From Tag to Protect: A Tag-Driven Policy Recommender System for Image Sharing

Anna Cinzia Squicciarini[1], Andrea Novelli[2], Dan Lin[3], Cornelia Caragea[4], and Haoti Zhong[5]

[1,2,5]Pennsylvania State University, E-mail: acs20@psu.edu, amn5567@psu.edu, hzz13@psu.edu
[3]Missouri S&T, E-mail: lindan@mst.edu
[4]University of North Texas, E-mail: ccaragea@unt.edu

***Abstract*—Sharing images on social network sites has become a part of daily routine for more and more online users. However, in face of the considerable amount of images shared online, it is not a trivial task for a person to manually configure proper privacy settings for each of the images that he/she uploaded. The lack of proper privacy protection during image sharing could raise many potential privacy breaches of people's private lives that they are not aware of. In this work, we propose a privacy setting recommender system to help people effortlessly set up the privacy settings for their online images. The key idea is developed based on our finding that there are certain correlations between a number of generic patterns of image privacy settings and image tags, regardless of the image owners' individual privacy bias and levels of awareness. We propose a multi-pronged mechanism that carefully analyzes tags' semantics and co-presence to derive a set of suitable privacy settings for a newly uploaded image. Our system is also capable of dealing with cold-start problem when there are very few image tags available. We have conducted extensive experimental studies and the results demonstrate the effectiveness of our approach in terms of the policy recommendation accuracy.**

I. Introduction

According to latest statistic reports [5], [27], online users are uploading 1.8 billion images in social networking websites, such as Facebook, Instagram, and Flickr, every day. Most of these photos reveal personal information (e.g., family photos, vacation photos), since they are mostly posted as a way of self-promotion, social networking and personal sharing. Along with each image, a "story" is often told, proxied by captions, tags, or comments. Tags, in particular, are now added to each image through users' annotations or by the website itself (e.g. Flickr automated-tags). It is worth noting that tags have been successfully used to facilitate a number of image-related tasks, ranging from information retrieval and search [2], [26], to content detection [12] and social discovery [9]. Yet, annotated images, if not well protected, could represent a privacy vulnerability [4], [29]. Unfortunately, current content management sites vary greatly in their levels of privacy protection. Further, recent studies [11] have consistently shown that despite the mechanisms in place to protect users' data, a considerable percentage of online users do not know how to properly set up their privacy settings or are unaware of the consequences linked with accidental disclosure of personal content.

978-1-5386-2487-6/17/$31.00

In order to help users configure proper privacy settings, several policy recommendation systems have been proposed (e.g. [24], [32], [11]). Some of these proposals require heavy user involvement [13] or rich user information [15], hindering adoption. Some recent methods require visual analysis of the image content itself [24], [32] which not only incur significant processing overhead but also need a large amount of data to ensure accuracy.

In this work, we propose a recommender system for image privacy, called T2P (Tag-To-Protect), to help people effortlessly set up the privacy settings for their online images. The proposed T2P system requires only a small amount of information (i.e., image tags) as input to produce accurate privacy recommendations. The key idea underlying T2P is based on the observation that there are certain correlations between a number of generic patterns of image privacy settings and image tags (also confirmed in related studies [24], [13]), regardless of the image owners' individual privacy bias and levels of awareness. Specifically, our study found that when enough instances of tags and tag combinations occurred with similar privacy preferences, they may become indicators of the privacy settings. The following is a simple example. Figure 1 (a), (b) and (c) are photos posted by different users in Flickr. These three photos have similar tags and they also share similar privacy settings. It is worth noting that our findings also resemble observations in recent research on image privacy recommendations [32], [24], [31] which point out that categories of images are linked to certain privacy preferences despite the individual bias.

Specifically, the proposed T2P system offers a multi-pronged mechanism that carefully analyzes tags' semantics and co-presence with privacy settings. When a user uploads a new image, the T2P system will first identify existing images with most similar sets of tags. Then, if there is a so-called dominant policy that exhibits the strongest privacy pattern observed among this group of images, this dominant policy will be recommended to the user. If there is not such a dominant policy, the T2P system proceeds to the fine-grained policy analysis via two means: (i) the vertical comparison which looks into the correlations between a tag and each access right included in the policy; and (ii) CoTag Graph Cohesiveness which models tag communities and the associated privacy patterns. The final recommendation could be either a single

(a) {*model*,*couture*,*fashion*,*mannequin*}

(b) {*model*,*fashion*,*glamor*}

(c) {*model*,*fashion*,*mannequin*}

Fig. 1: Flickr Images

policy or a list of ranked policies at users' choices. The contributions of our work is summarized as follows:

- The proposed T2P system releases the users from the burden of defining and customizing privacy policy settings to the largest extent, while limiting the amount of data and processing needed to provide a recommendation. Our proposed policy recommender algorithm utilizes solely the image tags to achieve high recommendation accuracy.
- The proposed T2P system is capable of handle the well-known and difficult cold-start problem. That is, although with slightly lower accuracy, T2P can provide appropriate recommendations even for images with zero or unknown tags.
- Our experiments produce early evidence of both overall accuracy of our approach as well as acceptability with real users, showing not only that we are able to recommend the "expected" policies, but also that our recommended policies are preferred as compared to other baseline methods.

II. Related works

Work on recommendations within social network settings is proliferating (e.g. [19], [16]). In particular, with respect to privacy, Besner et al. [1] pointed out that social network users want to regain control over their shared content but meanwhile, they feel that configuring proper privacy settings for each image is a burden. Similarly, related work suggests sharing decisions may be governed by the difficulty of setting and updating policies, reinforcing the idea that users must be able to easily set up access control policies [15], [22], [24], [14]. Some notable recent efforts to tackle these problems have been conducted in the context of tag-based access control policies for images [31], [13], showing some initial success in tying tags with access control rules. However, the scarcity of tags for many online images [25], and the workload associated with user-defined tags makes accurate analysis of the images' sensitivity based on this dimension only non-trivial, as we show in our work. Other work (e.g. [8], [15], [3], [22]) has focused on generic users' profile elements, and typically leveraged social context, rather users' individual content-specific patterns. In this paper, we focus on context as it is given by tags specifically linked to images, rather than context with respect to social network patterns and online user behavior.

