Models of User Engagement

Janette Lehmann[1], Mounia Lalmas[2], Elad Yom-Tov[3] and Georges Dupret[4]

[1] Universitat Pompeu Fabra Barcelona, Spain
janette.lehmann@gmx.de
[2] Yahoo! Research Barcelona, Spain
mounia@acm.org
[3] Yahoo! Research New York, USA
eladyt@yahoo-inc.com
[4] Yahoo! Labs Sunnyvale, USA
gdupret@yahoo-inc.com

Abstract. Our research goal is to provide a better understanding of how users engage with online services, and how to measure this engagement. We should not speak of one main approach to measure user engagement – e.g. through one fixed set of metrics – because engagement depends on the online services at hand. Instead, we should be talking of models of user engagement. As a first step, we analysed a number of online services, and show that it is possible to derive effectively simple models of user engagement, for example, accounting for user types and temporal aspects. This paper provides initial insights into engagement patterns, allowing for a better understanding of the important characteristics of how users repeatedly interact with a service or group of services.

Keywords: diversity of user engagement, models, user type, temporal aspect

1 Introduction

User engagement is the quality of the user experience that emphasises the positive aspects of the interaction, and in particular the phenomena associated with being captivated by a web application, and so being motivated to use it. Successful web applications are not just used, they are engaged with; users invest time, attention, and emotion into them. In a world full of choice where the fleeting attention of the user becomes a prime resource, it is essential that technology providers design engaging experiences. So-called *engagement metrics* are commonly used to measure web user engagement. These include, for example, number of unique users, click-through rates, page views, and time spent on a web site. Although these metrics actually 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. Major web sites and online services are compared using these and other similar engagement metrics.

User engagement possesses different characteristics depending on the web application; e.g. how users engage with a mail tool or a news portal is very different. However, the same engagement metrics are typically used for all types of web application, ignoring the diversity of experiences. In addition, discussion on the "right" engagement metrics is still going on, without any consensus on which metrics to be used to measure which types of engagement. The aim of this paper is to demonstrate the diversity of user engagement, through the identification and the study of *models of user engagement*. To this end, we analysed a large number of online sites, of various types (ranging from news to e-commerce to social media). We first show the diversity of engagement for these sites. To identify models of engagement, we cluster all sites using various criteria (dimensions) of engagement (e.g. user types, temporal aspects). Our results are two-fold. First, we can effectively derive models of user engagement, for which we can associate characteristics of the type of engagement. Second, by using various criteria, we gain different but complementary insights into the types of engagement.

The paper is organised as follows. Section 2 provides related work. Section 3 describes the data and engagement metrics used. Section 4 demonstrates the diversity of user engagement. Section 5 presents the methodology adopted to identify models of user engagement, and the outcomes. Section 6 looks at relationships between models, providing further insights into types of engagement. We finish with our conclusions and thoughts for future work.

2 Related Work

Approaches to measure user engagement can be divided into three main groups: self-reported engagement, cognitive engagement, and online behaviour metrics. In the former group, questionnaires and interviews (e.g. [7, 4]) are used to elicit user engagement attributes or to create user reports and to evaluate engagement. They can be carried out within a lab setting, or via on-line mechanisms (including crowd-sourcing). However, these methods have known drawbacks, e.g. reliance on user subjectivity. The second approach uses task-based methods (e.g. dual-task [8], follow-on task), and physiological measures to evaluate the cognitive engagement (e.g. facial expressions, vocal tone, heart rate) using tools such as eye tracking, heart rate monitoring, and mouse tracking [3].

Measures in the second group, although objective, are suitable for measuring only a small number of interaction episodes at close quarters. In contrast, the web-analytics community has been studying user engagement through online behaviour metrics that assess users' depth of engagement with a site. For instance, [5] describes *engagement metrics* that indicate whether or not users consume content slowly and methodically, return to a site, or subscribe to feeds. Widely used metrics include click-through rates, number of page views, time spend on a site, how often users return to a site, number of users, and so on. Only online behaviour metrics are able to collect data from millions of users. Although these metrics cannot explicitly explain why users engage with a service, they act as proxy for online user engagement: the higher and the more frequent the usage, the more engaged the user. Indeed, two millions of users accessing a service daily is a strong indication of a high engagement with that service. Furthermore, by varying specific aspects of the service, e.g. navigation structure, content, func-

Table 1. Engagement metrics used in this paper.

