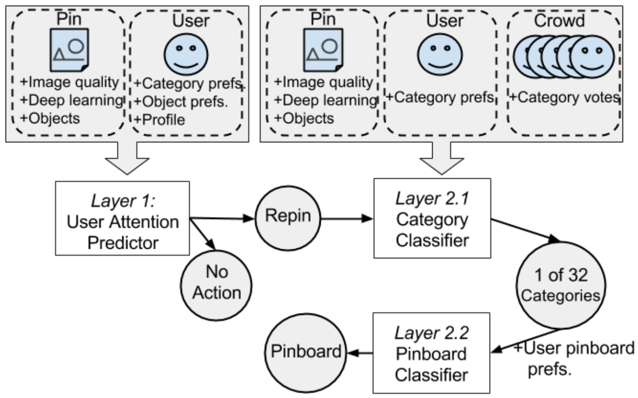


Figure 1: **Prediction cascade or pipeline to automate the manual curation of images on Pinterest**

robustly predict the category of an image. Second, users tend to specialise in a handful of categories; we use this to predict a personalised category, given the crowd "votes" for different categories. 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 can predict the specific pinboard onto which the image would be repinned, if it is known that the user has pinned an image onto some pinboard.

Thus, the remaining problem is predicting whether the user would pay any attention to an image and repin it (onto some pinboard). We use the *content* of the image 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 in the image, 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 (Figure 1). The first layer predicts whether the user will pay attention to a pin; the second predicts the category that the user will choose for the pin; and the third predicts 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%. On the easier problem of suggesting up to five possible pinboards for each pin and user, the pipeline of Figure 1 can achieve an accuracy of 75% – in other words, the pinboard manually chosen by the user appears in the top five predictions of the pipeline of classifiers.

The paper is structured as follows. §2 presents related work. §3 gives a brief background of Pinterest and our dataset. §4 demonstrates properties in the dataset that are useful to predict repins. 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 (Layer 1 in Figure 1). §7 then develops a two-stage multi-class classifier that predicts the board chosen by a user for a repin (Layers 2.1 and 2.2 in Figure 1). §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

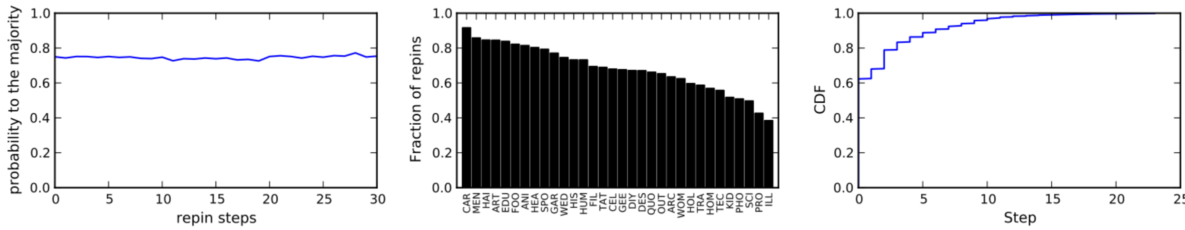
Concurrent with the rapid rise of Pinterest, there have been several studies of Pinterest as a content curation platform [11, 38, 36, 27, 8, 37, 26, 3]. Using a variety of approaches ranging from qualitative user studies [36, 38] and textual or content analyses [8, 26], to large-scale exploratory data analysis [38, 12, 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 [8, 26], the role of homophily and user specialisation [3], gender differences in activities [27, 8] and the motivations of content curators [38, 36, 12].

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.* [12, §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 *any of* the user's pinboards after 125 days (our model performs a multi-class classification into specific pinboards). Also, the best Han *et al.* models obtain an average precision of 0.015; we are able to obtain much higher values (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. This field has been well explored 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 [10] or popular [21]. Our paper contributes to this line of work, and aims to provide an understanding of how content curation happens, by collecting a large dataset of $> 200K$ images from Pinterest, and crucially, leveraging the category of the pinboards of these images to infer implicit label annotations.

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. [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. Here, we do the opposite: we use the actions of the crowd (agreement over the categories to assign a pin) to boost the performance of machine learning. Similar in spirit, but focused on a different task, is the model of Yin *et al.* [35], which takes actions of early voters as input and predicts eventual popularity of online items. An important difference is that this predicts a global trend whereas our focus is personalised predictions for individual users.

