conference on Information and Knowledge Management* (2011), CIKM, ACM, pp. 201–210.

[4] Carvin, A. Distant witness. CUNY Journalism Press, 2013.

[5] Catledge, L. D., and Pitkow, J. E. Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN systems 27*, 6 (1995), 1065–1073.

[6] Ceylan, H., Arapakis, I., Donmez, P., and Lalmas, M. Automatically embedding newsworthy links to articles. In *Proc. Conference on Information and Knowledge Management* (2012), CIKM, ACM, pp. 1502–1506.

[7] De Maeyer, J. Hyperlinks and journalism: where do they connect? In *Proc. Future of Journalism Conference* (2011).

[8] Dellarocas, C., Katona, Z., and Rand, W. Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Science 59*, 10 (2013), 2360–2379.

[9] Dezsö, Z., Almaas, E., Lukács, A., Rácz, B., Szakadát, I., and Barabási, A.-L. Dynamics of information access on the web. *Physical Review E* (2006).

[10] Diaz, F. Integration of news content into web results. In *Proc. Conference on Web Search and Data Mining* (2009).

[11] Drezner, D., and Farrell, H. The power and politics of blogs. *Public Choice 134*, 1–2 (2004).

[12] Dupret, G., and Lalmas, M. Absence time and user engagement: evaluating ranking functions. In *Proc. Conference on Web Search and Data Mining* (2013), WSDM, ACM, pp. 173–182.

[13] Fincham, K. Review: Storify (2011). *The Journal of Media Literacy Education 3*, 1 (2011).

[14] Flavin, C., and Gurrea, R. Users' motivations and attitude towards the online press. *Journal of Consumer Marketing 26*, 3 (2009), 164–174.

[15] Gibbs, W. J. Examining users on news provider web sites: A review of methodology. *J. Usability Studies 3*, 3 (May 2008), 129–148.

[16] Hu, Y., Farnham, S., and Talamadupula, K. Predicting user engagement on twitter with real-world events. In *Ninth International AAAI Conference on Web and Social Media* (2015).

[17] Keegan, B., Gergle, D., and Contractor, N. Hot off the wiki: Dynamics, practices, and structures in wikipedia's coverage of the tŌhoku catastrophes. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (New York, NY, USA, 2011), WikiSym '11, ACM, pp. 105–113.

[18] Kellar, M., Watters, C., and Shepherd, M. A field study characterizing web-based information-seeking tasks. *J. Am. Soc. Inf. Sci. Technol. 58*, 7 (May 2007), 999–1018.

[19] Kumar, R., and Tomkins, A. A characterization of online browsing behavior. In *Proceedings of the 19th International Conference on World Wide Web* (New York, NY, USA, 2010), WWW '10, ACM, pp. 561–570.

[20] Lehmann, J., Lalmas, M., Dupret, G., and Baeza-Yates, R. Online multitasking and user engagement. In *Proc. Conference on Information and Knowledge Management* (2013), CIKM, ACM, pp. 519–528.

[21] Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., and Kaltenbrunner, A. Reader preferences and behavior on wikipedia. In *Proc. Conference on Hypertext and Hypermedia* (2014), HT, ACM, pp. 88–97.

[22] Li, L., Zheng, L., Yang, F., and Li, T. Modeling and broadening temporal user interest in personalized news recommendation. *Expert Syst. Appl. 41*, 7 (June 2014), 3168–3177.

[23] Liu, J., Dolan, P., and Pedersen, E. R. Personalized news recommendation based on click behavior. In *Proc. Conference on Intelligent User Interfaces* (2010), IUI, ACM, pp. 31–40.

[24] Louis, A., Crestan, E., Billawala, Y., Shen, R., Diaz, F., and Crespo, J.-F. Use of query similarity for improving presentation of news verticals. In *VLDS* (2011), M. Brambilla, F. Casati, and S. Ceri, Eds., vol. 880 of *CEUR Workshop Proceedings*, CEUR-WS.org, pp. 62–67.

