Online Multitasking and User Engagement

Janette Lehmann
Universitat Pompeu Fabra
Barcelona, Spain
jnt.lehmann@gmail.com

Mounia Lalmas
Yahoo! Labs
Barcelona, Spain
mounia@acm.org

Georges Dupret
Yahoo! Labs
Sunnyvale, USA
gdupret@yahoo-inc.com

Ricardo Baeza-Yates
Yahoo! Labs
Barcelona, Spain
rbaeza@acm.org

ABSTRACT

Users often access and re-access more than one site during an online session, effectively engaging in *multitasking*. In this paper, we study the effect of online multitasking on two widely used engagement metrics designed to capture users browsing behavior with a site. Our study is based on browsing data of 2.5M users across 760 sites encompassing diverse types of services such as social media, news and mail. To account for multitasking we need to redefine how user sessions are represented and we need to adapt the metrics under study. We introduce a new representation of user sessions: tree-streams – as opposed to the commonly used click-streams – present a more accurate picture of the browsing behavior of a user that includes how users switch between sites (e.g., hyperlinking, teleporting, backpaging). We then discuss a number of insights on multitasking patterns, and show how these help to better understand how users engage with sites. Finally, we define metrics that characterize multitasking during online sessions and show how they provide additional insights to standard engagement metrics.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing; H.5.2 [**User interface**]: Evaluation/methodology

General Terms

Measurement

Keywords

user engagement, online multitasking, switching between sites, teleporting, hyperlinking, backpaging

1. INTRODUCTION

In the online industry, user engagement refers to the quality of the user experience associated with the desire to use a web application [18]. We focus on one aspect of user engagement, "stickiness", which is concerned with users regularly

spending time on a website. Stickiness is mostly measured through metrics assessing users' depth of interaction with a website. Widely-used metrics include click-through rates, time spent on a site (dwell time), page views, return rates, and number of unique users. Our work focuses on two such metrics, dwell time and page views, which we use as a basis to develop new metrics that incorporate an important user behaviour in online sessions: multitasking.

When users are performing a task on the web (e.g. planning a holiday), they may visit several sites (e.g. to compare offers from different travel sites, read reviews) within the session but also over several sessions before completing the task (e.g. to check offers over several days). In this paper, we are concerned with users accessing several sites within an online session, e.g., emailing, reading news, accessing a social network. A user may access several sites to perform a main task or he or she may actually perform several totally unrelated tasks in parallel (e.g., responding to an email while reading news). The latter is very comparable to our daily life, where we often handle several tasks in parallel and we switch between them [19]. This phenomenon can also be observed on the Web. We refer to both cases as *online multitasking*. In this paper, without distinguish between the two cases; our focus is the effect of accessing and re-accessing several sites within an online session on the browsing behaviour of users as measured by two commonly used engagement metrics. The metrics proposed in this paper, however, provide insights, about the type of multitasking.

Within a multitasking session, users can switch between sites[1] in several ways: *hyperlinking* (clicking on a link), *teleporting* (jumping to a page using bookmarks or typing an URL) or *backpaging* (using the back button on the browser, or when several tabs or windows are open and the user returns to one of them) [17, 25]. Backpaging implies that the navigation between sites is not linear, hence should not be modelled as standard linear click-streams. Referrer trees [16] were defined to model non-linear navigation. However, because engagement metrics are typically defined based on streams, and not trees, we "re-linearized" referrer trees into *tree-streams* and incorporate information about how users switch between sites.

Typical metrics that characterize the browsing activity of users when visiting a site are the number of page views and the dwell time during a visit on that site. Modifying what we understand by multitasking has naturally consequences on how these metrics are calculated and on the conclusions we can draw. We perform an extensive study of online multi-

---

[1]In this paper, we use site and website interchangeably.

tasking and its effect on these two metrics. We examine how often a site is visited within a session, how many sites are visited in the same session, the type of sites being considered (to capture the effect of the task, e.g., reading news vs. doing emails), how users switch between sites, and whether any of these influence the assessment of user engagement. One outcome is that multitasking is affecting the way users access sites and should be considered when measuring user engagement. We therefore define metrics that characterize multitasking during online sessions and show how they provide additional insights to standard engagement metrics.

Section 2 discusses related work. Section 3 describes the data used in our study. Tree-streams are formally defined and studied in Section 4. The nature of multitasking is investigated in Section 5. The new metrics are defined and evaluated in Sections 6 and 7, respectively. The paper ends with conclusions and thoughts for future work.

2. RELATED WORK

Web analytics and user engagement. A common web analytics application is the study of the characteristics of users browsing behavior. In the online web analytics industry is used to understand how users engage with a site and includes metrics such as click-through rate, time spent on a site (dwell time), page views, return rates, and number of users. These metrics, referred to as engagement metrics, assess users' depth of engagement with a site. Although they cannot explain why users engage with a site, they have been used as proxy for online user engagement. Indeed, the fact that, for example, two million users choose to access a site daily is an indication of a high engagement with that site. Major websites are compared on this basis. In this paper, we study multitasking within the remit of two such metrics, page views and dwell time.

Navigation. How users arrive or return to a page has been the focus of many studies e.g. [22]. For many years, users returned to a previously visited page via the back button [6]. Nowadays, the usage of tabs and tab switching has increased [7], and in fact overtaken the back button usage [4]. This way to navigate between pages is called parallel browsing or non-linear navigation, as pages can be (re)visited simultaneously. Tabs are particularly useful in online multitasking, as they allow users to pause one task to perform another [4]. Different models have been proposed to describe this behavior [1, 2]. In particular, referrer trees from server-side log data have been defined [16]. However, these cannot be used as they are to study engagement. Engagement metrics are calculated on a stream of interactions and there is no straightforward way to adapt them to the tree structure of more complex behaviors. In this paper, we propose to "linearize" referrer trees into *tree-streams*, which we then use as our model of the user interaction data.

Web task taxonomy. Diverse services (search engine, news portals, e-commerce) exist on the Web, enabling users to accomplish various tasks (e.g. shopping, searching, reading). Several approaches were developed to classify tasks and build taxonomies related to web search e.g. [13] and across all tasks in the Web [9, 12]. For instance, Kellar *et al.* [9] defined four types of web tasks, fact finding, information gathering, transaction, and browsing. Others have grouped web tasks according to the type of service, for example, into categories such as social media, search and shop-

ping [11, 10]. In this paper, we adopt a similar taxonomy, as previous work has shown that how users engage with a site is influenced by the type of site [11].

Multitasking. Online multitasking has been studied in the context of a web search session [13, 20]. For instance, it has been observed that multitasking happens in 81% of sessions [20]. Whereas a number of works provide evidence for multitasking during an online session [10], only Wang *et al.* [24] have studied this phenomenon in detail. Through an online survey, they show that 92% of the participants had online sessions where they accessed several sites, to perform between 2 to 8 tasks.

