Snap Decisions? How Users, Content, and Aesthetics Interact to Shape Photo Sharing Behaviors

Sanjay Kairam
Stanford University
Stanford, CA 94305 USA
skairam@cs.stanford.edu

Joseph 'Jofish' Kaye
John Alexis Guerra-Gomez
Yahoo Labs
San Francisco, CA 94103
jofish, jguerra@yahoo-inc.com

David A. Shamma
Yahoo Labs
San Francisco, CA 94103
aymans@acm.org

ABSTRACT
Participants in social media systems must balance many considerations when choosing what to share and with whom. Sharing with others invites certain risks, as well as potential benefits; achieving the right balance is even more critical when sharing photos, which can be particularly engaging, but potentially compromising. In this paper, we examine photo-sharing decisions as an interaction between high-level user preferences and specific features of the images being shared. Our analysis combines insights from a 96-user survey with metadata from 10.4M photos to develop a model integrating these perspectives to predict permissions settings for uploaded photos. We discuss implications, including how such a model can be applied to provide online sharing experiences that are more safe, more scalable, and more satisfying.

Author Keywords
Social media; photo-sharing; access controls; privacy settings; photography; user-generated content.

ACM Classification Keywords
H.1.2. User/Machine Systems: Human factors

INTRODUCTION
Participating in online social spaces lowers many of the physical, social, and geographic barriers usually associated with offline social interaction. These lowered barriers also amplify many of the challenges associated with social participation, especially around self-disclosure [26]. When information is shared offline, the individual disclosing it can usually estimate the audience as those within some physical distance; when sharing online, individuals can easily underestimate the size and composition of their potential audience [20].

Social media users often deal with concerns about self-disclosure through simple policies for sharing information about themselves and others [17, 18]. Often, the shock of regret caused by a privacy intrusion can incite adoption of more drastic policies for preventing over-sharing, including self-censoring or even complete withdrawal [28, 36, 38]. These types of coping strategies for dealing with over-sharing are especially important in the context of sharing photos, which can be particularly evocative or incriminating [3].

While the risks of over-sharing are well-documented, there are potential costs to under-sharing, as well. Sharing—photos in particular—has many observed benefits, including social interaction and enabling self-expression [6, 25, 35]. Sharing has been described by users as a critical element in their photo-management processes, both before and after the shift to managing photos digitally [6, 13]. Privacy is an important reason to limit self-disclosure, but it is just one concern that users are actively weighing against the potential benefits of sharing [11].

To aid users in making photo-sharing decisions maximizing these benefits while minimizing potential for harm, we must better understand how these decisions are made. Prior approaches have modeled photo-sharing decisions from two perspectives. One perspective has examined differences in user *dispositions*, tying these to demographics [9, 32, 33], social support [37], or both [21, 30, 31]. The other has modeled privacy decisions as *situational*, relating to the semantic or aesthetic qualities of content being shared [1, 3, 14, 19, 29].

In this paper, we analyze photo-sharing from these two perspectives; in addition, we consider the role that image aesthetics play in the selection of access controls, a topic previously unexplored in the social computing literature. Our findings provide nuanced insight into photo-sharing behaviors, uncovering site activity features that distinguish users with different high-level sharing practices and image content features which predict the use of specific permissions settings. These combined findings inform a model for predicting permissions settings for individual photos, which achieves very high accuracy ($> 94\%$) for particular subsets of photos.

Such a model has important practical implications. Several photo-management sites, such as Flickr and Google Photos, now offer automatic upload features, meaning the number of photos uploaded may easily outpace users' abilities to manually assign permissions. Imagine if, after uploading a batch of photos, the system could prompt users with subsets of photos along with suggested settings, which they could approve or dismiss with a single click.

CHI '16, May 07–12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/05 $15.00
DOI: http://dx.doi.org/10.1145/2858036.2858451

In the next section, we review some prior work relevant to sharing and privacy, particularly in the context of photos. We then present our analysis of permissions settings, which explores the separate roles played by user preferences, photo content, and image aesthetics. In each section, we operationalize our findings by identifying specific features we can mine from user and photo data, and we use these features to generate a combined model for predicting permissions settings for individual photos. We conclude by discussing observations about this model, as well as some practical and theoretical implications for our findings.

PRIOR PERSPECTIVES ON PHOTO PERMISSIONS

Prior work has modeled photo privacy decisions from two perspectives: high-level user preferences around sharing and content-level predictors for individual photos. We also address relevant literature on photo aesthetics and engagement.

User-Level Determinants of Sharing Behavior

Prior study of several social media systems has revealed that user demographics predict systematic differences in sharing behaviors. The role of gender in privacy decisions has been widely studied; female social network users have been shown to be more conscious of privacy issues [9] and more likely to maintain private [33] or friends-only [31] profiles. In the specific context of photo-sharing, women have been observed as sharing more content [21, 30], but with more restricted settings [21]. Prior work has also identified differences in sharing behaviors corresponding with age [30] and race [21].

Differences in sharing practices have also been linked to diverging user attitudes and dispositions. The perception of opportunities to share, build one's reputation, and receive social support appear to motivate users to contribute to electronic knowledge communities [37]. Prior study of Flickr, specifically, has shown that users who feel more committed to the site share more public photos, while those more oriented towards self-development share fewer [23]. An interview study by Miller and Edwards revealed two groups of users with distinct photo-sharing practices; 'Snaprs' were distinguished from 'Kodak Culture' users by increased photo organization practices, use of more professional cameras, and more sharing and interaction around photos [22].

The amount of social support that users receive offline and online also appears to moderate sharing behaviors. For individual pieces of content, increased social support appears to predict more public sharing patterns [21, 23, 30]. The awareness of a large potential audience for content, however, may cause users to be more wary of public sharing; in the context of Facebook, users with larger networks have been observed to be more likely to maintain a "friends-only" profile [31].

Content-Level Predictors of Photo Privacy

Applications for photo-sharing now provide several ways for users to manually add metadata. User-generated tags describing photo content have been used to build models predicting privacy settings with high accuracy [29]. These tags have also been leveraged to aid in the automatic construction of rule-based systems for photo access control [14].

The effort involved in manual tagging likely exceeds that required for setting permissions; it is thus appealing to consider how we might leverage photo metadata or algorithmically-generated features. Increased use of GPS-enabled cameras, such as cameraphones, means location information is increasingly available; prior research has found that, for many users, some locations are "more private" than others [1]. Other approaches have utilized content features generated through computer vision techniques. Adding SIFT visual/pixel features to a tag-based model, for instance, can provide a boost to accuracy in predicting permissions settings [29].

Computational Models of Image Aesthetics

Prior research has examined the role that image aesthetics can play in the consumption of shared photos. Engagement has been shown to correlate with crowdsourced aesthetics judgments [27] and with the use of particular image manipulations [2]. To our knowledge, however, there is no existing work specifically measuring the role that image aesthetics plays in sharing behaviors.

Joshi et al. provides a brief overview on aspects of image aesthetics, such as composition and use of color, which can be computationally identified [10]. In the context of image search, features such as good composition, shallow depth-of-field, and a uniform background predict whether users will like particular photos in search results [24]. Similar features have been shown to predict ratings of image aesthetics for specific genres of photography, such as portraiture [12].

