Predicting Pinterest: Automating a distributed human computation

Changtao Zhong, Dmytro Karamshuk, Nishanth Sastry
King's College London, UK
{changtao.zhong, dmytro.karamshuk, nishanth.sastry}@kcl.ac.uk

ABSTRACT

Everyday, millions of users save content items for future use on sites like Pinterest, by "pinning" them onto carefully categorised personal pinboards, thereby creating personal taxonomies of the Web. This paper seeks to understand Pinterest as a distributed human computation that categorises images from around the Web. We show that despite being categorised onto personal pinboards by individual actions, there is a generally a global agreement. We compare and contrast such manual categorisation with automatic image classification using state-of-the-art deep convolutional network, and show that each has its advantages. By combining both together with a user-specific model of interests and preferences, we show that manual curation actions of a majority of users can be predicted, there by *automating* a large fraction of Pinterest actions. Our model is able to both predict whether a user will repin an image onto her own pinboard, and also which pinboard she might choose.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Commercial Services, Data Sharing, Web-based services; I.4.9 [**Image Processing and Computer Vision**]: Applications; J.4 [**Social and Behavioral Sciences**]: Sociology

Keywords

Content Curation; Pinterest; Deep learning; User behaviours; Image analysis; Supervised learning

1. INTRODUCTION

Following on the heels of the information glut created by the user-generated content revolution, an interesting new phenomenon that has been termed *content curation* has emerged: Rather than *create* new content, content curation involves *categorising* and *organising* collections of content created by others. For instance, on Pinterest, perhaps the most prominent content curation site, users can collect and categorise images (and the URLs of the webpages that contain them) by "pinning" them onto so-called "pinboards". Similarly, delicious.com (formerly del.icio.us) allows users to categorise URLs by tagging them.

A crucial aspect of such sites is that content curated by one user is also (by default) made available to the rest of the users to curate. For instance, on Delicious, links of another user can be copied onto one's own list, by tagging them. Users on pinterest can copy images pinned by other users, and "repin" onto their own pinboards. Interestingly, such reappropriation and curation of content discovered by other users (termed as "repins") is by far the most common activity on Pinterest, constituting about 90% of user actions, as compared to directly discovering and pinning new images, which constitutes only 10% of actions[1].

Even if curating a content item that has already been categorised, users typically have re-adjust it for their own collections: For instance, in tagging systems, multiple studies [10, 28, 25] have recognised that inter-user agreement can be low, and different users may choose different tags for the same URL (although an overall consensus vocabulary may emerge per-URL, due to tag imitation or background knowledge [10]). Similarly, users' pinboards are highly personal and idiosyncratic represntations of their taste, and furthermore, users are free to choose to repin any image onto any pinboard. Consequently, curation on Pinterest remains a highly manual process as well.

Our goal in this paper is to make it easier to re-appropriate and re-categorise content for personal use: *Given a pin (image) and a user, we wish to predict whether the user would be interested in repinning the pin. Moreover, we wish to predict the pinboard onto which they would repin, and automatically suggest these to the user.*

Towards this end, we revisit the notion of agreement in the context of Pinterest. Unlike traditional social bookmarking, pinning on Pinterest does not involve creating an explicit vocabulary of tags to describe the image. However, each pinboard may be associated to one of 32 categories defined globally for all users by Pinterest. Thus, even though each pinboard may be exclusive to a user, the act of pinning *implicitly* categorises the image. In other words, *Pinterest users can be seen as performing a massive distributed human computation, categorising images found on the Web onto an (extremely coarse-grained) global taxonomy of 32 categories.*

We find that this lower dimension approximation of Pinterest has several desirable properties: First, for a given image, a remarkable ≈75% of repins tend to agree on the

[1] According to our dataset, collected originally in Jan 2013.

WWW 2015, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3469-3/15/05.
http://dx.doi.org/10.1145/2736277.2741671.

majority category, although the level of agreement varies with the category. This enables us to robustly predict the category of an image. Second, users tend to specialise in a handful of categories; we use this to learn whether a user would be interested in an image given its category. Third, most users appear to have one or two boards per category. Thus, given the category of an image, and the user, it is often trivial to predict the board to which the user would pin the image. Based on these observations, we are able to build classifiers that, given that the user has curated an image, can predict the pinboard onto which the image would be repinned, and automatically suggest these to the user.

We augment this by showing that the *content* of the image can be used to predict whether the user would be interested in repinning the pin. To build this predictor, we derive several thousands of image-content features (Table 1), ranging from basic visual and aesthetic features to features extracted from the layer right before the final classification layer of the state-of-the-art deep convolutional network in Caffe [18], and by recognising objects that may be embedded using the same convolutional neural network. Using these features, we construct a supervised machine learning model that is able to assign an image to the majority category. We also learn user preferences for these image features, and predict whether the image would be repinned by the user.

We compose these classifiers in a pipeline of three layers. The first to predict whether the user will pay attention to a pin; the second to predict the category that the user will chose for the pin; and the third to predict the pinboard chosen given the category. Together this pipeline or cascade of classifiers is able to predict curation actions on Pinterest with an accuracy of 69%.

The rest of this paper is structured as follows. In §2, we discuss related strands of research. §3 gives a brief background of Pinterest terminology and our dataset. §4 demonstrates that Pinterest users are highly specialised in the categories they are interested in, and generally agreement over category assignments. The rest of the paper develops a pipeline of predictors: §5 sets the stage, discussing the cascaded structure of the predictors and the features used. §6 develops a classifier to predict whether a user will pay any attention to a pin. §7 then develops a two-stage multi-class classifier that predicts the board chosen by a user for a repin. §8 puts it all together, showing that repins can be predicted, both in terms of whether users would be interested in repinning a pin and which of their boards they would place it onto. §9 ends by discussing implications for the wider research agenda.

2. RELATED WORK

Concurrent with the rapid rise of Pinterest, there have been several studies of Pinterest as a content curation platform [12, 38, 36, 27, 9, 37, 26, 3]. Using a variety of approaches ranging from qualitative user studies [36, 38] and textual or content analyses [9, 26], to large-scale exploratory data analysis [38, 13, 27, 3] and complex or multi-layer network studies [37], these works have shed light on a number of important aspects, such as the nature of its social network [38, 37], the relationship between Pinterest and other social networks such as Facebook [37] and Twitter [9, 26], the role of homophily and user specialisation [3], gender differences in activities [27, 9] and the motivations of the content curators [38, 36, 13].