Metrics	Description
Popularity	(for a given time frame)
#Users	Number of distinct users.
#Visits	Number of visits.
#Clicks	Number of clicks (page views).
Activity	
ClickDepth	Average number of page views per visit.
DwellTimeA	Average time per visit (dwell time).
Loyalty	(for a given time frame)
ActiveDays	Number of days a user visited the site.
ReturnRate	Number of times a user visited the site.
DwellTimeL	Average time a user spend on the site.

tionality, and measuring the effect on engagement metrics can provide implicit understanding on why users engage with the service. Finally, although this group of measures is really accounting for "site engagement", we retain the terminology "user engagement" as it is commonly used by the online industries. We look at models of user engagement based on this third group of metrics.

3 Metrics and Interaction Data

Engagement metrics The metrics used in this paper are listed in Table 1. As our aim is to identify models of user engagement, we restrict ourselves to a small set of widely reported metrics. We consider three types of engagement metrics, reflecting, *popularity*, *activity*, and *loyalty*. Popularity metrics measure how much a site is used, e.g. total number of users. The higher the number, the more popular the corresponding site. How a site is used is measured with activity metrics, e.g. average number of clicks per visit across all users. Loyalty metrics are concerned with how often users return to a site. An example is the return rate, i.e. average number of times users visited a site[5]. Loyalty and popularity metrics depend on the considered time interval, e.g. number of weeks considered. A highly engaging site is one with a high number of visits (popular), where users spend lots of time (active), and return frequently (loyal). It is however the case, as demonstrated next, that not all sites, whether popular or not, have both active and loyal users, or vice versa. It does not mean that user engagement on such sites is lower; it is simply different. Our conjuncture is that user engagement depends on the site itself.

[5] A user can return several times on a site during the same day, hence this metric is different to the number of active days.

Interaction data This study required a large number of sites, and a record of user interactions within them. We collected data during July 2011 from a sample of approximately 2M users who gave their consent to provide browsing data through the Yahoo! toolbar. These data are represented as tuples (timestamp, bcookie, url). We restrict ourselves to sites with at least 100 distinct users per month, and within the US. The latter is because studying the engagement of sites across the world requires to account for geographical and cultural differences, which is beyond the scope of the paper. This resulted in 80 sites, encompassing a diverse set of sites and services such as news, weather, movies, mail, etc.

4 Diversity in Engagement

Sites Figure 1 reports the normalized engagement values for the eight metrics and the 80 sites under study. All original values v_i of metric v are translated into an ordinal scale and then normalized (μ_v is the mean of the ordinal v_i values, and σ_v is the corresponding standard deviation value): $v'_i = (v_i - \mu_v)/\sigma_v$. The average value (ordinal) of an engagement metric becomes then zero. The y-axes in Figure 1 order the sites in terms of number of users (*#Users*). Finally, *MergeUE* is the linear combination of *#Users*, *DwellTimeA*, and *ActiveDays*.

We can see that sites differ widely in terms of their engagement. Some sites are very popular (e.g. news sites) whereas others are visited by small groups of users (e.g. specific interest sites). Visit activity also depends on the sites, e.g. search sites tend to have a much shorter dwell time than sites related to entertainment (e.g. games). Loyalty per site differs as well. Media (news, magazines) and communication (e.g. messenger, mail) have many users returning to them much more regularly, than sites containing information of temporary interests (e.g. buying a car). Loyalty is also influenced by the frequency in which new content is published (e.g. some sites produce new content once per week). Finally, using one metric combining the three types metrics (*MergeUE*) also shows that engagement varies across sites.

Metrics To show that engagement metrics capture different aspects of a site engagement, we calculate the pair-wise metrics correlations using Kendall tau (τ) rank correlation on the ordinal values. The resulting average intra-group correlation is $\tau = 0.61$, i.e. metrics of the same groups mostly correlate; whereas the average inter-group correlation is $\tau = 0.23$, i.e. metrics from different groups correlate weakly or not at all. This shows that the intuition we followed when we grouped the metrics is confirmed in practice.