In regards to image-specific protection, a recent body of work has focused on protection of image privacy (e.g., [24], [32]). To date, these works have focused on detecting image semantics by means of combinations of high-level and low-level visual features. Our approach aims at a content-driven privacy protection as well, without the heavy computational effort typically associated with image processing methods. As shown in our experimntal evaluation, the overall accuracy of our method is comparable, if not superior to existing state-of-the art methods.

Finally, a loosely related body of work is on recommendation of tags for social discovery [21], [18], [16] and for image classification [20], [34], [6], [26] in photo sharing websites like Flickr. In these works, the typical approach is for authors to firstly collect adequate images and then classify images according to visual and context-related features. After users upload their images, the server extracts features, then classifies and recommends relevant tags and groups. Here, we focus on the reverse approach, that is, from tags to recommendations of image privacy settings.

III. The T2P System

Our proposed Tag-To-Protect (T2P) system aims to automatically generate privacy policies for users who upload images to image sharing web sites. The T2P system is developed based on the study of the correlations between the image tags and the privacy settings. That is images with similar or significantly overlapping tags typically share similar privacy policies.

As shown in Figure 2, the T2P system is a multi-phase system which consists of the following components: (i) tag pre-processor; (ii) coarse-grain tag analyzer; (iii) fine-grained tag analyzer, and (iv) cold-start handler. In what follows, upon formally introducing policies and tags, we present the details of each component.

A. Policies and Tags

Our system takes as input two types of data: (i) image tags; and (ii) privacy policies, and then outputs the recommended privacy policies.

More specifically, given a set of k images $I = \{img_1, img_2, ..., img_k\}$. Each image img_i has a unique ID and a set of tags T_i= $\{t_1, t_2,...,t_{n_i}\}$, also denoted as $tags(img_i)$. Image tags are strings, which are usually contributed by users

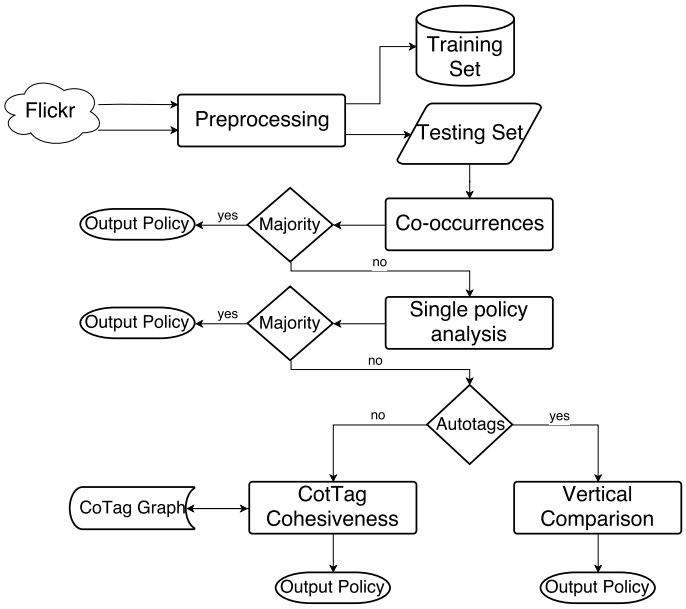


Fig. 2: System Overview

or taken from a pre-defined vocabulary. In addition to the user defined tags, some image sharing websites employ image recognition techniques to help generating tags to attach to images. These system automatically generated tags are called *autotags*. Frequently autotags may not always describe the image content accurately, due to the limitations of current autotag generation algorithms.

Each image is associated with a simple privacy policy as provided by many social websites. The privacy policy specifies who is allowed to conduct what action on the image. Although the specific syntax of a given access policy may vary according to the site's infrastructure, most settings offered in current sharing platforms can be abstracted as a tuple of the form P=$\{ID_{img_i}, (act_1, \ldots, act_n)\}$ denoting the following two items:

- ID_{img} is the image ID.
- $act_1, \ldots, act_n$ denote the list of allowed actions on the image. Each action may further take value within an interval, varying according to different levels of privacy.

For the sake of presentation, in what follows, we assume that the possible actions $act_1, \ldots, act_n$ are restricted to 'view' (v), 'comment' (c) and 'download' (d), as these are the most common operations possible on an image. We also assume, consistent with other work in this space [24], [13], [22] that the possible privacy values within these actions correspond to the following four levels of privacy:

- 0: (Private/You) The image owner him/herself or a special group of users that the image owner considers as fully trusted, can perform the action on the image;
- 1: (Friends and Family) Friends or relatives of the image owner can perform the action;
- 2: (Social Networks) Any user that has an account in the same social network site as the image owner can perform the action;
- 3: (Public/Anyone) Any Internet user can perform the action.

In what follows, when a policy is not associated with a given image, it is simply denoted as $p(v, c, d)$, with v, c, d being the privacy levels for the comment, view and download action, respectively. An example of a policy is given next.

Example 1. *Tom uploaded a photo whose id is "206A", and tagged it with tags "travel", "cruise", "rock climbing". He would like to allow his friends to view and comment, but nobody except himself can download the image. His privacy preference can be expressed by the policy* $\{($*"206A", (1,1,0)*$\}$

Note that this policy representation is claimed to be neither universal nor inclusive of all possible privacy labels to be associated to a given image. Yet, it is an abstract syntax generally used by popular social sites, and has been used successfully in a number of previous work. Note also that whether or not the policy options of above can be ordered and compared depends on the syntax adopted by the social media site. Accordingly, different types of accuracy measures can be developed.

B. Tag Preprocessor

It is generally recognized that user-provided image tags are noisy [26]. For example, in Figure 3 (b), there are many tags that seem to be only remotely connected with the image content (i.e. "iphone"), and that make it difficult to correctly interpret the image content. Therefore, in order to make meaningful privacy inference, the first step is to preprocess the tags to extract only semantically relevant and syntactically correct tags for the further consideration. In the following we present the details of our preprocessing method.