Other works do not explicitly refer to online multitasking, but provide useful insights. For instance, users access different sites during a session [10] and a large proportion of pages are visited more than once (revisitation rate around 81% in 2001 [14] and 73% in 2005 [6]). In addition, the frequency at which a page is revisited differs depending on user habits and the type of website [10, 17], or in other words, the web tasks a user accomplishes on the site. With respect to the latter, three types of revisitation have been identified, short-term (backtrack, undo), medium-term (re-utilize, observe) and long-term revisits (rediscover), where it was shown that 73% of revisits are short-term. This is in accordance with another study [6] that reported that 74% of revisits are performed within a session. All these provide a strong evidence that multitasking during online sessions exists and depends on the web tasks. However, since no metrics exist that explicitly account for multitasking, another focus of our work is the development of metrics that capture various aspects of online multitasking.

3. DATASET

We collected one month (July 2012) of anonymised interaction data (tuples of browser cookie, URL, referring URL and timestamp) from a sample of 2.5M users who gave their consent to provide browsing data through the toolbar of a large Internet company. These users represent a sample of users across the world who access that company and many other websites. Users with very low or high activity (lower and upper 5% of the distribution) were excluded, which resulted in a dataset of 785M page views. As we are working at site and not page level, we extracted from each page view the first level of the subdomain (e.g., wikipedia.org) that was visited. For larger portals (e.g., AOL, Google, MSN, Yahoo!) we considered the second level of the subdomain (e.g., mail.yahoo.com), as these sites provide numerous services (e.g., search, mail, news).

The most popular sites, measured by the number of users, were selected and categorised using two publicly available scheme: `http://www.dmoz.org/` and `http://dir.yahoo.com`. This resulted into a total of 11 distinct categories and 33 subcategories, which are listed in Table 1. The percentage of sites in these (sub)categories and a description of some of the subcategories are also shown. The categorised dataset contains 760 sites from 70 countries and regions, and accounts for 60% of the traffic in our original dataset. The sites cover a wide range of services (e.g., mail, news, shopping) sometimes catering for different subcategories of a given service (e.g., news about sport and finance). Studies have shown that the type of site influences engagement [11]. It is likely that online multitasking differs across sites of

Table 1: Site (sub)categories and percentages of sites in each (sub)category.

Cat.	Subcat.	%Sites	Description
news 22.1%	news	5.79%	
	news (soc.)	5.13%	society
	news (sport)	2.63%	
	news (enter.)	2.24%	music, movies, tv, etc.
	news (finance)	1.97%	
	news (life)	1.58%	health, housing, etc.
	news (tech)	1.58%	technology
	news (weather)	1.18%	
search 15.3%	search	12.63%	
	search (special)	1.58%	search for lyrics, jobs, etc.
	directory	1.05%	
service 11.6%	service	7.63%	translators, banks, etc.
	maps	3.03%	
	organization	0.92%	bookmarks, calendar, etc.
sharing 9.6%	blogging	3.55%	
	knowledge	3.55%	collaborative creation and collection of content
	sharing	2.50%	sharing of videos, files, etc.
navi 9.3%	front page	6.58%	
	front page (pers.)	1.84%	personalized front pages
	sitemap	0.92%	
support 8.7%	support	1.58%	sites that provide products and support for them
	download	7.11%	downloading software
shopping 7.9%	shopping	4.34%	
	auctions	2.11%	
	comparison	1.45%	sites to compare prices of products
leisure 5.7%	adult	2.76%	
	games	1.97%	
	entertainment	0.92%	sites with music, tv, etc.
mail 3.9%	mail	3.95%	
social 3.0%	social media	1.97%	
	dating	1.05%	
settings 2.9%	login	1.71%	
	settings	1.18%	profile setting, site personalization

different categories, and that these differences impact the understanding and interpretation of common metrics that assess the browsing behavior of users (as we already mentioned, in this study we focus on page views and dwell time).

4. NAVIGATION MODEL

Previous work [6, 25] showed that users commonly use the back button to revisit sites and frequently maintain several tabs open and switch between them. We will show that accounting for this behaviour provides additional insights on the effect of multitasking on metrics that measure online browsing behavior. To capture this type of navigation, we first define a new model called *tree-streams*.

4.1 Definition

Most studies investigating online behaviour model navigation as a linear click-stream; user interactions are ordered by timestamps and the accessed pages form a linear navigation path. This is illustrated in Figure 1 where (a) shows an example of log data and (b) shows the corresponding linear navigation path.

Click-streams based on standard server-side log data fail to capture some behaviours that we think are important
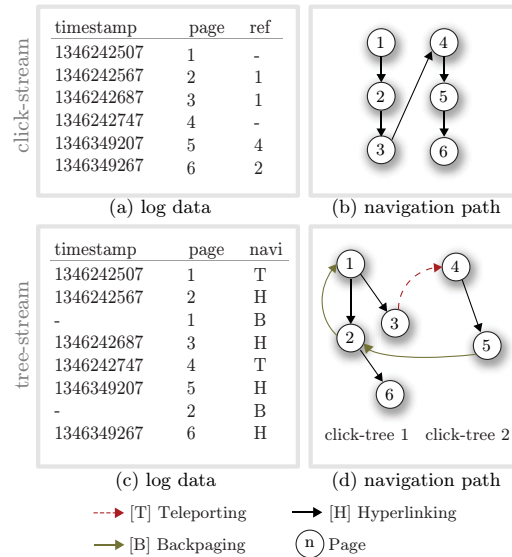

Figure 1: From click-streams to tree-streams.

when studying online multitasking. Users may return to an open tab or window or may use the back button to return to any previously visited and cached page. Since the revisit to a cached page does not require interaction with the web server, only the client-side log data records this type of navigation, and as such, is mostly non available. Therefore, we propose to model a part of client-side interactions using server-side log data, as these are more widely available.

Log data contain, beside the accessed page, the page the user is coming from (referral page): when a user is on a page and clicks on a hyperlink, the referral page is the page previously accessed. When no referral exists (as is the case with page 4), we deduce that the user jumped directly to the page using for instance a bookmark, a case of teleporting. Now looking at Figure 1(a), page 3 is accessed after page 2, but the referral is page 1. This implies that the user returned to page 1 before accessing page 3, probably using one or more tabs or the back button. The second visit to page 1 is not registered in the logs because browsers commonly cache the pages recently accessed to minimize bandwidth.

One way to consider the referrals when modeling user navigation are referrer trees, as proposed in [16, 23], where nodes represent pages and the links connect pages to their referrals. In Figure 1(d) we have two referrer trees, one starting from page 1 and the other from page 4, as both pages lack a referral. We "re-linearize" referrer trees into tree-streams as illustrated in Figure 1(c) and (d) by re-introducing the missing referrals. This is necessary as engagement metrics for evaluating user behaviour are calculated on streams and not trees. Besides, processing trees typically requires higher time and space complexity. Each pair of connected nodes is then labeled by one of the followings:

- **Teleporting [T]** Navigation without using a hyperlink (no referral is given) by entering a new URL, using a bookmark, etc.

