Image Ranking Based on User Browsing Behavior

Michele Trevisiol[*†] Luca Chiarandini[*†] Luca Maria Aiello[†] Alejandro Jaimes[†]

{trevi, chiarluc, alucca, ajaimes}@yahoo-inc.com

[*]Web Research Group
Universitat Pompeu Fabra
Barcelona, Spain

[†]Yahoo! Research
Barcelona, Spain

ABSTRACT

Ranking of images is difficult because many factors determine their importance (e.g., popularity, quality, entertainment value, context, etc.). In social media platforms, ranking also depends on social interactions and on the visibility of the images both inside and outside those platforms. In this context, the application of standard ranking methods is not clearly understood, and neither are the subtleties associated with taking into account social interaction, internal, and external factors. In this paper, we use a large Flickr dataset and investigate these factors by performing an in-depth analysis of several ranking algorithms using both internal (i.e., within Flickr) and external (i.e., links from outside of Flickr) factors. We analyze rankings given by common metrics used in image retrieval (e.g., number of favorites), and compare them with metrics based on page views (e.g., time spent, number of views). In addition, we represent users' navigation by a graph and combine session models with some of these metrics, comparing with PageRank and BrowseRank. Our experiments show significant differences between the rankings, providing insights on the impact of social interactions, internal, and external factors in image ranking.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering; H.3.5 [**Online Information Services**]: Web-based services

General Terms

Agorithms, Experimentation

Keywords

Image Ranking, Social Browsing, Flickr, BrowseRank

1. INTRODUCTION

Many social media platforms function as somewhat independent ecosystems in which users carry out a number of social activities. In Flickr, in particular, users can share content and participate in multiple communities by submitting their photos to groups, by joining groups, and by performing several types of actions over Flickr content (*e.g.*, comment, add notes, favorite, *etc.*). The result of this is that the way the content is consumed is strongly influenced by all of the different social navigation paths that lead to it: a photo on Flickr, for example, can be linked to from a user's favorite photo collection, from several groups, galleries, and via other mechanisms, including the "external" web (*i.e.*, URLs outside of the Flickr domain).

As more social media platforms emerge, one of the key questions is whether traditional ranking algorithms, that may not take into account the subtleties of navigation patterns driven by social connections, can be successful within those ecosystems. In particular, the problem we are interested in addressing is the general ranking of images in Flickr (*i.e.*, as in PageRank, we would like to rank all of the images or a subset of them, in order of importance). Such ranking can have many applications, including retrieval, and information discovery, among others.

The importance of images in Flickr, or of "nodes" in similar social media platforms might depend on a number of *internal* and *external* factors. For example, an image that is very popular in a group that has a cult following, may have been marked by many users as a favorite image. The image has a large number of favorites because people within the Flickr community, and in particular, in the specific cult, view and favorite the image. In contrast, an image of an important real world event (*e.g.*, the British Royal Wedding) may get a high number of views, not by belonging to groups, but instead because it is linked to by multiple external (*i.e.*, outside of the Flickr domain) media outlets, and get comparatively few favorite marks. One of the key questions is thus what the impact of those external and internal factors is on ranking and selection of content.

In this paper, motivated by the scenario described above, we investigate the factors that affect image ranking by performing an in-depth analysis of the results of several ranking algorithms taking into account both internal (*i.e.*, within Flickr) and external (*i.e.*, links from outside the Flickr domain) factors. In particular, we analyze rankings given by common metrics used in image retrieval (*e.g.*, number of favorites), and compare them with metrics based on page views (*e.g.*, time spent, number of views). More specifically, in order to take into account the structure of Flickr in terms of navigation paths to and from specific images, we represent the user's navigation by a graph and combine session models with some of these metrics. We implement PageRank, BrowseRank, and compare them with different rankings.

Our main contributions can be summarized as follows:

- We compare five different implicit and explicit image ranking methods to a number of features. Our analysis gives insights into what aspects each ranking method emphasizes.
- We introduce a variation of the BrowseRank algorithm in which navigation patterns are used to assign a different damping factor to each node in the graph.
- We analyze the connectivity patterns of a large BrowseGraph extracted from Flickr. Results point to structural peculiarities that differentiate browsing graphs from other complex graphs like social and similarity networks.

To our knowledge, this is the most detailed comparison between image ranking algorithms in terms of number of baselines and features considered, and it is the first attempt to use BrowseRank for an image ranking task.

The rest of the paper is structured as follows. In Section 2 we discuss related work. In Section 3 we present the features of our Flickr dataset and we describe how we extract user session information. In Section 4 we describe our modified BrowseGraph. In Section 5 we compare our method with other baselines, highlighting their qualities and shortcomings.

2. RELATED WORK

Our work focuses on query-independent ranking of images. When dealing with the Web graph, or any other corpus of interlinked resources, the most popular ranking algorithms are PageRank [25], HITS (Hypertext Induced Topic Selection) [17] and SALSA (Stochastic Approach for Link-Structure Analysis) [18]. In addition, there are many extensions to PageRank, such as BrowseRank by Liu *et al.* [22, 23], in which pages are weighted not only by the number of incoming and outgoing links, but also by the time that users spend on each page.

Besides the ranking algorithm used, the quality of the ranking is heavily influenced by the graph that is used to model the relations between documents. The difference between the standard hyperlinks graph of webpages and the graph of the browsing data was studied by Liu *et al.* [24, 22]. Authors compare the efficiency of PageRank in both graph types, showing that a link analysis algorithm performs better on a user browsing graph than on the whole hyperlink graph. Moreover, they show that a browsing graph generated from about 15 days of data is stable enough to be reliable. As a result of its good performance, several variations and improvements of BrowseRank have been proposed in the recent past [33, 9, 5].

It is important to report that PageRank-like algorithms applied to a complete network or to any of its subnetworks yield very different results. This problem is defined as the *PageRank local ranking problem* [2, 3]. In our case, this drawback can occur because the Flickr links network is a subset of the entire Web. Therefore, a ranking performed by PageRank-like algorithms applied to the entire Web can present very different results compared to the same algorithms applied to our subnetwork. However, we avoid this issue considering a first level of external links in addition to the Flickr network, *i.e.,* the set of websites from which the users enter in Flickr. Moreover, we use this sets of external links to compute the exact values of *stop* and *reset* probabilities in our version of PageRank and BrowseRank (see Section 4.1).

Web ranking algorithms have also traditionally been used to rank images by using metatada associated with the images and/or by including content-based analysis. A number of alternatives has been explored to improve the result of the Web ranking task, including visual diversification [28, 31], near duplicate detection [8], query expansion [12], visual position [7], faceted detection [29], and re-rank based on click data [13].