The three popularity engagement metrics show similar engagement type for all sites, i.e. high number of users implies high number of visits ($\tau = 0.82$), and vice versa. For the loyalty metrics, high dwell time per user comes from users having more active days ($\tau = 0.66$), and returning regularly on the site ($\tau = 0.62$). The correlation between the two activity metrics is lower ($\tau = 0.33$). There are no correlation between activity and, popularity or loyalty metrics. High popularity does not entail high activity ($\tau = 0.09$). Many site have many

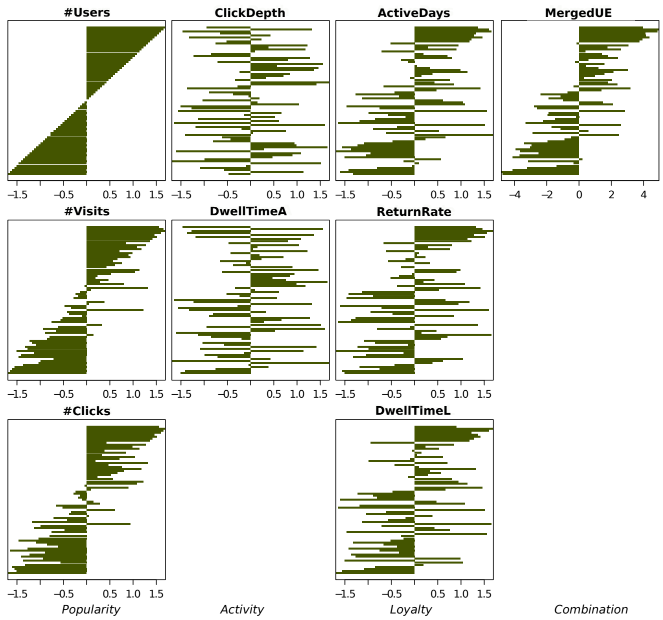


Fig. 1. Normalized engagement values per site (y-axes order sites by #Users).

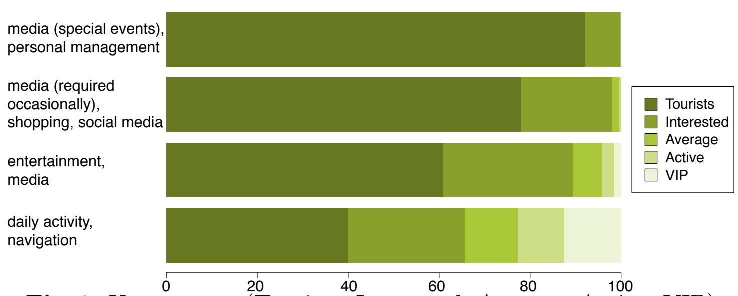


Fig. 2. User groups (Tourists, Interested, Average, Active, VIP).

users spending little time on them; e.g. a search site is one where users come, submit a query, get the result, and if satisfied, leave the site. This results in a low dwell time even though user expectations were entirely met. The same argument hold for a site on Q&A, or a weather site. What matters for such sites is their popularity. Finally, we observe a moderate correlation ($\tau = 0.49$) between loyalty and popularity metrics. This is because popular sites are those to which users return regularly. The same reasoning applies for the other metrics of these two groups.

Users Studies have shown that users may arrive in a site by accident or through exploration, and simply never return. Other users may visit a site once a month, for example a credit card site to check their balance. On the other hand, sites such as mail may be accessed by many users on a daily basis. We thus looked at how active users are within a month, per site. The number of days a user visited a site over a month is used for this purpose. We create five types of user groups[6]:

Group	Number of days with a visit
Tourists :	1 day
Interested :	2-4 days
Average :	5-8 days
Active :	9-15 days
VIP :	$\geq$ 16 days

The proportion of the user groups for each site is calculated, then sites with similar proportion of user groups are clustered using k-means. Four cluster were detected and the cluster centers calculated. Figure 2 displays the four cluster centers, i.e. the proportion of user groups per cluster. The types of sites in each cluster are shown, as illustration. We observe that the proportion of tourist users is high for all sites. The top cluster has the highest proportion of tourist users; typical sites include special events (e.g. the oscars) or those related to configuration. The second from the top cluster includes sites related to specific information that are occasionally needed; as such they are not visited regularly within a month. The third cluster includes sites related to e-commerce, media, which are used on a regular basis, albeit not daily. Finally, the bottom cluster contains navigation sites (e.g. landing page) and communication sites (e.g. messenger). For these sites, the proportion of VIP users is higher than the proportion of active and average users. The above indicates that the type of users, e.g. tourist vs. VIP, matters when measuring engagement.