- **Hyperlinking [H]** Navigation from one page to another using a hyperlink.

- **Backpaging [B]** Navigation to a cached page using tabs or back button. The page can be in the same referrer tree or part of another referrer tree.

As for click-streams, tree-streams are split into sessions, where a session ends if more than 30 minutes have elapsed between two successive page views. Finally, continuous page views of the same site are merged to form a site visit.

Tree-streams, as generated here from standard interaction data, do not contain all pages accessed via backpaging. A user may use the back button several times or return to several open tabs. Only the last page accessed in this manner and from which the user explicitly clicks on a link can be detected and included in the tree-stream. A more complete instrumentation would be required on the client side for all such pages to be detected.

Now we report some statistics. Compared with click-streams, tree-streams contain approximately 30% more page views. Moreover, 45% of the pages are accessed through hyperlinking, 31% through teleporting, and 24% through backpaging. In addition, 12% of the backpaged navigation land on a distinct tree, suggesting a task switch (or a different logical session according to the terminology in [16]). Other studies reveal similar figures. For instance, Obendorf *et al.* report in [17] that 43.5% of navigation is via hyperlinking and 14.3% is via backpaging using the back button (their work did not consider backpaging with tabs). However, a study from Huang *et al.* [7] showed that 11.3% of the page views include tab switches. Combining back button and tab usage of the two studies suggest that 24% navigation happens through backpaging, comparable to our findings. The proportion of teleporting reported in these studies is lower than what we observe in our dataset. However, these studies are based on a selective and comparatively limited set of 20 to 100 users. The log data used in our work includes millions of users with very different habits. Similarly to Kumar *et al.* [10] our dataset contains numerous sessions initiated by teleporting and only a few page views (see Table 2).

4.2 Backpaging and dwell time

A widely accepted metric is dwell time, which is the time users spend on a site during their visit. Calculating the exact dwell time is not obvious, since the time spent on the last page of a session is generally not known. Adding to this limitation, there is yet no consensus on how to identify when a session actually ends [15]. Multitasking makes it even more problematic to accurately calculate dwell time[2] because of backpaging. For instance, in Figure 1 the user backpaged from page 2 to page 1, from where he or she accessed page 3. No timestamp is associated with the user return to page 1, which makes it impossible to estimate accurately the time spent on pages 1 and 2. To mitigate this problem we approximate the dwell time as described next.

Let i and j be two pages. Assume that a user backpages from i to j. The time spent on these two pages is known and can be written as $t_{ij}^b = t_i^b + t_j^b$. What we do not know is t_i^b and t_j^b, the time spent on each page, i and j, respectively. We propose to estimate these values by the time spent in each page when accessed via teleporting or hyperlinking, which will be generally known. We denote these dwell time values

t_i^* and t_j^* and devised three methods to estimate them:[3] we averaged the times of (1) all visits on pages i and j, (2) only those visits where pages i and j were accessed in the same order, and (3) only those sessions in which the site containing page i (or j) was visited at least twice in the same session. The second approach focuses on the same page visit pattern (i then j) whereas the third considers sessions on a site where multitasking occurs.

We calculate for each approach the percentage (f) of cases for which t_i^* and t_j^* could not be computed because the pages were not visited through teleporting or hyperlinking. Since we analyze users browsing behaviour at site level, only backpaging between sites was taken into account. We restricted our analysis to sites viewed by at least 100 users. In total, we extracted around 17K page visit pairs.

We use linear regression to examine whether the estimations of t_i^b and t_j^b by t_i^* and t_j^* correlate with t_{ij}^b, using the linear equation $t_{ij}^b = x_0 + x_1 \cdot t_i^* + x_2 \cdot t_j^*$. A correlation of $r^2 = 0.43$ ($f = 0.00\%$) could be observed using all page views (approach 1), but the correlation increases ($r^2 = 0.50$, $f = 9.77\%$) when multitasking is considered (approach 3). The highest correlation ($r^2 = 0.52$) was obtained using the exact same pairs of page views (approach 2), but for $f = 29.89\%$ of cases t_i^* and t_j^* could not be defined. In all cases a low p-value was observed (p-value $< 2.2e^{-16}$). The two coefficients are smaller than 1, but almost the same (e.g. for approach 2, $x_1 = x_2 = 0.69$) indicating that pages are visited in the same manner when using backpaging, but the time per page visit is smaller. This validates using t_i^* and t_j^* as estimate of t_i^b and t_j^b, respectively.

In the rest of this paper, we use the third approach to estimate dwell time, because, although the correlation obtained is slightly lower than with the second approach, more dwell times could be approximated.

5. CHARACTERISTICS AND EFFECTS

We present a number of characteristics of multitasking which we observed in our dataset. We also show how multitasking influences the way users visit a site.

5.1 Multitasking in sessions

Our dataset contains 41 million sessions. The average number of sessions per user is 16.6 ($sd = 28.38$). We measure the activity during a site visit with two metrics, page views ($\#PageViews$) and dwell time ($DwellTime$). We define the degree of multitasking in an online session as the number of distinct sites visited during that session, denoted by $\#Sites$. In a session, a site may be revisited several times. Analogously to page revisitation rate defined in [21], we define the site revisitation rate or *recurrence rate* as:

$$RecRate = \frac{\#Visits - \#Sites}{\#Visits}$$

where $\#Visits$ is the total number of site visits during the session. Table 2 shows for sessions of increasing length, measured by page views, the multitasking, and site visit statistics. The average values across all sessions appears in the last row. On average, 10.20 ($sd = 18.85$) distinct sites are visited within a session, and for 22% ($sd = 0.26$) of the visits the site was accessed previously. The table shows that more

---

[2]http://www.kaushik.net/avinash/standard-metrics-revisited-time-on-page-and-time-on-site/

[3]For simplicity, we defined the average dwell time over all pages of a site, not for each page separately.

Table 2: Multitasking and site visit characteristics depending on the size of a session. Average and standard deviation are reported ($avg|sd$).

#PageViews $\leq$	Session		Multitasking					Site visits	
	%Sessions	Time [min]	#Sites	RecRate	Tele	Link	Back	#PageViews	DwellTime [min]
2	27%	0.65 | 3.05	1.10 | 0.30	0.00 | 0.00	–	–	–	1.10 | 0.30	0.59 | 2.92
4	37%	2.14 | 6.15	1.42 | 0.78	0.01 | 0.00	40%	60%	0%	1.28 | 0.65	1.51 | 4.72
8	50%	4.80 | 10.16	2.02 | 1.53	0.05 | 0.14	30%	52%	17%	1.49 | 1.10	2.37 | 6.09
16	65%	8.67 | 15.26	3.01 | 2.79	0.10 | 0.17	29%	47%	24%	1.69 | 1.69	2.88 | 6.99
32	79%	13.52 | 21.50	4.48 | 4.84	0.15 | 0.20	27%	45%	28%	1.87 | 2.41	3.02 | 7.49
64	90%	18.72 | 28.60	6.38 | 7.94	0.19 | 0.24	25%	44%	31%	2.03 | 3.34	2.93 | 7.70
128	97%	23.09 | 35.58	8.29 | 11.94	0.21 | 0.24	23%	44%	33%	2.18 | 4.48	2.78 | 7.78
256	99%	25.59 | 40.68	9.62 | 15.90	0.22 | 0.24	22%	44%	34%	2.29 | 5.78	2.66 | 7.84
512	100%	26.45 | 43.25	10.20 | 18.85	0.22 | 0.26	21%	44%	35%	2.36 | 7.03	2.59 | 7.98

sites are visited and revisited as the session length increases. Sessions with up to 16 page views consist on average 3.01 distinct sites with a recurrence rate of 0.10. By contrast, sessions with up to 256 page views have on average 9.62 different visited sites with a recurrence rate of 0.22.