Some authors have applied PageRank to image retrieval. A solution presented by Jean *et al.* [14, 15] based on the content of the images, is to extract the interest points from each image, classify the common ones, and create visual links between them. In that case, the hyperlinks are given by visual similarity among the images. Liu *et al.* [21] use a random walk algorithm for tag ranking in Flickr, to overcome the sparsity problem of the tags associated to the images inside the social network.

Much work has been made for image ranking inside a social network as Flickr, and in many cases they focus around the images with the highest number of favorites, as that is the clearest explicit action that users can make to show their interest for a specific photo. Pedro et al. [26] used the number of favorites in Flickr as relevance values for building and testing machine learning models. There are also studies that aim to detect favorite photos in Flickr or to predict the photos that a user is likely to favorite based on social, visual, and textual information [30].

Other papers investigate other explicit and implicit features that lead to similar results of the favorites. For instance, Prieur *et al.* [27] find a very high correlation between the number of favorites and the number of comments and views. Nevertheless, the reasons to select a picture as a favorite can be many[1] and they do not depend always on the user's interest. We want to show that using explicit features as favorites in an environment as Flickr, often leads to specific types of results.

The quality evaluation of any ranking of multimedia objects is not a trivial task due to the many quality dimensions at play. The attractiveness [10] and aesthetics [11, 16] of images retrieved always play an important role in user's satisfaction. Lerman *et al.* [20] for example, propose an automatic method to assign photo attractiveness values to photos by using textual and visual features. But in a specific environment such as an image-sharing social network, the relevance of the results may depend strongly on the social interactions. As many previous studies suggest [19, 4], social browsing and contact relationships are very important to model the interestingness of a resource in a socially connected environment.

In summary, although the literature in ranking is vast, we are not aware of work that specifically examines the ranking mechanisms we analyze in this paper and study those in relation to the image ranking task.

[1] `http://www.flickr.com/photos/pagedooley/6246688704/`

3. DATASET

For the purpose of this study, we took a sample of the *pageviews* of more than 10 million anonymous users from approximately two months of Flickr user log data, from August to October 2011. The pageviews are represented as plain text files that contain a line for each HTTP request satisfied by the Web server. For each pageview, our dataset contains the following fields:

$$\langle UserId, Time, ReferrerURL, CurrentURL, UserAgent \rangle$$

The *UserId* is a unique anonymized identifier computed from the Flickr userId in case of logged-in users and from a browser cookie otherwise. *CurrentURL* and *ReferrerURL* represent the current page the user is visiting and the page the user visited before arriving at the destination page. The *User-Agent* identifies the browser in use, and the *Timestamp* indicates when the page was visited. All of the data processing was anonymous and performed in aggregate. Flickr allows users to set specific pages to "private", but in our analysis we considered only public pages.

3.1 Pageview Filtering & Data Selection

In order to obtain a coherent dataset in terms of both timezone and activity, we focused on users who are located in the US by extracting the location of the IP address from the source of the HTTP request and filtering out non-US locations. We then removed traffic derived from Web crawlers by preserving only the entries whose User-Agent field contains a well-known browser identifier (*e.g.*, Firefox, Chrome). In spite of this filtering, there are cases in which the User-Agent field indicates that a legitimate browser was used, but the corresponding "users" have a very large number of pageviews. The frequency, however, suggests that such server requests could not have been made by humans, but instead were done automatically for malicious crawling. We therefore apply an additional filter by which we set a maximum threshold on the total number of pageviews per user. The threshold is set to remove a small percentage of the users (1% of the total amount). After applying the filtering steps described above, our sample contains approximately 309 Million pageviews.

3.2 Session Identification and Characteristics

Since user behavior varies over time, we group pageviews into *sessions*. A session consists of a sequence of events, typically requests made to the server by a user (*e.g.*, page views), over a specific period of time. To perform our analysis, we split the activity of a single user into different sessions when either of these two conditions hold:

- **Timeout:** the inactivity between two pageviews is longer than 25 minutes.

- **External url:** if a user visiting Flickr leaves the site and returns from a different domain, the current session ends even if previous visits are within the 25 minute threshold (we make the assumption that if a user is viewing a page on Flickr and visits another domain, then the session ends).

In the rest of the analysis we use the filtered dataset and sessions.

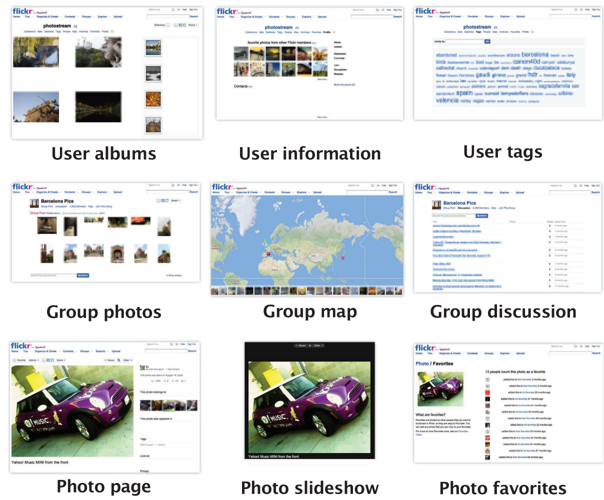

Figure 1: An example of pageviews that correspond to the entities of the browseGraph. Each row corresponds to a entity from top to bottom: user, group, photo

4. RANKING FROM BROWSEGRAPH

In this Section we explain how to create the BrowseGraph and we present a briefly analysis of the obtained graph. Furthermore, we describe the original BrowseRank algorithm and the modified version adapted to our domain.

4.1 Extraction of the BrowseGraph

The structure of a website is typically represented as a graph where nodes are pages and edges are the hyperlinks connecting them. In this model all the links have the same weight, disregarding how many times users go through them. The BrowseGraph [22, 23] is an alternative representation that captures the importance of the user navigation patterns by considering the actual transitions between one page to another, rather than hyperlinks.

The basic idea in our approach is that the navigation patterns within a social media platform have a strong impact on the importance of content. Therefore, we build a BrowseGraph based on our Flickr data: we create one node of the graph for each pageview that refers to one of three *entities* in Flickr, namely *users*, *groups*, and *photos*. Figure 1 shows examples of pages that are mapped to each entity, noting that all pageviews that show the same entity are condensed into a single node in the BrowseGraph. In other words, several types of pages are represented by a single node. For instance, since various URLs show the same image, all of those are mapped to a single node (fullscreen, slideshow, *etc.*).