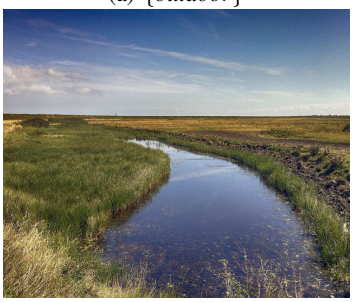
Consider a training data set TD in the form: $TD = \{(img_1, p_1),(img_2, p_2),..., (img_k, p_k)\}$, where img_i, p_i are images and policies IDs, respectively. Preprocessing is a cascading process that consists of three steps:

Step 1: Discard images that have too few or too many tags from the training dataset. This is because images with few tags do not provide enough information for policy inference. On the other hand, too many tags could increase noise and inconsistency, and hence is not suitable for policy inference either. For example, in Figure 3a, the image has a single tag "outdoor" which is neither a descriptive nor a contextual tag. In Figure 3b, the image is associated with a long list of tags which cause the confusion about the actual image content too. Therefore, we only consider images with a reasonable number of tags, falling in the range $[k_{min}, k_{max}]$, with $k_{min}, k_{max} \in \mathbb{N}$ (e.g. [3,6]). The effect of the choice of k_{min} and k_{max} are studied in the experiments. As the result of this round of filtering, we will obtain the following new training dataset TD'.

$$TD' = \{(img_i, p_i) \mid (img_i, p_i) \in TD \wedge \\ \wedge k_{min} < |T_i| < k_{max}\} \quad (1)$$

Step 2: Simplify each tag in TD' by reducing it to its root form based on a stemming algorithm [30]. For example, the

(a) {*outdoor*}



(b) {*contrast*, *interesting*, *uk*, *england*, *kent*, *close*, *detail*, *darkness*, *summer*, *iphone*, *sky*, *all*, *nature*, *focus*, *high*, *plant*, *iphoneography*}

Fig. 3: Images with an inconsistent list of tags.

tags "fish", "fishes" and "fishing" all stem into "fish", which carries the same meaning as the original tags. In this way, it helps to reduce the ambiguity during the subsequent tag comparison and analysis. The output from this step is a new training dataset TD''.

$$TD'' = \{(img'_i, p_i) \mid \forall (img_i, p_i) \in TD' : img'_i = simp(img_i)\} \quad (2)$$

Here, $simp$ function is defined as $simp(img) = img' \mid \forall t_i \in T : T' = \{stem(t_1), stem(t_2), ..., stem(t_n)\}$.

Step 3: Discard noisy tags. Noisy tags are tags that are linked with images associated with a set of extremely different policies. Typically (over 75% of the cases) these are adjectives. For example, a tag "red" may occur with an image of a red car and also an image of a violent scene with red blood. The first image is supposed to have a policy to allow anyone to view, while the second image a policy to give access to only a small group of authorized users. In this case, the tag "red" is not useful in determining the actual privacy protection level. In order to identify such types of noisy tags, we analyze the distribution of the policies associated with any tag (from any user). Specifically, we compute the frequency of the policy values associated with the 'view', 'comment' and 'download' operations, denoted as $freq.v$, $freq.c$ and $freq.d$, respectively. If the variance of any of the three frequencies of the tag is larger than a (empirically set) threshold θ, we will remove the tag from the training dataset. We note that our strategy is consistent with work on keyphrase extraction[28]. Keyphrase extraction is the problem of automatically extracting important phrases or concepts of a document. Keyphrases provide a high-level topic description of a document. While adjectives are typically helpful in sentiment analysis tasks, image tags are generally expressed by nouns and noun phrases. In keyphrase extraction tasks, phrases that end with an adjective and the unigrams that are adjectives are removed. By analogy, adjectives that are used inconsistently and therefore are uninformative are removed from tags.

C. Tag Analysis for Policy Recommendation

Our recommendation approach is based on the intuitive and yet powerful relationship between an image and its privacy settings, according to the presence of similar or semantically related groups of tags. It conducts policy recommendation from coarse-grained tag analysis to fine-grained tag analysis as elaborated in the following.

1) Coarse-grained Tag Analysis: The first level of the tag analysis checks the co-occurrences of a set of tags in the image tags of all the users. Tag sets are defined as follows.

Definition 1. *Let T be a set of tags $T = \{t_1, ..., t_n\}, n \geq 2$, and I a set of images. Let $tags(img)$ denote that tags associated with an image img. Tags in T form a co-occurrence, or are co-occurent, if and only if*

$$\exists\, img \in I \mid \forall i = 1, ..., n \ : t_i \in tags(img) \quad (3)$$

The reason to consider the co-occurrence of a set of tags is because of their joint descriptive power. For example, consider images tagged as $T_1 = (\text{"}model\text{"}, \text{"}fashion\text{"})$, $T_2 = (\text{"}model\text{"}, \text{"}fashion\text{"}, \text{"}style\text{"})$, and $T_3 = (\text{"}model\text{"}, \text{"}car\text{"})$. If we consider only a single tag at a time, the tag "model" is ambiguous since it could be either a fashion model or a car model. Instead, if we consider the set of tags together, T_1 and T_2 form a co-occurrence and they are more accurate in indicating that their images have similar content. Images with similar content are likely to be associated with similar policies. Based on this observation, we use co-occurrences of tags to infer the privacy policy as follows.

Given a newly uploaded image with a set of tags, we aim to find the policies which have been commonly associated with the similar set of tags to be recommended to this new image. Specifically, we enumerate the combinations formed by at least two tags of this new image, and check if the combination of the tags co-occur in other existing images. If a set of tags (T_j) co-occur in an image (img_j), we insert the image img_j's privacy policy p_j to the candidate policy list and increase this candidate policy's frequency $freq(p_j)$ according to a weight proportional to the number of times the co-occurrence is seen (with a minimum of 2). If at the end of the co-occurrence check, there are still some tags which are not included in any of the co-occurent tag sets, we insert the policies associated with these remaining tags into the candidate policy list too. When inserting these candidate policies, we increase their frequency only by 1 in order to give more weight to those co-occurent ones.

Finally, we will try to identify the dominant policy among the candidate policies. We say that a policy p is dominant among all other policies in a list L if and only if its frequency

is ϖ times higher as other policies. In the experiments, ϖ is set to 2. $freq(p) \geq \varpi \cdot freq(p')\forall p' \in L$

If such a dominant policy is identified, the system directly returns it as the recommended policy. Otherwise, it means the co-occurrences are not sufficiently strong to suggest a policy, and we proceed to the next level of tag analysis as presented in the subsequent section.

