Portrait of a Privacy Invasion

Detecting Relationships Through Large-scale Photo Analysis

Yan Shoshitaishvili, Christopher Kruegel, Giovanni Vigna
UC Santa Barbara
Santa Barbara, CA, USA
{yans,chris,vigna}@cs.ucsb.edu

ABSTRACT

The popularity of online social networks has changed the way in which we share personal thoughts, political views, and *pictures*. Pictures have a particular important role in the privacy of users, as they can convey substantial information (e.g., a person was attending an event, or has met with another person). Moreover, because of the nature of social networks, it has become increasingly difficult to control who has access to which content. Therefore, when a substantial amount of pictures are accessible to one party, there is a very serious potential for violations of the privacy of users.

In this paper, we demonstrate a novel technique that, given a large corpus of pictures shared on a social network, automatically determines who is dating whom, with reasonable precision. More specifically, our approach combines facial recognition, spatial analysis, and machine learning techniques to determine pairs that are dating. To the best of our knowledge, this is the first privacy attack of this kind performed on social networks.

We implemented our approach in a tool, called Creepic, and evaluated it on two real-world datasets. The results show that it is possible to automatically extract non-obvious, and non-disclosed, relationships between people represented in a group of pictures, even when the people involved are not directly part of a connected social clique.

1. INTRODUCTION

In the last decade, our society has experienced the explosive growth of two disruptive technologies: digital photography and the Internet. Rather than the bulky film cameras of yesterday with limited (and difficult to distribute) film, we now have small, integrated digital cameras with nearly limitless storage capacity.

Individuals are using these technologies to take an increasing number of photographs of themselves and each other using their personal devices. Because photographs are digital, and the devices taking them are increasingly connected to the Internet, these photographs frequently end up on social network sites. Currently, for example, Facebook alone receives more than 350 million new photo uploads every day [10], while Instagram receives 60 million uploads [11]. Snapchat, a photograph messaging service, has recently skyrocketed well past both platforms, with users sharing over 700 million photographs every day [23]. This trend is only likely to increase as curators of social networks are constantly removing impediments to sharing. For example, photographs taken on modern Android smartphones can be automatically uploaded and shared on Google+, Google's social network.

Increasing convenience to share photographs gives interested entities a valuable dataset to mine for insights into individuals. Social networks already use modern face recognition techniques to automatically determine which of a user's friends or other users of the network are present in their photographs. While this information is currently used to assist users in tagging photographs of their friends, it could also be used to gain insight into aspects of the users' social behavior beyond what the users would otherwise have volunteered. Specifically, our intuition is that the relationship status of a person can be determined just by examining photographs in which the person in question appears. We observed that people tend to physically organize themselves differently in different situations, depending on their relationships to the other people in those situations. While such behavior might vary with the cultures and ages of the subjects, there are norms within a given demographic group. Therefore, we suspected that by analyzing large corpora of photographs of social situations, we can detect the relationships contained within.

We designed and implemented a system, *Creepic*, that, given a large set of photographs, performs face detection and recognition to identify the people in each photo, extracts statistical features, and utilizes machine learning to determine which of the people are involved in romantic relationships. This system is automated and, once trained on a subset of people with a known relationship status, can predict the relationship status of other pairs within the group. This information could be used for a variety of purposes, and represents a violation of privacy akin to non-consensual web tracking or cyber-stalking, as prior research has shown that almost 75% of social network users do not willingly disclose their relationship statuses [12].

Our approach is not intended to be used by general social networks, such as Facebook, as such networks have access to other useful metadata about their users and should leverage this data in addition to user photographs. Instead, there are several scenarios in which entities would have access to a sufficient amount of photographs to be able to carry out such an attack:

1. Photo messaging and sharing services, such as Snapchat and Flikr, have become the most popular way to share photographs. These services lack the rich metadata accessible by general-purpose social networks such as Facebook, but have extensive access to user photographs. Such services can use our technique to determine their users' relationship status for advertising purposes.
2. On Android devices, photographs are stored in a phone's *internal storage*. However, this storage is also used for other data, and permission to access it is requested by a large percentage of Android applications. In fact, the privacy-focused application permission monitoring app, APeWatch [2], does not consider this permission to be privacy-relevant. The attacks described in this paper could be carried out by any sufficiently popular application with this permission, or any library (such as an advertising library) included in such applications. Specifically, an advertising network could easily implement a module that performs face recognition on photographs and uploads the resulting metadata, allowing the advertiser to determine a user's relationship status.

3. Facebook applications can request access to "photographs of friends" when Facebook users install them. This access can be used to amass a large corpus of photographs. In fact, this is exactly the method we use to create one of our datasets for evaluation, retrieving almost half a million photographs with only 34 authorized users. While Facebook provides a privacy setting, dubbed "Apps others use," that allows users to control data access granted to applications installed by their friends, it is unclear how widely known and used this setting is. Anecdotally, the authors were surprised to discover the existence of this setting (and equally surprised that it was set to allow photograph access for applications) shortly before the submission date.

Additionally, the techniques presented in this paper can be used to attack the privacy of users that share no photographs at all. If a user appears in a sufficient number of photographs (for example, if their friends grant access to *their* photographs to Creepic, and the user appears in those photographs), Creepic may still determine the user's relationship status. In fact, through the use of modern face recognition techniques, this could be possible even in cases where the victim *has no account* on the social network at all, by analyzing photographs uploaded by those friends who do have such accounts.

We will describe the implementation of our system and show its evaluation on two datasets. The first dataset was assembled by developing a Facebook application that executes our approach over the photographs of the user (who must approve our application's presence) and their friends (whose approval is not required). A total of 34 volunteers installed our application, granting us access to theirs and their friends' photographs. From these 34 volunteers, were able to retrieve 448,936 photographs, which contained a total of 241,189 users. To protect the anonymity and privacy of the users involved in the experiment, we obtained IRB approval for this experiment.

A second dataset was assembled using the data from Zimbio.com, a celebrity tracking website that comprises 1,536,243 photographs of 2,616 celebrities. This dataset includes paparazzi photographs of celebrities (whether alone or with friends or significant others) as well as photographs of official events, making it a well-balanced dataset for Creepic's analysis.

We show that in the case of both datasets, *Creepic* is able to identify the dating couples with a high degree of accuracy. Specifically, we can detect over 60% of photographed relationships, including relationships whose participants had not shared their relationship status, while keeping false positive low.

Because automated relationship detection is in its infancy, a human might outperform Creepic's detection on a small scale. However, Creepic is built to function at a scale that would likely be impossible to handle by a human analyst. This is the scale at which an organization would utilize such an approach, and Creepic demonstrates that such large-scale privacy attacks are possible.

In short, this paper makes the following contributions:

- We develop a new class of privacy attacks on a person's privacy: by analyzing photographs of a person, insight can be gleamed into his/her relationships with other persons. This attack can be carried out over social networks or through any means that grant access the to photographs of a user.
- We develop a system, called Creepic, that leverages machine learning techniques to analyze a set of statistical features of photographs, and determine which users in those photographs are involved in romantic relationships.
- We evaluate our system on two real-world datasets with two different types of data (Facebook user data and celebrity photographs) to show its applicability in varied environments and its viability as an attack on the privacy of users of social networks.

2. APPROACH

The Creepic system performs several steps to detect relationships in a dataset of photographs of a group of people.

Prerequisites. Creepic receives, as input, a dataset of photographs, in which it will detect relationships.

Face Detection and Recognition. Creepic processes the images in the dataset to detect and extract faces. Extracted faces are then processed by the face recognition component, which performs face recognition to identify what faces belong to what people. This step creates an understanding of who is in what photograph, and where in the photograph they are.

Timespan Splitting. To reason about changes in relationships over time (and avoid mis-classifications stemming from such changes), Creepic splits the provided dataset into multiple datasets, each containing a separate timespan of images.

Feature Extraction. Creepic processes the dataset to extract features relating to the presence and locations of individuals in each photograph.

Classification. The feature set is processed by a binary classifier which classifies each pair as either being in a relationship ("dating") or not being in a relationship ("non-dating").

A diagram of this process is presented in Appendix A, for the visual reader. In this section, we will present the design-level description of Creepic. Implementation details and determination of threshold values will be discussed in Section 3.

2.1 Prerequisites

Creepic's analysis requires several inputs. The first is a dataset of photographs of people, in which relationships will be detected. The minimum amount of photographs to make meaningful predictions depends on the *density* of people in the photographs of the dataset. If the there are too few photographs of a pair of people, Creepic will be unable to make an accurate prediction of their relationship status. We discuss the effects of the number of photographs of a given pair of people on detection rates in Section 3.2.

Creepic can leverage time-based features in its analysis, but this requires the photographs to be timestamped. Since digital cameras embed the time when the photograph was acquired as metadata in photograph, this information is usually widely available. While non-digital photographs (specifically, scans of physical photographs with misleading or missing timestamps) do sometimes get uploaded to social networks, the number of such photographs is negligible.

Additionally, due to the face recognition technology used, Creepic requires a set of pre-identified *reference photographs* for each person to perform face recognition. In the case of modern social networks, people are quite eager to tag themselves and their friends in the photographs they upload, and these photographs can be used as reference photographs.

Unless Creepic's classifier has previously been trained on labeled relationship data, it also requires labeled relationship statuses for at least a subset of the people present. In order to more efficiently use its time-based features, Creepic also requires information on the start and end dates of these relationships.

2.2 Face Detection and Recognition

The first step carried out by Creepic is the detection of faces in the dataset. Each photograph is processed and a set of faces identified and extracted for processing in future steps. We claim no innovation in the field of face detection: Creepic uses existing algorithms to carry out this step of the process.