The way users revisit sites, whether via teleporting, hyperlinking or backpaging, also varies depending on the session length. Whereas teleporting and hyperlinking are the most important mechanisms to re-access a site during short sessions (we have 30% teleporting and 52% hyperlinking for sessions with $\leq$ 16 page views), backpaging becomes more predominant in longer sessions. Similar results were reported in [17]. Both their and our study, show that tabs or the back button are often used to revisit a site, and that sites are often revisited during a session.

We also observe a relationship between the number of pages visited and the time spent on a page. For sessions with more than 32 pages views, the number of page views per site visit increases with the session length, but the time spent on each visit to the site tends to decrease. In other words, the longer the session, and the more the users are multitasking, the quicker they navigate between pages. This results in more page views but a lower dwell time per site visit. In addition, backpaging increases with session length and is associated with shorter time spent on a page (see the coefficient of the linear regression model in Section 4.2). This suggests that visitors use backpaging to access previously visited pages or sites quicker, and spend less time on pages or sites they are returning to.

5.2 Visit and inter-visit activity

We have illustrated the importance of multitasking during online sessions and its relation with the navigation behaviour of users. In this section, we show how multitasking affects the browsing activity of users on a site, which we measure as the time spent on the site, aka dwell time. We do so for four selected categories of sites: news (finance), news (tech), social media, and mail. We extract for each category a random sample of 10,000 sessions. Figure 3 shows various statistics regarding the number of visits during a session and the time between visits. We refer to the latter as "absence time" following the study described in [5].

Sites with the highest number of visits within a session belong to the social media category ($avg = 2.28$, $sd = 4.78$), whereas news (tech) sites are the least revisited sites ($avg = 1.76$, $sd = 1.59$). These two categories have an average absence time of $4.47min$ ($sd = 14.11$) and $3.95min$ ($sd = 14.16$), respectively, although the distributions are similar. The news (finance) sites have a skewer distribu-

tion, indicating a higher proportion of short absence time for sites in this category. Finally, mail sites have the highest absence time, $6.86min$ on average ($sd = 18.53$). However, when looking at the distributions of the absence time across all categories of sites, we see that the median is less than $1min$, and this for all categories. That is, many sites are revisited after a short break. We speculate that a short break corresponds to an interruption of the task being performed by the user (on the site), whereas a longer break indicates that the user is returning to the site to perform a new task.

Next, we look at how multitasking is related to the way sites are revisited within a session. For each site, we select all sessions where the site was visited at least four times. Figure 2 (Top) shows the average dwell time at the ith visit to a site, as a proportion of the total session length. The time spent on mail sites decreases at each revisit. The opposite is observed for social media sites. A possible explanation is that, for mail sites, there are less messages to read in subsequent visits, whereas for social media sites, users might initiate new conversations with friends that are online.

News (finance) is an example of category for which neither a lower or higher dwell time is observed at each subsequent revisit. We hypothesise that each visit corresponds either to a new task or a user following some evolving piece of information such as checking the latest stock price figures.

Figure 2 (Bottom) shows how users access a site at the ith visit as a percentage of the time they use teleportation, hyperlinking or backpaging. For all four categories of sites, the first visit is often through teleportation. Accessing a site in this manner indicates a high level of engagement, in particular in terms of loyalty, with the site, since users are likely to have bookmarked the site at some previous interaction with it. For instance, teleportation is more frequently used to access news (tech) sites than news (finance) sites.

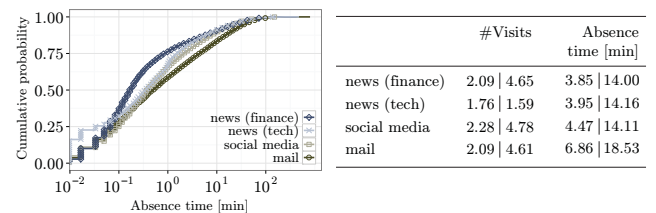|	#Visits	Absence time [min]
news (finance)	2.09 | 4.65	3.85 | 14.00
news (tech)	1.76 | 1.59	3.95 | 14.16
social media	2.28 | 4.78	4.47 | 14.11
mail	2.09 | 4.61	6.86 | 18.53

Figure 3: Site visit characteristics for four categories of sites: (Left) distribution of time between visits; and (Right) average and standard deviation of number of visits and time between visits ($avg|sd$).
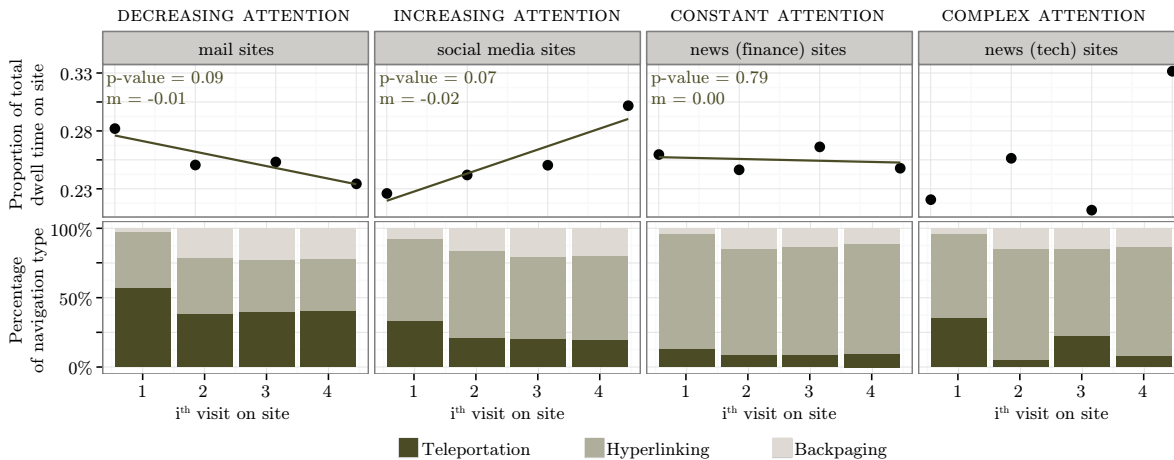
Figure 2: (Top) Visit patterns described by the average dwell time of the ith visit in a session. **(Bottom)** Usage of navigation types described by the proportion of each navigation type at the ith visit in a session.