The properties of crowd agreement over item categories is a crucial element in our design. This has been studied widely in the context of social tagging (e.g., on delicious, citeulike, bibsonomy, etc.). Wetzker *et al.* [34] found that even if people share a common understanding about content, they may tag differently. It is known [31, 38] that tagging behaviour is driven by users' motivation to either categorise content for global use and easier navigation, or to describe content for personal use and to improve its searchability. Users employ different kinds of vocabularies depending on their motivation: a more compact but also a more informative vocabulary for categorisation, and a more diverse, yet more fine-grained vocabulary for descriptions [31]. Drawing on these results, we use fine-grained features, obtained by recognising objects in images,



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

Figure 2: **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).

to learn users' personal preferences, and the coarse-grained global taxonomy of 32 pre-defined categories for obtaining agreement.

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, or copying an existing pin on Pinterest. Users can also like a pin or comment on it, but pining and 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. 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). The user who introduces an image into Pinterest is its *pinner*; others who copy onto their own pinboards are *repinners*.

Dataset. Our dataset is derived from a previous study [38], and includes nearly all activities on Pinterest between Jan 3–21 2013. It was obtained by repeatedly crawling Pinterest: To discover new pins, each of the 32 category pages was visited once every 5 minutes, and the latest pins of that category were collected. 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 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 to 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 game [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 $\mathrm{maj}_i(p)$ after users $1..i$ have repinned a pin is the category with the maximum count: $\mathrm{maj}_i(p) = \mathrm{argmax}_c \, count_i(p,c)$. The final majority category $\mathrm{maj}_\infty(p)$ is the majority category after all r repins have been made. The consensus or agreement level after r repins can be com-

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

puted as the fraction of pins in the final majority category after r repins: $agreement_r(p) = count_r(p, \mathrm{maj}_\infty(p))/r$.

Whereas other curation systems (such as social tagging on delicious.com) might push users towards agreement by suggesting tags [9], 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 $P[repin\_cat(p,i) = \mathrm{maj}_\infty(p)]$, that the ith repin of a pin agrees with the final majority category chosen for it. Confirming our intuition, Figure 2a 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 (≈75%) that the category implicitly chosen by a user agrees with the majority. Figure 2b shows that the average levels of agreement can vary across pins of various categories, from 91% to 43%; and in all categories except Illustrations and Products, 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 the pinners of a pin p, starting with the first pinner as *Repin 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 $\mathrm{maj}_k(p) = \mathrm{maj}_\infty(p),\ \forall k \geq a$. Figure 2c 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

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 3 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 4 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]Users can 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 assigned to it by the original pinner, even if they can read HTML.

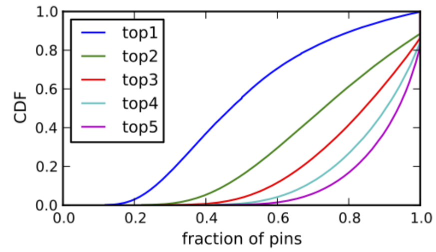


Figure 3: **Category concentration and user specialisation**. CDF of the fraction of users' pins in their top-k categories shows that each user specialises in a handful of categories.

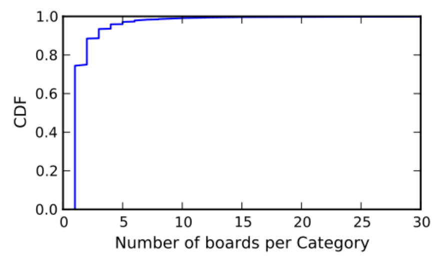


Figure 4: **Users' category choices can predict pinboard choices**. CDF of per-user number of boards per category shows that users tend not to have many boards in each category, implying that knowing the category assigned by a user, one can predict users' choice of pinboard for a pin.

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 3), 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 2, ≈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 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 2a 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 an image, whether the user will repin the image, and which pinboard it will be repinned to. In this section, we describe 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 (Figure 1). 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 recognise 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 the category c_i in which user u is going to put the repinned pin p: Formally, we build a model to predict $f_{2.1} : (p, u) \rightarrow \{c_1, c_2, ..., c_n\}$ where $\{c_1, c_2, ..., c_n\}$ is a set of user's u categories. Then in §7.2, we train a classifier $f_{2.2}$ to predict the pinboard given the category as selected by the user. Formally, we develop the model $f_{2.2} : (c, u) \rightarrow \{b_1, b_2, ..., b_n\}$. As expected from Figure 4, this turns out to be an almost trivial problem.