2) Fine-grained Tag Analysis: Arriving at this point means that a finer-grained analysis of images tags and privacy practices is needed. We propose two fine-grained tag analysis algorithms, namely *Vertical Comparison* and *CoTag Graph Cohesiveness*. The vertical comparison approach is good at dealing with images that have autotags, and the CoTag Graph approach is more effective for images with user created tags.

Unlike the co-occurrence approach whereby we consider the relationship between tag groups and entire policies, the vertical comparison approach looks into the relationship between tag groups and individual components in the policy. Recall that the policy used in this work considers three types of permissions on a shared image, which are "view", "comment" and "download", and each action is given a value in the policy to indicate the group of people who are allowed to perform the action (Section 3.1). For example, in a policy $p_4\{img_4, (3, 2, 1)\}$, the three values (3,2,1) are corresponding to three actions "view(v)", "comment(c)" and "download(d)" respectively. The value 3 for "view" means anyone on Internet can view $image_4$, the value 2 for "comment" means only the users in the same social network can comment on the image, and the value "1" means only the friends or family members of the image owner can download the image.

Taking each tag of the new image, we compute the frequency of the tag occurred with the specific value of each action in all existing policies. Let $freq_{act}(tag, act\_value)$ denote the frequency of the tag associated with action "act" which has the value "act_value". We can calculate the tag group frequency as shown in the following example.

Suppose that there are four privacy policies:

$p_1\{img_1, (3, 3, 3)\}, img_1(model, fashion)$
$p_2\{img_2, (3, 3, 1)\}, img_2(model, fashion, style)$
$p_3\{img_3, (2, 2, 1)\}, img_3(model, car)$
$p_4\{img_4, (3, 2, 1)\}, img_4(model, house)$

If a newly uploaded image has a tag "model", we can calculate the frequency of this tag in each policy component as follows:

$freq_v(model, 3) = 3, freq_v(model, 2) = 1$
$freq_c(model, 3) = 2, freq_c(model, 2) = 2,$
$freq_d(model, 3) = 1, freq_d(model, 1) = 3.$

After the calculation of the frequency of each tag in the new images, we then select the action values with the highest frequency to be included in the recommended policies. In case there is a tie, we choose the most restrictive action. Considering the above example, $freq_v(model, 3) = 3$, $freq_c(model, 2) = 2$, and $freq_d(model, 1) = 3$ will be selected. As a result, the recommended policy for the new image would be $p(3, 2, 1)$ which means the new image can be viewed by anyone in the Internet, commented by users with account in the same social network, and downloaded by only friends and family members.

The above vertical comparison is chosen if the new image has at least one autotag. When an image is associated with only user created tags, we propose a CoTag Graph Cohesiveness approach to predict the policy more effectively. In the CoTag Graph, we model the relationship among tags as a graph, whereby each node represents a tag, and each edge connects two tags that both appear with the same image. Moreover, we also assign the weight to each edge to indicate the number of times that the two connected tags occur together in the same image. Figure 4 shows an example CoTag Graph.

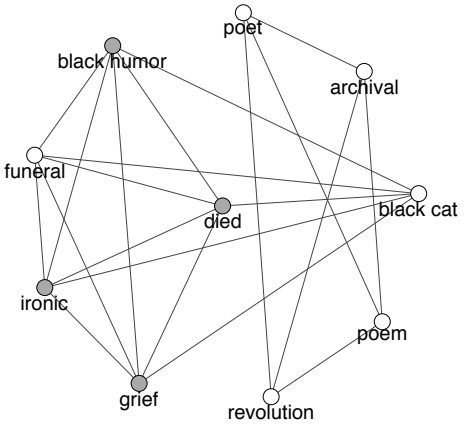


Fig. 4: An example of a CoTag Graph with two clusters.

The first step is to group tags in the CoTag graph according to the frequency of joint appearance on various images. For this, we employ the Louvain method [7] which has been proven to be an efficient heuristic method for identifying clusters in large networks.

The Louvian method is a greedy optimization method that attempts to optimize the "modularity" of a partition of the network. Modularity here is intended as a measure of the density of links inside communities compared to links between communities. The optimization includes two steps. First, the algorithm defines "small" communities by optimizing modularity locally. Next, it aggregates nodes belonging to the same community and builds a new network whose nodes are the communities. These steps are repeated iteratively until a maximum of modularity is attained and a hierarchy of communities is produced.

Once the tags are clustered, we select the most frequently used policy as the representative policy for each cluster. At the end of this process, every cluster has a representative policy and it is possible that some clusters share the same policy.

Next, we aim to find the cluster of tags that is most related to the tags in the newly uploaded image, and then use the representative policy of that cluster for the recommendation. The detailed steps are the following. First, for each tag t in the tag list of $tags(img_i)$ of a new image img_i, we search the

CoTag graph to find which cluster that the tag t belongs to. Since there may be multiple tags in the new image and hence multiple clusters may be returned from this step. In order to find the most related cluster, we define a cohesiveness value κ_t (Definition 2) to measure the tightness of the relationship between the tag t and the corresponding cluster.

Definition 2. *Let G be a CoTag Graph, and let V_G be its nodes (i.e. tags) set. The cohesiveness value κ_v ($v \in V_G$) is calculated as the maximum degree between each tag node in relation to its cluster, normalized with respect to the whole graph size as follows:*

$$\kappa_v = \frac{deg_C(v) \cdot |V_C|}{|V_G|} \tag{4}$$

where $deg_C(v)$ indicates the degree of node v, namely the number of edges incident to v, with respect to its cluster C and V_C is the set of vertices in the cluster.

After calculating the cohesiveness value κ_t for each tag associate with the new image, we select the tag (denoted as t_s) that yields the highest cohesiveness value. Then, the representative policy of the cluster that this tag t_s belongs to is presented as the recommended policy for the new image. Algorithm 1 outlines the main steps.