The motivation behind this representation is that we are interested in ranking *the photographs* and those photographs may appear in multiple places (e.g., the photo appears prominently in the photo page, in the slide show, and in the photo favorites layouts of Figure 1), but since we are not interested in ranking each of the individual pages, we group them into a single node.

Flickr contains other page categories (*e.g.*, personal settings and photo upload pages) which we do not consider so we refer to them as non-entities and we do not create

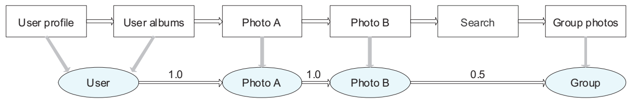

Figure 2: An example of a BrowseGraph. An example session is illustrated at the top, and the corresponding derived BrowseGraph is shown at the bottom. Gray arrows display the mapping between pageviews and browseGraph nodes.

Class	Examples	Ratio
search	search.yahoo.com, google.com	34.87%
social	facebook.com, tumblr.com	26.95%
mail	mail.yahoo.com, gmail.com	13.22%
aggregator	reddit.com, stumbleupon.com	7.76%
blog	blogspot.com, blogger.com	6.65%
photo	flickrhivemind.net, compfight.com	2.32%
microblog	twitter.com	2.26%
forum	discussion forums	2.00%
news	news.yahoo.com, cnn.com	1.67%
shop	ebay.com	0.85%

Table 1: Top ten most frequent external url classes in the dataset.

any nodes for them. The main reason to do so is that we are interested in the navigation between entities in Flickr. Therefore we need to discard some categories of traffic: *navigational* (*e.g.*, searching for photos), *configurational* (*e.g.*, changing settings, profile information) or *messaging*.

To build the browseGraph, we create the subgraph of each session and we then merge all subgraphs. Given a session $s = (p_1, p_2, ..., p_N)$ where p_n is a pageview, we map each entity pageview p_n to the vertex of the browseGraph and we connect them in the order they appear in the session. We then weight the arcs according to the number of non-entity pageviews between the source and the target. Intuitively, we would like to give the highest weight (namely 1) to the arcs that connect entities that appear in consecutive pageviews and a lower weight to pageviews that are more distant, to better express their actual proximity in the browsing activity. For example, in Figure 2, Photo A and Photo B are closer in the session than Photo B and Group. We do so by assigning the weight $w_{ij} = \frac{1}{NE(i,j)+1}$ where $NE(i, j)$ is the number of non-entity pageviews between i and j. We then compute the browseGraph by summing up all arcs with the same source and target.

A fragment of BrowseGraph is shown in Figure 2. The top row shows pageviews in a session, and the bottom part shows the resulting browseGraph. A vertex is created for each entity present in the session and gray arrows represent the mapping between pageviews and vertices. We can observe that the *Search* pageview, that displays the results of a query of the user and therefore does not refer to an entity, is not mapped to any BrowseGraph vertex but influences the weight of the corresponding edge of the BrowseGraph.

Modeling access from the Web (*i.e.*, domains different from Flickr) is important to detect the most frequently accessed entities from external sources. We therefore include in the graph also the nodes representing accesses to Flickr from external websites, derived from the *ReferrerURL* attribute of the first pageview of a session. However, since we are interested in understanding global navigation patterns, the full URL is too specific. For this reason we group external URLs in 17 classes as it was proposed by Chiarandini *et al.* [6]. Each class is detected by a set of regular expressions defined by manually inspecting the top 100 most popular referrer URLs. For example, the class "search" contains the URL of popular search engines (such as Yahoo!, Bing, and Google). The top ten most frequent classes are show in Table 1. The external URL classes cover around 99% of the total number of external URLs. For each class, we add a node to the BrowseGraph and we connect it to the nodes that correspond to the first entity of a session coming from the class.

	All	**Photos**	**Groups**	**Users**
#Nodes	49,275,691	46,569,946	183,996	2,521,749
$\langle k_{in} \rangle$	1.94	1.57	13.72	7.99
$\langle k_{out} \rangle$	1.94	1.51	13.69	9.05

Table 2: Browsegraph statistics, with detail on single node categories. $\langle k_{in/out} \rangle$ denote the average in- and out-degree.

	%Links			$\langle w \rangle$		
	Photos	Groups	Users	Photos	Groups	Users
Photos	0.6182	0.0098	0.1071	1.49	1.21	1.44
Groups	0.0114	0.0092	0.0057	1.54	1.44	1.65
Users	0.1332	0.0075	0.0979	1.48	1.41	1.32

Table 3: Flows and weights in browsegraph. Cells report the overall percentage of links flowing from a node type to another and the average weight $\langle w \rangle$ of edges according to the type of the endpoints.

4.2 Analysis of the BrowseGraph

The BrowseGraph we extract from the Flickr sessions has about 50 Million nodes 95 Million arcs, with in a very low graph density of around $3.8 \cdot 10^{-8}$. Statistics on average degree connectivity and graph size for the different Flickr entities are reported in Table 2. Higher degree of group and user nodes compared to photos suggests, as one might expect, that thematic groups and user profiles are hubs for the exploration of the website. The role of groups as navigation hubs is confirmed also by the inspection of the navigation flows between all of the possible pairs of node categories (Table 3), which shows that links from groups towards photos or users are on average used more often (*i.e.,* have higher weight) than other link types. Moreover, it appears that groups and user pages attract traffic from many sessions, but soon redirect this traffic to particular photos. This can be inferred by the fact the in-degree distribution for users and groups is heavier and broader than for photos, but the scenario is reversed when considering the distributions of the edge weights towards each node type (see Figure 3). In a nutshell, sessions end up in groups and user pages from anywhere in the network and from there they tend to converge to the most interesting or well positioned photos in the page.

Despite the important role of groups and user pages, the majority of arcs in the BrowseGraph are due to the navigation from one photo to another (62% of links, see Table 3). This is partially due to the disproportion in the cardinality of the three node categories (photo nodes account for 95%

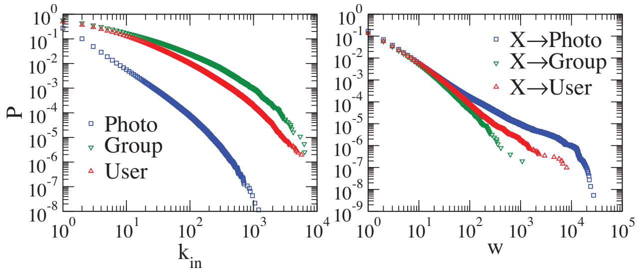

Figure 3: (Left) CCDF of the in-degree (k_{in}) for the three node types in the browsegraph. (Right) CCDF of arc weights for arcs terminating in nodes representing photos, groups or users.