Time Here, we show that depending on the selected time span different types of engagement can be observed. We use #Users to show this. Using the interaction data spanning from February to July 2011, we normalized the number of users per site (#Users) with the total number of users that visited any of the sites on that day. The time series for each site was decomposed into three temporal components: periodic, trend and peak, using local polynomial regression fitting [1]. To detect periodic behaviour we calculated the correlation between the extracted periodic component and the residual between the original time series and the trend component. To detect peaks, the periodic component was removed from the time series and peaks were detected using a running median.

Figure 3 shows graphically the outcomes for four sites (under examples). Possible reasons for a periodic or peak behaviours are given (under influence). Finally, sites for which neither periodic behaviour nor peak were found are given (under counter-example). The engagement pattern can be influenced by external and internal factors. Communication, navigation and social media sites tend to be more "periodically used" than media sites. Access to media sites tends to be influenced by external factors (important news) or the frequency of publishing

[6] The terminology and the range of days is based on our experience in how user engagement is studied in the online industry. For instance, a VIP user is one that comes on average 4 days per week, so we chose the value 16 days within a month.

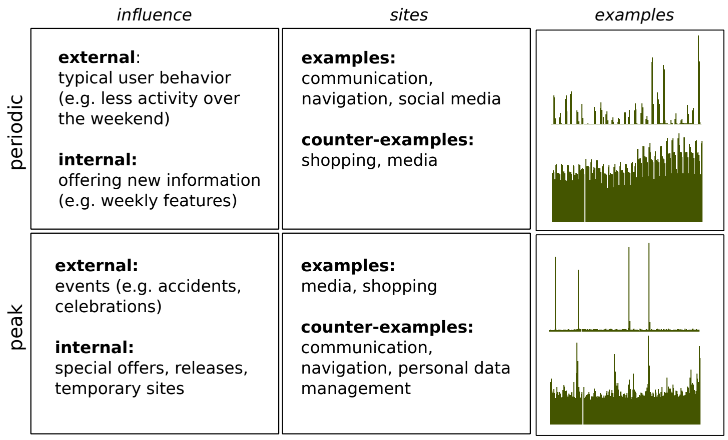


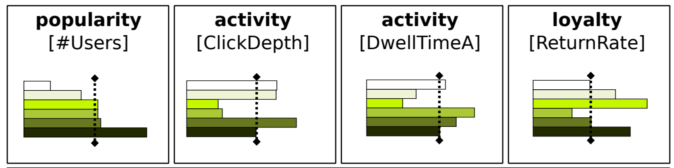
Fig. 3. Engagement over time using #Users (February – July 2011).

new information. Interesting is the fact that sites with a periodic behaviour tend to have no peaks and sites with peaks tend not to be periodic. Thus accounting for time is likely to bring important insights when measuring site engagement.

5 Models of User Engagement

The previous section showed differences in site engagement. We study now these differences to identify patterns (models) of user engagement. The base for all studies is a matrix containing data from the 80 sites under study. Each site is represented by eight engagement metrics. A metric can be further split into several dimensions based on user and time combinations. The values of each metric are transformed into an ordinal scale to overcome scaling issues. We clustered the sites using the kernel k-means algorithm [2], with a Kendall *tau* rank correlation kernel [6]. The number of clusters are chosen based on the eigenvalue distribution of the kernel matrix. After clustering, each cluster centroid is computed using the average rank of cluster members (for each metric). To describe the centroids (the models), we refer to the subset of metrics selected based on the correlations between them and the Kruskal-Wallis test with Bonferonni correction, which identifies values of metrics that are statistically significantly different for at least one cluster (compared to the other clusters).

Three sets of models are presented, based on the eight engagement metrics (general), accounting for user groups (user-based), and capturing temporal aspects (time-based). Although all dimensions could be used together to derive one set of models (e.g. using dimensionality reduction to elicit the important characteristics of each model), generating the three sets separately provides clear and focused insights into engagement patterns. When presenting each model, we give illustrative examples of the types of sites belonging to them. It is not our aim to explain why each site belongs to which model, and the associated implications.



	popularity	activity [ClickDepth]	activity [DwellTime]	loyalty
model m_{g6}	--			
model m_{g5}				
model m_{g4}		--	--	++
model m_{g3}		--	++	--
model m_{g2}		++		
model m_{g1}	++			

average ++ high -- low

Fig. 4. General models of engagement – Top panels display the cluster (model) centers. Bottom panels provide the corresponding model descriptions.