Algorithm 1 Graph cohesiveness algorithm

```
1:  function GRAPHCOHESIVENESS(tags, G)
    input: tags ← list, G ← graph
    output: maxpolicy ← policy
2:      maxweight ← 0
3:      maxpolicy ← NULL
4:      for each tag in tags do
5:          C ← cluster assigned to tag
6:          v ← vertex tag relative to the cluster C
7:          d ← degree of v in subgraph C
8:          localweight ← (d · C.size)/G.size
9:          if localweight > maxweight then
10:             maxweight ← localweight
11:             maxpolicy ← C.policy
12:     return maxpolicy
```

Example 2. *Consider a newly uploaded image with the following tag list: {"poem","died","funeral"}. Assume that the analysis of co-occurrences did not yield to any recommendation, and the CoTag graph cohesiveness technique is necessary. As shown in Figure 4, the image's tags are in two different clusters: "poem" and "funeral" in one, "died" in another. Moreover suppose that, policy with action set (2,2,2) is assigned to the first cluster and action set (1,1,1) to the second. The algorithm calculates the cohesiveness value for each node in the tag list, according to Equation 4. Let $\kappa_{poet} = 0.22$, $\kappa_{funeral} = 0.38$, and $\kappa_{died} = 0.42$ be these values. The policy assigned to the node with highest value is selected, in this case "died" and consequently, policy (1,1,1) is recommended for the new image.*

Dataset	#Imgs	# Tags per Img	Policies			
			Private	Family	Social	Public
PicAlert	2052	7.2	99	271	164	506
T2PData	3566	3.8	795	1299	627	845

TABLE I: Dataset statistics

Finally, we would like to discuss the reason for not using CoTag graph cohesiveness when autotags are present. The autotags are automatically generated by social networks which are usually very generic and popular tags such as "photo" to be associated with nearly every image. If a new image contains autotags, it is very likely the CoTag graph approach will select the cluster that the autotag belongs to due to the popularity of the autotags. However, autotags may not well capture the image owner's privacy preferences and hence the CoTag approach may not be less effective than the vertical comparison approach in this case.

D. Cold start Problem

As in any recommendation system, our approach must deal with the *cold start problem*. In the case of T2P, cold start arises when tags of a test image are not present in the training dataset, and therefore it is not possible to draw any inference. In order to address this issue, we adopt a "content-based" matching approach. We employ semantic analysis and word similarity [17], [10] to find tags in the available historical dataset that are semantically similar or conceptually connected with the ones originally associated to the image. Precisely, recommendations of privacy settings are generated according to two main steps. Given a image with a list of tags: 1) for every tag in tags list, compute a synonyms set and search for the presence of at least one synonym in the training set; 2) if it appears in the training set, add it to a new tags list, otherwise find and add to the tags list the tag in the training set that has the highest *similarity score* with the original one. In our experiments, synonyms and similarity scores are computed according to the lexical database WordNet [17].

Note that for our proposed method to work, we need at least a pair of tags linked with the image. This is not a concern, as image-uploading sites such as Flickr now support automatically-generated tags, that can be added to any image, and can be used to trigger our proposed cold-start method.

IV. Experimental Study

In this section, we first introduce our experimental settings including the data collection, and then we present the results of the T2P system's performance.

A. Experimental Settings

We test our approach on two datasets (see Table I). The first dataset is a sample from the "PicAlert" dataset [32], and it includes 2052 images, with 7523 corresponding tags. Each image has been assigned privacy policy through a Mechanical Turk study and these policies are used as ground truth.

We generated the second dataset, referred to as T2PData, through a user study. Participants were invited to select images

from the Flickr creative license repository. Each user had to select at least 60 images and up to 80 images. To prompt participants to provide photos for which they may have varied access control and privacy preferences, we provided them with a list of suggested photo categories, including: 1) up to 15 photos that they would not want to share with the general public; 2) up to 40 images that include people only; 3) up to 10 images that include life events (competitions, weddings, graduation ceremonies, etc.); 4) up to 20 pictures regarding traveling and/or their hobbies; 5) up to 15 pictures that they would only share with trusted friends. Participants were NOT asked to add tags to images that they selected. We keep the original tags provided by the image owners from Flickr. Participants were college students from non-STEM field who took the study as part of their generic research credit requirement. Participants were only asked to assign one appropriate policy for every image by treating the image as if it was their own photos. To minimize bias, no reference was made to the participants about the role of the image tags until after the experiment completion. In total, 77 participants completed the study and we have collected 3566 images associated with 13675 unique tags. Note that participants assigned images with slightly simplified access policies. Specifically, each image is given a policy that specifies what group of users are allowed to access the image but does not distinguish the exact action (e.g., view, download) on the image. Table I shows a summary of statistics of our datasets. In all experiments and for all approaches, we use two training sets: (i) 485 images sampled from PicAlert; and (ii) 306 images sampled from T2PData. We kept the ratio of policies consistent to the original datasets. Training and test sets were organized at per-image level.

In what follows, we first present the direct user study whereby we launch the T2P system and collect user feedbacks on the recommended policies. Then, we provide an anatomy of the recommended policy by comparing each component in the recommended policy with that in the collected ground truth and examine the similarity of them. Finally, we report the performance in the cold-start scenario.

B. Direct User Evaluation

In the direct user evaluation, we aim to study to what degree users are satisfied with the policies recommended by the proposed T2P system. We conducted two rounds of direct user evaluation. The study was advertised as an "Image Privacy Study", with detailed instructions and IRB pointer. Participants were recruited using Amazon Mechanical Turk Each and each participant was paid $1 for their work.

In the first round of user study, there are 31 participants. Each participant was presented 25 images taken from both the T2PData and PicAlert datasets. For each image, participants were asked to indicate their favorite privacy policy among a list of three. Each list contains the following three types of policies: one policy proposed by our T2P approach, one randomly generated policy and one policy generated by a decision tree algorithm. The decision tree algorithm is similar to that used in the most related work on tag-based policy recommendation [13]. Specifically, the decision tree classifier (C45) uses bags of words [33] to generate features from image tags, and then is trained for each type of policies. The max depth of the tree is set to 8 and the minimum sample that a leaf hold is 5, and Gini purity is used for splitting the tree. The three recommended policies are displayed in a random order for each image so that users do not know which one is generated by our proposed T2P to avoid any potential bias.