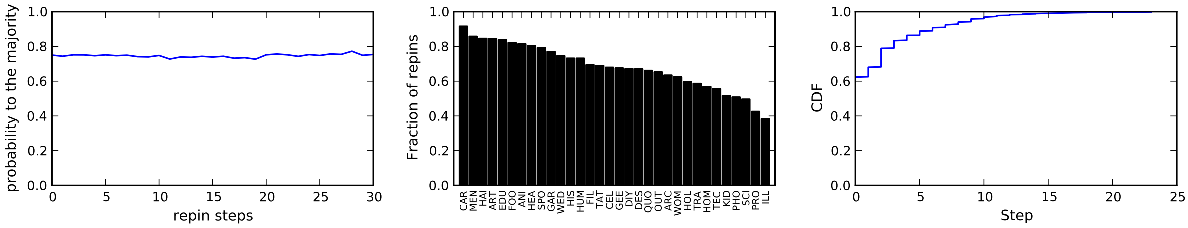
We believe this is the first work to extensively use a *machine learning approach* to automate content curation and thereby also try to obtain a *mechanistic understanding* of the *end-to-end process of social content curation* on Pinterest. Although Han *et al.* [13, §7] also explore pin prediction, the scope is a more limited problem of predicting 20 pins for each of 4667 users (in comparison, our dataset has 214K pins with 1.2M repins by 230K users) and checking whether these are in the user's pinboards after 125 days, without the multi-class classification into specific pinboards. Also, the best Han *et al.* models obtain an average precision of 0.015, which stands in contrast with the much higher values we obtain (c.f. Table 6). Also using a supervised machine learning approach is the preliminary work of Kamath *et al.* [20], who propose a model for recommending boards to Pinterest users.

This paper also contributes to image analysis and understanding. There has been a substantial body of work in this field over the years, but the availability of user-generated image data on social sites such as Flickr, and annotated image databases such as Imagenet [6], have enabled significant advances in the recent past, answering relatively sophisticated questions such as what makes an image memorable [17], interesting [11] or popular [21]. Our paper contributes to this line of research, and provides answers to the question of when an image is curated by a given user, by collecting a large dataset of $> 200K$ images from Pinterest, and crucially, leveraging the category of the pinboards of these images to infer implicit labels for most images.

Another related recent stream of research is concerned with using machine learning to predict the actions of a crowd. For instance, Blackburn and Kwak [1] predict crowdsourced decisions on toxic behavior in online games. Yin *et al.* [35] model latent personalities of those who vote to predict the popularity of online items. [19] trains a model that predicts individual decision and consensus vote for the Galaxy Zoo classification problem and uses this to optimise task assignment for actual crowdsourcing. Along these lines, here we develop a model that predicts content curation actions of the Pinterest community.

Social tagging phenomena have been widely studied in the context of bookmarking sites (delicious.com, citeulike.com, bibsonomy.org, etc.). In [34] authors revealed that even if people share a common understanding about content, they may tag differently. More than one study has suggested that tagging behavior is driven by users' motivation to either categorise content for easier navigation or describe content to improve its searchability[31, 38]. Consequently, users may employ different vocabularies depending on their motivation: a more compact but also a more informative vocabulary for categorisation, whereas a more diverse and a more fine-grained one for descriptions [31]. With respect to this flow of work we focus on the wide range of features that describe visual content, on the one hand, and aim to understand tagging activities when the taxonomy is fixed by the predefined set of categories, on the other hand. Previous studies have also found a high level of agreement

Interestingly, we find that although users' decisions are not supported by an automatic tag/category suggestion mechanism, such as on delicious.com or bibsonomy.com [34], This result is similar to the motivation of the ESP game [32], which *is designed* to induce agreement.



(a) Probability of majority pin (b) Levels of agreement in categories (c) No. pins for majority to appear

Figure 1: **Emergence of consensus in Pinterest**: (a) The category chosen by the *ith* pinner is independent of the category chosen by the previous $i-1$ pinners, and is the same as the category chosen by the majority of repinners with a remarkably high probability (≈0.75). (b) The average fraction of pinners in the majority can vary across category, ranging from 91% (cars and motor cycles, or CAR), to 43% (Illustrations and Posters, or ILL). All except PRO (45%, Products) and ILL have a majority > 50%. (c) Cumulative distribution function (CDF) of the probability that the majority category that emerges at the *ith* pin remains the majority category after all repins have occured. For over 60% of pins, the category of the very first pin determines the majority category; in over 90% of cases, a stable majority has emerged after just 5 repins (all pins have > 5 repins in our dataset)

3. PRELIMINARIES

We begin by briefly describing Pinterest, our terminology, and the dataset used in the rest of this paper:

Pinterest is a photo sharing website that allows users to organise thematic collections of images. Images on Pinterest are called pins and can be added in one of two ways. The first way is pinning, which imports an image from an external URL. The second is repinning from an existing pin on Pinterest. Users can also like a pin or comment on it, but repinning constitutes the vast majority (≈92% in our dataset) of actions.

Terminology: We use the terms *pin* and *image* interchangeably; a pin or image may be *repinned* multiple times. The user who introduces an image into Pinterest is called its *pinner*; others who re-appropriate it for their own pinboards are *repinners*. In this paper, we are mostly concerned with how users categorise and organise their images or pins. All pins are organised into collections called *pinboards* (or *boards*), which belong to one of the 32 globally specified *categories* (excluding one category, 'other', for non-categorised boards)

Dataset. The dataset we use comes from a previous study [38], and includes nearly all activities within Pinterest, during the period from January 3rd to 21, 2013. The data was collected as follows: To discover new pins, we visited each of the 32 category pages once every 5 minutes, and collected the latest pins of that category. Then, for every collected pin we visited the homepage of the pin every 10 minutes. A pin's homepage would list the 10 most recent repins and 24 most recent likes[2] which we collected and stored along with the timestamp of the crawling pass. Through these regular visits, we captured almost all the activity during our observation period. We estimate that the fraction of visits which resulted in missed activities stands at 5.7×10^{-6} for repins and 9.4×10^{-7} for likes. Further details of our dataset may be found in [38].