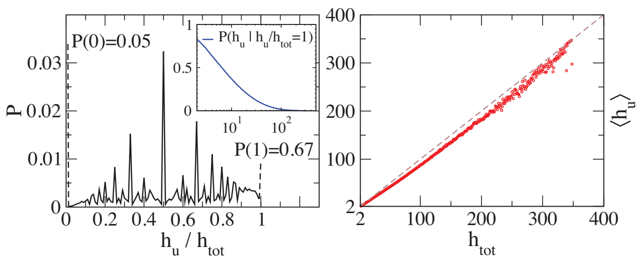

Figure 4: (Left) Distribution of ratio of session hops between two pictures belonging to the same owner (h_u) over the total number of hops (h_{tot}). The inset shows the CCDF of the number of hops for the sessions visiting only nodes of a single user, which constitutes the majority of cases. (Right) Average number of hops between pictures of the same owner $\langle h_u \rangle$ at fixed session length. Points lying almost on the diagonal mean very high correlation.

of nodes in the graph), but mainly it is the result of frequent navigation patterns. In fact, as shown in Figure 4, users very often browse photos of the same owner and this happens not only for short sessions, as highlighted by the rather broad distribution of session length for this case. This behavior is largely determined by the Flickr *photostream* at the top of every photo page, which shows a strip of 5 photos from the same owner, allowing easy navigation from one photo to another.

More generally, while surfing the Web, every user who visits a page eventually leaves following another link (or, more unlikely, end her session). As a result, a network created by the composition of such browsing patterns will have very high *balance* between in-and out-connectivity of nodes, as shown in Figure 5 (top). Such structural feature clearly differentiates navigation graphs from social graphs, in which popular individuals such as celebrities attract many connections and return a few back. The observed balance pattern gets slightly blurred only for very highly connected pages. Specifically, as shown in Figure 5 (bottom), well connected groups tend to have a higher in- than out-degree, while the distribution of the ratio of in-strength over out-strength for the most visited photos has a heavier tail towards values greater than one rather than towards zero. This confirms a scenario where the user navigation when not jumping from

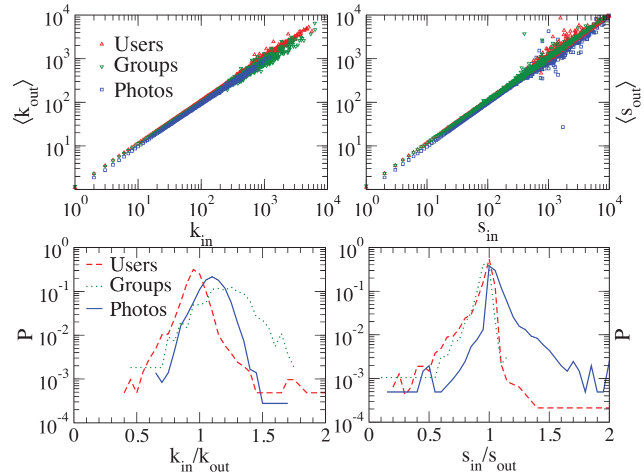

Figure 5: (Top) Average out-degree $\langle k_{out} \rangle$ and out-strength $\langle s_{out} \rangle$ at fixed values of in-degree k_{in} and in-strength s_{in}, for the three node types. Distribution of points almost perfectly aligned on the diagonal reveal the extremely high correlation between the amount of in and out session traffic which characterize navigation networks. (Bottom) Distribution of the ratio between in- and out-degree (in-and out-strength) for nodes with an in-degree (in-strength) higher than 500. The different skews of the distributions highlights the different roles of the three node types in browsing.

one photo to another, flows to hubs and gets redirected to popular photos.

Finally, a prevalent unidirectionality of browsing patterns can be evinced by the very low portion (0.17) of directed arcs $A \rightarrow B$ having a reciprocal $B \rightarrow A$, namely the reciprocation of the network. Again, this parameter is another footprint that discriminates navigation networks from social networks, which are on average highly reciprocated due to conventional social protocols.

4.3 Definition of Browserank

The BrowseGraph just outlined contains the information about user navigation paths and browsing behavior within Flickr. For example, the tendency to visit pictures in succession, moving directly from one photo to another, or exploiting the group and user nodes as hubs in order to select interesting photos and continue browsing.

Our goal is to use the computed BrowseGraph to rank entities inside Flickr. Since our BrowseGraph contains different types of nodes (photos, users, groups), not only photos are ranked. The obtained rank should capture well the global interest patterns leading the Web surfers to any of the entities considered. In this work we consider the rank for the photo nodes only, but in principle the rank scores obtained for user and group nodes can be used as well for different tasks.

Relying on the BrowseGraph structure alone may lead to a series of problems. Due to the low density of the graph (see Section 4.2), the ranking could be biased towards nodes with high degree (*e.g.*, a user with a large number of photos or spammers), regardless of the quality of the entities. Moreover, important node attributes such as the time spent on them or popularity would not taken into account. The ranking needs therefore to be adjusted using additional information. We applied and improved an algorithm that takes into account the time spent by the user on a page and uses this information to readjust the values returned by PageRank.

BrowseRank [22] is a ranking algorithm based on a continuous time Markov process model that exploits the link structure of the BrowseGraph. As opposed to the classic Markov process, BrowseRank takes into account the time that users spend on the page. In the context of Flickr, time spent on a photo could be a good indicator of interest by the user. Next, we describe the algorithm and the way in which we adapted it to the Flickr BrowseGraph.

4.3.1 Continuous-time Markov Model

As in [22], we use the Continuous-time Markov Model represented by the matrix $P(t) = [p_{nm}(t)]_{N\times N}$ where p_{nm} represents the transition probability from vertex v_n to v_m for time interval t.

The BrowseRank algorithm computes the stationary probability distribution $\{\pi_i\}$ by using the *transition rate matrix* $Q = [\frac{\partial}{\partial t}P(t)](0)$ and the Embedded Markov Chain (EMC). The EMC is a discrete Markov process derived from Q (for details see [22]). Given the stationary probability distribution of the EMC $\tilde{\pi}_i$, we can compute π_i using

$$\pi_i = \frac{\frac{\tilde{\pi}_i}{q_{ii}}}{\sum_{\{v_j\}} \frac{\tilde{\pi}_j}{q_{jj}}} \tag{1}$$

4.3.2 Embedded Markov Chain (EMC)

The EMC is a Markov Chain whose transition probabilities are based solely on the observed transitions between entities in the browseGraph $G = \langle\{v_i\}, \{e_{ij}\}, \{w_{ik}\}\rangle$ where $\{v_i\}$ is the set of vertexes, $\{e_{ij}\}$ is the set of edges and $\{w_{ij}\}$ the set of weights associated with the edges

In addition, for each node j, we compute the *reset probability* σ_j, *i.e.*, the probability of starting a new session in j as the number of sessions that start in j over the total number of sessions. Moreover, for each node j, we compute the *stop probability* α_j *i.e.* the probability of ending the session in j as the number of sessions that end in j over the total number of sessions that contain j. Both probabilities have been smoothed in order to avoid zero probabilities.