RESEARCH QUESTIONS AND DATA COLLECTION

This section provides a brief overview of our specific research questions about photo sharing, how sharing is supported in Flickr, and the data collected to answer these questions.

Research Goals

This paper focuses on the dual roles that user-level practices and image-specific features play in guiding decisions around photo permissions. Specifically, we pose the following research questions, answered in the subsequent sections:

RQ1 What high-level patterns do we observe in user sharing practices? To what extent can these user-level patterns be predicted from other measures of site activity?

RQ2 What image content features are particularly indicative of Public, Private, or Selective sharing?

RQ3 To what extent do image aesthetics influence the ultimate choice of permissions for specific photos?

We answer these questions as part of our larger goal to construct a model for predicting permissions settings for newly uploaded photos. We attach to this goal two additional criteria. First, rather than aiming for uniformly high accuracy across all photos, our model will ideally achieve exceptional accuracy for certain subsets of users and photos. An application suggesting permissions need not trigger for every photo, but when it does, it should be correct most of the time. Second, given our motivation of preserving user privacy, we aim to conduct our analysis in a manner that minimizes the need for human viewing, coding, or labeling of non-public photos.

Permissions Settings in Flickr

Flickr is a popular site for archiving and sharing digital photos; as of June 2015, more than 100 million photographers had uploaded over 10 billion images to the site [4, 5]. Flickr supports individuals with many use cases, from casual photographers to professionals; one way in which the site accomplishes this is by offering fine-grained access controls.

Flickr allows users to mark others as *Contacts*; these connections can further be classified into two groups, *Friends* and *Family*. These labels can be applied without endorsement or reciprocation from the target user. For an individual image, there are five possible settings: Public, Friends-only, Family-only, Friends-and-Family, and Private. Flickr offers settings to allow fine-grained control over specific aspects of photos, such as who can comment or apply tags; we focus here on the permissions setting applied to viewing the photo itself.

In this paper, we have collapsed the five permissions settings into three: *Public*, *Private*, and *Selective* (encompassing Friends, Family, and Friends-and-Family). Modeling permissions in this way allows our findings to generalize better to other mechanisms for sharing with specific groups, such as Circles on Google+ or Friend Lists on Facebook.

Data Collection

To address our research questions, we used a survey and photo elicitation study to collect qualitative data offering fine grained insight into user practices around photo-sharing. In addition, we supplement our survey analysis with data collected from Flickr logs.

Qualitative Data: Survey and Photo Elicitation

We recruited active Flickr users through Amazon Mechanical Turk, screening for active users (defined as those who had joined before 2015 and who had uploaded at least 25 photos). Those who matched these criteria received a link to an online application which administered the study.

Participants. We received 100 survey responses over a 5-day period. The task was posted to workers in the United States who had completed at least 1,000 tasks with a 95% approval rating. Respondents were paid $3.00 (except for ten who were paid $3.25); workers required approximately 20 minutes on average to complete the study, resulting in an hourly average rate of $9.00. After filtering out workers who had problems with the survey application or who provided incomplete responses, we were left with 96 responses.

The sample consisted of 47 women, 46 men, and 3 respondents who chose not to identify. Most respondents (86, or 89.6%) were from the USA. Ages ranged from 18–64, with the median age range being 25–34, and the majority (77, or 80.2%) were college graduates. Most (77, not the same as above) considered themselves casual or non-photographers as opposed to serious, semi-pro, or professional.

Measures. The survey application asked users to authenticate with Flickr to provide read-only access to their photos and account statistics. 20 photos were sampled from the user's account—balanced, to the extent possible, across the five possible permissions settings. Users could exclude any photos from the study, and 10 photos were randomly selected from those remaining and presented to the user, along with questions about the content, quality, and privacy of the image. Image metadata and user responses were saved for analysis; as the study dealt in part with private photos, however, we did not store the images themselves as part of our study dataset.

Quantitative Data: Flickr Usage Logs

We collected two datasets from Flickr usage logs to help in our analysis of the use of photo permissions settings.

Users. We first identified a sample of 638,930 Flickr users who had uploaded at least one photo to Flickr in January 2015. All users met minimum activity criteria, including at least 6 months membership prior to January 2015 and 25 prior photo uploads. For each user, we collected aggregate statistics describing several aspects of their on-site behavior prior to 2015, including photo upload and organization activity, social connections and interaction, and group participation.

Photos. For each of these users, we also collected statistics for a sample of photos uploaded during the month of January 2015, providing a dataset of 10.4 million photos. Data collected for each photo included metadata, such as information pulled from the photo's EXIF data, as well as additional features generated through computer vision techniques: tags describing image content and a score summarizing the overall aesthetics. As with the survey data, image content was not stored as part of our study dataset.

USER-LEVEL PATTERNS IN SHARING PRACTICES

We start by exploring user-level patterns in sharing practices. We examine what appear to be common high-level practices and identify features of site activity which may help in predicting use of these different practices.

The Prevalence of Default Practices

We first look at the responses given by survey participants to questions about their demographics and account usage. As part of the survey, users gave us access to their Flickr credentials, allowing us to collect data about their previously uploaded photos. We recovered the full history of permissions settings usage for 75 of these 96 users; we discuss below results from analyzing data from these users.

We represent each user's aggregated sharing activity using a vector representing the probabilities of assigning photos to each possible setting (Public, Private, or Selective). These probabilities are illustrated below in Figure 1. As the distributions of these probabilities were observed to be non-normal, we used non-parametric tests to examine relationships between the use of each setting and the demographic variables for which we collected information in the survey.

Using a Kruskal-Wallis rank sum test for each permissions setting, we observed no significant relationships between use of each setting and self-reported gender. As age, education, and photography expertise were reported using ordinal scales, we assessed the probability of significant relationships between these variables and use of each permission settings using Spearman Rank correlations (ρ).

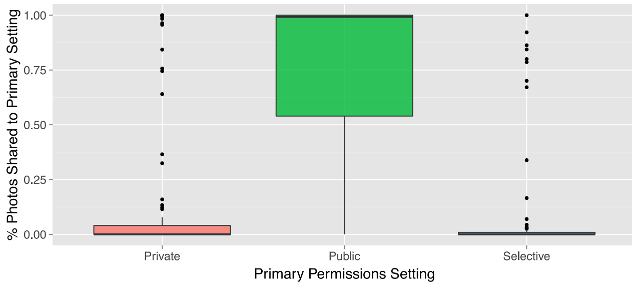


Figure 1. For each survey participant, we captured aggregated counts for permissions settings across all previously uploaded photos. This figure illustrates the distributions of permissions settings usage across users. For example, the percentage of photos uploaded as Selective by the average user is close to 0. We observe that most users upload all or most of their photos using a single setting.