5.2 Features

We tackle the above classification problems by incorporating both image- and user-related features (summarised in Table 1). In addition, we also use a feature derived from the "crowd" of other users repinning the image: the consensus agreement (§4.1) around the category of the image, as measured from the first five repins, serves as a powerful indicator of the category that other users might choose for the same image.

5.2.1 User-related features

We employ several user-related features which measure both the preferences as revealed by a user's repin history, and the user's extent of involvement with Pinterest, based on her profile.

User Image Preferences.

To describe users' preferences for particular types of content, we use two sets of features: *Category Preferences* and *Object Preferences*. The first set is a relatively coarse-grained approach wherein we measure how many images a user repins in each of the 32 different Pinterest categories. Object preferences are obtained by devising user-specific signatures from the visual features of those repins: We use the state-of-the-art approach in object recognition for images [6], 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. For each image, we can derive a 1000-long vector, representing Caffe's confidence levels in recognising the 1000 ImageNet classes it has been trained on. Then, for each user in our dataset, we derive a vector of 1000 features which represents the centroid of the vectors for each of their previous repins. This can be taken as an approximation of the user's preferences for different kinds of image objects.

User Profile Features.

We also take into account the extent of the user's activity on Pinterest by measuring different statistics from their profiles: number of followers, number of boards, number of repins, etc.

5.2.2 Image-related features

Image-related features are derived based on the content of the images. A more traditional approach is to use various metrics of the *quality* of the image. This is complemented by drawing a range of features using state-of-the-art object recognition for images.

Image Quality Features.

Firstly, we define 14 image quality features (*Image Quality-I and -II* in Table 1) to describe the content of an image. These include colour model-related *basic visual features* such as lightness, saturation, colourfulness, gray contrast, RMS contrast, naturalness, sharpness, sharp pixel proportion and left-right intensity balance. We also consider three *aesthetics-related features* that have been used previously in literature: simplicity, modified-simplicity and rule of thirds. Our goal is to assess how these image quality features can capture user preferences for a particular type of content.

We note that extraction of the majority of image quality 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 image quality features, except lightness, saturation, colourfulness and naturalness for which performance was not an issue. Our experiment on a random subset of 1000 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).

Object Recognition Features.

As described above, we train the Caffe library [18] using 1.3 million images annotated with 1000 ImageNet classes [6] and apply it to classify Pinterest images. Through this process, we extract two types of visual features: (1) *Deep neural network* features from the layer right before the final classification layer. These are known for a good performance in semantic clustering of images [7]; and (2) *Recognised objects* in the image, represented by a probability or confidence level for each of 1000 objects, as obtained from Caffe.

Pinner-related features.

In addition, we incorporate user-specific features of the original pinner of the image: Features such as number of followers are indicative of the status and reputation of the pinner, and might have a bearing on whether an image is repinned. Similarly, the taste and expertise of the original pinner may also indirectly influence the repinner, and are captured by the category preferences of the original pinner, and her activity levels (number of boards, repins etc.).

5.2.3 Crowd features

In addition to the above, driven by Figure 2c, we extract a simple but powerful feature: the majority category as seen after the first five repins. We then use these to predict the user-specific categories for repins beyond the first five (§7).