The transition probabilities of the EMC are computed in the following way:

$$emc_{ij} = \alpha_i \frac{w_{ij}}{\sum_{\{v_k\}} w_{ik}} + (1-\alpha_i)\sigma_j \tag{2}$$

Intuitively, Equation 2 indicates that as the user traverses node i of the graph, she may continue the navigation with probability α_i or randomly reset to any other node with probability $(1-\alpha_i)$. In case she continues, the transition probability is computed based on the observed transitions. In case she resets, the probability of ending up in node j is the reset probabilty σ_j. Equation 2 looks similar to the weighted PageRank algorithm [32] but we are able to exploit additional information that is not available to web crawlers. By having the number of sessions starting and stopping in a given node, we are able to estimate the specific reset and stop probabilities σ_i and α_i for every page i. The estimation of these parameters makes the random walk more realistic since it models the navigation of the user in a more accurate way. Equation 2 differs also from Equation 8 in [22] in the fact that we are not only estimating the reset probability, but also the stop probability. The additional advantage of this parameter estimation is that it avoids to manually set any parameter prior to running of the algorithm.

After computing the EMC transition probabilities, we compute the stationary probabilities $\{\tilde{\pi}_i\}$. Up to this point we have not taken into account the time spent by the user on the entities. It is indeed interesting to compare the performance of the ranking with and without this information. We will therefore save the $\{\tilde{\pi}_i\}$ and we will refer to them simply as the *PageRank*.

4.3.3 BrowseRank

As a final step, we include the information about the time spent by the user on the entities to improve the result of the previous section. For each vertex of the browseGraph v_i we compute the duration of the visits of users as follows:

- For each pageview p_n with timestamp t_n belonging to session s, we compute its duration d_n as the difference between the timestamp of the next pageview and its timestamp $d_n = t_{n+1} - t_n$. As we are not able to compute the duration of the last action of a session, we decided to discard it.

- We then compute the *aggregate durations* by summing up the duration of consecutive pageviews that refer to the same browseGraph vertex v_i.

- Finally, for each v_i we compute the sample mean $\bar{Z}_i$ and the sample variance S_i^2 of its aggregate durations.

We apply the addictive noise model [22] to cope with noise deriving from different connection speeds and we compute q_{ii} by solving the optimization problem in Equation 3.

$$\begin{aligned} \min_{q_{ii}} \quad & ((\bar{Z} - \frac{1}{q_{ii}}) - \frac{1}{2}(S^2 - \frac{1}{q_{ii}^2}))^2 \\ s.t. \quad & q_{ii} < 0 \end{aligned} \tag{3}$$

We can now solve Equation 1 to compute the value of BrowseRank for every node.

The BrowseRank algorithm is straightforward to parallelize in Map-Reduce. In terms of complexity, the most demanding step is the computation of the stationary probability distribution of the EMC $\tilde{\pi}_i$. Using the power method, the overall complexity of the algorithm is $O(N\log(1/\epsilon))$, with N number of edges in the graph and ϵ a given degree of precision [1].

5. EVALUATION

We compare the top 1,000 Flickr photos ranked using five different importance scores, specifically:

- **Favorites**: absolute number of favorite marks assigned to a photo. Favorites can be assigned only by Flickr users.

- **Views**: absolute number of views of the photo page (this includes users that are not logged in).

- **Time**: cumulative time spent by all of the visitors of a photo page.

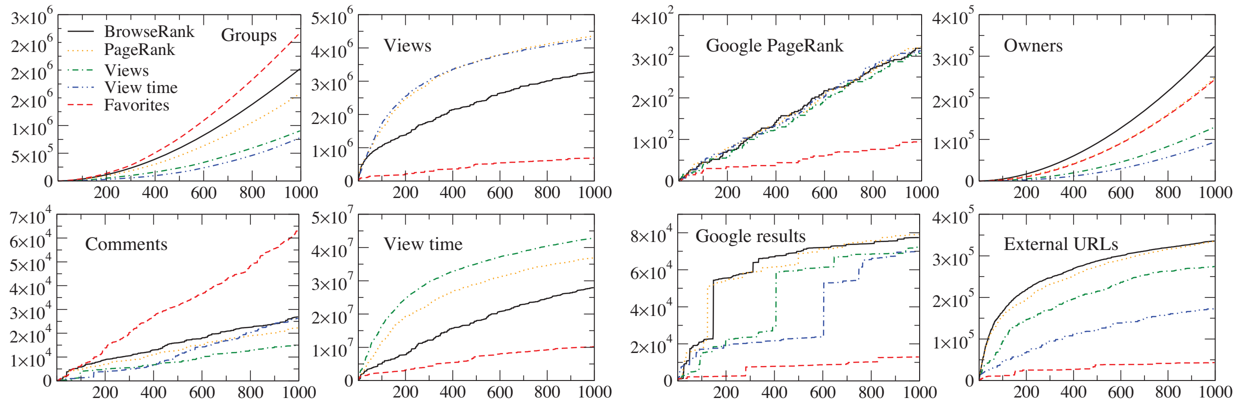

Figure 6: Comparison of the five ranking methods considered (BrowseRank, PageRank, Number of Views, View Time, Number of Favorites), according to eight features. Curves show the cumulative value of the feature up to the top $N \in [1, 1000]$ results in the ranking. The higher the curve, the more that feature is represented in the ranking. Columns from left to right depict results for number of distinct groups to which the photo in the ranking belongs, number of comments, number of views, view time, Google PageRank of the page containing the photo, number of Google results when using the page URL as search query, number of distinct owners of the photo, and number of external URLs from which the photo page has been reached. Number of views and view time are used as both ranking methods and features.

- **PageRank**: PageRank score of the photo page, with estimated start and stop probabilities as described in Section 4.3.2.
- **BrowseRank**: BrowseRank score of the photo page, with estimated start and stop probabilities as presented in Section 4.3.3.

The selected methods include a fairly general selection of explicit (Favorites), implicit (Views, Time) and centrality-based ranking techniques (Page/BrowseRank). The number of favorites has often been used as an evaluation baseline in Flickr photo ranking [26] as it is the most explicit indication of preference and the scores can easily be aggregated. Views and Time spent are also often used for ranking in photo sharing sites due to the ease of computation. Although quantitative correlations have been found between the visit count and the explicit user feedback on photos [20, 27], we show that all metrics behave in appreciably different ways.

