The Effect of Links on Networked User Engagement

Elad Yom-Tov
Yahoo! Research New York
eladtv@yahoo-inc.com

Mounia Lalmas
Yahoo! Research Barcelona
mounia@acm.org

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

Ricardo Baeza-Yates
Yahoo! Research Barcelona
rbaeza@acm.org

Pinar Donmez
Yahoo! Labs Sunnyvale
pinard@yahoo-inc.com

Janette Lehmann
Yahoo! Research Barcelona
janettel@yahoo-inc.com

ABSTRACT

In the online world, user engagement refers to the phenomena associated with being captivated by a web application and wanting to use it longer and frequently. Nowadays, many providers operate multiple content sites, very different from each other. Due to their extremely varied content, these are usually studied and optimized separately. However, user engagement should be examined not only within individual sites, but also across sites, that is the entire content provider network. In previous work, we investigated *networked user engagement*, by defining a global measure of engagement that captures the effect that sites have on the engagement on other sites within the same browsing session. Here, we look at the effect of links on networked user engagement, as these are commonly used by online content providers to increase user engagement.

Categories and Subject Descriptors

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

Keywords

Downstream engagement, networked engagement, links

1. INTRODUCTION

In the online industry, user engagement refers to the quality of the user experience that emphasizes the positive aspects of the interaction with a web application, and in particular the phenomena associated with wanting to use that web application longer and frequently. This definition is motivated by the observation that successful web applications are not just used, but they are engaged with [2].

On the other hand, many content providers, such as CNN or Yahoo!, operates multiple content sites, which are very different from each other. For example, Yahoo! has a multitude of sites, including sites on such diverse topics as finance, sports, celebrities, and shopping. Due to the varied content served by these sites, it is difficult to treat them as a single entity. For this reason, they are usually studied and optimized separately. However, content providers should address engagement on the different sites they manage as a whole, as sites can (and do) link to each other. For example, if a site does not have any links on its pages to other sites of the same provider, users will find it difficult to navigate to them, creating an engagement barrier. Conversely, linking to relevant content of the same provider will improve engagement.

Thus, the success of a site largely depends on itself, but also on how it is connected with other sites and how web traffic arrives to the site. In [5], we investigated *networked user engagement*, by defining a global measure of engagement that captures the effect sites have on the engagement on other sites. In this paper, we look at the effect of *links*, on networked user engagement, as these are commonly used to influence user engagement by e.g. web-masters and editors.

2. MEASURES OF USER ENGAGEMENT

Online behavior metrics aiming at assessing users' depth of engagement with a site have been used for many years [3, 4]. Although these metrics measure web usage, they are commonly employed as proxy for online user engagement: the higher and the more frequent the usage, the more engaged the user. In [5], we proposed a metric for measuring networked user engagement: how long users spend on a provider set of sites (a provider network) from a given site.

In this work, a site is an entity made of web pages that were put together to form a service (news, mail, etc). We define a session as all the pages visited by a user within 30 minutes or less from the first interaction with the site. A *provider* session (a Yahoo! session) corresponds to all contiguous pages of the provider sites visited within a session. Thus, a session is composed of one or more provider sessions.

Downstream engagement is defined as follows. Formally, we index the sites S visited during a session by $i \in 1, 2, \ldots, n$. For a site S_i, we use a binary indicator $\mathbb{1}_j$ that is 1 if S_j belongs to the provider *and* if the user did not visit any site not belonging to that provider between her/his visit from sites S_i and S_j. Otherwise it is 0. Let $t(S_i)$ be the time spent (dwell time) on site S_i. Downstream engagement $E(S_i)$ for sessions that contain one provider is defined as:

$$E(S_i) = \frac{\sum_{j>i} t(S_j)\mathbb{1}_j}{\sum_{j>i} t(S_j)}$$

Intuitively, this measures the fraction of time users spent on a content provider's sites (network) – without leaving – out

Table 1: Spearman correlations between number of links to the same site, links to another provider site, and link to a site outside the provider network, and three measures of engagement.

	e-commerce	news	women-interests
Downstream engagement			
Same site	0.03	-0.31	-0.27
Other provider sites	-0.09	0.20	0.22
Non-provider sites	-0.10	0.04	-0.25
Dwell time			
Same site	0.51	0.78	0.82
Other provider sites	-0.61	0.38	-0.68
Non-provider sites	-0.51	0.52	0.80
Site revisits			
Same site	-0.20	0.40	-0.16
Other provider sites	0.25	-0.52	-0.54
Non-provider sites	0.24	-0.34	-0.55

of the entire time they had available to spend browsing the Internet (the total session).