In this paper, we wish to understand how the features of the pin and the pinner affect the activity of repinning. Therefore, we focus only on users with more than 10 pins in our original dataset, and on pins which have been repinned at least 5 times, ending up with a set of 214,137 pins, 237,435 users and 1,271,971 repins for analysis.

[2]This setting has been changed in April 2013.

4. PREDICTABILITY OF REPINS

Curation on Pinterest is currently a highly manual procedure. Users select images that they like, and categorise it amongst one of several thematic collections or pinboards that they curate. Over 85% of respondents in a previous user study considered their pinning activity to be highly personal, akin to personal scrapbooking [38].

This paper, however, aims automate this procedure, as much as possible. To this end, we examine the extent to which properties of the pin, or the user, can assist in suggesting an appropriate pinboard of the user for the pin.

We first take a pin-centric view, and ask whether repins of other users can help, and show that users tend to strongly agree on the *category* that they implicitly assign to a pin, via the category of the pinboard they choose. Next, we take a user-centric view, and show that users tend to be highly specialised in their choice of pinboards, focusing on a handful of categories, and also typically have very few boards within each category. We conclude by highlighting the implications of these findings, which we make use of in subsequent sections.

4.1 Pinterest users agree on image categories

Pinboards are personal to each user, and pinboards of different users typically share at most a handful of pins, if at all. However, each pinboard may be assigned to one of 32 categories which have been pre-determined by Pinterest. Therefore, we may regard a repin as implicitly assigning one-of-32 labels to an image, reminiscent of ESP [32], a human computation task which greatly improved label prediction for images. We ask whether users agree on the category assignment for images in the context of Pinterest.

Formally, each repin by user u of a pin p to a pinboard b whose category is c is interpreted as an assignment of the category c to pin p by user u; we denote this as $repin\_cat(p, u) = c$. After users $1..i$ have repinned a pin, one can define the count of the category assignments of c to p: $count_i(p, c) = |\{k | repin\_cat(p, k) = c, \ \forall 1 \leq k \leq i\}|$ We define the majority category of an image or a pin as the category chosen by the maximum number of repinners[3]. In other words, the majority category $\text{maj}_i(p)$ after users $1..i$ have repinned a pin is the category with the maximum count: $\text{maj}_i(p) = \text{argmax}_c\, count_i(p,c)$. The final majority category $\text{maj}_\infty(p)$ is the majority category after all r repins have been made. The consensus or agreement level after r repins can be computed as the fraction of pins in the final majority category after r repins: $agreement_r(p) = count_r(p, \text{maj}_\infty(p))/r$.

Whereas other curation systems (such as social tagging on del.icio.us) might push users towards agreement by suggesting tags [10], in Pinterest, a pin does not come with a category of its own, and no category is suggested by the system or other users. Indeed, it is quite difficult on Pinterest to discover the category of a pinboard: Visitors to a pinboard's webpage can only determine its category through a meta-property in the HTML source code[4]. Even the owner of the board is only shown the category on the page for editing a board's details (not normally seen when the owner views her board). Because of this UI design decision in Pinterest, we expect that a user's choice of the Pinterest category to associate with a pin is made independently of other users or the system itself. Furthermore, the category choice is made only implicitly, as a result of an explicit choice made on which pinboard to place the image in. Thus, we expect this decision to be influenced by, for instance, whether the image being pinned fits thematically with the other images in the pinboard, and not by other users.

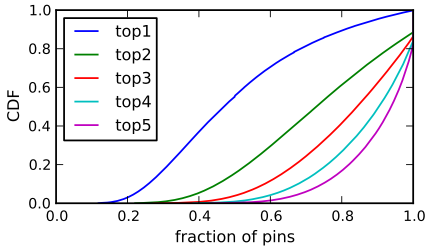
We first test our expectation that users individually decide on an image's category. We ask what is the probability $P[repin\_cat(p,i) = \text{maj}_\infty(p)]$, that the ith repin of a pin agrees with the final majority category chosen for it. Confirming our intuition, Figure 1a shows that the ith repinner's choice appears to be unaffected (either positively or negatively) by the choices of all the previous $i-1$ repinners. Furthermore, we see that there is a remarkably high chance ($\approx$75%) that the category implicitly chosen by a user agrees with the majority. Figure 1b shows that the average levels of agreement can vary across pins of various categories, from 91% to 42%; and in all categories except Illustrations, the final majority category has a clear majority of $> 50\%$ agreement.

Next, we ask how many pins it takes for the majority to emerge, *and stabilise*: Suppose we temporally order in ascending order the pinners of a pin p, starting with the first pinner as 1. We wish to know the number of repins required (smallest pinner number a) at which the majority category is the final majority category, and the consensus on the majority is unchanged by all subsequent pins. Formally, we want the smallest pin a such that $\text{maj}_k(p) = \text{maj}_\infty(p),\ \forall k \geq a$. Figure 1c shows the cumulative distribution of the number of repins a required for stable agreement to emerge. In over 60% of images, this happens with the very first pin. After 5 repins, over 90% of images have reached consensus on the final majority category.

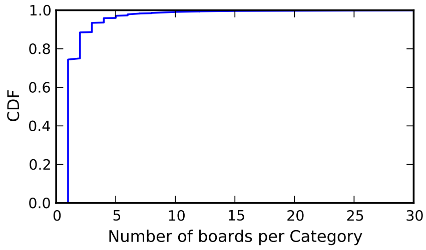
4.2 Pinterest users specialise in few categories

[3]Note that we do not require >50% of pinners agree on a category, although this often happens.

[4]Users often repin images from the homepage of Pinterest, and may not even visit the board of the original pin. Thus they may not know the category even if they can read HTML



(a) Category concentration



(b) Num. boards per category

Figure 2: **User specialisation** (a) CDF of the fraction of users' pins in their top-k categories shows that each user specialises in a handful of categories. (b) CDF of number of boards per category shows they have users tend not to have many boards in each category, implying that categories can predict users' choice of pinboard for a pin.

Having looked at a pin-centric view on predictability, we now look for user-specific patterns that can aid us in categorising their content automatically. Again we focus on categories. We first look at the distribution of categories per user and find that users are highly specialised: Figure 2a considers the fraction of a user's pins which are in the top-k categories of the user. This shows, for example, that about half the users have nearly half their pins in pinboards belonging to their the top category, and 80% users have *all* their pins in pinboards belonging to their top-5 categories. This indicates a high degree of specialisation.