5.1 Popularity, Interestingness, and Diversity

When comparing different picture sets, image quality is just one of the parameters. In particular, when images are embedded in dynamic social environments, the interest people have in particular photos can be determined (or influenced) more by the social dynamics of a community (*e.g.*, a group in Flickr) than by the inherent quality of the photos themselves. Similarly, interest can originate externally (*i.e.*, many photos in Flickr are linked from outside of Flickr) and thus be important independently of their aesthetic qualities (*e.g.*, photos of important events).

Given that several factors can be taken into account in considering a ranking of images, we identified four importance macro-notions and we list some quantitative *features* for each of them. All of the features were then used as evaluation parameters to compare the rankings.

- **Internal popularity**. Popularity of a photo inside the Flickr community. Popularity does not necessarily imply quality, but directly expresses the interest of users in a particular item. Features describing photo popularity are the number of *comments* the picture receives and the number of internal Flickr *groups* in which it appears[2].
- **External popularity**. We consider measures of external popularity: the *number of search results* obtained from a Google search using the photo page URL as a query, the *Google PageRank* of that URL, and the number of browsing sessions originating from an *external URL* that visit the photo page as the first Flickr page[3].
- **Collective attention**. Users not logged into Flickr as well as Flickr users who do not actively give feedback on photos, implicitly express their interest in specific photos by visiting the pages that contain them and by spending time on them. Therefore, we use the total *number of views* of a photo and the cumulative *time spent* on the photo as an aggregate measure of attention that a generic Web user, whether or not logged into Flickr, devotes to that image.
- **Diversity**. One of the applications of ranking a large-set of photos might be to display the most interesting ones. In this case, a very homogeneous set of pictures may result appealing to some user categories but are less likely to attract a wide public. Assuming that photos belonging to the same user are on average more homogeneous than pictures taken from different users,

[2]Although placing an image in multiple groups does not automatically make it popular, one can argue that photos that appear in multiple groups can be considered to be more popular because they have wider exposure

[3]For several queries, the Google search results were similar to those obtained by other search engines, so we used them as a representative metric

	$\frac{\lvert photos_{tag}\rvert}{\lvert photos\rvert}$	$\lvert tags\rvert$	$\lvert set(tags)\rvert$	$\langle tags\rangle$	H
BrowseRank	0.73	**7913**	**4347**	**7.93**	**11.23**
PageRank	0.75	7129	3583	7.39	10.57
Favorites	0.53	4164	2936	5.98	10.81
Time	0.80	6192	2245	6.20	9.31
Views	**0.83**	6523	2113	7.14	7.14

Table 4: Statistics on the tag diversity for the top 1000 photos in the rankings. Columns report, from left to right: fraction of tagged photos, number of tags, number of distinct tags, average number of tags per photo, and entropy H associated to the tag frequency distribution. Entropy is given in number of bits ($\log_2$). Highest values are highlighted in bold.

diversity can be estimated by the number of different *photo owners*. Additionally, an analysis on the diversity of the corpus of *tags* of the photos can be a measure of the variety of concepts represented.

We use the number of views and the view time as both ranking metrics and evaluation parameters to draw a more complete analysis of other rankings. We could have done the same for the number of comments, but we omitted to use it as a ranking metric because its performance was very similar to the Favorites. Cumulative values of each of the features defined are shown in Figure 6. To give a long-range overview of the behavior of the different ranking strategies, we show the feature values for the top 1k photos. Nevertheless, since many applications need much shorter ranked lists, we report that the relative position of the different curves is nearly unchanged for the top 20 and top 100 photos, for all the metrics considered.

Results reveal that most Favorites have good internal popularity, being the top metric in both number of groups and comments, but behave worse than any other metric in terms of external popularity and collective attention. In contrast, photos with top BrowseRank scores are less popular internally (even though their scores are comparable to favorites up to the top 100) but they attract relatively more collective attention and position above any other metric when counting external relevance and owner diversity. PageRank behaves worse than BrowseRank except for collective attention. Finally, Page views and Time perform reasonably well for external popularity and by definition in collective attention, but surprisingly the ranked photos have relatively little popularity in groups and receive few comments.

Diversity in terms of tag categories is explored separately in Table 4. The richness of the annotation corpora from the five rankings are evaluated in terms of number of (distinct) tags appearing in the corpus or on single photos. Furthermore, we computed the entropy on the tag frequency distribution as a measure of uncertainty of the type of tags attached to a randomly selected photo. BrowseRank clearly outperforms all other metrics.

5.2 Quality From Visual Inspection

Assessing the quality of photos by visually inspecting them is a challenging task due to the intrinsic subjective component of the evaluation. However, to gain insights into how different quantitative features impact the type of images shown, we show the top 10 images for all of the 5 ranks considered (Figure 7).

Albeit any manual classification is ultimately arbitrary,

	Art	Events	Series	Peculiar
BrowseRank	2,3,7,8	1,4,5	6	9,10
PageRank	4,5,7,9	2,3,6	1,8	10
Favorites	1,3,5,6	2,9	4,7,8,10	-
Time	5,9	10	1,2,3,4,6,7,8	-
Views	9	2,5	1,3,4,6,7,8,10	-

Table 5: Manual classification of top 10 ranked photos into four categories representing high-quality artistic images, natural and social events, picture series, and peculiar or fun images. Image numbers refer to Figure 7.

we partition the photos in four well-recognizable, high-level categories that help to better understand the nature of the top photos. The pictures shown are assigned to one of the following categories: 1) *artistic* high-quality landscapes or portraits 2) major natural and social *events*, 3) part of specific photo *series* or serial events, and 4) *peculiar* or curious shoots. Classification of each image is reported in Table 5.

At first glance, the Time and Views rankings are dominated by a majority of photos depicting scared visitors to a horror house[4]. Traces of the same series, plus a couple of pictures from a humorous calendar series are present also among the top Favorites; besides that, artistic pictures are prevalent, followed by two photos related to breaking news. BrowseRank and PageRank have an almost identical set of pictures, in different order. They both contain as many artistic images as Favorites but more images related to trending topics or natural events. Series-related pictures are present (*i.e.*, horror house and mosaics of electronic games characters) but just as singletons. Photos of peculiar art installation or entertainment activities complete the rank.