5.1 General Models

We look at models of user engagement, without accounting for user type or temporal aspect. We refer to them as "general models". Our eight metrics generate six "general" models of user engagement, visualized in Figure 4. As the three popularity metrics exhibit the same effect, only *#Users* is reported. The same applies for the loyalty metrics, i.e. only *ActiveDays* is reported. The two activity metrics yield different behaviours, hence are both shown.

In **model $\mathbf{m_{g1}}$**, high popularity is the main factor; by contrast, low popularity characterizes **model $\mathbf{m_{g6}}$**. Media sites providing daily news and search sites follow **model $\mathbf{m_{g1}}$**; whereas **model $\mathbf{m_{g6}}$** captures interest-specific sites. The main factor for **model $\mathbf{m_{g2}}$** is a high number of clicks per visit. This model contains e-commerce and configuration (e.g. profile updating) sites, where the main activity is to click. By contrast, **model $\mathbf{m_{g3}}$** describes the engagement of users spending time on the site, but with few click and with low loyalty. The model is followed by domain-specific media sites of periodic nature, which are therefore not often accessed. However when accessed, users spend more time to consume their content[7]. Next, **model $\mathbf{m_{g4}}$** is characterized by highly loyal users, who spend little time and perform few actions. Navigational sites (e.g. front pages) belong to **model $\mathbf{m_{g4}}$**; their role is to direct users to interesting content in other sites, and what matters is that users come regularly to them. Finally, **model $\mathbf{m_{g5}}$** captures sites with no specific engagement patterns.

[7] Looking further into this, it seems that the design of such sites (compared to mainstream media sites) leads to such type of engagement, since new content is typically published on their front page. Thus users are not enticed to reach (if any) additional content in these sites. This is the sort of reasoning that becomes possible by looking at models of user engagement, as investigated in this paper.

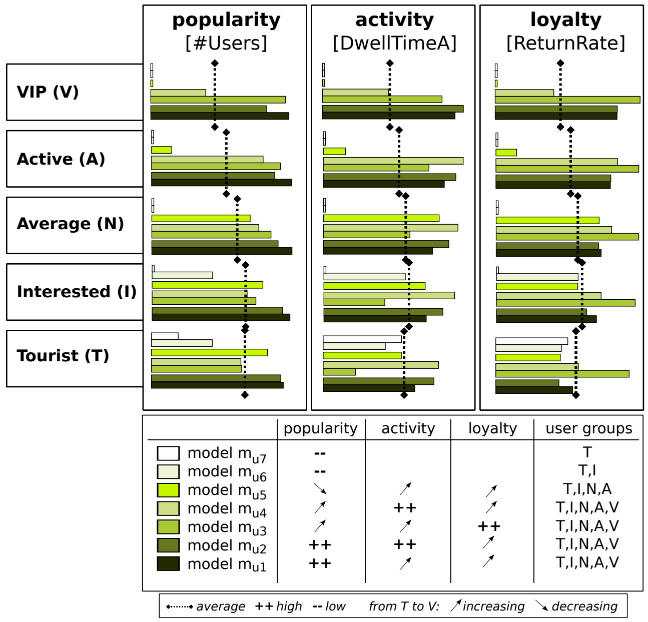


Fig. 5. User-based models of engagement – Top panels display the cluster (model) centers. Bottom panels provide the corresponding model descriptions.

5.2 User-based Models

We investigate now models of user engagement that account for the five user groups elicited in Section 4. The eight metrics were split, each into five dimensions, one for each user group, i.e. VIP to Tourists. This gives 40 engagement values per site. A site without a particular user group get 0 values for all metrics for that group. We obtain seven "user-based" models (clusters), visualized in Figure 5. We only report the results for one metric of each group (#Users, *DwellTimeA* and *ReturnRate*), as these are sufficient for our discussion.