After the first visit, backpaging is increasingly used to access a site. This is an indication that users leave the site by opening a new tab or window, and then return to the site later to continue whatever they were doing on the site. Finally, users still revisit a site mostly through hyperlinking, suggesting that links still have an important role in directing users to a site. For instance, news (finance) sites are mostly accessed through links; users are directed to sites of this category via a link.

The browsing activity for news (tech) sites (here measured by dwell time) is fluctuating. Either no patterns exist or the pattern is complex, and cannot easily be described. However, when looking at the first two visits or the last two visits, in both cases, a higher dwell time can be observed in each second visit. This may indicate that the visits belong to two different tasks, and each task is performed in two distinct visits to the site. Teleportation is more frequent at the 1st and 3th visits, which confirms this hypothesis.

We presented two main findings in this section. First, the time between two visits, or the absence time, can be used as an indication as to whether a user returns to a site to continue on a previously started task or to start a new task. The latter case is a sign of loyalty, as users return to the site to accomplish some new tasks.

Second, the activity pattern at the subsequent visits to a site provides additional information about how users engage with the site. We have identified four main patterns of *user attention to the site* (decreasing, increasing, constant and complex) and given examples of sites belonging to each in Figure 2. For instance, an increase in dwell time (increasing attention) can reflect "stickiness": users are increasingly engaged with the site during the session. In the next section we define two metrics that capture these multitasking characteristics.

6. MEASURING MULTITASKING

In the previous section, we showed that on average 22% of the visits re-access sites previously visited within the same session, and that revisitation has an effect on user activity on the revisited sites. Therefore, only considering the user activity – in terms of dwell time and page views – within a visit provides a partial view of user engagement. In this section, we propose two metrics that cover different aspects of how users access a site during a session.

Page views during a visit and dwell time are commonly used metrics representing the browsing activity of a user. We therefore adapt them to our more complete definition of a session. Since these measures play somewhat interchangeable roles, it is convenient to associate a subscript to each one: we will use pw to refer to (the number of) page views and dt to the dwell time during a visit.

Visit and session activity. We also consider two other common metrics, $Visit_m$ and $Session_m$. $Visit_m$ is either the dwell time ($m = dt$) or the number of page views ($m = pw$) calculated at visit level. $Session_m$ is analogous, but calculated across the whole session. For instance, $Session_{dt}$ is the total time spent on a site during the full online session. $Session_m$ in some way accounts for the fact that users multitask when online.[4] In the rest of this paper, we compare our two proposed metrics to these.

6.1 Cumulative activity

Section 5.2 showed that looking at absence time provides some insights into how users engage with a site. We make the following assumption on how to interpret the time between site visits: if the next visit is shortly after the previous one, we consider that the two visits belong to the *same* task. If the time between the two visits is long, the user is returning to the site to perform a different task.[5] In the case of a search engine for example, a short absence time could refer to the same search task, whereas a long one indicates a different search. The latter case is related to the loyalty of users to a site, whereas the former is more related to the actual activity of the user on the site, an activity that was briefly interrupted. Drawing from the web search context again [8], a short absence time – often taken to be less than 30 minutes in the literature – is indicative of a query re-formulation or

---

a re-orientation of the original task, whereas a long absence time indicates that the two sessions are unrelated.

We use the following metric to express this:

$$CumAct_{m,k} = \log_{10}(v_1 + \sum_{i=2}^{n} v_i \cdot iv_i^k)$$

Here v_i corresponds to the browsing activity during the ith visit (page views or dwell time) and iv_i is the browsing activity between the $(i-1)^{th}$ and i^{th} visit (page views or absence time). With our definition of $CumAct_{m,k}$, a long absence or high number of page views between visits to a site is an indication of a high loyalty to that site: the user is returning to the site to perform some new tasks. The value of $CumAct_{m,k}$, referred to as the *cumulative activity* metric, increases with the time spent between visits to the site.

The exponent k is used to scale the iv_i parameter; a high value increases the importance of between-visit activity iv_i. When $k = 0$, the focus is solely on the visits to the site, and $CumAct_{m,0}$ corresponds to the total activity on the site during the whole session (as in $Session_m$).

Selection of k. The exponent k can take several values. Using the dwell time as proxy for the browsing activity, we study what these values imply. To verify that the proposed cumulative activity metric adds new insight, we must ensure that it does not capture the same information as the common metrics $Visit_{dt}$ and $Session_{dt}$. We could compare the scores the different metrics attribute to different sites, but it is unclear how these scores should be normalized to be comparable. We therefore compare instead the rankings of sites using $Session_{dt}$, $Visit_{dt}$ and $CumAct_{dt,k}$ for various values of k with the Kendall tau coefficient (τ).

In addition, also using τ, we compare the rankings of sites using two successive values of k, that is $CumAct_{dt,k}$ and $CumAct_{dt,k-0.5}$. The objective is to determine the extent to which accounting for absence time affects the ranking of sites; a high τ value means that the sites are ranked similarly even though the importance associated with the absence time has increased by 0.5. All comparisons are shown in Figure 4(a).

While the value of k is low, corresponding to a low importance given to absence time, the cumulative activity metric correlates highly with $Visit_{dt}$ and $Session_{dt}$. The measurement is dominated by site visits.

Increasing k to 3 causes τ to decrease (for instance we see that $\tau(CumAct_{dt,3}, Session_{dt}) = 0.30$), whereas the correlation between the rankings defined by successive values of k and $k-0.5$ increases (we have $\tau(CumAct_{dt,3}, CumAct_{dt,2.5}) = 0.97$). This shows that accounting for absence time captures the effect of multitasking when measuring user browsing activity. However, values of k higher than 3 lead to minor differences in the τ values, indicating that absence time does not bring new perspectives on multitasking. We therefore fix k to 3 for the rest of our study.

6.2 Activity pattern

We know from Section 5.2 that the browsing activity during a site visits varies and that these variations depend partly on the type of sites (e.g. news vs. mail). Motivated by this, we define two measures that describe how the dwell time or number of page views is changing from visit to visit. We aim to capture a number of assumptions, which we illustrate in the context of dwell time. We interpret an increase
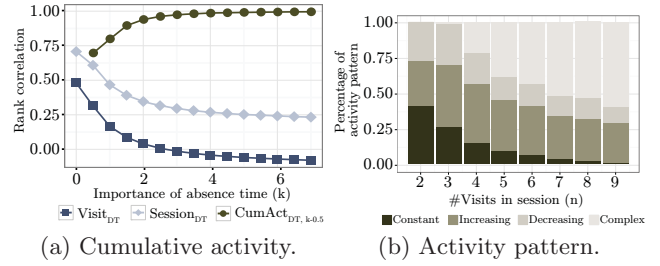

(a) Cumulative activity. (b) Activity pattern.