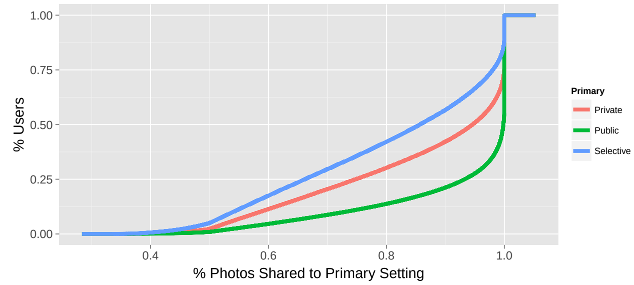


Figure 2. This figure illustrates the use of default practices for the 639K Flickr users in our sample. Users are divided according to their most used (*primary*) setting. For each line, the y-value shows the proportion of users who share at most $x\%$ of their photos using that setting. For example, if we look at where the green line marked 'Public' crosses $x = 0.9$, we see that about 15% of users who adopt 'Public' as their primary setting upload fewer than 90% of their photos using this setting.

We observed relationships between respondents' self-reported age and their use of the Public ($\rho = -0.25, p < 0.05$) and Selective ($\rho = 0.32, p < 0.005$) settings, indicating that older users shared a smaller proportion of photos as Public and a larger portion using the Selective options. We also observed a strong positive relationship between self-reported education and use of the Selective setting ($\rho = 0.538, p < 0.01$). We observed no relationships between self-reported photography expertise and aggregate permissions setting use.

In analyzing the data, however, a clear pattern emerged: users tended to favor heavily one setting—which we call the *primary* setting—over others. 43/75 of users surveyed (57.3%) assigned their primary setting to over 99% of their photos; in cases where such a large proportion of photos are uploaded to a single setting, we refer to to this as a *default practice* (which we distinguish from the actual default upload setting which some users may set using the web interface). In order to assess the prevalence of default practices among the larger population of Flickr users, we analyzed permissions distributions collected for the 638,930 users described earlier.

Figure 2 shows empirical cumulative density plots summarizing the extent to which users have adopted default practices around each permissions settings. This chart can be interpreted as follows: the line marked "Public" represents all users for whom Public is their primary setting. If we choose a point along the x-axis, then the corresponding y-value represents the proportion of these users who upload at least $x\%$ of their photos using the Public setting. We can interpret the lines marked "Private" and "Selective" in a similar manner.

The sharp increase in each line towards the right of the graph illustrates the prevalence with which users adopt these default practices. We specifically observe that 318896/638930 (49.9%) of users upload at least 99% of their photos to their primary settings, showing that the behavior we observed among survey users is fairly typical. We further observe that 231740/638930 (36.2%) of users upload *all* (100%) of their photos using a single setting. Users who adopt default practices are also more likely to use Public as their primary setting. For all users in our sample, 65.2% chose the Public option most often for their photos. For users with default practices, 78.3% used the Public option most often.

Default Practices: Intentional or Not?

For some participants, there is good reason to believe that heavy reliance on a default practice is completely intentional. To some extent, this is confirmed by responses from the survey given to the prompt: *Please describe any general policies or practices you might have when it comes to choosing access controls for your photos on Flickr.*

Some users demonstrate clear motivations to share photos widely. P28 [Public, 100%] responded *"All my photos were set to public view and I released my rights to the public domain. Flickr was, first and foremost, about sharing my photos with the world."* P41 [Public, 100%] stated *"I want to make everything I post to flickr publicly available for noncommercial use. I try to post high quality photos with relevant tags and minimal information about my home address or family details."* P85 [Public, 100%] said *"I don't care who views my photos but I certainly care how my photos are used. So I make sure the copyright settings are at maximum."* These users are not just aware of their practices; they are savvy about licensing and legal issues related to sharing online.

For other users, there is reason to believe that these default practices stem from convenience. Many participants in the survey described self-censoring behaviors, or limiting what they upload to prevent photos from being seen by unintended parties. P59 [Public, 100%] stated *"I only share only what I would be comfortable with anyone seeing, so there's no reason for me to limit access."* P13 [Public, 100%] similarly stated *"I don't put anything on the internet that I don't want others to see."*

Some users' responses indicated potential ignorance about how photos are being shared, such as P5 [Public, 100%] *"I usually make sure everything is friends only"*, P43 [Public, 100%] *"Always private"*, and P36 [Selective, 92%] *"Most of the time I don't find the time to set privacy for each photo, so many of them might default to the most private setting."*

Discussion

In this subsection, we observed that roughly 50% of users tend to use the same setting for sharing most or all of their photos. These findings about sharing echo observations about the usage of filters for consuming content, namely that many users tend not to change their practices or to do so infrequently [16]. Survey respondents' descriptions of their sharing practices echoed findings from prior work on other types of social media systems about self-censoring [28, 38] and balancing reasons to limit sharing against motivations to distribute content more widely [11].

One practical consequence of our findings is that for certain sub-groups of users, it may be possible to predict future permissions settings with extremely high accuracy. However, as observed, some users may be sharing mistakenly in a way that differs from their preferences. In the next subsection, we examine the extent to which we can predict users' preferred sharing practices from other indicators of on-site behavior.

Predicting Permissions Patterns from Site Behavior

We start by identifying characteristics of on-site behavior which may distinguish users who adopt different default practices for sharing. We incorporate these into a model for predicting the use of these different practices. By predicting permissions practices from other measures of site behavior, we may be able to identify users whose predicted patterns of activity differ from their actual practices.

Behavioral Features

Our review of the literature on user-level differences in sharing behaviors pointed to several classes of behavioral features which may distinguish users with different overall sharing practices. We specifically choose 25 variables which cover various aspects of participation in Flickr but likely generalize to other social photo-management platforms.

User Tenure. Prior study of Flickr users has shown that photo-sharing frequency declines as site tenure increases [23]. We measure the time since a user's first upload as, `Member Age`.

User Contribution. One feature which distinguished 'Snaprs' from 'Kodak Culture' was the frequency of content creation [22]. We characterize direct user contributions through `Monthly Photo Uploads` and `Upload Frequency` and indirect contributions through `Monthly Outgoing Comments` and `Unique Comment Recipients`.

Camera Usage. 'Snaprs' also utilized more professional cameras than 'Kodak Culture' photographers [22]. We inferred camera type using photo EXIF data to generate three features: `Proportion Point & Shoot`, `Proportion DSLR`, and `Proportion Cameraphone`.

Photo Organization. 'Snaprs' organized their photos more actively, as well [22]. Photo-sharing sites such as Flickr offer multiple ways for users to organize photos, including albums for browsing (`Album Count`) and tags for retrieval (`Tags per Photo`, `Proportion Photos Tagged`).

Social Connections. Previous findings from several studies emphasized the complex role that one's social network may play in guiding sharing behaviors [21, 23, 30, 31]. We capture a user's network size through `Contact Count` and network composition through `Proportion Friends`, `Proportion Family`. We also capture closeness of ties using `reciprocal` relationships (the proportion of a user's contacts who also designated the user as a contact, as Family, or as a Friend).

Social Photo Interaction. 'Snaprs' were also observed as interacting more around photos [22]. We characterize basic incoming social interaction through `Views per Photo` and `Faves per Photo`. We capture incoming commenting behavior through `Incoming Comments per-Photo`, and `Unique Commenters`. We note that even private photos can receive views, faves, and comments, since photos can be made available to others via secondary mechanisms, such as share links and Flickr groups. We will revisit this issue in our discussion of the model results.