Table II reports the results of the user study. As we can see that, the T2P policy is favored by much more users than the policies generated by the random and decision tree approaches. In particular, in 47.9% of the cases, users preferred the T2P policy, while only in 30.9% and 21.2% of the cases, users prefer random policy and the decision tree policy respectively. Furthermore, we note that in the case where a user chose a policy different from the one proposed by the T2P, 62.3% (denoted as T2P(similar) in the table) of these choices are policies which has only component slightly different to our recommended policy. For example, the user chose a policy to allow friends only while the T2P recommends to share with family members. This means that, even if a user did not choose a T2P recommendation, their preferred policy is very similar to the one proposed by our system. The effectiveness of the T2P approach is attributed to the correct analysis of the relationship between image tags and privacy preferences.

Method	**Top choice**
T2P	47.9%
Random	30.9%
Decision Tree	21.2%
T2P(similar)	62.3%

TABLE II: User Study – Choosing the Favorite Policies

In the second user study, we recruited 40 users and each was provided 20 images. For each image, the participant was asked to rank a list of three policies. These three policies are the top three candidate policies generated by the T2P for this image. The three policies are displayed in a random order rather than their rankings to the participants. Our goal is to see how often the user would prefer our top ranked policies, and how the T2P ranking order would fare as compared to the users provided rankings. The results are reported in Table III. As we can see, in nearly 80% of cases, our T2P policy ranking has at least one match with the user ranking. In 26.2% cases, our ranking matches the user's ranking exactly. About half of the cases, our top ranked policies are also the users' top choices. In the next section, we will further examine that by providing a list of ranked policies, the chance of users accepting our proposed policies become higher.

C. Anatomy of the T2P Policies

We now examine the recommended policies generated by T2P in-depth. We directly compare the recommended policies with the ground truth that we collected for the two datasets: PicAlert and T2PData. It is challenging to determine how close the recommended policy is to the ground truth (beside exact match) because the values in the policy components are

Rank#1	Rank#2	Rank#3	Ratio
✓	✓	✓	26.2%
✓	✗	✗	17%
✗	✓	✗	22.8%
✗	✗	✓	12%

TABLE III: Ranking of policies ✓is a policy which ranking of the user matches the user's ranking. Item, ✗is a ranking which does not match user's preference

categorical. Therefore, we adopt the following two accuracy metrics in the evaluation:

- **Distribution Distance (DD)**: We compare distributions of component values in the policy sets. In detail, we first compute the frequency distribution of our recommended policy set in terms of four permissions: $f0$ for "self", $f1$ for "friend", $f2$ for "social network", $f3$ for "public", where f0+f1+f2+f3=1. Next, we obtain a separate frequency distribution from ground truth (denoted as GT) or another heuristic. We quantify the "distributions' distance" (denoted as DD) between these distributions by taking the square root of the sum of the squares of the policy-component-wise differences. The smaller the distance, the more similar the distributions, and hence more accurate the policies are.
- **Policy Dominance** ($\subset$): This metric indicates policy strictness [24], [23]. We consider a natural order of possible actions as dictated by the cardinality of users presumably in each group, ranging from 0 in self, to millions, in *public*. We say that p_a dominates p_b, or $p_a \subset p_b$ if all the permissions for all the possible actions for policy p_a target smaller groups than the permissions for p_b. In this setting, there are three kinds of errors. The permission can be i) more restrictive than intended by the user (i.e., the recommended policy dominates the ground truth); or ii) less restrictive than intended by the user, or c) incomparable. The incomparable case occurs when policy p_a is more restrictive than policy p_b in some actions but less restrictive in other actions.

For comparison, we not only continue to use the tag-based recommendation approach (i.e., the decision tree algorithm as described in the previous section), we also implement a state-of-the-art image-processing based algorithm, i.e., Radial Basis Function (RBF) kernel SVM classifier. The decision tree algorithm is the same as described in the previous section. The SVM classifier is trained over a set of low-level visual features, and uses cross-validation to obtain optimized parameters. In our case, the feature set is a bag-of-words over SIFT features [35]. SIFT features are employed as they have shown to be extremely successful for image semantic analysis and classification [32], [24]. We perform a 10-fold cross validation to test the classifier's performance over the PicAlert dataset.

Table IV reports the distribution distance between the policies generated by the three approaches and the ground truth. Note that the tested distributions are all statistically different (Tukey HSD inference $p < 0.005$). As we can see that, the T2P has the smallest distribution distance (1.384) to the ground truth policies, which means the T2P policies are most similar to the ground truth policies than those generated by the decision tree and the SVM approach.

Method	Distribution Distance from Ground Truth
T2P	1.384
Decision Tree	1.850
SVM	1.9276

TABLE IV: Distribution Distance (DD)

To take a further look at these distributions, Figure 5 shows the percentage of each value for the "view" action in the whole policy set. We can observe that the T2P has the most similar percentage on each policy option (i.e., "only-you", "friend-fam", "social net" and "public") to the ground truth than the other two approaches. This indicates that the T2P not only yields the smallest overall Distribution Distance but also performs equally well in each policy component. The biggest discrepancy between T2P and the ground-truth labels are seen in the friends and family category, possibly due to the slight under-classification of social network and public category.

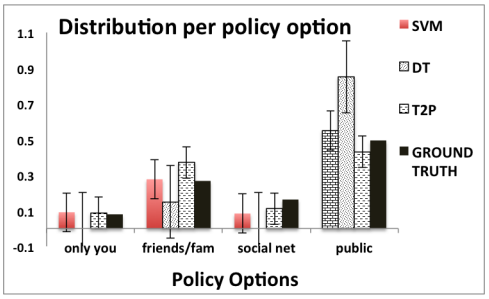


Fig. 5: Percentage of each policy option for "View" Action"

Since in Figure 5, the SVM seems to be quite similar to the ground truth in the "view" action, we further compare the performance of the T2P and the SVM in terms of all policy components ("view", "comments", "download") in Figure 6. In Figure 6, each bar is color-coded to denote the percentage of results in three cases: (i) the recommended policy is equal to GT, (ii) the recommended policy is dominated by GT; and (iii) the recommended policy dominates GT. We can observe that although the SVM looks similar to the ground-truth policies in the previous test, its indeed contain fewer number of equal policies than the T2P. Another interesting observation is that the T2P approach generates more policies that are less restrictive than the ground truth compared to the SVM approach. This could be useful in the real applications in that the T2P provides a slightly larger group of users for the image owner to further narrow down.