Figure 4: Cumulative activity for different importance values of the absence times (k), and activity pattern for different numbers of visits in a session (n).

in dwell time as a user getting increasingly more engaged with the site at each revisit, whereas we hypothesise that a decrease in dwell time corresponds to a user shifting his or her attention away from the site, arguably because the focus moves to some other task on another site. Finally, a constant dwell time is interpreted as the user repeatedly visiting the site to perform the same type of task. We do not attempt to automatically identify which patterns apply to which types of sites (we leave this for future work); our focus is the provision of measures that account for such patterns.

We modify the measure of "inversion number", a common measure of sortedness, to express the above:

$$inv_{m,n} = \sum_{i,j}^{n}\{v_i + v_j \mid i < j \ and \ v_i <= v_j\}$$

Here v_i and v_j with $i < j$ correspond to the browsing activity during the ith and j^{th} visit, respectively. Whereas the original inversion number determines the number of (v_i, v_j) pairs that do not exhibit a natural order $(v_i > v_j)$, our measure counts how often an increase or no change in browsing activity is observed $(v_i <= v_j)$. Moreover, $inv_{m,n}$ considers the extent to which the browsing activity changes when comparing the ith to the jth visit of a site, where for instance, an increase of dwell time from $v_i = 10sec$ to $v_j = 12$ secs is considered less important than one from $v_i = 10$ secs but to $v_j = 2min$. We use n to refer to the number of visits to a site within a session. As shown in Section 5.2, how a site is visited during an online session depends on how often the site is revisited.

We next normalize $inv_{m,n}$ between -1 and 1 to define a measure that models the shift of attention in the browsing activity during a session:

$$AttShift_{m,n} = 2\frac{inv_{m,n} - minInv_{m,n}}{|maxInv_{m,n}| - |minInv_{m,n}|} - 1$$

Here $minInv_{m,n}$ and $maxInv_{m,n}$ are the inversion numbers after re-ordering the site visits according to, respectively, decreasing and increasing values of the browsing activity on the site (quantified using dwell time or page views). When $AttShift_{m,n}$ equals to 1, the attention is shifting towards the site (*increasing attention*). When $AttShift_{m,n}$ equals to -1, we have the opposite; the attention is shifting away from the site (*decreasing attention*). Finally a value of $AttShift_{m,n}$ close to 0 means no identifiable patterns; we refer to this as *complex attention*.

Finally, the browsing activity at each visit may remain constant, e.g. same dwell time at each visit. $AttShift_{m,n}$

cannot capture this, so we define an additional measure to express this:

$$AttRange_{m,n} = \frac{\sigma(V_{m,n})}{\mu(V_{m,n})}$$

where $\sigma(V_{m,n})$ is the variance and $\pi(V_{m,n})$ is the average browsing activity of all visits V on a site during a session. $AttRange_{m,n}$ is a normalized variance and a value near 0 indicates that the browsing activity exhibits only small fluctuations; we refer to this as *constant attention*.

Selection of n. We aim to understand how the activity pattern changes depending on the number of visits in a session. We use dwell time as the measure of browsing activity. First, we use $AttRange_{dt,n} < 0.1$ to express that the attention remains constant. For all values above 0.1, we use $AttShift_{dt,n}$ to identify the type of activity pattern, namely, increasing, decreasing or complex attention. We say that $AttShift_{dt,n} < -0.25$ indicates a decrease of attention, $AttShift_{dt,n} > 0.25$ means that the attention is shifting towards the site at each subsequent visit, and any value between -0.25 and 0.25 indicates a complex attention. The values 0.1 and 0.25 are chosen arbitrarily, and are sufficient to analyse the effect of n on the activity patterns.[6] Figure 4(b) reports the percentage of sites that belong to each type of activity patterns for different number values of n.

The more often sites are visited in a session (the higher the value of n), the less their activity pattern is either constant or decreasing. In addition, the number of sites that follow an increasing attention pattern does not decrease as much. The percentage of sites with a complex activity pattern however increases. We hypothesise that this is because the site is visited to perform separate tasks, but when a task is carried out in several visits, these visits may follow some specific (not complex) activity patterns (as illustrated with news (tech) sites in Figure 2). In the rest of this paper, we fix n to 4, as the sites are distributed equally over the four types of activity patterns.

7. EVALUATION

The newly proposed multitasking metrics, $CumAct_{m,3}$, $AttShift_{m,4}$, and $AttRange_{m,4}$ provide new insights into the browsing behavior of users. We contrast these with the more standard engagement metrics $Visit_m$ and $Session_m$. For space reason and because the results are similar, we only present the outcomes for dwell time ($m = dt$).

7.1 Site rankings

To compare metrics, we rank sites according to each metric and then we evaluate the similarity between these rankings. Admittedly, if two metrics produce the same ranking, they are equivalent and hence redundant. The Kendall tau coefficient (τ) is used to compare the rankings. We also compare the ranking of sites per category. In Web Analytics only sites of the same category are usually compared against each other (e.g. mail sites, news sites, etc). Indeed, for instance, search sites and social media sites provide different services, and it is important to know how the same type of service is engaged with by users across the sites offering that service. The rank correlation over all sites and the average rank correlation per category are reported in Table 3.

[6]For instance, a higher value than 0.1 would lead to more patterns classified as constant. We leave for future work the detailed study of activity patterns.

Table 3: Kendall tau rank correlation coefficient between five activity metrics. (Top) Rank correlation between all sites. (Bottom) Average rank correlation per site category and its standard deviation.

	$Visit_{dt}$	$Session_{dt}$	$CumAct_{dt,3}$	$AttShift_{dt,4}$
τ between all sites				
$Session_{dt}$	0.57			
$CumAct_{dt,3}$	-0.04	0.24		
$AttShift_{dt,4}$	0.09	0.22	0.02	
$AttRange_{dt,4}$	-0.01	-0.01	-0.26	0.19
τ between all sites per category (avg|sd)				
$Session_{dt}$	0.6|0.2			
$CumAct_{dt,3}$	0.1|0.2	0.4|0.2		
$AttShift_{dt,4}$	0.1|0.2	0.1|0.2	0.1|0.2	
$AttRange_{dt,4}$	-0.1|0.1	-0.1|0.2	-0.1|0.2	0.1|0.1

It is apparent that accounting for multitasking, with the new metrics $CumAct_{dt,3}$, $AttShift_{dt,4}$, and $AttRange_{dt,4}$, leads to different conclusions with regard to engagement within the entire session. The same can be said when looking at engagement across the entire session ($Session_{dt}$). The metrics $Visit_{dt}$ and $Session_{dt}$ have the strongest positive correlations ($\tau(Visit_{dt}, Session_{dt}) = 0.57$). However, the correlations are not high enough to suggest that one of them is redundant. The time a user spends on a site during a single visit or during the whole session does not correlate with the cumulative activity metric (τ is -0.04 and 0.24, respectively) or the two activity pattern metrics (τ is between -0.01 and 0.22). We can also see that the cumulative activity metric ranks sites differently compared to the two activity pattern metrics (τ is 0.02 and -0.26). This shows that all metrics convey different information about the browsing activity of users when multitasking is involved. The same can be observed when comparing sites within a same category. We therefore conclude that all metrics capture distinct aspects of how users engage with a site during a session.