After extracting faces from the dataset, Creepic executes a face recognition algorithm to associate each extracted face to a person. This step uses the reference photographs provided to the system as input. Again, existing face recognition algorithms are leveraged to achieve this.

The output of this step is a list, for each photograph, of what people appear in the photograph and where in the photograph their faces are located.

2.3 Timespan Splitting

After the face recognition step, Creepic's dataset contains a set of people, the photographs in which those people appear, and information on the coordinates of their faces in those photographs. To enable it to reason about changes in relationships, Creepic splits this dataset into a series of sliding-window timespans containing images between certain dates. For example, if using a sliding window of 6 months with an interval of 3 months, a dataset spanning a year will be split into a timespan of photographs from January through June, one from March through September, one from June through December, and so forth.

For the remainder of the analysis, the same pair, in two different timespans, is considered to be two separate pairs. This means that, except for the timespan-level features described in Section 2.4.2, the timespans are analyzed in isolation. This allows Creepic to reason about changes in relationship status over time.

For timespan, Creepic identifies pairs of people that appear together in a sufficient amount of the timespan's photographs. Any pair that does not meet this threshold is discarded from the rest of the analyses for the given timespan.

It is feasible that a dataset might not include timestamp information for photographs. In the absence of this information, and the resulting inability to split the images into timespans, this step is omitted, and Creepic is unable to reason about changes of relationships over time.

2.4 Feature Extraction

Creepic's feature extraction component processes one timespan at a time to extract features for each pair of people in that timespan. Every timespan produced Creepic extracts two classes of features: image-level features, which utilize information in the photographs and spatial relationships between the faces of people present in those images, and timespan-based features, which reason about how the relationship of the pair of people changes over time.

Creepic generates these features by processing the timespans sequentially and generating a set of image-level features and a set of timespan-level features for every pair in the timespan's pair list. For each pair, the average of image-level features across all images in which they appear is recorded.

To guide the section of our features, we studied the situations under which a photograph containing a dating pair of people might be taken. We determined two orthogonal attributes of such situations: privacy and intimacy.

The first is the *privacy* of a situation. A private situation is one in which only a person's friends are present (for example, a private party). We have observed that photographs in these settings tend to be taken close to the subjects in question, and that most of the faces in the photograph are tagged or easily identifiable by face recognition. Conversely, a public situation is one in which a person and their friends attend a public event, and strangers are present in the photo. These photographs tend to be zoomed out, and some of the faces present will neither be tagged nor matchable via face recognition.

The second attribute that we have determined is the *intimacy* of a situation [1]. An intimate situation is one in which the pair is photographed without their friends (for example, a date) while a *social* situation is one in which the pair and their friends are photographed together. Specifically, one style of photograph that we encountered in photographs from the former situation is a "self-shot" or "selfie". In the absence of other people to take the photograph, one member of the pair holds the camera at arm's length and photographs them and their significant other. These photographs have distinguishing properties: they are zoomed in because the camera has to be held in someone's hand, and the faces of the couple cover a large portion of the photograph.

These attributes can be combined into the following four situations:

Private intimate situations. This is a type of situation in which the dating pair is together, with no other people present.

Public intimate situations. In this situation, the dating pair is together, but in a public place. We expect other, often untagged people to appear in photographs in this situation.

Private social situations. Private social situations comprise private parties and gatherings with friends. We expect most people in such photographs to be tagged, and for the photographs to be close-ups of the subjects. We expect a couple in this situation to mostly be close to each other, and do not expect another individual to frequently be physically between them.

Public social situations. A public social situation is one where a group of friends spends time together in a public setting, such as a club or restaurant. We expect to see many untagged faces in such photographs, and the images to be wider-angle to capture the action.

Because these situations may overlap slightly in terms of what we expect to see in a photograph, and our features are geared toward identification of dating pairs who participate in all situations, several of our features overlap.

2.4.1 Image-level Features

Creepic extracts several features based on information at the image level. It is important to note that the raw number of images that a pair of people shares does not explicitly factor in our computation. Since Creepic tracks only the average values, it is not biased toward pairs who take, or have taken, more photographs of themselves than other pairs.

Ambient light level. With the intuition that certain relationships might generate more photographs of events taking place in the evening, such as restaurant visits and parties, we determine the ambient light level of each photograph by converting the photograph to grayscale and taking the average color value.

Face distance. We calculate the absolute distance between the centers of the pairs of faces. This is driven by the intuition that people who are dating will appear closer together in photographs.

Face distance, size-adjusted. To accommodate zoomed-in photographs of pairs (where the absolute distance between the centers of faces will thus be large), we divide this distance by the average of the widths of the faces.

[1] We do not mean sexual intimacy, but rather, couples simply spending time alone.

Face-size ratio. Some faces are mis-detected to be close to each other due to perspective. The face of a person in the foreground can appear very close to the face of a person in the background, though the two are actually far apart. To compensate, we calculate a feature of the ratio between the areas of the two faces in the image.

Pair face-area coverage. To differentiate close-up and wide-angle photographs of pairs, we calculate the percentage of the image that is covered by the pair of people in question. This feature is inspired by "self-shots" of people, which are popular on social networks.

Total face-area coverage. A normalized measure of the area of the image that is covered by *all* of the people in the photograph is also calculated. This is done as a differentiation between private and public social situations, where the latter is expected to have a higher percentage of a photograph covered by faces.

Tagged count. To differentiate between intimate and social situations, we track the count of tagged (or recognized by the face recognition step) people in a photograph.

Untagged count. We expect photographs in public situations to have a larger amount of untagged, and unrecognized, faces. The intuition behind this feature is to identify how often the pair of people are in such situations.

Vertical positioning. Differentiating between pairs that are dating and pairs otherwise related (for example, the relation of a parent and child) requires special attention. People tend to feature in many photographs of themselves with their young children, which otherwise confuses some of the other features. Since parent-child photographs tend to have the parent's face higher in the photograph than the child's face, the relative vertical position of the centers of the faces helps differentiate these pairs from dating pairs.

Betweeners count. The position of people in a photograph in social situations can be an important indicator of relationship status. We observed that a pair of individuals who are dating tends to have fewer people in between them in a photograph than one that is not. We determine the number of people between two people by evaluating if another face rectangle intersects the lines between the corners of the face rectangles of the people in question.

Competitor count. We define a "competitor" as any face that is closer to either of the faces of the pair of people in question than the two are to each other, expecting that a dating pair will have less competitors, on average, than a non-dating pair.

An example photograph with extracted features is reproduced in Appendix C.

The features listed above are not meant to predict a relationship status in isolation; instead, they are meant to be utilized together. As such, it is easy to form a counter-argument against a single feature: for example, the "Ambient Light Level" feature might fail if a couple is afraid of the dark (or, possibly, lives on a beach), and thus always takes photographs in the daytime. While this might break that particular feature, the couple may still behave as expected in regards to the other features, enabling Creepic to detect their relationship. Of course, a couple might abandon convention and fail to be described by any of our features, in which case Creepic would be unable to detect this couple.

Another important consideration is corner cases of the features. For example, one might argue that the "Face distance, size-adjusted" features could be confused by two faces that are close together, but extremely thin. In such a case, this feature would detect the faces as being very far away from each other, and introduce noise into the analysis. It must be kept in mind that these features deal with human faces, not abstract geometric constructs, and human faces have limited range of sizes that tends to not trigger such corner cases. In other words, human beings do not have inch-wide faces.

2.4.2 *Timespan-level features*

If a dataset contains more than one timespan, the differences between a pair in two timestamps can give more insight into this pair's relationship status and its change over time. To detect these changes, Creepic extracts a set of features based on a pair's presence and features in prior timespans.

Betweeners entropy. If there are people physically separating a pair, the *entropy* of these people (i.e., how stable the identity of the betweener is) can be a crucial clue to the dating status of the pair. We expect that dating pairs will be either together or, barring that, separated by a varying set of people. However, if a given pair is frequently separated by one or a small set of individuals, those individuals might be the dating partners of one member of a pair.

Competitor entropy. Similar to the betweener entropy, we compute the entropy of the Competitors detected across all photographs in a timespan.

Frequency of association. With the intuition that a dating pair will be more likely to have photographs together in a previous timespan, we include as a feature the percentage of prior timespans in which the pair in question has more than the threshold number of photographs.

Drift features. Creepic computes what we term a "drift" feature for each other feature. Drift feature are, for each other feature, the ratio of the feature between the current timespan and the prior timespan in which the pair appeared in enough photographs. The goal behind drift features is to detect a change in a relationship over time as a pair transitions into or out of an intimate relationship.

2.5 Classification

Creepic uses a binary classifier to determine, for each timespan, which pair of people are dating. The classifier operates on the features extracted in the previous step.

The prediction itself is based on a Support Vector Machine, a common binary classifier used in machine learning. Creepic trains an SVM with a labeled subset of relationship data or reuses a previously-trained SVM from a different dataset with labeled relationships.

SVM training If a trained SVM is not provided, Creepic splits the data into a training and testing set. From each timespan, it selects a subset of pairs are known, from the labeled relationship data, to have been dating during that timespan. If timestamps are not available for the start and end of labeled relationships, dating pairs are assumed to be dating throughout all timespans. For the non-dating portion of the training set, Creepic selects pairs according to the following criteria: for all pairs (A,B) such that (A,B) is not in the dating set, Creepic selects (A,B) for the non-dating set if there exists a person C such that (A,C) or (B,C) are in the dating set. In other words, to avoid training on unlabeled relationships, Creepic includes pairs assumed not to be dating *because one member of that pair is dating someone else*.

Training data for all timespans is then combined into a single training set. The same pair, in two different timespans, is considered to be two separate pairs. This is done so that Creepic can recognize relationships in different points in time. Specifically, in the absence of timestamps on relationship data, noise might be introduced into the training set, as statistics from before or after the people were dating will make it into the features of dating pairs.