We use two other measures, *dwell time*, and *number of revisits*, which provide complementary insight into the effect of links on downstream engagement. Dwell time is the total time a user spends visiting contiguous pages within a site; and the number of revisits is, for a given a site, the number of non-contiguous visits to that site within a session.

3. DATA

Our study required user behavior data for computing the three measures, and site data, for analyzing links. To collect site data, we extracted the front page of three popular Yahoo! sites once every hour. The three sites correspond to news, e-commerce and women's interest sites. The site data was used to generate, for each site and per hour, the links pointing to within the same site, to other sites on the Yahoo! network, and to sites outside the Yahoo! network.

User data was collected from an anonymized sample of users who gave their consent to provide browsing data. A total of 19.4M sessions were recorded from approximately 265,000 users. These users represent a sample of the Yahoo! user base, who spend a certain fraction of their browsing time on Yahoo! sites. The user data was used to measure average dwell time, average downstream engagement, and average number of revisits for each of the pages at each time and date. The user and front page datasets were joined by site, date and time.

4. RESULTS

Table 1 shows the correlations between the three types of links (their numbers) and the three engagement metrics, for our three sites.

With respect to the news site, if the aim is to keep users on the site, then having links to the site itself helps (i.e. more news stories to consume). It also seems that when users are on the news site, they are there to read news (if appropriately enticed). Nonetheless, they can be directed to other sites on the network. Providing users with external links does not affect downstream engagement, but it does for dwell time. Our hypothesis, which requires verification, is that too many choices cause users to remain where they are. Other reasons are plausible: it might be that adding links to external sites lead the user to believe that the news is well documented and trustworthy.

With respect to the e-commerce site, links have little effect on downstream engagement, whereas they have on dwell time. Dwell time can be increased by having more links into the site itself, and no links to other and external sites. Overall, it seems that, as expected, users of the e-commerce site have a defined intent. Attempts to direct them to additional sites, is pointless, for example, by showing extra links. In fact, it hurts the time spent on the site.

For the women-interests site studied here, having links to other sites in the provider network can help increasing downstream engagement, but is not recommended if the aim is to keep users on the site. We observe the opposite effect when looking at links to the site itself, which is logical. Having links outside the provider network can hurt downstream engagement, but does the opposite in terms of dwell time (likely for a similar reason as stated before for news). Overall, this site, and likely other similar ones, are good sites where a provider has some good leeway to influence user engagement within a site and across sites.

We discuss the results with respect to revisits. Interestingly, links on the same site for both the e-commerce and women-interests sites are negatively correlated with this measure. Looking at dwell time, it may simply mean that users on these sites spend time on them, navigating through the links within these sites, having then less opportunities to return within a session. This observation is different for news, where the correlation is positive. It is however not clear how to interpret this correlation looking both at dwell time and downstream engagement.

Links to other or external sites for the news and women-interests sites are negatively correlated with revisits; users who leave these sites do not return to them. However, the same type of links has the opposite effect on the e-commerce site: although they do not help dwell time and downstream engagement, they seem to help revisits. It could simply be that users will anyway return to the site regularly, because they have specific intent when visiting the site initially, even if they divert to other sites.

To conclude, our work demonstrates the importance of optimizing web sites not only to engagement of users on each site (e.g. [1]), but also to directing them towards other sites of the same content provider. Future work will analyze the effect of content as well as the structure of web pages to improve intra-site engagement as well as better determining the boundaries of user sessions.

5. REFERENCES

[1] C. Liu, R. W. White, and S. Dumais. Understanding web browsing behaviors through weibull analysis of dwell time. *SIGIR*, 2010.

[2] H. L. O'Brien and E. G. Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *JASIST*, 59(6), 2008.

[3] E. T. Peterson and J. Carrabis. Measuring the immeasurable: Visitor engagement. Technical report, WebAnalyticsDemystified, 2008.

[4] K. Rodden, H. Hutchinson, and X. Fu & Measuring the User Experience on a Large Scale: User-Centered Metrics for Web Applications. *CHI*, 2010.

[5] E. Yom-Tov, M. Lalmas, G. Dupret, R. Baeza-Yates, P. Donmez and J. Lehmann. Networked User Engagement, Submitted for Publication, 2012.