We next consider how users choose to create pinboards. Figure 2b shows that most users have one or two pinboards in each category. Thus, it appears that users are mostly following the coarse-grained taxonomy developed by pinterest, and are in fact simply categorising images in this space, rather than on highly personalised pinboards.

4.3 Implication: Board prediction is easy

The results of §4.1 strongly suggest that most repinners agree on the category to assign to a pin, and furthermore, this majority category can be deduced just by observing the first few (say 5) repins. Secondly, since users' pins in different categories are highly skewed (Figure 2a), users' own personal favourite categories can be predicted as a choice for the category used. In examining the corpus of repins, we find that (consistent with Figure 1), that $\approx$87% of repins (after the first five repins) are made to the majority category. A further 4.7% of repins are made not to the majority category, but to the category in which the user has most of her pins. Thus, we expect that predicting the category of a particular pin based on these powerful signals can be an easy problem, and exploit these in §7.1.

Further, §4.2 suggests that users tend to have very few boards per category. Thus, once the category is predicted, we expect to be able to predict the actual pinboard chosen by the user as well. Finally, for the few cases when the user's pins are not in the majority category, we propose to use the fact that users specialise in a few categories to predict the correct category and thereby the board used. We explore the above two strategies in §7.2.

We might also conjecture that the high levels of agreement seen for a pin's category may in fact be a result of the high degree of user specialisation within categories: Since users choose to repin in very few categories, the fact that the user has paid any attention to a pin and repinned it is a strong indicator that the pin belongs to the given category. This may help explain the result of Figure 1a that nearly 8 out of 10 repinners agree on the category for a pin.

5. PREDICTING PINTEREST: AN OUTLINE

In the rest of this paper, we will use the notions of user category specialisation and agreement uncovered from the data, together with a number of user- and image-content related features to develop a model for predicting users' repins. Our ultimate goal, as stated earlier, is to automatically suggest, given a user and a pin, whether the user will repin and which pinboard it will be replicated to. In this section, we describe the the features used, and our outline model for predicting content curation actions as a cascade of predictors. Later sections will use the dataset described in §3 to validate the different parts of the model.

5.1 Curation as a cascade of predictors

We model the process of curation as a cascade of predictors. A content curation action involves a user u who "repins" a pin onto one of her pinboards. The first prediction problem (§6) is to take a user u and a pin p, and predict an action $f_1 : (p, u) \rightarrow \{noaction, repin\}$. Next, the *repin* involves a further user-specific decision as to which pinboard the pin should be placed in (§7). We may formulate this problem as a classification task where a machine learning classifier f_2 is trained to recognize pinboard b_i to which user u is going to put repinned pin p, i.e., $f_2 : (p, u) \rightarrow \{b_1, b_2, ..., b_n\}$ where $\{b_1, b_2, ..., b_n\}$ is a set of user's u pinboards. However, taking cue from §4.1 and §4.2, we split this task into two. First, we predict the *category* that the user might implicitly choose for the image (§7.1), i.e., we train a classifier $f_{2.1}$ to recognise category c_i to which user u is going to put repinned pin p, i.e., $f_{2.1} : (p, u) \rightarrow \{c_1, c_2, ..., c_n\}$ where $\{c_1, c_2, ..., c_n\}$ is a set of user's u categories, using the majority predicted after the first five repins as a strong input feature. Then in §7.2, we train a classifier $f_{2.2}$ to predict the pinboard given the category: $f_{2.2} : (c, u) \rightarrow \{b_1, b_2, ..., b_n\}$. As expected from Figure 2b, this turns out to be an almost trivial problem.

5.2 Features

We tackle the above classification problems by incorporating both pin- and user-related features. Pin-related features include content/image-specific features, various metadata about the pin such as who the pinner is, the consensus agreement (§4.1) around the category of the pin, as measured from the first five repins, etc. User-related features of the repinner being considered includes user metadata and statistics, personal predisposition of the user to specialise in and repin content of particular categories (§4.2), as well as image-specific features of previously pinned images. All features are summarized in Table 1.

5.2.1 Basic Visual Features

Firstly, we define 14 basic visual features to describe the content of an image. The colour based features includes: lightness, saturation, colorfulness, gray contrast, RMS contrast, naturalness, sharpness, sharp pixel proportion and left-right intensity balance. Additionally, we consider three aesthetics features: simplicity, modified-simplicity and roles of thirds. More details can be found in Table 1. Our goal is to assess how these basic visual features can capture user preferences for a particular type of content.

We note that extraction of the majority of visual features requires significant computational resources on the scale of 151K images. Therefore, we used a dataset of down-scaled images (i.e., with width equal to 200 pixels) to extract all visual features, except for lightness, saturation, colorfulness and naturalness for which performance was not an issue. Our experiment on a random subset of 1K images showed that the Pearson's correlation coefficients between the features extracted from the original and rescaled images were over 0.90 across all features, suggesting that the error introduced by using the down-scaled images is reasonably small (an average absolute error of 0.01).

5.2.2 Deep Learning Features

Next, we use the state-of-the-art approach in image classification [7], deep convolutional networks [22], to extract a set of more fine-grained visual features. More specifically, we train a deep convolutional network using Caffe library [18] based on 1.3 million images annotated with 1000 ImageNet classes and apply it to classify Pinterest images. This way we are able to extract two types of visual features: (1) features from the layer right before the final classification layer. These are known for a good performance in semantic clustering of images [8]; and (2) objects detected from the 1000 final Image Object classes.

5.2.3 Defining User Preferences

To describe users' preferences for particular type of content we analyze visual features of the images they repin: we measure how many images they repin in different Pinterest categories. In addition, we devise user-specific signatures from the visual features of those repins: for each user in our dataset we measure a vector of 1000 features which represent a centroid of Image Object features from their previous repins. Last but not least, we take into account the level of user activity on Pinterest by measuring different statistics from their profiles: number of followers, number of boards, number of repins, etc.

5.2.4 Other pin-related features

In addition to the above, driven by Figure 1c, we extract a simple but powerful feature: the majority category as seen after the first five repins. We then use these to predict repins beyond the first five (§7). In addition, we use all user-specific features of the original pinner: Features such as the number of followers, are indicative of the status and reputation of the user, and might have a bearing on whether/not an image is repinned. Similarly the taste and expertise of the original pinner may also indirectly influence the repinner, and are captured by the user's image object centroid, and activity levels in terms of number of boards, repins etc.