Since the number of dating and non-dating pairs in real-world data is disparate (specifically, there are considerably more of the latter), the training set ends up being dominated by non-dating pairs and the SVM's prediction suffers. To balance this, we limit the number of non-dating pairs to a set threshold, depending on the number of dating pairs.

After the SVM is trained, it can be used to predict relationships in a dataset.

SVM Prediction. The trained SVM is used to predict relationships among pairs in all timespans of the dataset. Each pair is classified as either a dating pair or a non-dating pair.

Result Filtering. Creepic attempts to filter the results to reduce false positive rates. One frequent cause of false positives is the misdetection of a group of friends as a complete (or near-complete) graph of dating pairs. When encountering such situations, Creepic leverages the fact that relationships tend to be monogamous. More formally, Creepic removes a dating classification from all pairs that match the following criteria: for each pair (A,B) that are predicted as a dating pair, Creepic removes this classification if there exists a person C such that (A,C) are predicted as dating or (B,C) are predicted as dating. This filtering technique helps Creepic maintain a lower false positive rate, at the expense of a lower rate of true positives (i.e., missing polyamorous relationships). However, splitting pairs across timespans still allows Creepic to detect a given person dating different people at different times.

3. IMPLEMENTATION

We made several implementation choices in the course of implementing Creepic. These choices mostly have to do with the technologies that we chose for our implementation, and specific values we determined to be optimal for various thresholds.

3.1 Tools

While open image-processing tools are available, few achieve good results in face detection and face recognition. Thus, we faced a technical challenge in adapting the tools that are available to our use, and keeping results sufficiently precise.

3.1.1 Face Detection

We utilized the OpenCV [18] library to implement the face detection capabilities in Creepic. OpenCV implements face detection in the form of Haar cascade classifiers [25], and provides Python bindings that we utilize in Creepic. As a measure to reduce the number of misdetection of non-face object as faces, we first use OpenCV's face detection cascade[2], and then we use OpenCV's eye detection cascade[3] on the detected face. This successfully filters out some of the false face detections at the cost of detecting a lower amount of actual faces.

3.1.2 Face Recognition

Some datasets, such as our Facebook dataset, contain metadata that provides coordinate information of tagged faces. Creepic can utilize this information directly by matching it against the detected faces in a photograph. This allows for the reuse of existing capabilities in certain social networks.

When such information is not available, Creepic utilizes face recognition methods offered by the Facerec library [24]. We apply additional filtering depending on what information is available for the dataset to reduce noise in Creepic's analysis. For example, our celebrity dataset from Zimbio contains information as to which celebrities are present in which photographs (but not where their faces are located). In this case, we limit the nearest-neighbor classifier to consider only faces of the specified celebrities when recognizing faces in an image. In the case of a standard social network, the face recognition can be limited to the faces of a users' friends to save computing resources, although this would exclude analysis of users who do not participate in the social network.

3.1.3 Machine Learning.

We use LIBSVM [6] for the machine learning component of Creepic. After feature extraction, the features are passed to LIBSVM for training and relationship prediction.

3.1.4 Distributed Computation.

To support the amount of data that must be processed, our implementation of Creepic comprises one master node and a set of 42 worker node, with 4GB of memory per node.

The master node handles the initial retrieval of the dataset, including communications to Facebook to retrieve the Facebook dataset and Zimbio to retrieve the celebrity dataset. It then dispatches tasks to the worker nodes, who download the photographs themselves, carry out the face detection and face recognition (the latter, being a memory-intensive process, is why each node needs 4GB of memory), and perform the feature calculation. The results from those tasks are aggregated by the master, which carries out the final SVM training and prediction.

3.2 Thresholds

Creepic's implementation requires values for two thresholds: the minimum number of photographs in which a pair must appear, within one timestamp, to be considered for the analysis, and the ratio of dating and non-dating pairs to include in the training set. The former is aimed at reducing noise in the features, due to statistical instability of pairs with fewer photographs, and the latter is to avoid overwhelming the SVM classifier with either dating or non-dating pairs.

By analyzing Creepic's results with different values of these thresholds, we determined a threshold of 5 minimum images per pair per timestamp, and a non-dating/dating pair ratio of 2 for training the classifier. We present our protocol for determining these thresholds in Appendix D

4. EVALUATION

To determine the effectiveness of our approach, we evaluated our implementation of Creepic on two real-world datasets. We present the results in this section, and discuss their interpretations and implications in Section 5.

4.1 Datasets

We chose a Facebook dataset, obtained by requesting volunteers to install a Facebook application that we developed for this purpose, and a celebrity dataset, acquired from the celebrity tracking website Zimbio.com. These datasets show the application of our approach under two different circumstances: one with the full cooperation of the users being analyzed, and one carried out without the subjects' awareness. Additionally, the two datasets feature different types of photographs: personal photographs of people and their families and friends, and professional photographs of celebrity events.

4.1.1 Facebook Dataset

Our Facebook dataset was aggregated by asking volunteers to install a custom Facebook application that was written by us. This

[2] We use the "haarcascade_frontalface_alt2.xml" cascade for face detection.

[3] We use the "haarcascade_eye.xml" cascade for eye detection.

application can retrieve the volunteer's relationship status and the relationship status of all of their friends using the Facebook API. Of course, these statuses are only used for verifying Creepic's results, not for the detection itself. Our application can also access the list of photographs in which the user or their friends are tagged. Facebook provides a list of tags for each photograph in the form of the coordinates of face centers and a reference to the associated user. However, because most of our features require the encompassing rectangle of the face, which Facebook does not provide, we must still perform our face detection step on the photographs themselves. Rather than carry out face recognition on the detected faces, we simply check which tag locations intersect with which face rectangles, and mark those rectangles as belonging to the user associated with the tag.

A total of 34 users volunteered to install our application. From those 34 users and their 11,942 unique friends, we retrieved 448,936 photographs with a total of 241,189 unique tagged users. After matching detected faces to provided tags, we identified 370,526 faces in 217,220 photographs. 86,689 of these photographs contained more than one identified face.

The demographics of the Facebook dataset may be found in Appendix B.

For the purposes of training and cross-validation, we leveraged the relationship statuses of the volunteers and their friends. Unfortunately, due to limitations in Facebook's Graph API, we were unable to retrieve the relationship *history* of the users or the timestamp of the start of their current relationships. Because of this, the relationships in the Facebook dataset contain no timestamps, and each relationship is assumed to persist throughout every timespan of the dataset.

We processed this dataset in 8 hours, the bulk of which was spent on face detection. Feature extraction was performed in 30 minutes, and training and classification in less than 2 minutes. It is important to stress that the processed photographs included *all* of the photographs that the volunteers and their friends had ever uploaded to Facebook. Adding new photographs to this dataset would only incur processing time for face detection and feature extraction on the additional photographs (which is negligible compared to the full set), as well as the full 2-minute cost of the training and classification.

Ethics. We designed the experiment as carefully as we could to avoid endangering the privacy of either the volunteers or their friends. Our volunteers were informed about the type and amount data that we would be accessing and the purpose toward which this data would be accessed, but care had to be taken since the volunteers' friends did not explicitly consent to the experiment (although our use of their data is covered under the privacy policy that they agreed to upon joining Facebook and, in fact, many Facebook application regularly access photographs of friends). Photographs accessed by the system are only used for face detection and then discarded. The ages of users are retrieved only to understand the demographics of the dataset, and the statuses of users are retrieved during each experiment and discarded after being used for training and cross-validation. All data transmission took place over SSL-protected channels. Other than the data of which users appear in which photographs and Creepic's predictions, no privacy-sensitive data was ever stored. Even the photograph metadata and Creepic's predictions are purged after every experiment is concluded, and Creepic's predictions are never explicitly revealed to the volunteers, to preserve the privacy of their Facebook friends.

We obtained IRB approval for this experiment (including the unconsented access to friends' photographs), and are thankful to the IRB for assisting us in every step of designing this experiment protocol in a privacy-aware way.

4.1.2 Celebrity Dataset

We retrieved the celebrity dataset from Zimbio.com, a website that tracks celebrity rumor and hosts paparazzi and publicity photos of celebrities. Each image in the dataset is tagged with the celebrities that appear in that image, but not with information about the coordinates of their faces. To determine the coordinates of each celebrity's face, we utilized our face detection to extract faces from every image. As reference photographs for each celebrity, we used all images in which that celebrity was the only one tagged by Zimbio.com and in which we detected only one face. After training our face recognition on these reference photographs, we performed our face recognition step against the rest of the dataset.

The dataset comprises 1,536,243 photographs of 2,616 celebrities and the dating history of each celebrity. After face detection and recognition, we successfully identified a total of 1,135,551 celebrity faces in 1,018,866 of the images. Out of these, 107,085 images contain more than one celebrity face, with a total of 223,770 celebrity faces being recognized in such images. The images with only a single celebrity were sampled for representative photographs. To keep memory usage of the face recognition component manageable, we randomly sampled 75 representative photographs for each celebrity.

While one might expect a dataset of celebrity photographs to be biased toward relationship photographs, we found that this is not the case. Celebrities are photographed in a wide range of settings, including many professional environments such as movie sets or other social gatherings. This provides a balanced dataset for our analysis.

Relationship labels in this dataset contain the start and end date of the relationships, generally accurate to the month. 917 of the celebrities have at least one labeled relationship, with a total of 2,684 relationships in the dataset.

The face detection step for the photographs took 20 hours, with an additional 8 hours required for face recognition on the images containing more than one celebrity. Feature extraction was performed in 2 hours, with the training and classification requiring less than 10 minutes. Since face detection, recognition, and feature extraction are completely parallelizable, increasing the amount of worker nodes would linearly decrease the required number of time for all but the training and classification steps.