7.2 Case study

We examine how users engage with a site, in terms of their browsing activity, and how these differ across sites. We also concentrate on the importance of considering multitasking. We cluster sites characterized by their $Visit_{dt}$, $Session_{dt}$, $CumAct_{dt,3}$, $AttShift_{dt,4}$, and $AttRange_{dt,4}$ values to identify how similar they are in terms of user behavior. We use the kernel k-means algorithm [3], with a Kendall tau rank correlation kernel. Five clusters were identified. The cluster centroid, the percentage of sites in a category belonging to the cluster (CA column), and the percentage of the sites in the cluster (CL column) are shown in Figure 5. Only the most representative site categories per cluster are listed. We selected categories for which most of the sites are belonging to the cluster (highest CL values) with the condition that at least 5% of the sites in the cluster belong to this category ($CL >= 5\%$). We normalized $Visit_{dt}$, $Session_{dt}$, and $CumAct_{dt,3}$ using the z-score,[7] hence the figure shows the extent to which the standard deviation of a metric value is above or below the mean.

[7]The other metrics are already normalized.

C1 and C2. Sites belonging to clusters $C1$ (e.g. mail, maps, news) and $C2$ (e.g. auctions, front pages, shopping) have a high activity per visit ($Visit_{dt}$) and during the whole session ($Session_{dt}$). A low cumulative activity $CumAct_{dt,3}$ indicates that users are returning after a short absence time, which suggests that users do not return to these sites to start new tasks (they return to continue whatever task they started earlier on). In addition, two opposite activity patterns can be observed. For sites in cluster $C1$, a low value of $AttShift_{dt,4}$ means that dwell time decreases at each subsequent visit (e.g. users spend less time reading news articles than earlier on in the session), which suggests that the focus of the user is moving to some other site(s). For sites in cluster $C2$, the dwell time increases from visit to visit (high $AttShift_{dt,4}$), but the low value of $AttRange_{dt,4}$ indicates that the increase is small. This shows that the attention is shifting slowly towards the site (e.g. users want to make some purchases on a shopping site, hence becoming more focused in this task).

C3 and C4. Sites belonging to clusters $C3$ and $C4$ (e.g. front pages, search, auctions) are sites for which the multitasking effect is significant. For sites in cluster $C3$, the dwell time per visit ($Visit_{dt}$) is lower than on average, but the dwell time per session ($Session_{dt}$) is higher. This means that measuring engagement at visit level only can lead to incorrect conclusions, as users may return several times to the site during a session. The cumulative activity metric ($CumAct_{dt,3}$) is the highest in both clusters compared to the other clusters, indicating that users perform several tasks on these sites within a session (e.g. they have several search tasks). This is further accentuated by the low values of $AttShift_{dt,4}$ and $AttRange_{dt,4}$, which suggests that the visits are not connected with each other, as no simple activity pattern could be observed and the attention remains the same (users dedicate the same time to the site as earlier).

C2 and C3. Comparing the site categories of clusters $C2$ and $C3$ shows that the browsing behavior can differ between sites of the same category. Users visit auction, front page, and shopping sites that belong to cluster $C2$ only once to perform their task, indicated by a high dwell time per visit ($Visit_{dt}$) and a low cumulative activity ($CumAct_{dt,3}$). The opposite can be observed for sites that belong to cluster $C3$, where users return several times to the site – even after a longer absence time – and perform new tasks.

C5. Cluster $C5$ contains sites that are characterized by a short dwell time per visit ($Visit_{dt}$) and over the whole session ($Session_{dt}$); users do not spend a lot of time on these sites. Moreover, the low cumulative activity ($CumAct_{dt,3}$) indicates that the time between the visits is also short. Service and download sites belong to this cluster. This kind of site is visited to perform a precise and defined task, e.g. download a new app and hence the results make intuitive sense. However, the high values of $AttShift_{dt,4}$ and $AttRange_{dt,4}$ reveal that, if users return to the site, they spent more time than before (e.g. first users search for the app they want to download using, for instance, a search site, and then they finally download the app).

In this section, we showed that our proposed metrics reflect how users engage with a site, in terms of their browsing behaviour during the entire session, in the presence of multitasking. We identified different types of behavior and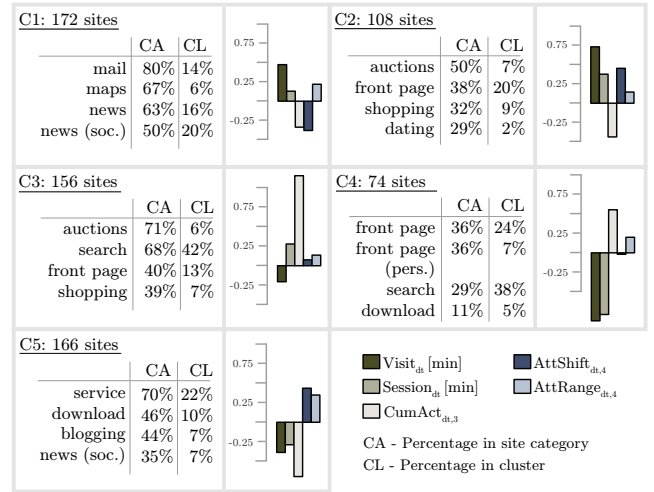 observed that the behavior can differ significantly even between sites of a same category. Some sites are visited to perform a single task during a session. The time between visits is then short and the visits are connected with each other while user attention shows a particular trend, either shifting towards or away from the site. Other sites are characterized by steady return of users even after long absence times. In this case the visits tend not to be connected, and the activity pattern is complex: only parts of the visits belong to a same task.

Figure 5: Site clusters and browsing characteristics.

C1: 172 sites	CA	CL
mail	80%	14%
maps	67%	6%
news	63%	16%
news (soc.)	50%	20%

C2: 108 sites	CA	CL
auctions	50%	7%
front page	38%	20%
shopping	32%	9%
dating	29%	2%

C3: 156 sites	CA	CL
auctions	71%	6%
search	68%	42%
front page	40%	13%
shopping	39%	7%

C4: 74 sites	CA	CL
front page	36%	24%
front page (pers.)	36%	7%
search	29%	38%
download	11%	5%

C5: 166 sites	CA	CL
service	70%	22%
download	46%	10%
blogging	44%	7%
news (soc.)	35%	7%

$Visit_{dt}$ [min] $AttShift_{dt,4}$
$Session_{dt}$ [min] $AttRange_{dt,4}$
$CumAct_{dt,3}$

CA - Percentage in site category
CL - Percentage in cluster

8. CONCLUSIONS AND FUTURE WORK

This paper studies online multitasking and its effect in measuring user engagement. We focus on one aspect of user engagement, "stickiness", which is concerned with users spending time on a site. Stickiness is mostly measured with metrics assessing users' depth of interaction with a site. We focused on two such metrics, page views and dwell time. Our study is based on the online interactions of a large sample of users and their online browsing activity on 760 sites.