6. PREDICTING USER ATTENTION

Here we develop the first step of the cascade pipeline described in §5.1. We analyze the features which drive user attention towards a given pinned image and predict whether the user will take an action on it or not. Specifically, we consider two classes of signals: those of the pinned image and the user. A pin p is described by the set of visual features V_p, attributed to the content of the pinned image and profile features H_p of a user who published the image. A user u, in turn, is described by the set of her profile features H_u. We formalize the problem of predicting user attention as a classification problem where we train a binary classifier $f_1 : (p, u) \rightarrow \{repin, noaction\}$ to predict a binary output $\{repin, noaction\}$ for a given input (p, u). For the purpose of this analysis we have chosen a Random Decision Forest classifier [5] known for a good prediction performance in image classification problems [2].

6.1 Generating negative samples

One of the challenges in training a model for the system with the absence of explicit negative feedback (as there is no "dislike" button in Pinterest) is to generate realistic negative training samples. The fact that a pin was not repinned by a user does not necessarily mean they would not have liked to repin it. It might have been the case that a repin did not happen simply because the user didn't have a chance to see the pin, and had she seen, she might have taken an action on it. To account for this variance when generating negative samples, we assume that the pins which are published just before the time a user is taking an action are more likely to be noticed by the user. Thus, for a user u who took actions at times $\{t_1, t_2, \ldots, t_n\}$ we randomly select n negative samples among pins that were published in the time intervals of *one hour before the time of the actions* [6] and which were not repinned by the user. Note, that this approach is justified by the fact that over 80% of repins happen in interval of one hour since previous repin of the same image (Figure 3), suggesting that pins which have not been curated in an hour long interval are likely to be replaced on the home page (or category boards) by more recent activities and therefore would be less likely noticed by a user.

6.2 Validation

To assess performance of the proposed model we split the dataset into three consecutive time intervals: We use all pins from the first interval to learn user's preferences; all repin activity from the second one to train the model and all repins from the third one to test the model. Further, we consider two different experimental settings: when only category preferences of a user are taken into account and when both category preferences and visual object preferences are considered together. To this end, we are able to assess the extent to which Pinterest categories can capture specialisation of individual users. The results of the experiments are summarized in Table 2, and feature importances in Table 3.

Firstly, we note that, consistent with §4.2, the prediction performance is high (Accuracy of 0.66 and Precision of 0.70) even when only category preferences of individual users are considered. From Table 3 we also note that Category Preferences of users along with Deep Learning and Image Objects features are the most important to predict user attention in this scenario. However, when we add User Object Preferences performance of the prediction algorithm improves by 7-18% across all considered metrics (Accuracy, Precision, Recall, F1-Score) and, similarly, Category Preferences features are replaced by the Object Preferences in the feature importance rank. This suggests that a set of more sophisticated visual features extracted from the images that a user has previously repinned can better capture user specializa-

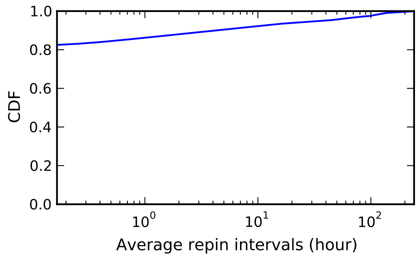


Figure 3: **Repins are concentrated in time**: CDF of average time intervals between repins shows that successive repins for a pin tend to happen quickly.

Metric	Category Prefs.	Object Prefs.
Accuracy	0.66	0.77
Precision	0.70	0.83
Recall	0.64	0.69
F1-Score	0.66	0.75

Table 2: **Performance of User Attention Prediction**: Given an image, the task is to predict whether the user will pay it any attention, i.e., will repin it or not. Two different settings are considered: when only user preferences for categories are known and when user preferences for visual objects are also taken into account.

Feature Type	Category Prefs	Object Prefs
Object Preferences	–	0.40
Deep Learning	0.37	0.32
Visual Objects	0.21	0.26
User Category Prefs	0.32	0.005
Pinner Category Prefs	0.005	0.005
Pinner Profile	0.001	0.001
User Profile	0.09	0.001
Basic Visual	0.001	0.001

Table 3: The relative importance of different classes of features for User Attention Prediction measured as expected fraction of the samples that a feature contributes to in the constructed Random Decision Forest.

[5] We used RF implementation from the *SKLearn* package with $\sqrt{n_{features}}$ split and 500 estimators (other values from 10 to 1000 were also tested, but 500 showed the best tradeoff between speed and prediction performance).

[6] We tried time windows of other sizes, ranging up to six hours before the time of repin. The precise time window size does not appear to affect prediction performance.