Group Participation. Like many content-sharing sites, Flickr provides opportunities to engage with users outside of one's defined contacts. We capture community participation with `Groups Joined`, `Average Group Size`, `Proportion Groups Public`, and `Proportion Groups Founded`.

Identifying Users with Default Practices

We now explore the extent to which these behavioral features can help us to identify users who adopt default practices ($>99\%$ photos shared to a particular setting). Our motivation here is two-fold. First, identifying these users effectively means identifying groups of users for whom we can likely generate highly accurate predictions about future permissions settings. Second, understanding what behavioral features are associated with users in these different categories could lend insight into the types of users who adopt each.

We constructed three logistic regression models using the 25 features proposed above. All variables representing counts and averages (except for `Member Age`) were log-transformed due to the non-normality of the distributions, and all variables were re-scaled to have a mean of 0 and a standard deviation of 1. For each model, the outcome predicted is whether a user shares 99% or more of his or her photos using the corresponding setting. While creating balanced samples might improve prediction accuracy, we train each model using all 638,930 users; logistic regression is robust to class imbalances unless one class is exceedingly rare, and including all users better approximates the proposed applications of the model.

We computed the reduction in deviance for each model relative to a null model, and use chi-square statistics to assess statistical significance. For the Default-Public model, we have ($\chi^2(271806, 25), p < 2 \times 10^{-16}$), for the Default-Private model, we have ($\chi^2(206308, 25), p < 2 \times 10^{-16}$), and for the Default-Selective model, we have ($\chi^2(27599, 25), p < 2 \times 10^{-16}$), meaning that all models provide significantly more explanatory power than the null models. The regression coefficients for the three models are summarized in Table 1. We highlight some observed patterns below.

We observe that the members who rely on each of these "default practices" are newer to the site than those who do not, with the strongest effect for the Default-Private strategy. It is possible that certain strategies, such as self-censoring by keeping all photos private, are adaptive for users who are newer and less familiar with the site.

Variable	Public		Private		Selective	
	β	p	β	p	β	p
(Intercept)	-0.94	***	-4.89	***	-4.69	***
Member age	-0.26	***	-0.51	***	-0.01	–
Monthly uploads	-0.06	***	-1.33	***	0.74	***
Upload frequency	0.11	***	0.03	–	-0.14	***
Monthly comments	-0.26	***	0.49	***	-0.17	***
Unique recipients	0.01	–	-0.32	***	-0.02	***
Proportion P&S	-0.01	***	-0.03	***	0.04	***
Proportion DSLR	0.01	***	-0.02	–	0.05	***
Proportion phone	-0.63	***	0.09	***	-0.22	***
Album count	-0.33	***	0.07	***	0.06	***
Tags per photo	0.00	–	-0.20	***	-0.05	***
Proportion tagged	0.14	***	0.02	***	0.01	–
Contact count	-0.28	***	-0.11	***	0.16	***
Proportion friends	-0.07	***	-0.11	***	0.20	***
Proportion family	-0.31	***	-0.20	***	0.40	***
Reciprocal contacts	0.08	***	0.02	–	0.06	***
Reciprocal friends	-0.06	***	0.02	–	0.04	***
Reciprocal family	-0.12	***	-0.04	***	0.08	***
Views per photo	1.63	***	-1.77	***	-0.11	***
Faves per photo	0.35	***	-2.22	***	-0.59	***
Comments per photo	-0.09	***	-0.21	**	1.56	***
Unique commenters	0.01	–	0.67	***	-1.63	***
Group count	-0.22	***	0.91	***	-0.00	–
Average group size	-0.17	***	-0.07	**	-0.09	**
Proportion public	0.25	***	-0.19	***	-0.34	***
Proportion founded	-0.13	***	-0.10	***	0.04	***

Table 1. β coefficients for the three logistic regression models categorizing users as Default-Public, Default-Private, or Default-Selective. (* : $p < 0.01$), (: $p < 0.005$), (*** : $p < 0.001$). The sign of the β coefficient indicates the direction of the relationship. For example, users who take more photos on camera phones are more likely to adopt Default-Private and less likely to adopt Default-Public or Default-Selective strategies.**

Default-Public users upload more photos using DSLR's and fewer photos using Point & Shoot or phones. Default-Private users upload more photos from phones. These findings match our expectations that users with more "professional" cameras will adopt more public sharing patterns. The results are mixed regarding photo organization. Members who adopt a Default-Public strategy make fewer albums, but tag a higher percentage of their photos than those who do not. Members who adopt Default-Private and Default-Selective strategies engage in both types of organization for more of their photos.

Larger contact networks are associated with increased likelihood of adopting a Default-Selective strategy and decreased likelihood of adopting a Default-Private or Default-Selective strategy. Individuals with larger networks are likely to cultivate them, in part, with the goal of using the available selective sharing features. Observations about incoming interaction around shared photos are mostly as expected, given that public photos are more readily available to a larger audience. We do note the interesting observation, however, that users who adopt Default-Selective strategies receive more comments per photos, even with their limited audience.

Finally, we observe an interesting pattern in the use of groups by these different types of users. Joining more groups decreases the likelihood of adopting a Default-Public strategy while increasing the likelihood of adopting a Default-Private strategy, possibly running counter to our expectations. It is possible that groups represent a secondary mechanism for managing selective sharing by allowing users to make photos accessible to particular subsets of users. Interestingly, founding more groups makes users less likely to adopt a Default-Private strategy. In this case, typically-private users may be seeking out existing groups of users who share their interests.

Logistic regression models provide probabilities of assigning a user to the negative or positive class; by choosing different thresholds, we can assign more or fewer users to each class, thus obtaining different accuracy scores. Measuring the area under the receiver-operator curve (AUC) provides a means of summarizing accuracy across different threshold choices. We assess model performance by measuring the area under the receiver-operator curve (AUC), which can vary from 0.5 to 1, obtaining: $\text{AUC}_{\text{Public}} = 0.856$, $\text{AUC}_{\text{Private}} = 0.959$, and $\text{AUC}_{\text{Selective}} = 0.859$. These scores indicate good to excellent performance in separating out users who adopt default practices from the rest of the population.

Discussion

We have identified sub-groups with predictable sharing patterns covering 50% of the user population; we have further shown that we can identify these users effectively using features mined from other aspects of site behavior. For these users, we can likely generate, with high accuracy, predictions for all new photos without much additional information. Some users mis-classified by the model may actually represent users who are sharing in a manner which differs from their desired settings; we could consider prompting these users to ensure that they are sharing as intended.

We noted some interesting observations from the model. Default-Selective users receive more comments per photos, though their photos are available to fewer users overall. We similarly observe an association between group participation and adopting a Default-Private strategy. These users may be sharing selectively, using personal lists or interest groups, as a strategy to ensure that interested parties see more relevant content and thus engage more, as observed in prior study of selective sharing motivations [11].

A potential concern with this model is the inclusion of the social interaction variables—faves, views, and comments; these are surely influenced by the chosen permissions setting for a particular photo. As mentioned previously, however, photos can be shared on Flickr via additional mechanisms, such as share links and groups. Versions of these models generated without these four features performed similarly, with AUC > 0.8 for each. Furthermore, our observation that Default-Selective users receive more comments per photo highlights that the use of more restrictive permissions settings does not completely dictate the amount of attention that a photo receives.