5.3 Discussion

The overall scenario emerging from the comparison shows that different metrics promote different types of photos. Rankings based on explicit feedback, namely Favorites, boost pictures that are well spread across Flickr groups and that receive attention from active Flickr users, but that may not have great impact outside of Flickr. Top rated images tend to belong to a small set of owners and convey a lower semantic variety than the pictures from centrality-based rankings. Artistic photos made by professionals are prevalent.

BrowseRank and PageRank, instead, overshadow a bit the very popular content *inside* Flickr to provide images with higher semantic variety and with apparently stronger interest from a broader part of the Web (outside of Flickr). This includes popular photos on trendy social events or pictures about popular fun facts or peculiar subjects. A positive side-effect of this is that photos that are related to popular memes just inside Flickr (*e.g.*, horror house pictures) are downgraded, and tend to disappear from the top ranking. Moreover, being based on the data from the navigation log only, centrality rankings are fully *implicit*. They do not need an active user base commenting or voting on the images. This means that BrowseRank and PageRank are effectively more able to pick up diverse image collections, and produced more balanced lists by considering external links to the photos. Such algorithms can be profitably parallelized, making efficient their computation even for big social media sites like Flickr.

[4]See http://www.nightmaresfearfactory.com/

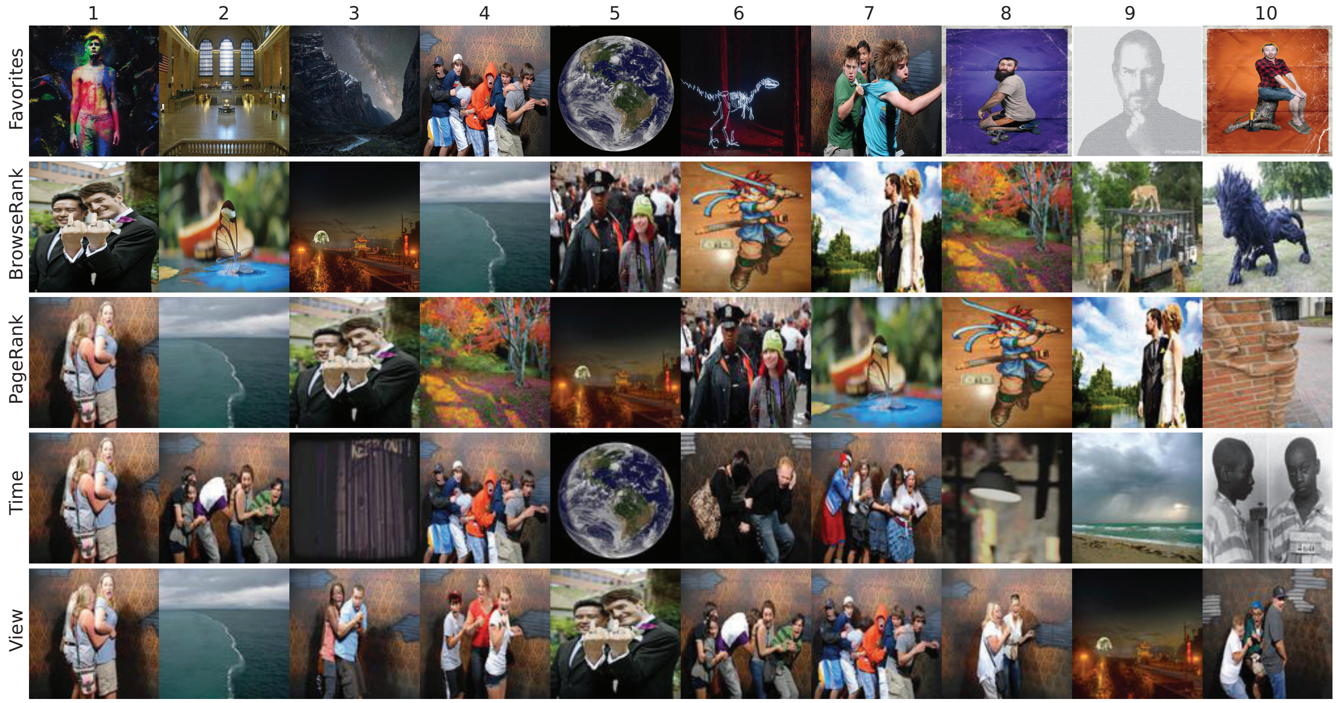

Figure 7: Top 10 photos for the five ranking strategies considered. Pictures include: (F2) shot of an empty railroad station during a hurricane in US, (F4 and similar) pictures of visitors to a horror house, (F8,F10) fun calendar series, (F9) memorial potrait of Steve Jobs, (B1) portrait in support of gay marriage, (B4) rare natural phenomenon of water masses at different densities melting one into another (the photo was broadcast by several news media), (B5) arrests during the "Occupy Wall Street" movement demonstrations, (B6) mosaics of a popular electronic-game character, part of a wider series, (B9) close lion encounters tourist van, (B10,P10) art installations, (T10) mugshot of the youngest African-American sentenced to death in the US, (F1,B2, and more) artistic portraits, landscapes or photoart.

Simpler metrics such as time spent and number of views have the advantage of an easy computation, but overall they perform poorly compared to others, at least in terms of diversity of the results.

6. CONCLUSIONS AND FUTURE WORK

The problem of general ranking of images in social photo sharing services has not a widely-accepted solution and differences between different strategies have not been explored in depth so far.

To shed light on this matter, we compared five possible ranking strategies in Flickr: explicit feedback (number of favorites), implicit user information (number of views and time spent viewing), and graph-centrality methods (PageRank and BrowseRank) applied to the BrowseGraph, namely the graph of the user browsing sessions. In particular, we contribute to the definition of a customized version of the BrowseGraph that is limited to the navigation patterns inside the boundaries of the considered service, but that takes into account also the entry points of users navigating to Flickr from other domains. The purpose of such model is to express the complexity of navigation patterns in a meaningful way that captures the importance that images have outside of the social media platform being considered. Unlike previous work in PageRank-based algorithms, we estimate a different damping factor for each page from the user session information.

A comparison between rankings was performed on a large Flickr dataset along several axes including the internal and external popularity of ranked images, the overall attention that they attract from Web users, their diversity in terms of ownership, and semantic categories, and their visual appearance. Results show that the ranking based on explicit user feedback behaves better than simple implicit methods. Favorite-based ranking boosts mainly professional artistic photos that are very popular inside Flickr but they are limited in variety and have low impact on the external Web. On the contrary, centrality-based methods, BrowseRank in particular, promote images that have attracted interest of external Web services like news media and produce more diverse rankings, minimizing the noise due to serial but relatively uninteresting photos periodically popping out in Flickr.