	features	dim	Description
User features (U or P)	**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	User preferences for object classes as recognised by the Deep Neural Network Caffe [18] (see below for details) is computed as the centroid of the Caffe-generated object classes of all images (re)pinned by the user during a week-long training period in our dataset (3–9 Jan 2013).
	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.
Image features (I)	**Image Quality-I: Basic Visual Features**		
	Lightness	2	Derived directly from HSL colour 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 colour space. The average and standard deviation of all the pixels are used.
	Colourfulness [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 $yb = \frac{R+G}{2} - B$, and μ, σ denote mean and standard deviation respectively.
	Gray contrast [5]	1	This is measured as the 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.
	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 colour 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 as described in [16]. The final naturalness score is computed as the weighted average: $NS = \sum_i NS_i \times NP_i$.
	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.
	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 (we use a threshold value of 2, following [4]). The sharp pixel proportion is the fraction of all pixels that are sharp.
	Intensity balance [4]	1	This measures how different the intensity is on the left side of the image compared to the right. We use OpenIMAJ [13], an open-source library for multimedia content analysis, to produce the intensity histograms of the left and right portions of the image and evaluate the similarities between them using Chi-squared distance.
	Image Quality-II: Aesthetic Features		
	Simplicity-1 [23]	1	Simplicity in a photograph is a distinguishing factor in determining whether a photograph is professional. To compute this, the RGB channels are quantized respectively into 16 different levels and a histogram (H) of 4096=16^3 bins are generated for each RGB combination, and each pixel is put into the appropriate bin. Simplicity is defined as: $(\lvert S \rvert / 4096) \times 100\%$, where $S = \{i \mid H(i) \geq \gamma h_{max}\}$, where $H(i)$ is the count of pixels in the ith bin, h_{max} is the maximum count in any bin, and $\gamma = 0.01$.
	Simplicity-2 [4]	1	A modified version of Simplicity-1. Instead of evaluating the simplicity of the whole image, Simplicity-2 extracts the subject region of a photograph and what remains is the background region. The colour distribution of the background is used to evaluate simplicity as above.
	Rule of Thirds [4]	1	This is a 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".
	Object Recognition Features		
	Deep Neural Network (DNN) [7, 22]	4096	We use Caffe [18], an open-source implementation of deep convolutional neural networks [22] to train an eight-layer convolutional network on 1.3 million images annotated with 1000 classes, taken from the ImageNet [6, 22] image classification challenge. Then we extract 4096 features from the layer right before the final (following [7]), and use these as an abstract indication of objects in the image.
	Recognised Objects [22]	1000	We use the deep convolutional network described above to recognise object classes in Pinterest images and use them as 1000 Image Object features.

Table 1: **List of features used in the cascade of classifiers used to predict user curation actions on Pinterest**. The dimension (dim.) column gives the number of scalar values in a feature. User-specific features are used both to describe the user who is repinning the image (U) as well as the original pinner (P) who introduced the image on Pinterest. We also use the majority category as computed by the crowd of (the first 5) users pinning an image as a feature in §7. Image features (I) are based both on indicators of image quality, as well as object recognition using a Deep Convolutional Neural Network. User preferences among the recognised object classes is also captured as the user-specific feature "Object Preferences".

6. PREDICTING USER ATTENTION

Here we develop the first step of the cascade pipeline described in §5.1. We analyse 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. The user u is described by the set of features U in Table 1, which depend on her category and object preferences as well as statistics of her user profile on Pinterest. The pin p is described by the set of image features I in Table 1, which may be attributed to the content of the pinned image. These are augmented by a set of features drawn from the Pinner who published the image, as various characteristics of the original Pinner, such as her "taste" of images, and how influential a user she is on Pinterest, may affect repinnability of the image.

We formalise 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 the pin, she might have taken an action on it. To account for this variance when generating negative samples, we assume that pins published just before the time a user is taking an action are more likely to have been noticed by the user. Thus, for a user u who had n repin actions, we randomly select n negative samples[6], by choosing one pin published within *one hour of the time of each of the actions*, and which were not repinned by the user.

Note, that this approach is justified by the fact that over 80% of repins happen within one hour of the previous repin of the same image (Figure 5). A possible reason for this could be that pins which have not been curated recently are likely to be replaced on the Pinterest homepage (or category-specific pages) by more recent activities and would be less likely noticed by a user. We are not interested in the exact cause of this phenomenon; rather we wish to exploit it to generate negative samples of pins that may have been seen by a user but were not repinned. We also tried time windows of other sizes, ranging up to six hours before the time of repin. The precise window size does not appear to affect prediction results.

6.2 Validation

To assess the performance of the proposed model, we split the dataset into three consecutive time intervals: We use all pins from the first interval to learn users' 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. 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 (e.g. Accuracy=0.66) even when only user category preferences are considered (i.e., user object preferences are not considered). From Table 3 we also note that Category Preferences of users along with the image-related DNN and Recognised Objects features are the most important to predict user attention in this scenario. However, when we add User Object Preferences, prediction performance improves by 7-18% across all considered metrics (Accuracy, Precision, Recall, F1-Score). In this scenario, only ImageNet-related features are required, and none of the other sets of features contribute to more than 0.5% of samples. This suggests that the fine-grained object-related preferences are much more ef-

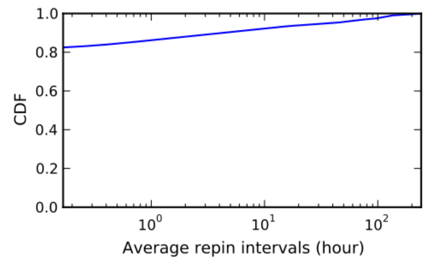


Figure 5: CDF of average time intervals between repins of an image showing that **repins are concentrated in time**.