The first two models, **model $\mathbf{m_{u1}}$** and **model $\mathbf{m_{u2}}$** are characterised by high popularity across all user groups. Activity is high across all user groups for **model $\mathbf{m_{u2}}$**, whereas it increases from Tourist to VIP users for **model $\mathbf{m_{u1}}$**. Finally, both models are characterised by an increase in loyalty from Tourist to VIP users. Popular media sites belong to these models. The next two models, **model $\mathbf{m_{u3}}$** and **model $\mathbf{m_{u4}}$**, exhibit the same increase in popularity from Tourist to VIP users. High loyalty across all groups and an increase in activity from Tourist to VIP users further characterise **model $\mathbf{m_{u3}}$**. Sites falling in this model include navigation pages (e.g. front pages). High activity across all user groups apart for VIP and an increasing loyalty from Tourists to Active users is an important feature of **model $\mathbf{m_{u4}}$**, which typically include game and sport sites. Interestingly, **model $\mathbf{m_{u4}}$** is characterised by a low number of VIP users, compared to the three previous models.

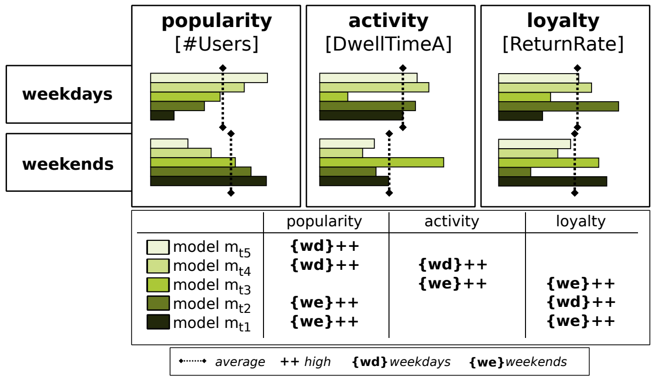


Fig. 6. Time-based models of engagement – Top panels display the cluster (model) centers. Bottom panels provide the corresponding model descriptions.

Third, **model $\mathbf{m_{u5}}$** model caters for the engagement of Tourist, Interested and Average users. Loyalty increases going from Tourist to Average users, which makes sense as loyalty is used to determine the user groups. More interestingly is that activity augments the same way, whereas popularity decreases. Shopping and social media sites belong to this model. Finally, **model $\mathbf{m_{u6}}$** and **model $\mathbf{m_{u7}}$** are concerned with the low engagement (popularity) of Interested and Tourist users, and only Tourist users, respectively. They correspond to sites on very particular interests or of a temporary nature; as such popularity for these two groups of users is low compared to other models. Moreover, **model $\mathbf{m_{u7}}$** indicates that when on site, the activity of Tourist users is not negligible. By contrast, **model $\mathbf{m_{u6}}$** highlights a higher activity of Interested users than Tourist users.

5.3 Time-based Models

We look now at models of user engagement that account for the temporal aspect. For simplicity, we consider two time dimensions, *weekdays* and *weekends*. Each site becomes associated with fourteen metrics; seven of our engagement metrics are split into these two time dimensions (*ActiveDays* is not used, as it has a different time span). To elicit the differences in engagement on weekdays vs. weekends, we transformed the absolute engagement values into proportional ones, e.g. the proportional *ReturnRate* is $\text{ReturnRate}_{weekdays}$ / ($\text{ReturnRate}_{weekdays}$ + $\text{ReturnRate}_{weekend}$). The same methodology as that used for the other types of models was then applied. This led to the identification of five "time-based" models of engagement (clusters), shown in Figure 6.

We can see that **model $\mathbf{m_{t1}}$** and **model $\mathbf{m_{t2}}$** describe sites with high popularity on weekends; loyalty is also high on weekends for **model $\mathbf{m_{t1}}$**, whereas it is high on weekdays for **model $\mathbf{m_{t2}}$**. Both models characterize sites related to entertainment, weather, shopping and social media. The loyalty in **model $\mathbf{m_{t2}}$** is more significant on weekdays, because it contains sites for daily use, whereas **model $\mathbf{m_{t1}}$** contains sites relating to hobbies and special interests. Second, **model $\mathbf{m_{t3}}$** characterizes sites that are highly active, and to which users return frequently on weekends. Sites following this model include event related media sites (e.g. sport), search and personal data management (e.g. calendar, address book). Finally, **model $\mathbf{m_{t4}}$** and **model $\mathbf{m_{t5}}$** are similar as they both are characterised with high popularity during weekdays, and **model $\mathbf{m_{t4}}$** is further characterised by high activity during weekdays. The models are followed by sites related to daily and particular news and software; **model $\mathbf{m_{t4}}$** exhibits higher activity because it contains sites used for work issues.