We further test the relationship between the number of tags assigned to an image and the prediction accuracy. We classified our test datasets according to the number of tags per image, into four groups: images with 2 to 4 tags, with 5 to 7 tags, with 8 to 10 tags and with 11 to 14 tags. As shown in Figure 7, T2P performs best (i.e. it has the higher number of perfect matches with the ground-truth labels) when an image is labeled

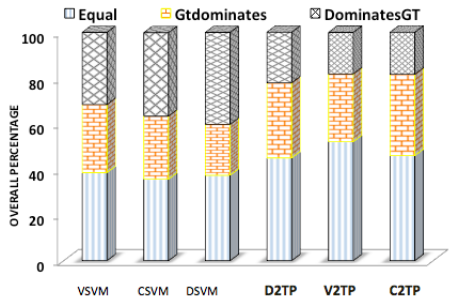


Fig. 6: Comparison with SVM, whereby VSVM denotes "view" SVM, CSVM "Comment" SVM and DSVM "Download SVM. VT2P denotes "view" T2P, CT2P "Comment" T2P and DT2P "Download" T2P.

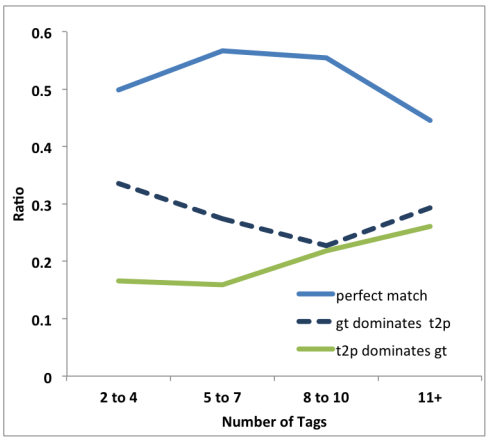


Fig. 7: Accuracy by number of tags per image

by 5 to 10 tags. When there are too many tags per image, the prediction accuracy decreases as our policies seem to be less restrictive than the original labels. This indicates that a large set of tags may include noisy labels that introduce too many heterogeneous patterns. This finding therefore supports the need of our pre-processing and label filtering strategy reported in Section III-B.

#Tags	PicAlert			T2PData		
	Precision	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
2-4	0.44	0.46	0.45	0.96	0.64	0.77
5-7	0.63	0.72	0.67	0.96	0.65	0.77
8-10	0.52	0.82	0.64	0.93	0.65	0.76
RBF-SVM	0.56	0.58	0.57	0.38	0.79	0.521

TABLE V: Accuracy values with simplified privacy options

Finally, we consider a variant test scenario by using the binary policies. We measured precision and recall on a simplified binary version of the policies, for both our method and our baseline model over SIFT features (see above). The binary representation of a policy means we consider *Only you* and *Family & Friends* as *Private*, and *Social Network* and *Anyone* as *Public*. Simplified policies are used to test the adaptability of our approach in different policy settings. Our findings are reported in Table V. In general, our approach performs better in the T2PData dataset.

D. Cold Start Evaluation

To assess performance of our method in case of "cold-start" instances, we involved 18 users for a total of 110 images. All the tested images had associated tags with no policies associated to them. We used test images from the T2PData dataset, and the PicAlert dataset for semantic similarity analysis, and vice versa, we used PicAlert images for test images and the T2P dataset for similarity analysis (78 and 32 images, respectively from T2PData and PicAlert).

Each study participant was asked to select one of four privacy policies, where one was a T2P policy, one was SVM, one was randomly generated, and one was created using the decision tree algorithm. Policies were presented in random order to avoid biases.

The policy proposed by T2P is predominant with respect to the others, obtaining by far the highest preference, followed by SVM. We further checked whether the choice of selecting T2P over other policies was related to the overall privacy inclinations of the users. We found that our method performs best for users who claim stronger privacy awareness - according to their responses to pre-session questionnaire. For users with stronger privacy concerns, our method is preferred over 58% of the times, whereas it is less successful for users who claim not to care (33.6%). This may be due to the relative strictness of our policies, as compared to other methods.

V. Limitations

Our T2P system has some limitations that affect generalizability of our approach:

- *Personalized recommendation:* Our T2P system does not take user-specific differences into account, and hence it would recommend the same policy for images with the same tag set, irrespective of the user who owns the image. This may work well in general cases for many users. Yet, users with very different privacy preferences compared to the norm may be unsatisfied with this approach. To address these special privacy needs, some extensions to our T2P system, such as adding weights to image tags according to individual user's privacy preferences, can be added in the future.
- *Policy representation:* Our policy representation was carefully chosen based on what options are provided by popular social network sites as well as previous research studies in this area [13], [24], [22]. In the future, we plan to study how to integrate more expressive policy languages and how they may affect the policy prediction accuracy and acceptability.
- *Experimental bias:* There are potential experimental biases which may affect the results of our empirical studies. This is because the experimental results are unavoidably limited by the data we collected from our participants, along with the photos and image tags that we used for our analysis. Yet, by asking users to provide a set

of heterogeneous images from the Flickr repository we have avoided possible fears of sharing personal content. Further, we let our participants choose among a large set of images, allowing them to pick the ones that most resonate with their preferences.

VI. Conclusions

We presented an effective privacy recommender system for images, which requires only small amount of information i.e., the image tags to recommend privacy policies for new images shared online. The proposed system is multi-pronged, and therefore able to tackle various scenarios including the challenging cold-start problem. Although some limitations exist, our empirical results showed that the majority of users expressed their satisfaction with regards to the proposed privacy policies.

VII. Acknowledgements

Portions of Dr Squicciarini's work was supported by National Science Foundation Grant 1453080 and Grant 1421776. Dr Caragea's work was supported by National Science Foundation Grant 1421970.

References

[1] A. Besmer and H. Lipford. Tagged photos: concerns, perceptions, and protections. In *CHI '09: 27th international conference extended abstracts on Human factors in computing systems*, pages 4585–4590, New York, NY, USA, 2009. ACM.

[2] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu. Can all tags be used for search? In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 193–202. ACM, 2008.

[3] J. Bonneau, J. Anderson, and L. Church. Privacy suites: shared privacy for social networks. In *Symposium on Usable Privacy and Security*, 2009.