	features	dim	Description
Image features	**Deep Neural Network Features**		
	Deep Convolutional Network [8, 22]	4096	The deep convolutional neural network from the ImageNet [7, 22] image classification challenge. We use Caffe [18], an open-source implementation of deep convolutional networks to train an eight-layer convolutional network on 1.3 million images annotated with 1000 ImageNet classes. Then we extract 4096 features from the layer right before the final (following [8]).
	Image Objects [22]	1000	We use the deep convolutional network described above to predict the objects of Pinterest images and use them as 1000 Image Object features.
	Basic Visual Features		
	Lightness	2	Derived directly from HSL color space. The average and standard deviation of the lightness values of all the pixels in the image are derived as lightness features.
	Saturation	2	Derived directly from HSL color space. The average and standard deviation of all the pixels are used.
	Colorfulness [14]	1	A measure of a image's difference against gray. It is calculated in RGB as [14]: $\sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} + 0.3\sqrt{\mu_{rg}^2 + \mu_{yb}^2}$, where $rg = R - G$ and $yb = \frac{R+G}{2} - B$.
	Gray contrast [5]	1	It measures relative variation of lightness across the image in HSL colour space. It is defined as the standard deviation of the normalised lightness $\frac{L(x,y)-L_{min}}{L_{max}-L_{min}}$ of all image pixels.
	RMS contrast [30, 33]	1	Defined by the standard deviation of all the pixel intensities relative to the mean image intensity or $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$.
	Naturalness [16]	1	A measure of the degree of correspondence between images and human perception of reality. It is described by grouping the pixels with $20 < L < 80$ and $S > 0.1$ in HSL color space according to their H (hue) values into three sets: Skin, Grass and Sky. The naturalness score NS_i, $i \in \{Skin, Grass, Sky\}$, and the proportion of pixels NP_i are derived from the image. Then the final naturalness score is: $NS = \sum_i NS_i \times NP_i$.
	Sharpness [30]	1	A measure of the clarity and level of detail of an image. Sharpness can be determined as a function of its Laplacian, normalized by the local average luminance in the surroundings of each pixel, i.e., $\sum_{x,y} \frac{L(x,y)}{\mu_{xy}}$, with $L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$, where $\mu_x y$ denotes the average luminance around pixel (x, y).
	Sharp pixel proportion [4]	1	Photographs that are out of focus are usually regarded as poor photographs, and blurriness can be considered as one of the most important features for determining the quality of the photographs. The photographs are transformed from spatial domain to frequency domain by a Fast Fourier Transform, and the pixels whose values surpass a threshold are considered as sharp pixels (t = 2). The sharp pixel proportion is the fraction of sharp pixels of total pixels.
	Intensity balance [4]	1	It measures how different the intensity is on the left side of the image compared to the right. Two sets of histograms are produced for the left and right portions of the image. The histograms are later converted into chi-square distributions to evaluate the similarities between them, i.e., $\mid \sqrt{\sum_{i=1}^{k}(E_{left} - E_{right})} \mid$.
	Aesthetic Features		
	Simplicity-1 [23]	1	Simplicity in a photograph is a distinguishing factor in determining whether a photograph is professional or not. For a image, the RGB channels are quantized respectively into 16 different levels and the histogram (H) of 4096 bins is generated for the photographs. The simplicity feature is defined as: $(\parallel S \parallel / 4096) \times 100\%$, where $S = \{i \mid H(i) \geq \gamma h_{max}\}$ and $\gamma = 0.01$.
	Simplicity-2 [4]	1	This is the modified version of Simplicity-1. Instead of evaluate the simplicity of the whole image, Simplicity-2 extracts the subject region of a photograph and what remains is the background region. It use the color distribution of the background to evaluate the simplicity of the photograph.
	Rule of Thirds [4]	1	It is the most well-known photograph composition guideline. The idea is to place main subjects at roughly one-third of the horizontal or vertical dimension of the photograph. It is measured by how close the main subjects are placed near these "power points".
User features	**User Image Preference Features**		
	Category Preferences	32	Users preferences towards different Pinterest categories, described by the fraction of images users have (re)pinned into each category since they signed up on Pinterest.
	Object Preferences	1000	Using a week long sample of data (3-9 Jan, 2013), we calculate centroid of Image Objects of individual (re) pins in a user's history and use that to describe user's preferences in visual content items.
	User Profile Features		
	Pinboard count	1	Represents the number of personal pinboards of a user.
	Secret board count	1	Calculates the number of users' private pinboards, only accessible by the owner.
	Pin count	1	Represents the total number of pinned and repinned images.
	Like count	1	Measures the number of images a user has liked.
	Follower count	1	Accounts for the number of users who follow the user under consideration.
	Following count	1	Represents the number of users who are followed by the current user.

Table 1: List of features used for machine learning automation of predicting curation actions on Pinterest. (Dim. or dimension gives number. of scalar values in a feature).

Features	ACC
User Category Preference (**Prefs.**)	0.42
Pin **Major** Category+Prefs.	0.85
Visual features+Prefs. (∗)	0.77
Major+Visual+Prefs. (∗)	0.88
Random	0.19

Table 4: **Performance of Repin Category Prediction**: Given a user and an image repinned by her, the task is to predict which category she will repin the image into. The table shows performance in terms of Accuracy (ACC). Stars (∗) indicate χ^2 feature selection is applied to select 200 most relevant features for the classification.

tion than Pinterest categories. With respect to other considered features, i.e., profile of the user and original pinner, our results suggest their minor importance when combined with user preference and image content features.

7. CATEGORY AND BOARD PREDICTION

In this section we elaborate our model by introducing the pinboard classifier which aims to predict a user's choice of a board for a repined image. We recall that a Pinterest user may have several different pinboards each assigned to one of 32 globally defined categories. Each of the images in the datasets are thus implicitly labeled by users with one of the 32 categories. However, the choice of a repin category for individuals may also differ from the consensus vote, allowing personal preferences in image categorization. We take this into account by combining the output of the consensus vote for a pin with users' personal preferences for categories. Finally, in Figure 2b we showed that a choice between pinboards within the same category (a user may have several boards within a category) is almost trivial as most of the users (i.e., over 75%) have only one board per category and a minor adjustment to the category prediction model can be applied to account for board preferences.

7.1 Category prediction

We design a multi-class Random Forest classifier to learn which category a user will repin a given image into. Specifically, we consider two classes of signals: those of the consensus vote and the user preferences. We predict the consensus vote of a pin p using two methods:

- **Major**: Predicts repin categories according to the consensus vote as seen after the first five repins. This strategy is based on §4.1, and is expected to perform well on ≈85% of repin activities.
- **Visual**: Similar with §6, we use the visual features (including deep learning, visual object and basic visual features) extracted from images to predict the consensus vote of images.

We capture user's preferences for categories by estimating empirical probabilities $p_u(c_1)$, $p_u(c_2)$, ..., $p_u(c_{32})$ that a user u will repin an image into categories $c_1, c_2, ..., c_{32}$. We use the three sets of scores to estimate the influence of personal and consensus vote individually and also combine them.

To evaluate performance of the proposed model we split the dataset into two consecutive time intervals: We use all pins from the first interval to train the model and all repins from the second one to test the model. The results of those experiments are summarised in Table 4.

Firstly, we note that the prediction performance is high (with Accuracy=0.42) even with only user category preference, comparing with randomly selecting categories among user's categories (with Accuracy=0.19). When we add visual features and majority vote information, the performance is improved dramatically (Accuracy improved from 0.42 to around 0.80). Also the strategy based on the majority vote considerably outperforms the results of the visual features-based predictions (with 8% accuracy improvement) suggesting that even very advanced image classification techniques (which yield Accuracy@1 = 0.77) can hardly bit performance of the crowd-sourced image classification (with Accuracy@1 = 0.86). This is confirming our previous finding in §4.1 that Pinterest users exhibit high level of agreement on the taxonomy of images. Finally, we also note that the prediction performance is better when majority vote is considered in combination with individual users' preferences.