4.2 Feature Selection

Between the image-level and timespan-level features described in Section 2.4, Creepic trains its SVM on 28 features. In order to better understand the relative importance of these features, we computed the information gain provided by each feature for both datasets. The results are presented in Table 1.

It is interesting to observe that different features are of different importance for the two datasets. The difference in feature weights between the datasets has two main implications.

The first implication is that relationships in different situations are suited to detection by different types of features. For example, the most important feature for the Celebrity dataset is the frequency of association between two people. However, this feature does not figure very heavily in the Facebook dataset. The reverse is true of, for example, the average number of competitors: it is the most important feature in the Facebook dataset, but is less important in the Celebrity dataset. Upon analysis of the datasets, we concluded that this is because of the pattern of photographs in the Celebrity dataset: Celebrities are most frequently photographed either with coworkers (for example, other actors in a film) at official events with whom they do not necessarily associate on a regular basis or, alternatively, with their significant others (generally by paparazzi), with whom they associate regularly. On the other hand, the Facebook dataset contains more photographs of everyday life, in which association

Feature	FB Gain	Zimbio Gain
Ambient Light Level	0.01	0.01
Face distance	0.13	0.02
Face distance, size-adjusted	0.42	0.12
Face-size ratio	0.03	0.05
Pair face-area coverage	0.21	0.11
Total face-area coverage	0.08	0.01
Tagged count	0.65	0.67
Untagged count	0.65	0.32
Vertical positioning	0.04	0.08
Betweener count	0.84	0.56
Competitor count	1.0	0.43
Betweener entropy	0.67	0.29
Competitor entropy	0.56	0.56
Frequency of association	0.15	1.0
Ambient Light Level (d)	0.04	0.293
Face distance (d)	0.04	0.25
Face distance, size-adjusted (d)	0.06	0.25
Face-size ratio (d)	0.04	0.29
Pair face-area coverage (d)	0.04	0.29
Total face-area coverage (d)	0.04	0.29
Tagged count (d)	0.04	0.32
Untagged count (d)	0.11	0
Vertical positioning (d)	0.04	0.25
Betweener count (d)	0.29	0.56
Competitor count (d)	0.56	0.25
Betweener entropy (d)	0.29	0.15
Competitor entropy (d)	0.06	0.35
Association frequency (d)	0.16	0.76

Table 1: Normalized information gain for Creepic's features for the Facebook and Celebrity datasets. (d) signifies drift features.

frequency is less useful in differentiating between daters and non-daters.

The second implication of these differences is that Creepic is able to adapt its analysis to detect relationships in both datasets. To avoid overfitting and focus on the most relevant features, we discard any features with an information gain less than 0.1 during the machine learning step. This leaves us with 13 features for the Facebook dataset and 21 features for the Celebrity dataset.

4.3 Results

We used Creepic to predict relationships for both the Facebook and the Celebrity dataset. In our experiments, we utilized a timespan of 140 days, stepping 70 days for each successive timespan and resulting in an overlap of 70 days between consecutive timestamps. The Celebrity dataset had 84 such timespans, while the Facebook dataset had 41. As mentioned in Section 2.3, pairs are considered different across different timespans. The Celebrity dataset contained 1,569 dating pairs and 7,936 non-dating pairs which had more than 5 photographs within a timespan, while the Facebook dataset contained 532 dating and 8,730 non-dating pairs with more than 5 photographs within a timespan.

For each experiment, we evaluate Creepic via a 10-fold cross-validation of the dataset by randomly sampling 10% of the dating pairs for the dating portion of the training set. This 10-fold cross-validation was performed 10 times, for a total of 100 classification runs per experiment. We then aggregated the 100 classification runs and computed the Matthews correlation coefficient (MCC, a value signifying the quality of a binary classifier, with 0 being random and 1 being perfect classification [16]) for the entire set. The resulting detection rates, along with the Matthews correlation coefficient are presented in Table 2.

4.3.1 Prediction rate

For both datasets, Creepic achieved a true positive rate of over 60% for both datasets, while maintaining a false positive rate of 11.8% for the Facebook dataset and 8.3% for the Celebrity dataset. In this context, a true positive is the detection of a pair of people by Creepic in which not only are both people in a relationship but, specifically, they are in a relationship with each other. That is, in a complete social graph, the possible number of such pairs in a dataset is $\frac{N^2-N}{2}$, with N as the number of people in the dataset. Likewise, a false positive is a pair of people who are not in a relationship with each other, even if one or both of these people are in a relationship. That is, some of the people in the False Positive pairs *could* be dating, they are just not dating the other member of the pair.

4.3.2 Baseline comparisons

We implemented several "naive" solutions as a baseline to compare to Creepic: randomly guessing a partner for every person, random guesses for each pair, random guesses knowing the number of pairs that are dating, and a classifier based on the number of photographs in which a pair appears alone.

Randomly guessing a partner, out of the people that appear with them in photographs, poses some problems. For example, the average person is present in photographs with with 26.2 other people in our Facebook dataset, and this method would achieve a 3.8% true positive rate. In the Celebrity dataset, the average person is present in photographs with 8.9 other people, with the resulting true positive rate of 11.2%. However, this approach is user-centric as opposed to being pair-centric and, other than the true positive rate, is hard to directly compare against our approach since it is hard to reason about how it would determine non-dating pairs.

Randomly classifying each pair into dating or non-dating will, of course, produce a 50% true positive and 50% false positive rate in both datasets. Being random, these results have a Matthews correlation coefficient of 0. It is important to note that, aside from failing to actually predict anything, this approach would yield multiple predicted dating partners per person, while Creepic achieves a higher true positive rate and much lower false positive rate while predicting only a single dating partner per person.

A random detection of K pairs as dating, where K is the number of true relationships in the dataset, yields a true positive and false positive rate of just 16.5% for the Celebrity dataset and a true positive and false positive rate of just 5.7% for the Facebook dataset. Again, both results have a Matthews correlation coefficient of 0.

We also carried out a naive analysis based on the number of photographs in which people appear alone, classifying any pair with more than 3 "Selfie“ photographs as dating. This "Selfie" experiment resulted in 37.9% true positive rate and 47.6% false positive rate (with an MCC of -0.072, indicating near-randomness) for the Celebrity dataset and a 35.0% true positive and 15.3% false positive rate (with a non-random, but still low, MCC of 0.124) in the Facebook dataset. It is notable that this naive classifier achieves moderately different false positive rates for the two datasets. From manual analysis of the "Selfie" photographs, this seems to be due to a combination of different social behavior between celebrities and people in the Facebook dataset. Creepic automatically adapts to these differences between the two datasets by recognizing that the relevance of its features differs between them, and this adaptability is one of our system's strength.

4.3.3 Identifying unlabeled relationships

To evaluate Creepic's ability to detect unlabeled relationships,

Experiment	Timespans	Labaled dating pairs	Labeled nondating pairs	True positives	False positives	MCC
Facebook - Creepic	41	532	8,730	61.6%	11.8%	0.516
Facebook - Random K Sample	41	532	8,730	5.7%	5.7%	0.0
Facebook - Random Guess (per pair)	41	532	8,730	50%	50%	0.0
Facebook - Selfies	41	532	8,730	35.0%	15.3%	0.124
Celebrity - Creepic	84	1,569	7,936	65.8%	8.4%	0.592
Celebrity - Random K Sample	84	1,569	7,936	16.5%	16.5%	0.0
Celebrity - Random Guess (per pair)	84	1,569	7,936	50%	50%	0.0
Celebrity - Selfies	84	1,569	7,936	37.9%	47.6%	-0.072

Table 2: The averaged results of 10 executions of 10-fold cross-validation by Creepic on the Facebook and Celebrity datasets, along with the baseline comparisons. Total pairs and dating pairs are aggregated for all timespans. MCC is Matthews correlation coefficient.

we manually analyzed the reported false positives in the Celebrity dataset and enlisted our volunteers' help in further analysis of the Facebook dataset.

A random sampling of false positives in the Celebrity dataset revealed that a tenth of these detections were, in fact, actually dating pairs. For such pairs, Zimbio.com was simply missing their relationship information from the database, which caused us to mark the result as a false positive. We verified such pairs by finding descriptions of the relationships in question on other celebrity tracking websites.

For the Facebook dataset, we further analyzed false positives by providing random samplings of Creepic's dating and non-dating pair selections, mixed with random pairings of the volunteers' friends, to our volunteers and asking them to designate which of the pairs were dating. In order to preserve the privacy of the volunteers' Facebook friends, we did not reveal which of the pairs were randomly created, which were Creepic's dating predictions, and which were non-dating predictions. This phase of the experiment, along with the rest our Facebook experiment, was carefully designed with guidance from the IRB to preserve the privacy of the volunteers and, especially, their friends. Our volunteers' responses revealed an 8% occurrence of actual, but unlabeled, relationships among the false positives reported by Creepic.

This result has two implications:

- Creepic's true positive and false positive rates are lower and upper bounds, respectively.
- Creepic is actively violating user privacy by detecting relationships which are not actively publicized by the people involved.

4.3.4 Characteristics of detected pairs

It seems intuitive that Creepic will have different degrees of success with pairs that appear in different numbers of photographs. To gain insight into how this affects Creepic's detection rate, we computed the detection rates of relationships according to the number of photographs of that pair that are present in our datasets. We found that Creepic required many more photographs for the Celebrity dataset than for the Facebook dataset. Specifically, Creepic can achieve a true positive rate over 50% with an average of 12 photographs of a given pair of users for the Facebook dataset. A similar rate for the Celebrity dataset requires 100 photographs. This is visualized in Figure 1).