PHOTO-LEVEL PATTERNS IN SHARING DECISIONS

We now move to photo-level features which guide specific sharing decisions. We first consider how permissions settings vary according to the semantic content in a particular image, and then consider the role played by image aesthetics.

The Role of Image Content

In this section, we explore the extent to which particular types of algorithmically-identifiable content may correspond with the use of certain permissions settings. Finding reliable associations between content and permissions would enable much easier classification for certain types of images.

Leveraging Algorithmically-Generated Tags

We discussed previously approaches which made use of user-generated tags to generate predictions about privacy settings [14, 29]; we observed, however, that manual tagging likely required as much attention and effort as choosing privacy controls.

Popular photo-sites, such as Flickr and Google Photos, have started adopting computer vision techniques for automatically generating tags describing the content of images. Flickr, specifically, has applied automatic tagging to all photos in its corpus [34] using a method building on convolutional neural networks [15] and trained entirely on public Flickr photos.

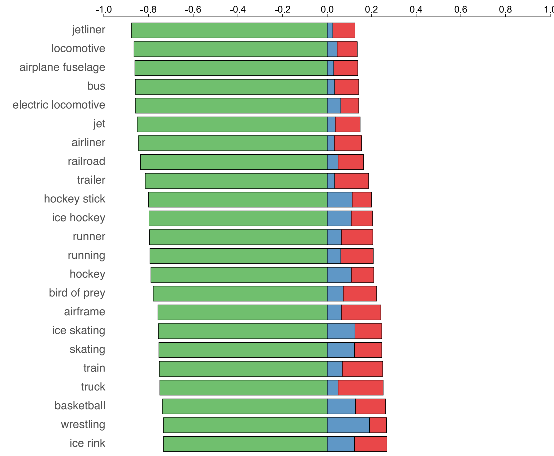
Identifying Tags Associated with Different Settings

We started by collecting automatically-generated tags for each of the 10.4M photos in our dataset. Using this dataset, we were able to take each of the 1,712 unique auto-tags which were applied to these photos and generate a frequency distribution across photos. Here, we discuss observations regarding the 499 auto-tags which were applied to at least 0.1% of our sample of photos.

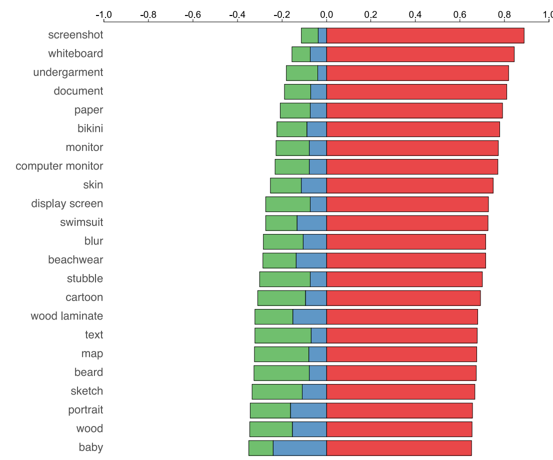
The 10 most common automatically-generated tags for photos in our sample were, in descending order: `people`, `face`, `outdoor`, `indoor`, `groupshot`, `nature`, `monochrome`, `blackandwhite`, `sport`, `portrait`. In this list are many of the same tags previously reported as most common in public photos [34]. We do see that photos of people (e.g., `people`, `face`, `groupshot`, `portrait`) and tags representing stylistic features (e.g., `monochrome`, `blackandwhite`) are more prevalent, indicating some of the trends we might see when shifting our focus to photos uploaded as Private or Selective.

Figure 3(a) shows tags ranked by the likelihood that a photo to which they are applied is public. Here, we see a few themes emerge. This ranking illustrates several types of photos which may be particularly likely to be shared publicly. One category is photos of large vehicles (e.g., `jetliner`, `locomotive`, `airplane fuselage`, `bus`). For photos tagged with any of these keywords, we could predict with fairly high certainty ($\geq 85\%$) that the public setting will be used. Another category indicative of public sharing is sports (e.g., `hockey stick`, `ice hockey`, `runner`, `running`).

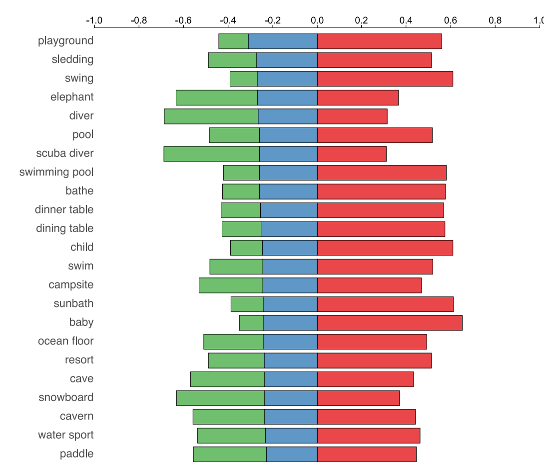
Figure 3(b) shows tags strongly associated with private photos. Unsurprisingly, we see a category associated with photos which may be revealing (e.g., `undergarment`, `bikini`, `skin`, `swimsuit`). We also observe another category of likely private photos that we might not have anticipated, one dealing with screenshots and images of text (e.g., `screenshot`, `whiteboard`, `document`, `paper`). For these two classes of photos, we could likely predict with high accuracy that the user intends to apply the private setting.



(a) Tags ranked by the likelihood that photos will be set as Public. Themes in the top tags include large vehicle and sports imagery.



(b) Tags ranked by the likelihood that photos will be set as Private. Themes in the top tags include revealing or text imagery.



(c) Tags ranked by the likelihood that photos will be set as Selective. Top themes include family activity or home imagery.

Figure 3. Distribution of permissions settings for photos tagged with specific computer-generated tags. Each chart shows tags ranked by the likelihood that photos with those tags will be shared with each setting. Bars from left to right represent Public, Selective, and Private settings.

Finally, in Figure 3(c), we see the tags which best predict use of the selective setting. Unsurprisingly, most of these tags have to do with children or family activities. Photos auto-tagged with `playground` or `sledding`, for instance, are assigned the selective setting at a far higher rate ($> 30\%$) than average. We also observe several keywords which indicate photos taken within the home (e.g., `dinner table`).

Discussion

We have identified commonly-applied computer-generated tags which are predictive of different permissions settings. This investigation provides insight into the types of photos that users take and how their intended uses for these photos may vary according with the content. The Selective sharing of family photos and Private sharing of physically revealing photos both match intuitions about common use cases for these settings. The tendency for vehicle and sports photos to be marked as Public may indicate, for instance, photographers who are sharing photos in interest groups. The observed behavior of privately archiving screenshots and images of text content may be higher than expected, possibly representing an under-served use case for photo-sharing sites.