A@k	Major+Visual+Prefs.	Prefs.	Random
1	0.73	0.35	0.15
2	0.84	0.52	0.27
3	0.89	0.63	0.37
4	0.92	0.70	0.46
5	0.94	0.76	0.53

Table 5: Result of pinboard prediction as assessed by the Accuracy@k metric.

7.2 Pinboard prediction

Finally, we look at the way users select pinboards under each category. From Figure 2b we observe that the vast majority of users (75%) has only one board under each category, suggesting that in the most of the cases the problem of choosing a pinboard for a repined image is similar to that of choosing a category. Nevertheless, here we testify how the board prediction performance is changing if we account for multiple user boards. Practically, we use the Merge strategy introduced in the previous section and extended it for the board level. We compute empirical probabilities of a user to put an image into board of a category and combine them with the category prediction result to calculate the prediction score for a given pinboard.

Thus, the output of our predicting method is a list of pinboards ranked by the corresponding predicted scores. We evaluate prediction power of our method by calculating accuracy at a given cut-off K of the top predicted categories (Table 5). Formally, we define $Accuracy@K$ as a fraction of the experiments in which the ground truth pinboard was successfully predicted among the $top@K$ of the prediction list.

Comparing the ($Accuracy@1$) results of the Random benchmark between the Table 5 and Table 4 we note that pinboards prediction is just slightly more difficult problem than that of predicting categories. The results of the board prediction with the user preference alone and with all features reflect those of the category prediction with an average decrease of 10% in performance. We also note that prediction performance for the Top-5 pinboards goes over a mark of 94%, an observation which can be useful in design of board recommendation applications.

8. END-TO-END PERFORMANCE

To test the end-to-end performance of the proposed methods, we devise a cascaded-predictor which sequentially combines individual classifiers introduced in the previous sections, i.e., separately trained User Attention and Pinboard classifiers. We estimate the overall prediction performance of the system by calculating the accuracy, precision and recall of the proposed cascade predictor. These metrics are calculated as an outcome of an imaginary multi-class classifier $f : (u, p) \to \{noaction, b_1, b_2, ..., b_n\}$ where $b_1, b_2, ..., b_n$ denote users' pinboards. We also measure $Accuracy@K$, $Precison@K$ and $Recall@K$ at different cut-offs K of the $top@K$ pinboards predictions. We note that the testing set for these experiments is sampled such that the fraction of non-action and repin cases is set to 1:1, assuring that the number of positive and negative cases in attention prediction experiments are equal.

The results of the experiments presented in Table 6 suggest that the end-to-end performance remains on a high level of $Accuracy@1 = 0.68$ for the $Top@1$ pinboard prediction and further increases to $Accuracy@5 = 0.75$ for the $Top@5$ users' pinboards. Since we need to predict among multiple users' boards, we define precision and recall by distinguishing between correct or incorrect classification of a user's board (defined as true/false positives) and correct or incorrect prediction of no action (defined as true/false negatives). From Table 6, we report that the end-to-end precision remains on a level of 0.68, and reaches 0.77 for predicting among the $Top@5$ users' pinboards, suggesting an overall high level of predictability of individual curation actions on Pinterest.

	@1	**@2**	**@3**	**@4**	**@5**
Accuracy	0.69	0.71	0.73	0.74	0.75
Precision	0.60	0.70	0.72	0.76	0.77
Recall	0.50	0.58	0.62	0.63	0.64

Table 6: End-to-end results.

9. DISCUSSION AND CONCLUSIONS

Social bookmarking and curation is becoming increasingly important to the Web as a whole: Pinterest for instance has become an important source of referral traffic, second only to Facebook amongst all the major social networks [29]. Furthermore, Pinterest referral traffic is valuable for e-commerce sites, being 10% more likely to result in sales, with each sale being on average \$80, double the value of sales from Facebook referrals [15]. Therefore, understanding Pinterest can result both in fundamental insights into the nature of content curation, as well as commercially valuable applications such as recommending items that users are willing to buy, and optimising marketing campaigns for brand awareness and recall. Understanding what makes users curate an image could also help other applications such as improving the relevance of image search results.

This work takes first steps towards this research agenda by showing that although Pinterest users are curating highly personalised collections of content, they are effectively participating in a crowdsourced categorisation of images from across the web, by their choices of pinboards. By exploiting the fact that user pinboards can have an associated category, we reinterpret the act of pinning as a distributed human computation that categorises images from across the Web into the 32 categories recognised on Pinterest. When viewed through this perspective, it becomes readily apparent that there is overwhelming agreement among users on how to categorise images. Additionally, we see that users tend to specialise in a handful of categories, and tend not to have several boards in the same category. Furthermore, even within their favourite categories, their attention is skewed towards the top 1-5 categories.

Based on these observations, we developed a cascade of predictors, that, given a pin and a user, is able to predict whether the user would repin it, and if so, to which of her pinboards. The three layers of the cascade could be conceived as providing a possible mechanistic understanding of content curation on Pinterest. Although one could possible conceive of alternate mechanistic views, the behaviour and performance of the predictors we built serve to illuminate some of the various factors involved in content curation: As can be expected, the first decision of whether the user repins the pin at all, depends to a large extent on the visual features of the image, suggesting that appearances are highly important on Pinterest. In particular, we found that object features extracted using a state-of-the-art deep convolutional network yielded up to 18% improvement across all considered metrics (Accuracy, Precision, Recall, F1-score) highlighting that object recognition may play a central role in understanding what a user is interested in. The next layer in the cascade, predicting what category a user is likely to assign to a pin, is dominated by one factor: that most users agree on the category. Indeed, by looking at the first five repins, we are able to predict other repins with $\approx 87\%$ accuracy, and although several other types of features, ranging from visual features of the image to features of the user were examined, none were able to improve over this single feature, even after merging features using multiple methods. The final layer of predicting the board given the category turned out to be an almost trivial problem, underscoring that users, rather than showing complicated behaviours, appear to be "operating" just in the 32-dimension approximation of Pinterest global categories.