Metric	Without Obj Prefs.	With Obj 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., whether the user will repin it or not. Two different settings are considered: when only user preferences for categories are known (i.e., without knowing user's object preferences), and when user preferences among the 1000 object classes recognisable by Caffe [18] are also taken into account.

Feature Type	Without Obj. Prefs.	With Obj. Prefs.
Object Preferences (U)	–	**0.40**
DNN (I)	0.37	0.32
Recognised Object (I)	0.21	0.26
Category Prefs (U)	**0.32**	**0.005**
Category Prefs (P)	0.005	0.005
Profile Features (U)	0.09	0.001
Profile Features (P)	0.001	0.001
Image Quality-I &-II (I)	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 Random Decision Forests constructed for the two scenarios of Table 2. Feature classes correspond to Table 1, and are ordered in descending order of importance when object preferences are used. It may be observed that user category preferences become unimportant when more fine-grained object preferences are taken into account.

[5]We used the Random Forest 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]This means there are the same number of the positive and negative samples in our training set. We also evaluate our model in imbalanced cases (with positive/negative = 10/90) and observe similar performance (accuracy and AUC).

fective at capturing user preferences than the coarse-grained category preferences.

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. In this section, we first develop a model to predict which category a user will choose, given that she will repin the image. We then refine this and predict which pinboard is used, if the category chosen by the user is known. The latter problem is trivial, as users tend to have very few pinboards per category (Figure 4). The former problem is aided enormously by deriving a straightforward category indication from the actions of the crowd (§4.1).

7.1 Category prediction

We design a multi-class Random Forest classifier[7] to learn which category a user will repin a given image into. Specifically, we consider three classes of signals:

User Because of user specialisation, we expect that most of the repin activities of the user is restricted to only a few categories, and furthermore, even amongst these categories, there may be a skewed interest favouring certain categories over others. Thus, given that the user has repinned an image, we can expect her underlying skew of interest amongst different categories to be a reasonable guess for the category chosen for the repinned image. Thus the user category preferences in Table 1 can be interpreted as the empirical probabilities $p_u(c_1)$, $p_u(c_2)$, ..., $p_u(c_{32})$ that a user u will repin an image into categories $c_1, c_2, ..., c_{32}$.

Image The decision of the repinner on which category to assign can be modulated by the features of the image, the objects in the image, and how closely the objects in the image match the interests of the user. Therefore, in this class of signals, we include all the Image features (I) (c.f. Table 1) as well as the matching object preferences of the user.

Crowd There is also a good agreement over category assignment amongst different users (c.f. §4.1). Thus, beyond any preferences that the user may have, the consensus or crowd vote on the category, e.g., as seen after the first five repins for the image, is a heuristic for the category that might be assigned by a given repinner.

As before, to evaluate performance of the proposed model, we split the dataset into three consecutive time intervals: We use all pins from the first interval to learn users' object preferences (this is common to §6 and is not repeated). We train the model based on activities in the second interval, and all repins from the last one are used to test the model.

The results of the experiments are summarised in Table 4. Firstly, we note that the prediction performance is quite high even when only using users' skew in their category preferences (Accuracy=0.42, compared with the baseline accuracy=0.19 obtained by randomly selecting a category among user's categories). When we add deep

[7]Again, we used the Random Forest implementation from the *SKLearn* package with $\sqrt{n_{features}}$ split and 500 estimators. We consider different combinations of features (as specified in Table 4), using χ^2 feature selection as indicated in Table 4.

Features	**ACC**
User	0.42
Image+User	0.77($*$)
Crowd+Image+User	0.88($*$)
Crowd+User	0.85
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.

learning-based image recognition and image quality features to modulate user preferences amongst different categories, we see a dramatic improvement in accuracy to 0.77. Further adding information about the crowd-indicated category gives us an extremely accurate model with an accuracy of 0.88.