The Role of Image Aesthetics

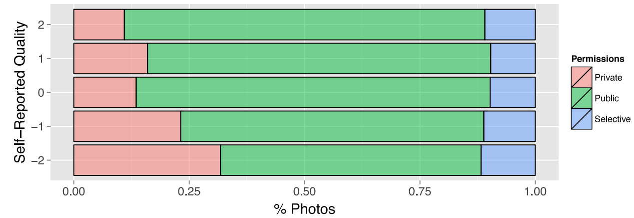
Frequently, we upload multiple photos of an event or even of a specific shot, leaving us with a cluster of images representing similar content. In many cases, the decision about which of these photos to share relates to the quality of the images and how they reflect on our skills as a photographer. In this subsection, we first examine the extent to which users' self-assessments of image quality predict their choice of permissions settings. We then explore the feasibility of using computer-assessed aesthetics scores as part of a model for predicting permissions settings.

Self-Assessed Quality and Permissions

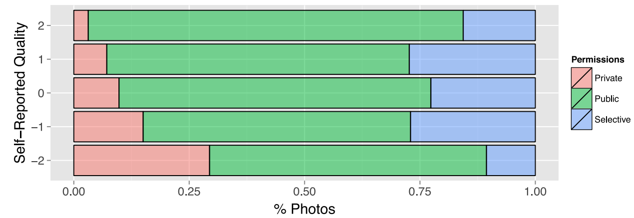
In order to explore the extent to which this is true more generally for Flickr users, we again looked to data from the photo-elicitation portion of our survey study. For each photo, we asked participants to provide a self-assessment of the quality of each photograph relative to others they have taken.

Specifically, they used a 5-point Likert scale to indicate their level of agreement with the following statement: *This image would rank among the higher quality photographs I have taken (in terms of composition, clarity, brightness, etc.)*. In addition, rather than showing participants the actual permission setting for these recently-uploaded photos, we asked them to guess which setting had been used, allowing us to compare these guesses against the actual chosen setting.

Figure 4(a) illustrates how actual permissions settings chosen for photos vary with users' self-assessed quality judgments. A χ^2 test of independence was performed to examine the relation between this self-assessed quality judgment and the actual privacy setting (again simplified to Public, Private, and Selective), revealing a significant relationship, $\chi^2(8, N = 916) = 24.02$. We observed that, as self-assessed quality increases, the choice of setting shifts from Public to Private. The probability of assigning a photo as Selective remains constant as self-assessed quality changes.



(a) Actual permissions settings by self-assessed quality score.



(b) Guessed permissions settings by self-assessed quality score.

Figure 4. Distribution of actual and guessed permission settings by self-assessed quality score. Most of the highest quality photos (score = 2) are shared publicly; as quality decreases, an increased proportion are shared privately. This trend is the same for guessed permissions settings, indicating the effect may be robust over time.

Figure 4(b) similarly illustrates how the guessed permissions settings vary with the same quality judgments. A χ^2 test revealed a similar significant relationship between self-assessed quality and the privacy setting that users guessed for photos, $\chi^2(8, N = 916) = 49.50$. Trends regarding estimated use of the Public and Private setting were similar to those for actual use, though amplified somewhat. We observe an interesting pattern where photos perceived to have slightly lower or slightly higher quality than average were more likely to be shared selective.

Computer-Assessment of Photo Aesthetics

Because it would be difficult to collect quality assessments from users about all of their photos, we were interested in exploring the feasibility of using computationally-generated aesthetics judgments to help model photo quality. Using a method similar to the one used to generate automatic tags, Flickr has also developed a model for photo aesthetics, which assigns each image a score from 0–1. This model was originally trained on a dataset of public photos, allowing us to infer aesthetic judgments over all photos without the need for human viewing or labeling of non-public photos.

We collected aesthetics scores and permissions settings for 9.95M of the photos in our dataset. As the distribution of aesthetics scores was bounded and non-normal, we again utilized a Kruskal-Wallis rank sum test. We observed a significant relationship between this computer-generated aesthetic judgment and the choice of permission setting assigned by the photo owner ($H(2) = 159690, p < 2 \times 10^{-16}$).

Figure 5 illustrates how the distribution of aesthetics scores (visualized using a square-root transformation) varies across permissions settings. We can observe a clear pattern in which Public photos are assessed as more aesthetically pleasing than Private or Selective photos.

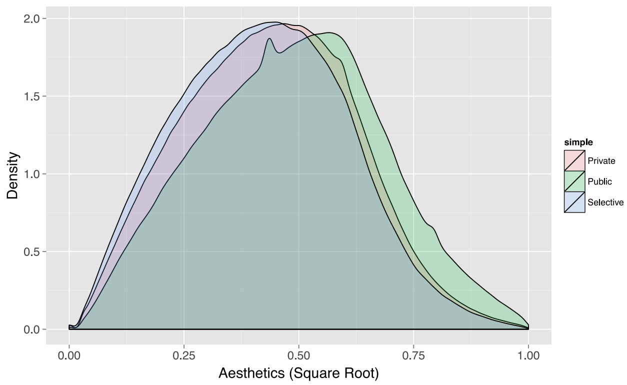


Figure 5. Distribution of aesthetics scores for photos shared with each permission setting. We observe that the distribution of Public photos is shifted towards higher aesthetics scores than Private or Selective. A square-root transformation is used to help the reader see the differences, but the relationship is significant with or without this transformation.

Discussion

We observe differential usage of permissions settings as a function of self-assessed photo aesthetics judgments, for both actual and perceived permissions setting usage, implying a persistent influence on sharing behavior. We also explored the feasibility of using algorithmically-assessed photo judgments, observing that publicly shared photos have significantly higher aesthetics scores. This finding indicates that computer-assessed photo aesthetics could provide a useful feature in identifying cases in which users deviate from their usual practices in choosing permissions settings.

DEVELOPING A MODEL FOR PHOTO SHARING

In the previous sections, we identified features summarizing user behavior, image content, and aesthetics which may be predictive of permissions choices for individual photos. All of the features identified can be mined from user's site behavior or algorithmically computed from the user's images.

Method

Our model uses a decision tree classifier, which learns simple rules in order to separate the data (in this case, photos) into increasingly small buckets. For each bucket, the aim is to find a rule which best distinguishes between classes for the items in that bucket; these items are further split into two smaller buckets using this rule. This approach provides simple, human-understandable classification rules. In addition, it allows us to identify subsets of data for which we can make particularly accurate predictions. We limit our tree to 5 levels, which allows data to be separated into 32 different paths.

Our model looks at 101 features for each photos, comprising 25 user-level features (the behavioral features identified earlier), 75 content features (binary features for the top public-dominant, private-dominant, and selective-dominant tags, 25 each), and the aesthetics score. Our data consists of photos taken in January 2015 by the 639K users previously analyzed. We restrict our analysis to photos which have at least one of the tags in the feature set, and we sample to ensure we have exactly one photo per user, leaving us with a set of 67,211 photos [52.2% Private, 35.5% Public, and 12.3% Selective].

While overall accuracy is not our goal, we establish two baselines. The first, predicting the majority-class or assuming that every photo is private, achieves 52.2% accuracy. The second baseline is provided by our survey participants. If the reader recalls, we asked participants to guess the permissions settings *for their own photos*. Overall, they achieved 67.1% accuracy; we set this as our second, more ambitious baseline.