As future work, we plan a refinement of general image ranking by combining different strategies and the production of personalized image ranking by processing the local browsing information occurring in the logical proximity of the activity of a user. The application of BrowseRank to query-dependent image ranking has to be studied as well. Possible developments include the application of our customized BrowseRank to other domains like news services, knowledge directories or general-purpose social media. Regarding the exploration of the inherent BrowseRank properties, interesting research lines include the analysis of the impact that temporal patterns have on the quality of the ranking and a thorough theoretical analysis of the effectiveness of our customized BrowseGraph model in the estimation of the global PageRank of a document in the Web.

7. ACKNOWLEDGMENTS

This research is partially supported by European Community's Seventh Framework Programme FP7/2007-2013 under the ARCOMEM and Social Sensor projects, by the Spanish Centre for the Development of Industrial Technology under the CENIT program, project CEN-20101037 "Social Media", and by Grant TIN2009-14560-C03-01 of the Ministry of Science and Innovation of Spain. Furthermore we would like to thank Paolo Boldi for his helpful suggestions.

8. REFERENCES

[1] M. Bianchini, M. Gori, and F. Scarselli. Inside PageRank. *ACM Transactions on Internet Technology*, 5(1):92–128, 2005.

[2] P. Boldi, M. Santini, and S. Vigna. Do Your Worst to Make the Best: Paradoxical Effects in PageRank Incremental Computations. *Algorithms and Models for the Web-Graph*, pages 168–180, 2004.

[3] M. Bressan and L. Pretto. Local computation of PageRank: the ranking side. In *Conference on Information and Knowledge Management*, pages 631–640. ACM, 2011.

[4] A. Capocci, A. Baldassarri, V. Servedio, and V. Loreto. Friendship, collaboration and semantics in Flickr: from social interaction to semantic similarity. In *MSM*, pages 8:1–8:4. ACM, 2010.

[5] M.-F. Chiang, W.-C. Peng, and P. Yu. Exploring latent browsing graph for question answering recommendation. *World Wide Web Internet And Web Information Systems*, pages 1–28, 2011.

[6] L. Chiarandini, M. Trevisiol, and A. Jaimes. Discovering Social Photo Navigation Patterns. In *ICME*. IEEE, 2012.

[7] F. Chierichetti and R. Kumar. Optimizing Two-Dimensional Search Results Presentation. *Science*, pages 257–266, 2011.

[8] O. Chum, J. Philbin, and A. Zisserman. Near duplicate image detection: min-hash and tf-idf weighting. In *BMVC*, volume 3, page 4, 2008.

[9] B. Gao, T.-Y. Liu, Y. Liu, T. Wang, Z.-M. Ma, and H. Li. Page importance computation based on Markov processes. *Information Retrieval*, 14(55):488–514, 2011.

[10] B. Geng, L. Yang, C. Xu, X. Hua, and S. Li. The role of attractiveness in web image search. In *Multimedia*, pages 63–72. ACM, 2011.

[11] R. Halonen, S. Westman, and P. Oittinen. Naturalness and interestingness of test images for visual quality evaluation. In *SPIE*, volume 7867, page 34, 2011.

[12] E. Hoque, O. Hoeber, and M. Gong. Evaluating the Trade-Offs between Diversity and Precision for Web Image Search Using Concept-Based Query Expansion. In *WI-IAT*, volume 3, pages 130–133. IEEE, 2011.

[13] V. Jain and M. Varma. Learning to Re-Rank: Query-Dependent Image Re-Ranking Using Click Data. In *WWW*, pages 277–286. ACM, 2011.

[14] Y. Jing. Pagerank for product image search. In *WWW*, pages 307–315. ACM, 2008.

[15] Y. Jing and S. Baluja. VisualRank: applying PageRank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1877–90, Nov. 2008.

[16] H. Katti, K. Bin, T. Chua, and M. Kankanhalli. Pre-attentive discrimination of interestingness in images. In *ICME*, pages 1433–1436. IEEE, 2008.

[17] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[18] R. Lempel and S. Moran. The stochastic approach for link-structure analysis (SALSA) and the TKC effect. *World Wide Web Internet And Web Information Systems*, 33:387–401, 2000.

[19] K. Lerman and L. Jones. Social Browsing on Flickr. In *ICWSM*, pages 1–4. ACM, 2007.

[20] K. Lerman, A. Plangprasopchok, and M. C. Wong. Personalizing Results of Image Search on Flickr. In *ITWP*. AAAI, 2007.

[21] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In *WWW*, pages 351–360. ACM, 2009.

[22] Y. Liu, B. Gao, T.-Y. Liu, Y. Zhang, Z. Ma, S. He, and H. Li. Browserank: letting web users vote for page importance. In *SIGIR*, pages 451–458, New York, NY, USA, 2008. ACM.

[23] Y. Liu, T.-Y. Liu, B. Gao, Z. Ma, and H. Li. A framework to compute page importance based on user behaviors. *Information Retrieval*, 13(1):22–45, 2009.

[24] Y. Liu, M. Zhang, and S. Ma. User browsing graph: Structure, evolution and application. In *WSDM (Late Breaking-Results)*. ACM, 2009.

[25] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. *World Wide Web Internet And Web Information Systems*, 54(2):1–17, 1998.

[26] J. S. Pedro, P. Street, and S. Sheffield. Ranking and Classifying Attractiveness of Photos in Folksonomies. In *WWW*, pages 771–780. ACM, 2009.

[27] C. Prieur, D. Cardon, J.-S. Beuscart, N. Pissard, and P. Pons. The Stength of Weak cooperation : A Case Study on Flickr. *arxiv.org*, 65(8):610–613, 2008.

[28] R. van Leuken, L. Garcia, X. Olivares, and R. van Zwol. Visual diversification of image search results. In *WWW*, pages 341–350. ACM, 2009.

[29] R. van Zwol and E. Al. Faceted exploration of image search results. In *WWW*, page 961. ACM, 2010.

[30] R. van Zwol, A. Rae, and L. Garcia Pueyo. Prediction of favourite photos using social, visual, and textual signals. In *Multimedia*, pages 1015–1018. ACM, 2010.

[31] O. Wu, W. Hu, and B. Li. Group ranking with application to image retrieval. In *Conference on Information and Knowledge Management*, pages 1441–1444. ACM, 2010.

[32] W. Xing and A. Ghorbani. Weighted PageRank algorithm. In A. A. Ghorbani, editor, *CNSR*, pages 305–314. IEEE, 2004.

[33] B. Zhou, Y. Liu, M. Zhang, Y. Jin, and S. Ma. Incorporating web browsing activities into anchor texts for web search. *Information Retrieval*, 14(3):290–314, 2010.