6 Relationship between Models

We checked whether the three groups of models describe different engagement aspects of the *same set* of sites or that they are largely unrelated. We calculate the similarity between the three groups using the *Variance of Information*. The outcome is shown in Table 2 (5.61 is the maximal difference). We observe the highest (albeit low) similarity between the general and user-based models. The user- and time-based models differ mostly. Overall, all groups of models are independent i.e. they characterize different if not orthogonal aspects of user engagement, even though the matrices used to generate them are related.

We cannot show here all the relationships between each model of each group. Instead, we discuss two cases. For **model $\mathbf{m_{g1}}$**, a general model characterizing popular sites, 38% of its sites belong to **model $\mathbf{m_{u1}}$** (high popularity and increasing activity and loyalty from tourists to VIP users), and 31% follow **model $\mathbf{m_{u5}}$** (no VIP users, decreasing popularity and increasing activity and loyalty from tourists to active users). We now look at the user-based **model $\mathbf{m_{u2}}$** characterizing sites with high popularity and activity in all user groups and an increasing loyalty from Tourists to VIP users. Sites following this model are split into two time-based models, **model $\mathbf{m_{t2}}$** (50%) (high popularity on weekends and high loyalty on weekdays), and **model $\mathbf{m_{t3}}$** (50%) (high activity and loyalty on weekends). This comparison provides different angles into user engagement, allowing to zoom into particular areas of interests, e.g. further differentiating the "high loyalty" associated with **model $\mathbf{m_{u2}}$** into weekdays vs. weekends.

Table 2. Intersections of the models – cluster similarities.

	General	User	Time	(Range [0,5.61])
General	0.00	3.50	4.23	
User	3.50	0.00	4.25	
Time	4.23	4.25	0.00	

7 Conclusions and Future Work

Our aim was to identify models of user engagement. We analysed a large sample of user interaction data on 80 online sites. We characterised user engagement in terms of three families of commonly adopted metrics that reflect different aspects of engagement: popularity, activity and loyalty. We further divided users according to how often they visit a site. Finally, we investigated temporal behavioural differences in engagement. Then using simple approaches (e.g. k-means clustering), we generated three groups of models of user engagement: general, user-based and time-based. This provided us different but complementary insights on user engagement and its diversity. *This research constitutes a first step towards a methodology for deriving a taxonomy of models of user engagement.*

This paper did not study why a site follows one engagement model. However, while analysing our results, we observed that sites of the same type (e.g. mainstream media) do not necessarily belong to the same model(s) of engagement. It would be interesting to understand the reasons for this, e.g. is it the type of content, the structure of the site, etc? Furthermore, other aspects of user engagement should be considered. Accounting for user demographics (e.g. gender, age) and finer-grained temporal aspects (e.g. time of the day) are likely to bring additional and further insights into modelling engagement. Incorporating geographical location will bring perspectives related to culture and language. Finally, we must revisit engagement metrics. Indeed, the description of models often referred to only some of the metrics employed. A major next step will be to map the most appropriate metrics to each model of engagement.

Acknowledgements Janette Lehmann acknowledges support from the Spanish Ministry of Science through the project TIN2009- 14560-C03-01.

References

1. R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I. Terpenning. A seasonal-trend decomposition procedure based on loss. *Journal of Official Statistics*, 6:3–73, 1990.
2. I. Dhillon, Y. Guan, and B. Kulis. A unified view of kernel k-means, spectral clustering and graph cuts. Technical report, 2004.
3. J. Huang, R.W. White, and S.T. Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *CHI*, 2011.
4. H. OBrian, E. Toms, K. Kelloway, and E. Kelley. The development and evaluation of a survey to measure user engagement. *JASIST* 61(1):50-69, 2010.
5. E.T. Peterson and J. Carrabis. Measuring the immeasurable: Visitor engagement. Technical report, Web Analytics Demystified, 2008.
6. S. Sabato, E. Yom-Tov, A. Tsherniak, and S. Rosset. Analyzing system logs: a new view of what's important. In *USENIX workshop on Tackling computer systems problems with machine learning techniques*, 2007.
7. J. Sauro and J. S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *CHI*, 2009.
8. P. Schmutz, S. Heinz, Y. Métrailler, and K. Opwis. Cognitive load in ecommerce applications: measurement and effects on user satisfaction. *Advances in Human-Computer Interaction*, 2009:3:1–3:9, 2009.