Most studies investigating online behaviour model user navigation with linear click-streams. Users may return to a site via an open tab or window or use the back button to return to any previously accessed page, from which they access the next page. Both cases generate additional page views that are not part of standard linear click-streams. Referrer-trees overcome this but cannot be used in a straightforward way to calculate engagement metrics. We "re-linearize" referrer-trees into tree-streams, which are like click-streams, but offer a richer representation of the interaction data. We are able to study how users visit websites, whether through teleporting, backpaging or hyperlinking, bringing additional insights about online multitasking.

We have shown that online multitasking exists, as many sites are visited and revisited during a session. We also demonstrated that multitasking influences the way users access sites and that this depends on the site under consideration. Metrics that describe the browsing activity - such as dwell time and page views during a single visit or the whole session - do not account for multitasking, that is how the visits differ from each other and what users are doing while not on the site. Therefore, we defined two new metrics, each

aiming at capturing specific aspects of the browsing activity with a site during a session when multitasking is involved.

The cumulative activity metric $CumAct_{m,k}$ accounts for the activity between site visits. A user revisiting a site may means that he or she is returning to the site to continue with the same task; in this case, the "visit" to the site is actually splits into several sub-visits, and should be viewed as one visit. The longer the time between two visits, the higher the likelihood that the user is returning to that site to perform a similar type of task, albeit a new one (e.g. a new search for a different information need). This somewhat reflects the loyalty of the user to the site, and the cumulative activity metric increases the importance of visits that preceded a longer break of activity on that site.

The activity pattern metrics, attention range $AttRange_{m,n}$ and attention shift $AttSift_{m,n}$, provide information about what is happening on a site at each revisit, in terms of time spent or page views. These metrics allowed us to identify the sites for which the attention of the user is "shifting" towards or "away" from the site. In the former, the user becomes more focused in the task being performed by visiting that site, whereas in the latter, the user is slowly doing something else, on some other sites.

All metrics were compared, in terms of how they correlate, and we could clearly see that our proposed metrics indeed provide different insights into how users engage with a site at the visit-level and session-level, acknowledging that users are going to other sites during a session.

Future work. Our next steps are two-fold. First, our definition of a task is simplistic and should be extended. For example, to book a holiday, a user may visit several sites within the same session and one site many times across several sessions. It will be important to re-align our findings and conclusions with a more general definition of task. Moreover, we need to further study when a user is returning to a site to continue with the same task (continue reading the news articles) or to start a totally different task (a new unrelated information need), as these two cases should be treated differently. Accounting for this will likely lead to more advanced models of user attention to a site, an important research direction to follow. Second, we focused on two metrics used to assess the browsing activity within a visit. We did not consider, for example, bounce or click-through rates. It will be important to extend our study to a larger set of metrics, including those related to other aspect of user engagement, such as loyalty and popularity [11]. In addition, the concept of page view is not well defined with dynamic changes to a web page such as in Ajax, so we would like to address this issue too. Finally, we did not take user demographics into account and how multitasking on sites differ in different countries.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] G. Bonnin, A. Brun, and A. Boyer. Taking into account tabbed browsing in predictive web usage mining. In *SOTICS*, 2011.

[2] F. Chierichetti, R. Kumar, and A. Tomkins. Stochastic models for tabbed browsing. In *WWW*, 2010.

[3] I. Dhillon, Y. Guan, and B. Kulis. A unified view of kernel k-means, spectral clustering and graph cuts. Technical report, University of Texas, Dept. of Computer Science, 2004.

[4] P. Dubroy and R. Balakrishnan. A study of tabbed browsing among Mozilla Firefox users. In *CHI*, 2010.

[5] G. Dupret and M. Lalmas. Absence time and user engagement: evaluating ranking functions. In *WSDM*, 2013.

[6] E. Herder. Characterizations of user web revisit behavior. In *LWA*, 2005.

[7] J. Huang and R. W. White. Parallel browsing behavior on the Web. In *HT*, 2010.

[8] B. J. Jansen, A. Spink, and V. Kathuria. How to define searching sessions on web search engines. In *WebKDD*, 2007.

[9] M. Kellar, C. Watters, and M. Shepherd. A goal-based classification of web information tasks. *ASIST*, 2006.

[10] R. Kumar and A. Tomkins. A characterization of online browsing behavior. In *WWW*, 2010.

[11] J. Lehmann, M. Lalmas, E. Yom-Tov, and G. Dupret. Models of user engagement. In *UMAP*, 2010.

[12] S. E. Lindley, S. Meek, A. Sellen, and R. H. R. Harper. "It's simply integral to what I do": enquiries into how the Web is weaved into everyday life. In *WWW*, 2012.

[13] C. Lucchese, S. Orlando, R. Perego, F. Silvestri, and G. Tolomei. Identifying task-based sessions in search engine query logs. In *WSDM*, 2011.

[14] B. McKenzie and A. Cockburn. An emprical analysis of web page revisitation. In *HICSS*, 2001.

[15] D. Mehrzadi and D. G. Feitelson. On extracting session data from activity logs. In *SYSTOR*, 2012.

[16] M. Meiss, J. Duncan, B. Gonçalves, J. J. Ramasco, and F. Menczer. What's in a session: Tracking individual behavior on the Web. *CoRR*, 2010.

[17] H. Obendorf, H. Weinreich, E. Herder, and M. Mayer. Web page revisitation revisited: Implications of a long-term click-stream study of browser usage. In *CHI*, 2007.

[18] H. L. O'Brien and E. G. Toms. What is user engagement? A conceptual framework for defining user engagement with technology. *JASIS*, 2008.

[19] J. Rubinstein, D. Meyer, and J. Evans. Executive control of cognitive processes in task switching. *Exp. Psychol. Hum. Percept. Perform.*, 2001.

[20] A. Spink, M. Park, B. J. Jansen, and J. Pedersen. Multitasking during web search sessions. *Inf. Process. Manage.*, 2006.

[21] L. Tauscher and S. Greenberg. How people revisit web pages: Empirical findings and implications for the design of history systems. *Hum.-Comput. Stud.*, 1997.

[22] S. K. Tyler and J. Teevan. Large scale query log analysis of re-finding. In *WSDM*, 2010.

[23] M. Viermetz, C. Stolz, V. Gedov, and M. Skubacz. Relevance and impact of tabbed browsing behavior on web usage mining. In *Web Intelligence*, 2006.

[24] Q. Wang and H. Chang. Multitasking bar: prototype and evaluation of introducing the task concept into a browser. In *CHI*, 2010.

[25] H. Zhang and S. Zhao. Measuring web page revisitation in tabbed browsing. In *CHI*, 2011.