[25] Marshall, C. C. The gray lady gets a new dress: A field study of the times news reader. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (New York, NY, USA, 2007), JCDL '07, ACM, pp. 259–268.

[26] McCreadie, R., Macdonald, C., and Ounis, I. News vertical search: when and what to display to users. In *Proc. Conference on Research and Development in Information Retrieval* (2013), SIGIR, ACM, pp. 253–262.

[27] Milne, D., and Witten, I. H. Learning to link with wikipedia. In *Proc. Conference on Information and Knowledge Management* (2008), CIKM, ACM, pp. 509–518.

[28] O'Brien, H. L. Exploring user engagement in online news interactions. *American Society for Information Science and Technology (ASIS&T) 48*, 1 (2011), 1–10.

[29] O'Brien, H. L., and Toms, E. G. What is user engagement? a conceptual framework for defining user engagement with technology. *J. Am. Soc. Inf. Sci. Technol. 59*, 6 (Apr. 2008), 938–955.

[30] Oghina, A. *Optimizing engagement in online news*. PhD thesis, University of Amsterdam, 2013.

[31] Opgenhaffen, M., and d'Haenens, L. The impact of online news features on learning from news: a knowledge experiment. *Journal of Internet Science 6*, 1 (2011), 8–28.

[32] Pew Research Center. Understanding the participatory news consumer. http://www.pewinternet.org/̃/media/Files/Reports/2010/PIP_Understanding_the_Participatory_N 2010.

[33] Pew Research Center. In changing news landscape, even television is vulnerable. http://www.people-press.org/files/legacy-pdf/2012 News Consumption Report.pdf, 2012.

[34] Reese, S. D., Rutigliano, L., Hyun, K., and Jeong, J. Mapping the blogosphere: Professional and citizen-based media in the global news arena. *Journalism 8*, 3 (2007), 235–261.

[35] Roos, J. M. *Hyper-Media Search and Consumption*. PhD thesis, Duke University, 2012.

[36] Saez-Trumper, D., Castillo, C., and Lalmas, M. Social media news communities: gatekeeping, coverage, and statement bias. In *Proc. Conference on Information and Knowledge Management* (2013), CIKM, ACM, pp. 1679–1684.

[37] Sato, Y., Yokomoto, D., Nakasaki, H., Kawaba, M., Utsuro, T., and Fukuhara, T. Linking topics of news and blogs with wikipedia for complementary navigation. In *Social Software: Recent Trends and Developments in Social Software*. Springer, 2011, pp. 75–87.

[38] Sellen, A. J., Murphy, R., and Shaw, K. L. How knowledge workers use the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2002), CHI '02, ACM, pp. 227–234.

[39] SON, J.-W., KIM, A.-Y., AND PARK, S.-B. A location-based news article recommendation with explicit localized semantic analysis. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2013), SIGIR '13, ACM, pp. 293–302.

[40] STEENSEN, S. Online journalism and the promises of new technology. *Journalism Studies 12*, 3 (2011), 311–327.

[41] THE POYNTER INSTITUTE. Survey: Americans turn to established media for breaking news, mobile. http://www.poynter.org/latest-news/top-stories/190586/new-data-show-shifting-patterns-as-people-seek-news-across-platforms/, 2012.

[42] TUROW, J., AND TSUI, L. *The Hyperlinked Society: Questioning Connections in the Digital Age.* The University of Michigan Press, 2008.

[43] WAGNER, E. J., LIU, J., BIRNBAUM, L., AND FORBUS, K. D. Rich interfaces for reading news on the web. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2009), IUI '09, ACM, pp. 27–36.

[44] YADAMSUREN, B., AND ERDELEZ, S. Online news reading behavior: From habitual reading to stumbling upon news. *American Society for Information Science and Technology (ASIS&T)* (2011).