Results

We performed cross-validation by running the model 5 times, each time training on 90% of the data and then computing performance statistics using the 10% which was held-out for testing. The model achieved 73.2% accuracy, on average, beating the two baselines we established. We then trained a final model on the full data sample.

Using the model trained on all the data, we were able to follow paths of decision rules to find subsets of photos for which we could predict the permissions setting with particularly high accuracy. Specifically, we identified sub-groups totaling 12,547 photos (18.6%) across which we could predict permissions with 94.2% accuracy.

Using rules which contributed to this high-accuracy set as examples, we discuss some of the types of rules learned by the classifier below. We use the terms 'low' or 'high' in discussing variable values in reference to the average value, μ, for proportion variables and $\log(\mu)$ for count variables, following the transformations discussed earlier. For tags, rules capture the presence or absence of a tag, and for the aesthetics score, the rules reference the raw score.

Rules incorporating only behavior

Some of the rules demonstrated what we observed earlier; some photos can be classified accurately using just behavioral features describing the users who took them. The tree separated out photos taken by users whose had used cameraphones less frequently and whose photos had received fewer photos in the past; of these photos, 94.6% were private.

Rules combining behavior and content

Some rules made predictions about photos based on the intersection of user behavior and content features. One such rule separated out photos taken by users who received more views per photo than average and who had a higher proportion of contacts marked as Family. While these photos were difficult to classify using only user-level information, we could categorize them as Private with 92.7% accuracy if we knew they had been automatically tagged with tags *text* and *screenshot*.

Rules combining behavior, content, and aesthetics

Finally, we observed some rules which combined all three types of features. For users whose photos received more views on average and who used a cameraphone at least moderately, we could predict that a photo that they uploaded was public with 93.7% accuracy if it received an aesthetics score ≥ 0.1785 and wasn't tagged as a *portrait*.

DISCUSSION

Our initial motivation for conducting this investigation into photo-sharing behaviors was to build a model capable of predicting photo permissions. We have developed a model which combines features corresponding to user behavior, photo content, and image aesthetics. For our sample of photos, this model performs relatively well in predicting settings for new photos based on past behavior. This model achieves higher overall accuracy, in fact, than users in our survey did when guessing the settings that they had recently applied to their own photos.

The models used to compute the tags and aesthetics scores for images were both trained on datasets of public photos. Taken together, this means that every component in the model we present below has been developed without human labeling of privately or selectively shared photos.

Design Implications

In the beginning of the paper, we imagined an application which could suggest permission settings for a batch of recently uploaded photos. We observed that for such a system, it may be preferable to identify subsets of users or photos for which accurate predictions can be made than to try to generate predictions for all photos. Our final model gives us some latitude in building such a system.

We have identified rules which allow us to isolate specific users who upload most of their photos to a single setting. Our system might consider prompting these users to toggle a default setting for all newly uploaded photos, if they haven't done so already. Given our observation that some users are sharing with a setting different from the one intended, the system could also provide a check for these users.

We have also identified rules which combine user behavior and photo content. In these cases, when we see particular users upload particular types of photos, we could provide targeted recommendations for permissions settings for these photos, which users could accept or reject. For large batch uploads, this could eliminate the large amount of work required to manually search through photos and toggle settings.

Finally, we have identified rules that represent the intersection of user practices, photo content, and aesthetics. Photographers often take multiple shots of a scene or an event; with increased use of photography apps and editing software, users are now creating multiple versions of their images before deciding to share, applying transformations and filters which they hope will make the photo more engaging [2]. Auto-uploading all of these versions means that sharing decisions can quickly multiply. We imagine a scenario where an algorithm could potentially sift through a set of photos with similar content from a user and make suggestions about which to share by incorporating aesthetic judgments.

Theoretical Implications

Flickr is far from a typical social media system; in contrast with text communication, photos must be captured, kept, uploaded, and shared before they are attended to by others. In this way, sharing photos is more like planning an exhibition (as described in Hogan [8]) than an improvised performance (as described in Goffman [7]). Studying photo-sharing in Flickr provides a useful comparison with the extensive existing work on sharing in more traditional systems, such as Facebook, which mix these types of social performances. We observe that even in the context of these structured exhibitions, many users still manage access controls using broad strokes rather than fine-grained strategies.

The specific features used in our models draw on findings from several different areas related to sharing, privacy, and social networking. Our models of user practices around permissions settings, for instance, distill findings from several small-scale, mostly qualitative studies [21, 22, 30] into specific, quantifiable metrics and demonstrates that these can be used to effectively predict differences in sharing behaviors.

Future Work

While there are many directions we would like to take this work, there are two which clearly emerge from our findings. First, for users who adopt default practices, we expressed some uncertainty about the extent to which these practices are intentional. We observed some specific cases from the users that we surveyed in which their actual practices differed from their stated preferences. In future work, we would like to see how users engage with interfaces summarizing their sharing patterns, and perhaps comparing them to predicted sharing patterns aggregated over similar users.

The second area we would like to explore concerns secondary means of selectively sharing content. At least in the case of Flickr, interest groups can provide a secondary means of sharing photos with a wider audience without making it completely public. Another similar mechanism is share links, used in systems such as Google Docs as well as Flickr, to provide specific individuals access to content. Fully understanding selective sharing practices will involve studying the full ecosystem of options for providing access to content.

CONCLUSION

In this paper, we have analyzed behavior around photo sharing from three perspectives: user practices, photo content, and image aesthetics. We found that many users share most of their photos using a single permission setting, and that we can identify these users from aspects of their site behavior. We uncovered particular aspects of photos which correspond with the use of permission settings. We have observed that photo aesthetics measurably influences access control decisions. We have used these findings to generate a model which can produce high-accuracy predictions for particular subsets of users and photos.

This work presents a detailed investigation into photo-sharing, a key aspect of online social participation. We believe that our findings can directly inform the design of systems which can make photo-sharing a safer and more satisfying experience. We are intrigued to see how the findings here might generalize to the sharing of other types of content.

ACKNOWLEDGMENTS

We would like to thank the members of Yahoo's HCI Research group and members of the Flickr engineering team for their support and feedback on this project.

REFERENCES

1. Shane Ahern, Dean Eckles, Nathaniel S. Good, Simon King, Mor Naaman, and Rahul Nair. 2007. Over-exposed?: Privacy Patterns and Considerations in Online and Mobile Photo Sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 357–366. DOI: http://dx.doi.org/10.1145/1240624.1240683

2. Saeideh Bakhshi, David Shamma, Lyndon Kennedy, and Eric Gilbert. 2015. Why We Filter Our Photos and How It Impacts Engagement. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10573

3. Andrew Besmer and Heather Richter Lipford. 2010. Moving Beyond Untagging: Photo Privacy in a Tagged World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1563–1572. DOI: http://dx.doi.org/10.1145/1753326.1753560

4. Flickr. 2015a. Find every photo with Flickr's new unified search experience. (2015). http://blog.flickr.net/en/2015/05/07/flickr-unified-search/.