[4] Bullguard. Privacy violations, the dark side of social media, 2014. http://www.bullguard.com/bullguard-security-center/internet-security/social-media-dangers/privacy-violations-in-social-media.aspx.

[5] Business Insider. We are now posting a staggering 1.8 billion photos every day on social media every day, 2014. http://www.businessinsider.com/were-now-posting-a-staggering-18-billion-photos-to-social-media-every-day-2014-5.

[6] H.-M. Chen, M.-H. Chang, P.-C. Chang, M.-C. Tien, W. H. Hsu, and J.-L. Wu. Sheepdog: group and tag recommendation for flickr photos by automatic search-based learning. In *Proc. of the 16th ACM International Conference on Multimedia*, pages 737–740. ACM, 2008.

[7] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Generalized louvain method for community detection in large networks. In *Proc. of Intelligent Systems Design and Applications Conference (ISDA)*, pages 88–93. IEEE, 2011.

[8] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 351–360, New York, NY, USA, 2010. ACM.

[9] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao. Tag-based social image search with visual-text joint hypergraph learning. In *Proc. of the 19th ACM international conference on Multimedia*, pages 1517–1520. ACM, 2011.

[10] L. Han, A. Kashyap, T. Finin, J. Mayfield, and J. Weese. Umbc ebiquity-core: Semantic textual similarity systems. In *Proc. of the Second Joint Conference on Lexical and Computational Semantics*, volume 1, pages 44–52, 2013.

[11] B. Henne, C. Szongott, and M. Smith. SnapMe if you can: Privacy threats of other peoples' geo-tagged media and what we can do about it. In *Proc. of the Sixth ACM conference on Security and privacy in wireless and mobile networks*, pages 95–106. ACM, 2013.

[12] L. Hollenstein and R. Purves. Exploring place through user-generated content: Using flickr tags to describe city cores. *Journal of Spatial Information Science*, 2010(1):21–48, 2010.

[13] P. Klemperer, Y. Liang, M. Mazurek, M. Sleeper, B. Ur, L. Bauer, L. F. Cranor, N. Gupta, and M. Reiter. Tag, you can see it!: using tags for access control in photo sharing. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 377–386. ACM, 2012.

[14] K. Liu and E. Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Trans. Knowl. Discov. Data*, 5:6:1–6:30, December 2010.

[15] A. Mazzia, K. LeFevre, and E. Adar, April 2011. UM Tech Report #CSE-TR-570-11.

[16] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, and G. Chen. Travel recommendation using geo-tagged photos in social media for tourist. *Wireless Personal Communications*, 80(4):1347–1362, 2015.

[17] T. Pedersen, S. Patwardhan, and J. Michelizzi. Wordnet:: Similarity: measuring the relatedness of concepts. In *Demonstration papers at HLT-NAACL 2004*, pages 38–41. Association for Computational Linguistics, 2004.

[18] A. Plangprasopchok and K. Lerman. Exploiting social annotation for automatic resource discovery. *CoRR*, abs/0704.1675, 2007.

[19] A. Salehi-Abari and C. Boutilier. Preference-oriented social networks: Group recommendation and inference. In *Proc. of the 9th ACM Conference on Recommender Systems*, pages 35–42. ACM, 2015.

[20] J. San Pedro and S. Siersdorfer. Ranking and classifying attractiveness of photos in folksonomies. In *Proc. of the 18th International conference on World wide web*, WWW '09, pages 771–780, New York, NY, USA, 2009. ACM.

[21] N. Sawant. Modeling tagged photos for automatic image annotation. In *19th ACM International conference on Multimedia*, MM '11, pages 865–866, New York, NY, USA, 2011. ACM.

[22] M. Shehab and H. Touati. Semi-supervised policy recommendation for online social networks. In *Proc. of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 360–367. IEEE Computer Society, 2012.

[23] A. C. Squicciarini, D. Lin, S. Sundareswaran, and J. Wede. Privacy policy inference of user-uploaded images on content sharing sites. *IEEE Transactions on Knowledge and Data Engineering*, 27(1):193–206, 2015.

[24] A. C. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede. A3P: adaptive policy prediction for shared images over popular content sharing sites. In *Proc. of the 22nd ACM Conference on Hypertext and Hypermedia*, pages 261–270. ACM, 2011.

[25] H. Sundaram, L. Xie, M. De Choudhury, Y. Lin, and A. Natsev. Multimedia semantics: Interactions between content and community. *Proceedings of the IEEE*, 100(9):2737–2758, 2012.

[26] J. Tang, S. Yan, R. Hong, G.-J. Qi, and T.-S. Chua. Inferring semantic concepts from community-contributed images and noisy tags. In *Proc. of the 17th ACM international conference on Multimedia*, pages 223–232. ACM, 2009.

[27] The Atlantic. How many photographs of you are out there in the world? http://www.theatlantic.com/technology/archive/2015/11/how-many-photographs-of-you-are-out-there-in-the-world/413389/, year=2015.

[28] X. Wan and J. Xiao. Single document keyphrase extraction using neighborhood knowledge. In *AAAI*, volume 8, pages 855–860, 2008.

[29] H. Xu, H. Wang, and A. Stavrou. Privacy risk assessment on online photos. In *Research in Attacks, Intrusions, and Defenses*, pages 427–447. Springer, 2015.

[30] Yatsko's Computational Linguistics Laboratory. Y-stemmer. http://yatsko.zohosites.com/y-stemmer.html.

[31] C. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt. Providing access control to online photo albums based on tags and linked data. *Social Semantic Web: Where Web*, 2, 2009.

[32] S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova. Privacy-aware image classification and search. In *Proc. of the 35th international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '12, pages 35–44, New York, NY, USA, 2012. ACM.

[33] Y. Zhang, R. Jin, and Z.-H. Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, 2010.

[34] N. Zheng, Q. Li, S. Liao, and L. Zhang. Which photo groups should I choose? A comparative study of recommendation algorithms in Flickr. *J. Inf. Sci.*, 36:733–750, December 2010.

[35] H. Zhou, Y. Yuan, and C. Shi. Object tracking using sift features and mean shift. *Computer vision and image understanding*, 113(3):345–352, 2009.