Because of the collective efforts of large numbers of Pinterest users, we have amassed an extensive annotated set of images (over 1.2 million category annotations for 214K pins). Although there is a great deal of agreement in general, individual users may have slightly different views about categorisation, and similarly, the categories may not be mutually exclusive (e.g., one user might classify an image of the Eiffel tower into an "Art" pinboard while another might choose "Architecture", both of which are standard Pinterest categories). We find that by incorporating user preferences, we are able to further improve the performance, evidence that users are "personalising" these categories by reinterpreting and reaggregating them through pinboards. Thus, this arrangement of allowing users the freedom to create pinboards suiting their own purposes, whilst at the same time associating the pinboards to a small number of standard categories appears to strike a good balance between the rigid taxonomies of an ontology and the free-for-all nature of so-called folksonomies [24], enabling a meaningful global categorisation (with enough power to predict image categories based on their visual features), whilst at the same time allowing user flexibility. We believe the image-based information in this dataset will be valuable in diverse other applications, and plan to make it available for researchers.

Acknowledgements

This research was partially supported by the UK Engineering and Physical Sciences Research Council via Grant No. EP/K024914/1, and by the Henry Lester Trust.

10. REFERENCES

[1] Blackburn, J., and Kwak, H. Stfu noob!: predicting crowdsourced decisions on toxic behavior in online games. In *WWW* (2014).

[2] Bosch, A., Zisserman, A., and Munoz, X. Image classification using random forests and ferns.

[3] Chang, S., Kumar, V., Gilbert, E., and Terveen, L. G. Specialization, homophily, and gender in a social curation site: findings from pinterest. In *CSCW* (2014).

[4] Che-Hua Yeh, Yuan-Chen Ho, Brian A. Barsky, and Ming Ouhyoung. Personalized photograph ranking and selection system. In *MM* (2010).

[5] Cheng, H., Zwol, R. v., Azimi, J., Manavoglu, E., Zhang, R., Zhou, Y., and Navalpakkam, V. Multimedia features for click prediction of new ads in display advertising. In *KDD* (2012), ACM.

[6] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *CVPR* (2009).

[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In *CVPR* (2009).

[8] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. DeCAF: A deep convolutional activation feature for generic visual recognition.

[9] Gilbert, E., Bakhshi, S., Chang, S., and Terveen, L. I need to try this?: a statistical overview of pinterest. In *CHI* (2013).

[10] Golder, S. A., and Huberman, B. A. Usage patterns of collaborative tagging systems. *Journal of information science* (2006).

[11] Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., and Gool, L. V. The interestingness of images. In *ICCV* (2013).

[12] Hall, C., and Zarro, M. Social curation on the website pinterest.com. *proceedings of the American Society for Information Science and Technology* (2012).

[13] Han, J., Choi, D., Chun, B.-G., Kim, H.-c., and Choi, Y. Collecting, organizing, and sharing pins in pinterest: Interest-driven or social-driven?

[14] Hasler, D., and Suesstrunk, S. E. Measuring colorfulness in natural images. In *Electronic Imaging 2003* (2003), International Society for Optics and Photonics.

[15] Hayes, M. How pinterest drives ecommerce sales. Available from `http://www.shopify.com/blog/6058268-how-pinterest-drives-ecommerce-sales`, 2012.

[16] Huang, K.-Q., Wang, Q., and Wu, Z.-Y. Natural color image enhancement and evaluation algorithm based on human visual system. *Computer Vision and Image Understanding* (2006).

[17] Isola, P., Xiao, J., Torralba, A., and Oliva, A. What makes an image memorable? In *CVPR* (2011).

[18] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* (2014).

[19] Kamar, E., Hacker, S., and Horvitz, E. Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (2012).

[20] Kamath, K. Y., Popescu, A.-M., and Caverlee, J. Board recommendation in pinterest. In *UMAP Workshops* (2013).

[21] Khosla, A., Sarma, A. D., and Hamid, R. What makes an image popular? In *WWW* (2014).

[22] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[23] Luo, Yiwen, and Tang, Xiaoou. Photo and video quality evaluation: Focusing on the subject. In *ECCV* (2008).

[24] Mathes, A. Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication* (2004).

[25] Noël, S., and Beale, R. Sharing vocabularies: tag usage in citeulike. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2* (2008), British Computer Society.

[26] Ottoni, R., Las Casas, D., Pesce, J. P., Meira Jr, W., Wilson, C., Mislove, A., and Almeida, V. Of pins and tweets: Investigating how users behave across image-and text-based social networks. In *ICWSM* (2014).

[27] Ottoni, R., Pesce, J. P., Las Casas, D. B., Franciscani Jr, G., Meira Jr, W., Kumaraguru, P., and Almeida, V. Ladies first: Analyzing gender roles and behaviors in pinterest. In *ICWSM* (2013).

[28] Rader, E., and Wash, R. Influences on tag choices in del. icio. us. In *CSCW* (2008).

[29] Rose, K. Pinterest is sneaking up on twitter, facebook, and google. New York Magazine, Available from `http://nymag.com/daily/intelligencer/2014/05/pinterest-is-sneaking-up-on-twitter-and-facebook.html`, 2014.

[30] San Pedro, J., and Siersdorfer, S. Ranking and classifying attractiveness of photos in folksonomies. In *WWW* (2009).

[31] Strohmaier, M., Körner, C., and Kern, R. Why do users tag? detecting users' motivation for tagging in social tagging systems. In *ICWSM* (2010).

[32] Von Ahn, L., and Dabbish, L. Labeling images with a computer game. In *CHI* (2004).

[33] Webster, M. A., and Miyahara, E. Contrast adaptation and the spatial structure of natural images. *JOSA A* (1997).

[34] Wetzker, R., Zimmermann, C., Bauckhage, C., and Albayrak, S. I tag, you tag: translating tags for advanced user models. In *WSDM* (2010).

[35] Yin, P., Luo, P., Wang, M., and Lee, W.-C. A straw shows which way the wind blows: ranking potentially popular items from early votes. In *WSDM* (2012).

[36] Zarro, M., Hall, C., and Forte, A. Wedding dresses and wanted criminals: Pinterest. com as an infrastructure for repository building. In *ICWSM* (2013).

[37] Zhong, C., Salehi, M., Shah, S., Cobzarenco, M., Sastry, N., and Cha, M. Social bootstrapping: how pinterest and last. fm social communities benefit by borrowing links from facebook. In *WWW* (2014).

[38] Zhong, C., Shah, S., Sundaravadivelan, K., and Sastry, N. Sharing the loves: Understanding the how and why of online content curation. In *ICWSM* (2013).