We analyzed the datasets themselves for insight into this difference and concluded that the celebrities in our dataset frequently have their pictures taken with other celebrities (co-stars, fellow artists, etc.), rather than simply with their friends or significant others, and this introduces a higher amount of noise into our analysis, requiring more photographs of the pairs to compensate.

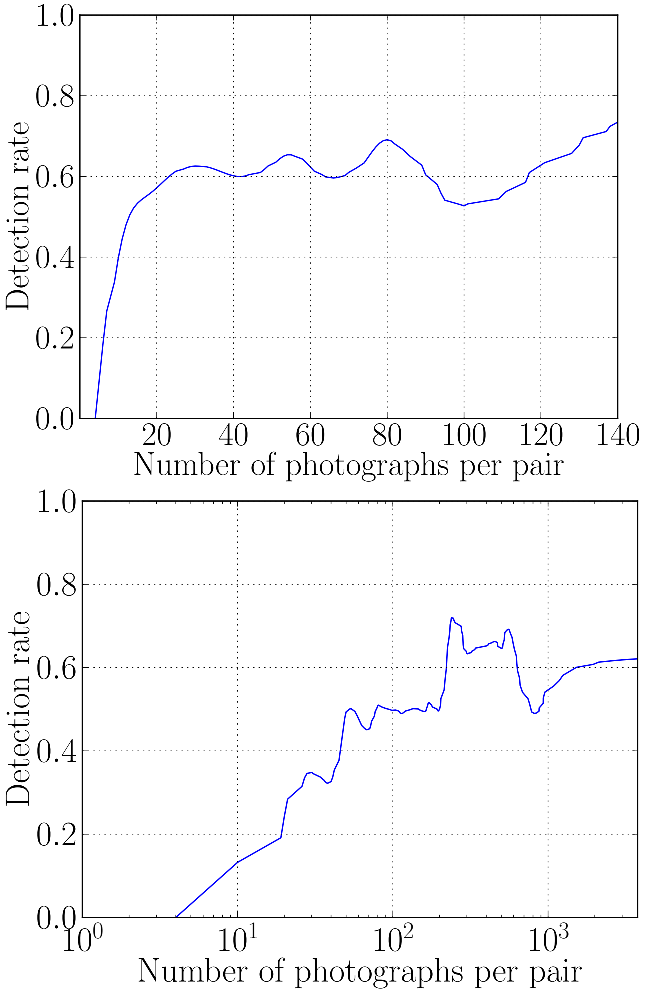


Figure 1: Creepic's detection rate for the Facebook (above, linear X axis) and Celebrity (below, logarithmic X axis) datasets according to the amount of photographs of that pair are present in the dataset.

4.3.5 Relationship detection timing

Since the Celebrity dataset contains start and end dates for Celebrity relationships, we were able to evaluate Creepic's relationship change detection on this dataset. To do this, we kept track of the first and last timespan in which Creepic detected a pair of celebrities as dating. We then calculated the detection accuracy for both the start and the end of the relationship. For the start of the relationship, we tracked the difference in time between the start of the timespan in which Creepic first detected the relationship and the timestamp of

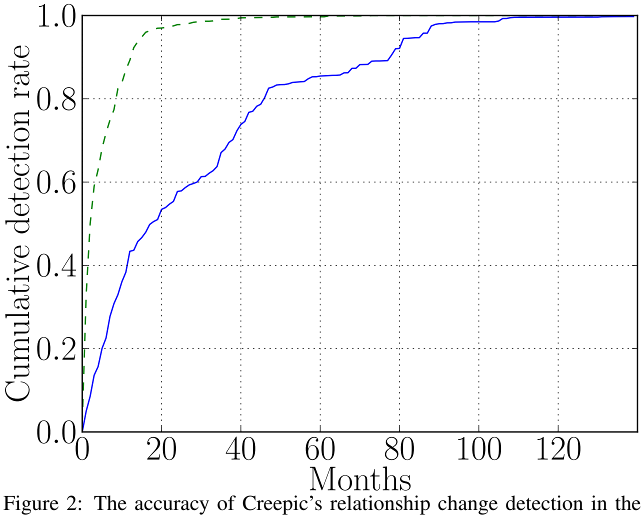


Figure 2: The accuracy of Creepic's relationship change detection in the Celebrity dataset. The X axis tracks the accuracy in months, and the Y axis provides the cumulative amount of relationships whose start (solid lines) and end (dotted lines) times are detected with an accuracy within that many months.

the first photograph that Zimbio.com had after the listed start of the relationship in the dataset. Similarly, for the end of the relationship, we tracked the difference in time between the end of the timespan in which Creepic last detected the relationship and the timestamp of the last photograph that Zimbio.com had before the listed end of the relationship in the dataset. Of course, the interval of the sliding window imposes an upper limit on the level of granularity that Creepic can achieve.

We found that Creepic excels at detecting the termination of relationships, even in cases where the parties involved continue to be photographed together (which is, for example, quite common in celebrity relationships). Over 80% of breakups were detected within 9 months. However, relationship start detections were less promising. Creepic made over 50% of its detections over 18 months after the start of the relationship. In analyzing the data, we concluded that this is due to the nature of many celebrity relationships. Except for those celebrities that are famous enough to be pursued heavily by paparazzi, the significant others of most celebrities begin to commonly appear in photographs after the relationship is more established, and might only appear in several snapshots before that. We detail the detection accuracy in Figure 2.

The Celebrity dataset, with its start and end dates and large range of relationship durations, can also provide insight into the *types* of relationships that Creepic best detects. We calculated the detection rate of relationships, grouped by the duration of the celebrity relationship, and found that Creepic excelled at detecting relationships lasting longer than 4 years, detecting over 60% of such relationships. For relationships shorter than 4 years, Creepic's detection rate was proportional to the relationship duration, becoming lower for shorter relationships. However, Creepic still detected over 20% of relationships lasting over 6 months. The full range of detection rates is presented in Figure 3.

4.4 Precision/Recall Trade-off

Because the number of non-dating couples dominates the number of dating couples, even a small false positive rate leads to a large drop in precision [4]. For example, with a true positive rate of 65.8% and a false positive rate of 8.4% for the Celebrity dataset, and assuming a 10% occurrence of unlabeled dating relationships among the non-dating pairs, Creepic's precision (the chance that a detected dating relationship is actually a dating relationship) is 35.5%.

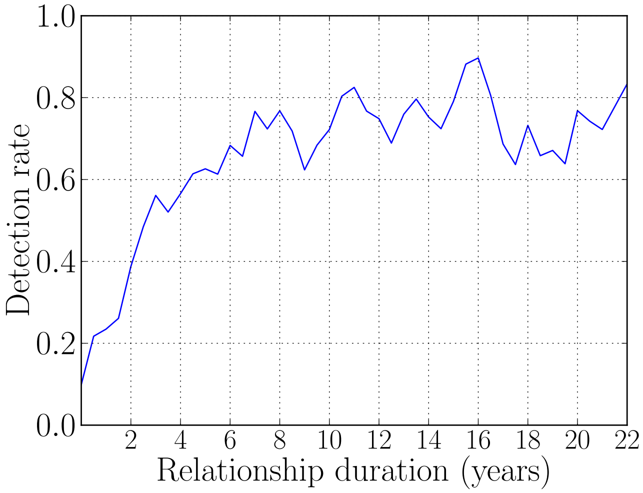


Figure 3: Creepic's detection rate for different durations of relationships in the Celebrity dataset.

This represents a typical trade-off in machine learning techniques: by modifying the importance of a true positive over a false positive, we can maximize Creepic's precision at the expense of recall. For example, for the Facebook dataset, we can achieve a precision of over 80% if we accept a 30% recall rate (30% true positive rate and 0.4% false positive rate), and a 100% precision (i.e., no false positives) with a recall of 23%. This trade-off is diagrammed, in the form of Creepic's ROC curve, in Figure 4.

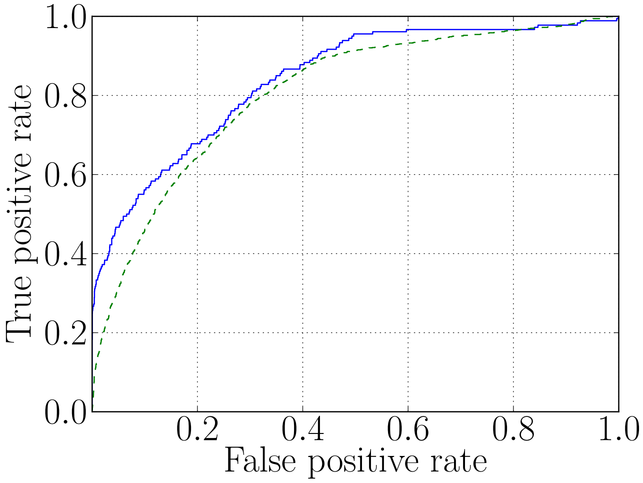


Figure 4: The ROC curve for Creepic's relationship classification. The dotted line is for the Celebrity dataset and the solid line is for the Facebook dataset. The AUC for the former is 0.81 and for the latter 0.84.

Thus, if high recall is critical, Creepic could be used as a prefilter to make it feasible for a human analyst to analyze large datasets (raising the precision with the help of a human analyst while maintaining a high recall), or in a fully automated fashion (sacrificing recall for a higher precision). In the former case, the reduction in the dataset greatly increases the feasibility of human analysis. In the case of the Facebook dataset, Creepic's default detection settings reduce the dataset that needs to be verified to 1,357 pairs from the original 9,262. For the Celebrity dataset, this reduction is from 9505 to 1669. A further discussion of this effect, with measurements of dataset reduction for different values of precision and recall and a presentation of our ROC curve, is presented in Appendix E.