Given that the image features we consider are based on a state-of-the-art deep learning library, it is interesting to compare the performance of image-related features with a similar signal derived from the crowd. Table 4 shows that even by just using the user preferences among categories together with crowd-derived category information, we can obtain an accuracy of 0.85 (compared with 0.77 for Image+User features), suggesting that crowdsourced image categorisation is more powerful than current image recognition and classification technology.

7.2 Pinboard prediction

Next, we look at the way users select pinboards under each category. From Figure 4 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 repinned image is similar to that of choosing a category. Nevertheless, some users may have more than one pinboard per category. To account for this, we combine the empirical probabilities of a user choosing a given category for an image (previously computed in §7.1) with the empirical probability of the user putting any given image into that pinboard (computed as the fraction of the user's images in the given pinboard).

This gives us a prediction score for each pinboard, allowing us to compute a ranked list of pinboards. 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 measure the performance using $Accuracy@K$, defined 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 in Table 5 with the random benchmark in Table 4, we note that pinboard prediction is just a slightly more difficult problem than that of predicting categories. The accuracy results of board prediction reflect those of the category prediction with an accuracy decrease of 15% (from 0.88 for category alone, to 0.73 for the more fine-grained board prediction). We also note that prediction performance for the Top-5 pinboards achieves 94% accuracy, which suggests that recommending a small set of possible boards for each pin can be a feasible implementation route for our method.

8. END-TO-END PERFORMANCE

To test the end-to-end performance of the proposed methods, we devise a cascaded-predictor (Figure 1) which sequentially combines individual classifiers introduced in the previous sections, i.e.,

A@k	Crowd+Image+User	Random
1	0.73	0.15
2	0.84	0.27
3	0.89	0.37
4	0.92	0.46
5	0.94	0.53

Table 5: Accuracy@K of predicting Top@K pinboards for pins.

the 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 $f : (u, p) \rightarrow \{noaction, b_1, b_2, ..., b_n\}$ 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 $Accuracy$ $@1 = 0.69$ 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 see that the end-to-end precision remains on a level of 0.60, 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 for the cascade of predictors (Figure 1).

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 more relevant image search.

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 alternate constructions may be possible, the behaviour and performance of the predictors we built serve to illuminate some of the 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. In particular, the fine-grained object-related preferences extracted using a state-of-the-art deep neural network were much more effective at capturing user preferences than coarse-grained category preferences, 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 of an image, we are able to predict other repins with $\approx 85\%$ accuracy, and although several other types of features, ranging from visual features of the image to features of the user were examined, only a marginal improvement of $\approx 3\%$ could be made over this single feature. We also found that this crowdsourced category was more powerful than image object recognition, at predicting personalised user category choices.

The final layer of predicting the board given the category turned out to be an almost trivial problem, suggesting 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 at `http://bit.ly/pinterest-dataset`.

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. *ICCV* (2007).

[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 *ACM 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] 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. *arXiv preprint 1310.1531* (2013). Subsequently published in ICML 2014.

[8] GILBERT, E., BAKHSHI, S., CHANG, S., AND TERVEEN, L. I need to try this?: a statistical overview of pinterest. In *CHI* (2013).

[9] GOLDER, S. A., AND HUBERMAN, B. A. Usage patterns of collaborative tagging systems. *Journal of information science* (2006).

[10] GYGLI, M., GRABNER, H., RIEMENSCHNEIDER, H., NATER, F., AND GOOL, L. V. The interestingness of images. In *ICCV* (2013).

[11] HALL, C., AND ZARRO, M. Social curation on the website pinterest.com. *Proceedings of the American Society for Information Science and Technology* (2012).

[12] 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? *SIGMETRICS* (2014).

[13] HARE, J. S., SAMANGOOEI, S., AND DUPPLAW, D. P. Openimaj and imageterrier: Java libraries and tools for scalable multimedia analysis and indexing of images. In *ACM MM* (2011).

[14] HASLER, D., AND SUESSTRUNK, S. E. Measuring colorfulness in natural images. In *Proc. SPIE 5007, Human Vision and Electronic Imaging VIII* (2003).

[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 1408.5093* (2014). Subsequently published in ACM MM 2014.

[19] KAMAR, E., HACKER, S., AND HORVITZ, E. Combining human and machine intelligence in large-scale crowdsourcing. In *AAMAS* (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 *NIPS* (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 *BCS-HCI* (2008).

[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).