5. Flickr. 2015b. Thank You, Flickr Community! (2015). http://blog.flickr.net/en/2015/06/10/thank-you-flickr-community/.

6. David Frohlich, Allan Kuchinsky, Celine Pering, Abbe Don, and Steven Ariss. 2002. Requirements for Photoware. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work (CSCW '02)*. ACM, New York, NY, USA, 166–175. DOI: http://dx.doi.org/10.1145/587078.587102

7. Erving Goffman. 1959. *The presentation of self in everyday life*. University of Edinburgh.

8. Bernie Hogan. 2010. The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. *Bulletin of Science, Technology, & Society* 30, 6 (2010), 377–386. DOI: http://dx.doi.org/10.1177/0270467610385893

9. Mariea Grubbs Hoy and George Milne. 2010. Gender differences in privacy-related measures for young adult Facebook users. *Journal of Interactive Advertising* 10, 2 (2010), 28–45. DOI: http://dx.doi.org/10.1080/15252019.2010.10722168

10. Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Quang-Tuan Luong, James Z Wang, Jia Li, and Jiebo Luo. 2011. Aesthetics and emotions in images. *Signal Processing Magazine, IEEE* 28, 5 (2011), 94–115. DOI: http://dx.doi.org/10.1109/MSP.2011.941851

11. Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. 2012. Talking in Circles: Selective Sharing in Google+. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1065–1074. DOI: http://dx.doi.org/10.1145/2207676.2208552

12. Shehroz S. Khan and Daniel Vogel. 2012. Evaluating Visual Aesthetics in Photographic Portraiture. In *Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging (CAe '12)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 55–62. http://dl.acm.org/citation.cfm?id=2328888.2328898

13. David Kirk, Abigail Sellen, Carsten Rother, and Ken Wood. 2006. Understanding Photowork. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 761–770. DOI: http://dx.doi.org/10.1145/1124772.1124885

14. Peter Klemperer, Yuan Liang, Michelle Mazurek, Manya Sleeper, Blase Ur, Lujo Bauer, Lorrie Faith Cranor, Nitin Gupta, and Michael Reiter. 2012. Tag, You Can See It!: Using Tags for Access Control in Photo Sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 377–386. DOI: http://dx.doi.org/10.1145/2207676.2207728

15. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger (Eds.). Curran Associates, Inc., 1097–1105.

16. Cliff A.C. Lampe, Erik Johnston, and Paul Resnick. 2007. Follow the Reader: Filtering Comments on Slashdot. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1253–1262. DOI: http://dx.doi.org/10.1145/1240624.1240815

17. Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3217–3226. DOI:http://dx.doi.org/10.1145/1978942.1979420

18. Airi Lampinen, Sakari Tamminen, and Antti Oulasvirta. 2009. All My People Right Here, Right Now: Management of Group Co-presence on a Social Networking Site. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. ACM, New York, NY, USA, 281–290. DOI:http://dx.doi.org/10.1145/1531674.1531717

19. Kevin Lewis, Jason Kaufman, and Nicholas Christakis. 2008. The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. *Journal of Computer-Mediated Communication* 14, 1 (2008), 79–100. DOI: http://dx.doi.org/10.1111/j.1083-6101.2008.01432.x

20. Eden Litt. 2012. Knock, knock. Who's there? The imagined audience. *Journal of Broadcasting & Electronic Media* 56, 3 (2012), 330–345. DOI: http://dx.doi.org/10.1080/08838151.2012.705195

21. Eden Litt and Eszter Hargittai. 2014. Smile, snap, and share? A nuanced approach to privacy and online photo-sharing. *Poetics* 42 (2014), 1–21. DOI: http://dx.doi.org/10.1016/j.poetic.2013.10.002

22. Andrew D. Miller and W. Keith Edwards. 2007. Give and Take: A Study of Consumer Photo-sharing Culture and Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 347–356. DOI: http://dx.doi.org/10.1145/1240624.1240682

23. Oded Nov, Mor Naaman, and Chen Ye. Motivational, Structural and Tenure Factors that Impact Online Community Photo Sharing. In *International AAAI Conference on Web and Social Media (ICWSM '09)*.

24. Pere Obrador, Xavier Anguera, Rodrigo de Oliveira, and Nuria Oliver. 2009. The Role of Tags and Image Aesthetics in Social Image Search. In *Proceedings of the First SIGMM Workshop on Social Media (WSM '09)*. ACM, New York, NY, USA, 65–72. DOI: http://dx.doi.org/10.1145/1631144.1631158

25. Anne Oeldorf-Hirsch and S. Shyam Sundar. 2010. Online photo sharing as mediated communication. In *annual conference of the International Communication Association, Singapore*.

26. Leysia Palen and Paul Dourish. 2003. Unpacking "Privacy" for a Networked World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 129–136. DOI:http://dx.doi.org/10.1145/642611.642635

27. Rossano Schifanella, Miriam Redi, and Luca Maria Aiello. 2015. An Image Is Worth More than a Thousand Favorites: Surfacing the Hidden Beauty of Flickr Pictures. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10547

28. Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The Post That Wasn'T: Exploring Self-censorship on Facebook. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 793–802. DOI: http://dx.doi.org/10.1145/2441776.2441865

29. Anna C. Squicciarini, Cornelia Caragea, and Rahul Balakavi. 2014. Analyzing Images' Privacy for the Modern Web. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT '14)*. ACM, New York, NY, USA, 136–147. DOI: http://dx.doi.org/10.1145/2631775.2631803

30. Michael A. Stefanone and Derek Lackaff. 2009. Reality Television as a Model for Online Behavior: Blogging, Photo, and Video Sharing. *Journal of Computer-Mediated Communication* 14, 4 (2009), 964–987. DOI: http://dx.doi.org/10.1111/j.1083-6101.2009.01477.x

31. Fred Stutzman and Jacob Kramer-Duffield. 2010. Friends Only: Examining a Privacy-enhancing Behavior in Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1553–1562. DOI: http://dx.doi.org/10.1145/1753326.1753559

32. Monika Taddicken. 2014. The âĂŸPrivacy ParadoxâĂŹ in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. *Journal of Computer-Mediated Communication* 19, 2 (2014), 248–273. DOI: http://dx.doi.org/10.1111/jcc4.12052

33. Mike Thelwall. 2008. Social Networks, Gender, and Friending: An Analysis of MySpace Member Profiles. *J. Am. Soc. Inf. Sci. Technol.* 59, 8 (June 2008), 1321–1330. DOI:http://dx.doi.org/10.1002/asi.v59:8

34. Bart Thomee, David Ayman Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L. Li. 2016. YFCC100M: The New Data in Multimedia Research. *Commun. ACM* 59, 2 (2016). DOI: http://dx.doi.org/10.1145/2812802

35. Amy Voida and Elizabeth D. Mynatt. 2005. Six Themes of the Communicative Appropriation of Photographic Images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 171–180. DOI: http://dx.doi.org/10.1145/1054972.1054997

36. Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. "I Regretted the Minute I Pressed Share": A Qualitative Study of Regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS '11)*. ACM, New York, NY, USA, Article 10, 16 pages. DOI: http://dx.doi.org/10.1145/2078827.2078841

37. Molly McLure Wasko and Samer Faraj. 2005. Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly* 29, 1 (2005), pp. 35–57. http://www.jstor.org/stable/25148667

38. Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for My Space: Coping Mechanisms for Sns Boundary Regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 609–618. DOI: http://dx.doi.org/10.1145/2207676.2207761