5. DISCUSSION

Creepic's results are significant: Creepic is not just guessing whether a given person is dating or not. Instead, it is determining *who* that person is dating from the pool of other users. Additionally, Creepic does this with no information other than what it extracts from images of these people, and achieves a low false positive rate in its detection. It does this while adapting to the different social behaviors that may be prevalent in provided datasets, such as the different social behaviors between celebrities (in the Celebrity dataset) and Facebook users (in the Facebook dataset). Finally, we are able to detect the timing of the start and end of relationships with reasonable accuracy, giving rise to additional uses of the information that our system produces.

While the datasets we chose came from Facebook and Zimbio.com, the approach behind Creepic (and, indeed, its implementation) is generic, and can be used for any social network with photographs, including Instagram, Google+, Flickr, or any of the myriad of other options.

Limitations. A core technical limitation of Creepic is the limited accuracy of public face detection and recognition techniques, which can be as low as 64% [20]. This problem is further exasperated by social network users, as a user might jokingly tag some inanimate object as one of their friends, introducing noise into the face recognition stage and further reducing the effectiveness of Creepic's detection. If a face detection technique with higher accuracy were available, Creepic would be able to achieve better results.

There are several situations under which Creepic's approach would fail to properly identify relationships. To begin with, our statistical analysis could be manipulated by a person with prior knowledge of our approach by uploading images specifically tailed to poison our features. While there are countermeasures that Creepic could take, our approach assumes that the population being analyzed is not attacking the analysis system we developed.

Additionally, Creepic's analysis was carried out on a dataset comprising mostly members of Western culture. While it's possible that it would be less effective due to different behaviors of other cultures, we designed Creepic to be adaptable to different importance of different features among cultures.

Countermeasures. One of the troubling implications of Creepic is the difficulty of evading its analysis. A core issue is the difficulty of controlling access to photographs. Controlling the distribution of photographs requires a good working knowledge of the privacy settings of any used social network. For example, if user A grants access to his photographs to user B on Facebook, and uses Facebook's current default privacy configuration, any application that B has installed can request permission (from B as opposed to A) to access these photographs. Since providing these permissions is a requisite for using many Facebook applications, many users acquiesce. In fact, we used this exact functionality to carry out an evaluation of our approach, and the majority of the relationships we detect are due to photographs that we retrieve from the friends of the user who participated in the evaluation and, in many cases, even from the friends of those friends.

The number of photographs that we were able to collect from 34 volunteers and their friends suggests that privacy options to combat such photograph distribution, though they exist, seem to be seldom used. Educating users about this possibility, or making the default privacy settings of social networks more strict, would reduce the ability of external entities to carry out this attack. However, any photographs that a user uploads to a social network are still vulnerable to being used, by that social network, for relationship detection.

More troublesomely, even if an individual decides to avoid social networks altogether, their privacy can still be invaded. Unless they forbid their friends from taking any pictures of them at all, analysis of their social behavior could still be carried out based on photographs uploaded by their friends. Asking friends to not take any pictures of one's self can adversely affect one's social life, and we expect that few people would decide to follow this route, instead leaving themselves open to photograph-based privacy invasion.

That being said, there are some possible countermeasures that can be taken to evade this analysis, representing an attack against different phases of Creepic's analysis:

1. A browser extension could be developed to encrypt photographs on upload to a social network, and decrypt them when they are viewed. The encryption keys could then be privately shared between a user and their friends. This has the drawback of requiring all involved parties to install the extension, but would effectively foil Creepic by making it impossible to extract faces on a large scale.

2. Creepic's face recognition step could be resisted by intentionally mistagging individuals in photographs. By introducing enough noise into the training data of the face recognition step, an individual could severely hamper Creepic's prediction rate for their relationships.

3. An interesting countermeasure would be the generation of fake images, containing many people in random positions and situations, to pollute the features extracted by Creepic during feature extraction. This would then lead to random predictions from the classifier due to the amount of noise in the analysis.

All of these countermeasures introduce some level of discomfort for users, ranging from the requirement to install additional software to the presence of random or mistagged photographs among the users' albums, but would limit Creepic's effectiveness in relationship detection.

Further Implications. The attack against user privacy introduced in this paper has several important implications. First, the privacy of people who are not active members of a social network, or possibly who are not even aware that pictures of them are being taken, could be violated. This is demonstrated by the celebrity dataset: even without the celebrities having uploaded their own photographs, we were able to detect relationships by leveraging paparazzi photographs. This shows how the inclusive, share-friendly nature of social networks has implications that go beyond the world of the network's members. In the context of a social network, as long as a person's friends are uploading photographs of them, even without the person participating on social networks or being tagged in photos, that person is subject to the privacy violations exposed in this paper.

Second, the centralization of large corpuses of pictures and the collection of meta-information (tagging, relationship status, location) that is ubiquitous in social networks opens even more opportunities for privacy violations, as it provide basic ground truth information that can be used to develop classifiers with even higher precision and recall.

Third, one could foresee the use of techniques similar to the ones described here to derive other types of relationships or associations (e.g., co-conspirators in a terrorist plot or decision-making power within an organization). The intelligence and police community is already actively using the information available on social networks in order to track down terrorists and criminals, as it has been demonstrated in the follow-up to the Boston bombings of April 15, 2013.

The approach presented in this paper shows that analysis of this kind can be performed automatically at a very large scale, with high accuracy. Such an approach could feasibly be used to help target criminal and terrorist investigations.

Specifically, by extending the approach to include *temporal* features as well as spacial features, Creepic might be applicable to *video* analysis. Automatically correlating individuals, from the same vantage point but over different points in time, on a security camera feed could present opportunities for intelligence gathering and security surveillance. For example, in the UK, where video surveillance is extremely common [13], automatic analysis and correlation of individuals in such video could be used to analyze crime patterns or (more pessimistically) for increased monitoring of citizens.

Even without extending Creepic's approach to other media, it can likely be used to gleam other information from user photographs. For example, an advertiser might be interested in how "social" a user is, and Creepic could estimate that by identifying the number of strangers in the users' photographs. Likewise, a more in-depth profile of a user could be built by analyzing what time of day a user takes more photographs or analyzing the situations in which the user appears with other people (i.e., the number of people, the frequency of such photographs, etc). Finally, if location information is embedded in the photographs or could be determined through some other mean (such as recognition of landmarks), a user's affluence level might be determinable based on how frequently they travel.

All of these ideas are different views of a central idea: the high incidence of photography in the modern world leaves behind tangible, measurable, and useful footprints, and these footprints can be used to achieve a view into our habits and preferences.

6. RELATED WORK

Sociometry. Work has been done in extracting romantic relationships from the structure of a social graph, such as the graph of friendship on a social network [15]. This work, carried out by Backstrom et al., appears to achieve a high recall rate, but their analysis does not include the false positive rate of detection, and so it is unclear how feasible the approach is in practice. Additionally, Creepic has the capability to detect relationships of people that are not actually included in the social graph, as long as they appear in enough photographs. This ability, along with the additional vectors of image acquisition by interested parties, make Creepic unique in sociometric systems.

A recent blog post from Facebook has revealed that they are, indeed, interested in predicting the relationship status of users [9]. Their approach is based on the emotional level of messages that users, who are starting relationships, send each other on social networks. The blog post does not discuss prediction accuracy (instead, they discuss the characteristics that they see in relationships). This approach to relationship detection differs from Creepic's in several ways. To begin with, private messages between users are required, which likely means that the privacy invasion would be limited only to users that approve such access (unlike, as with Creepic, their friends as well). Additionally, users that do not participate in social networks cannot be targeted, since they have no messages to target.

Image Processing. In this paper, we utilize several recent advances in image processing to detect and recognize faces in images. Initially, face detection utilized machine learning techniques such as neural networks [21] and support vector machines [19]. Face detection techniques have grown increasingly sophisticated and performant over the past decade, and fast detection using Haar classifiers [25] is supported by the open-source OpenCV computer vision library [18], which we utilize in our implementation.

Face *recognition*, the capability to identify that two given faces belong to the same individual, is a considerably more difficult problem. Modern techniques need to be "trained" with a series of photos representing each subject in question before being able to recognize subjects in photos. In an early approach to face recognition, *Eigenfaces* [22] were generated by performing a mathematical process called Principal Component Analysis on a set of photos of a single subject. The Eigenfaces approach, however, suffers in the presence of lighting and angle changes. A more modern approach, *Fisherfaces* [5], is more resilient to such changes.

Social Network Privacy. The growth of online social networks, and the increase in the personal information that users share over them, has prompted an large deal of research into social network privacy [7, 12, 14].

One relevant advance in this space focuses on extracting additional information about individual users. For example, multiple groups have developed techniques to extract additional information about users of a given social network by correlating a given user's profile with his or her profiles on other social networks [1, 3]. Such techniques can aid in the creation of a dataset for Creepic to analyze. For example, starting with a list of Facebook users, this technique could be used to acquire the users' photographs from other social networks, such as Instagram. An increased amount of photographs would, in turn, increase Creepic's effectiveness. We leave the exploration of this effect as future work.

Researchers are also thinking about the privacy of the *photographer*, not just of the subjects of a photograph. Nagaraja, et al. propose techniques to leverage view synthesis algorithms to obscure the location of a photographer taking possibly controversial or sensitive photographs [17].

Finally, it is important to discuss what consequences Creepic's results can have on real-world social network users, other than the aforementioned advertising and direct privacy-invasion. Davis et al. have investigated the use of available location data, coupled with friendship status, to determine the (withheld) geographic locations of users based on the (publicly-offered) locations of their friends [8]. Creepic's results essentially add a weight to the friendship information used by this system, and could thus be used to increase the system's efficiency.

7. CONCLUSION

In this paper, we presented Creepic, an approach for the automatic detection of romantic relationships between pairs of people by leveraging image analysis. We described the design and implementation of Creepic, and presented the results of experiments that quantify its prediction accuracy. We also discussed the important privacy implications that Creepic has on the modern world of social networking, given its high success rate at predicting relationships.

As digital media continues to increase in popularity and future devices make photography and video monitoring increasingly ubiquitous, the concerns raised by the results achieved in this paper will become overwhelmingly important to address. While deducing a person's relationship status would most likely not radically harm or impact that person, future advances along the same lines could allow outsiders to derive increasingly private or important information from publicly available data such as photographs.

8. REFERENCES

[1] A. Acquisti, R. Gross, and F. Stutzman. Faces of facebook: Privacy in the age of augmented reality. 2011.

[2] APeWatch. App Permission Watcher. `http://apewatch.de/`.

[3] M. Balduzzi, C. Platzer, T. Holz, E. Kirda, D. Balzarotti, and C. Kruegel. Abusing social

networks for automated user profiling. In *Recent Advances in Intrusion Detection*, pages 422–441. Springer, 2010.

[4] M. Bar-Hillel. The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3):211–233, 1980.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

[6] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[7] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107(52):22436–22441, 2010.

[8] C. A. Davis Jr, G. L. Pappa, D. R. R. de Oliveira, and F. de L Arcanjo. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15(6):735–751, 2011.

[9] Facebook. The Formulation of Love. `https://www.facebook.com/notes/facebook-data-science/the-formation-of-love/10152064609253859`.

[10] Facebook, Ericcson, and Qualcomm. A Focus on Efficiency. `https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-ash3/851590_229753833859617_1129962605_n.pdf`.

[11] GigaOM. Instagram reports 200M users. `http://gigaom.com/2014/03/25/instagram-reports-200m-users-20b-photos-shared/`.

[12] P. Gundecha, G. Barbier, and H. Liu. Exploiting vulnerability to secure user privacy on a social networking site. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 511–519. ACM, 2011.

[13] T. Independent. How average Briton is caught on camera 300 times a day. `http://www.independent.co.uk/news/uk/this-britain/how-average-briton-is\-caught-on-camera-300-times-a-day-572781.html`.

[14] C. Jernigan and B. F. Mistree. Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10), 2009.

[15] J. K. Lars Backstrom. Romantic partnerships and the dispesion of social ties: A network analysis of relationship status on facebook. In *Proceedings of the 2014 ACM conference on Computer supported cooperative work*. ACM, 2014.

[16] B. W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

[17] S. Nagaraja, P. Schaffer, and D. Aouada. Who clicks there!: anonymising the photographer in a camera saturated society. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 13–22. ACM, 2011.

[18] OpenCV. OpenCV. `http://opencv.org`.

[19] E. Osuna, R. Freund, and F. Girosit. Training support vector machines: an application to face detection. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 130–136. IEEE, 1997.

[20] B. Rouberol. Comparison of face detection tools. `http://bit.ly/1olWs4n`.

[21] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1):23–38, 1998.

[22] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[23] T. Verge. The new Snapchat brilliantly mixes video and texting. `http://www.theverge.com/2014/5/1/5670260/real-talk-the-new-snapchat-makes-texting-fun-again-video-calls`.

[24] P. Wagner. Facerec. `https://github.com/bytefish/facerec`.

[25] P. I. Wilson and J. Fernandez. Facial feature detection using haar classifiers. *Journal of Computing Sciences in Colleges*, 21(4):127–133, 2006.

APPENDIX

Appendices

A. CREEPIC APPROACH DIAGRAM

Taking a training set of photographs (called *Representative Photographs*) of people in a social network and labeled relationships between some of those people, Creepic processes a set of *unlabeled* photographs and identifies relationships between the individuals present. We present a diagram of this approach in Figure 5.

B. FACEBOOK DATASET DEMOGRAPHICS

We were able to retrieve age and gender information for 5,089 users in our Facebook dataset. 29 such users were under the age of 18, 140 between 18 and 20, 1,086 between 21 and 24, 2,254 between 25 and 29, 923 between 30 and 34, 359 between 35 and 44, 168 between 45 and 54, 78 between 55 and 64, and 52 were 64 years old and older. 3,003 (59.1%) of these users were male, and 2,086 (40.1%) female. Figure 6 details the average number of photographs in which a user is tagged, grouped by age range. We found that, in our dataset, users between 18 and 29 years of age dominated the tagged photographs. However, all age groups had outliers up to over 100 photographs per user.

These measurements reveal that our dataset does contain a bias. In fact, a report published in August 2012 found that the gender distribution of Facebook users is 60% female and 40% male. According to the same report, the average user of Facebook is 40 years old, as opposed to 28 in our dataset. However, in the absence of an unbiased dataset, we have not been able to investigate the effect that such a bias has on relationship detection.

People in different age groups tend to appear in photographs with a different number of other people. The average number of people in photographs in which a user appears (even those uploaded by others), by age of the user, is shown in Figure **??**.

C. EXAMPLE PHOTO FEATURES

For improved clarity, Figure 7 provides example output of the feature extraction step for an example photograph.

D. THRESHOLDS

Creepic's implementation requires values for two thresholds: the minimum number of photographs in which a pair must appear, within one timestamp, to be considered for the analysis, and the ratio of dating and non-dating pairs to include in the training set. The first

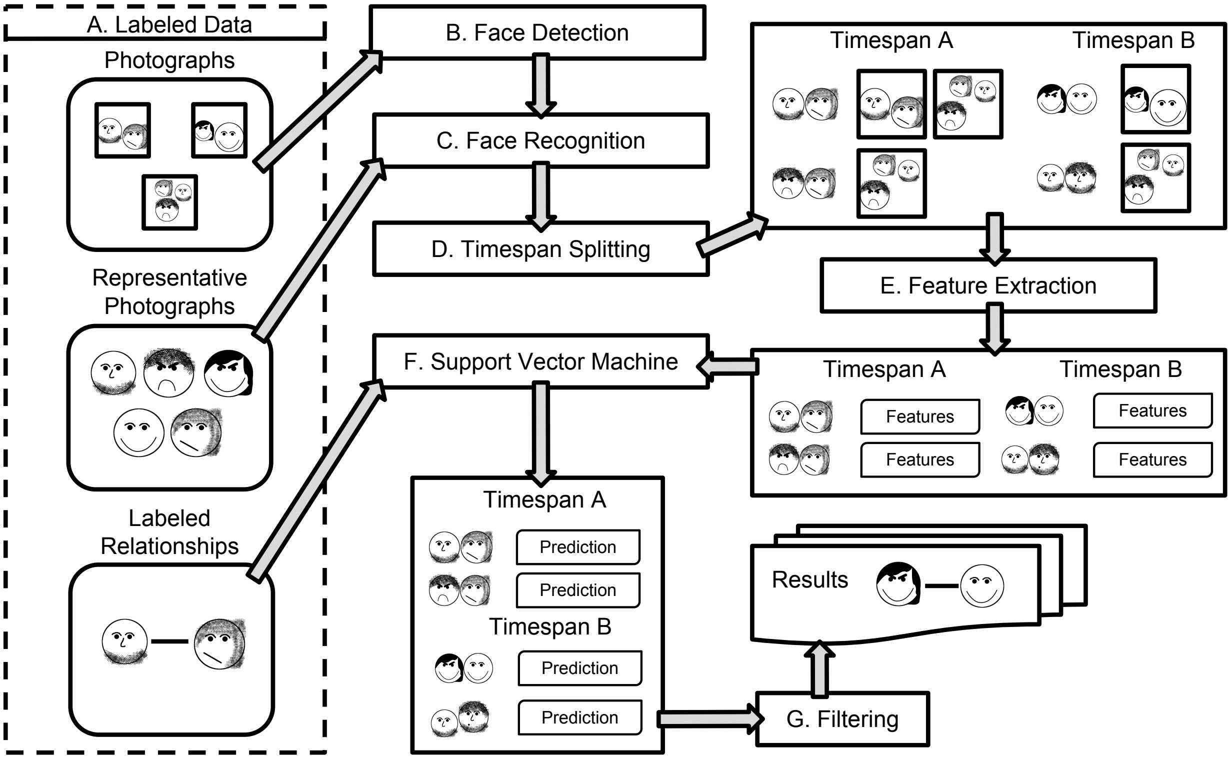


Figure 5: A diagram of Creepic's approach. Given labeled input data, Creepic carries out face detection and recognition, splits the data into timespans, extracts features, trains an SVM, and predicts dating relationships. Steps A - G correspond to subsections A - G of Section 2.

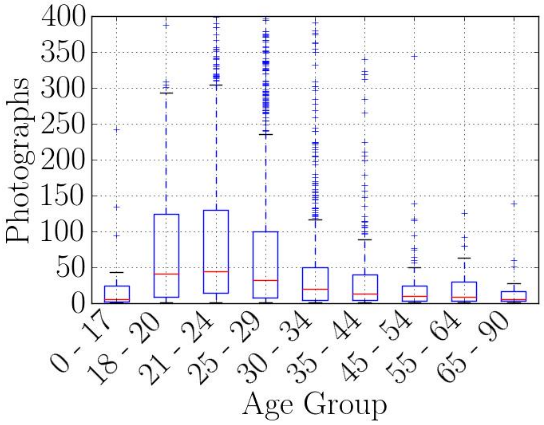


Figure 6: The average number of photographs in which a user appears (Y axis), by age of the users (X axis), in the Facebook dataset.

threshold is the minimum number of photographs in which a pair must appear, for any given timespan, before the pair's features are calculated. Setting this value too low increases the amount of noise that enters the Creepic's machine learning step, and erodes the prediction accuracy by causing overfitting on the training set. Alternatively, setting the threshold too high causes a large number of users to be discarded, and results in too little data with which to train the SVM, ultimately causing underfitting. This threshold can vary depending on the average amount of photos per pair in a dataset.

To choose the proper value for this threshold, we computed the Matthews correlation coefficient [16] of Creepic's results given different minimum values. Matthews correlation coefficient is a measure, between -1 and 1, of the quality of prediction for a machine learning system. A value of 1 corresponds to perfect prediction, -1 to perfect mis-prediction, and 0 to a random performance. The results for Creepic's Facebook dataset are presented in Figure 8. Creepic's prediction quality levels out when using a threshold value of 5, which we utilize for our experiments.

Another threshold that needs to be determined is the ratio of dating and non-dating pairs to include in the training set when training the SVM. This has a direct effect on the prediction accuracy, as the improper configuration of this threshold results in one class (i.e., non-dating pairs) dominating the other class (i.e., dating pairs) in the analysis. For example, we found that if the training set contains more than 4 times as many non-daters as daters, the non-dating space dominates the dating space and true positives plummet as the SVM labels every pair as non-dating. On the other hand, if the training set contains an equal or smaller amount of non-daters compared to daters, the dating couples dominate the training set and the classifier produces an unreasonable number of false positives.

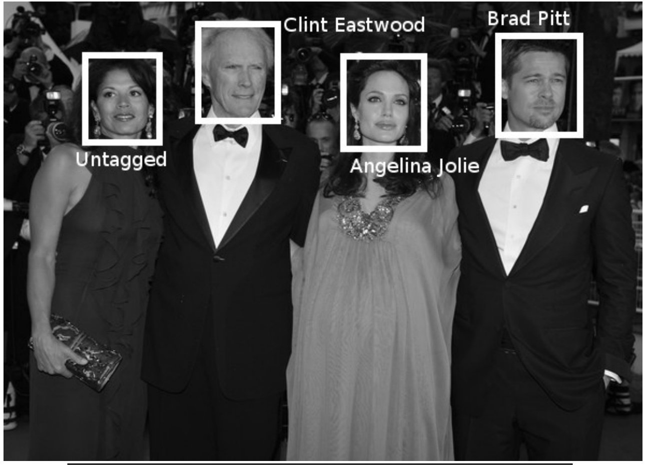
We again used the Matthews correlation coefficient to compute the optimal value for this threshold, as shown in Figure 9, which was determined to be 2.

E. STATISTICAL TRADE-OFFS

As with most binary classifiers, Creepic exhibits an implicit trade-off between performance and recall. By applying different importance to true positives over false positives, Creepic achieves different rates of precision and recall.

This can be used to configure Creepic as either a prefilter for human analysis (where a human analyst ensures a high precision, but uses Creepic to create a dataset with a much higher ratio of dating

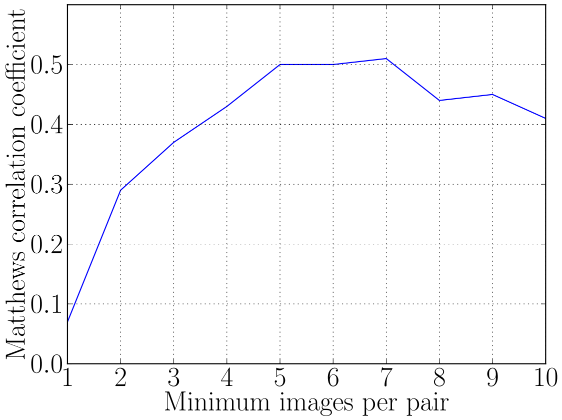
Figure 7: Feature extraction example.



Feature	Value
Ambient light level	62
Face-distance	0.25
Face distance, size-adjusted	2.01
Face-size ratio	0.94
Pair face-area coverage	0.05
Total face-area coverage	0.10
Tagged count	3
Untagged count	1
Vertical positioning	0.05
Betweeners count	0
Competitor count	1 (Clint Eastwood)

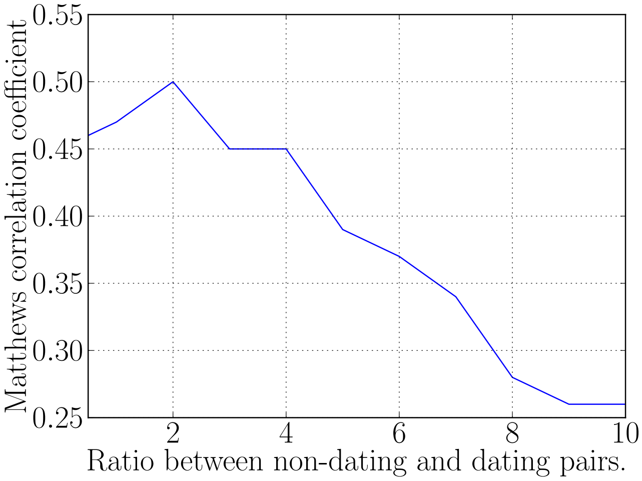
An example of feature extraction for an example photograph of Angelina Jolie, Clint Eastwood, and Brad Pitt. The pair being analyzed is Brad Pitt and Angelina Jolie. The squares mark the detected faces, and all spatial measurements are normalized to a total picture width of 1.0 and total picture

Figure 8: Minimum photographs per pair.



The Matthews correlation coefficient of Creepic's results with different thresholds for the minimum photographs a pair must have before being processed.

Figure 9: Dating/non-dating training ratio.



The Matthews correlation coefficient of Creepic's results with different thresholds for the ratio between dating and non-dating pairs in the training set.

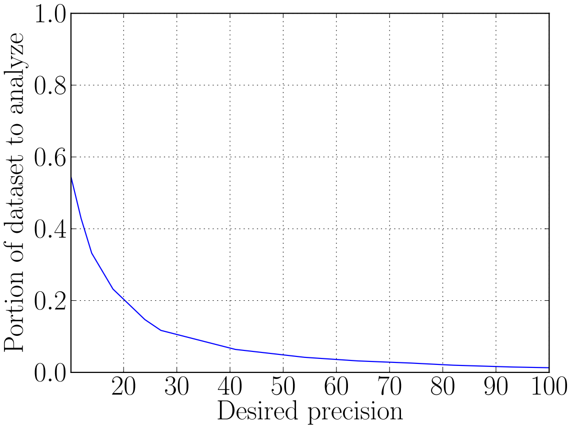


Figure 10: The reduction in the dataset for further manual analysis with a desired minimum precision (Facebook dataset).

couples) or a high-precision automated analysis system (although at the expense of recall). The various precision and recall values coincide to different points on Creepic's ROC curve, which we present in Figure 4. Depending on the desired application, Creepic returns a different number of predictions, with a different percentage of correct predictions.

The resulting reduction of the dataset, for desired precision and recall values, is presented in Figures 10 and 11 for the Facebook

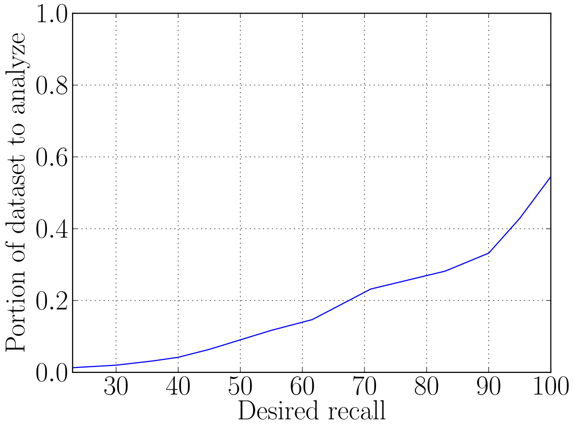


Figure 11: The reduction in the dataset for further manual analysis with a desired minimum recall (Facebook dataset).

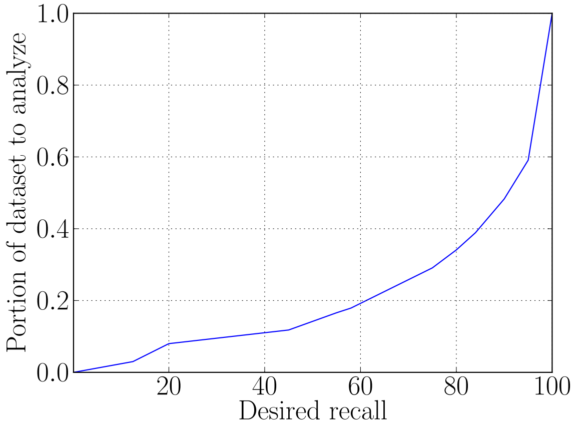


Figure 13: The reduction in the dataset for further manual analysis with a desired minimum recall (Celebrity dataset).

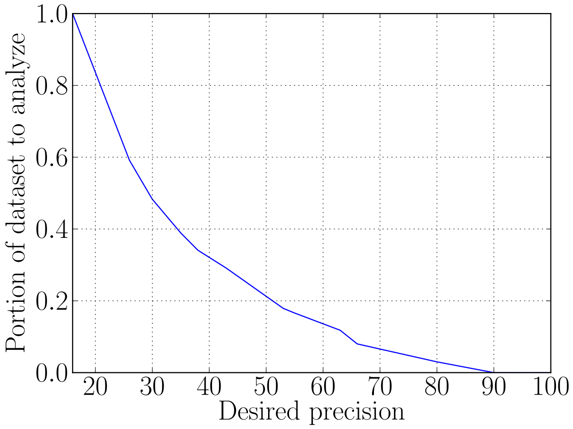


Figure 12: The reduction in the dataset for further manual analysis with a desired minimum precision (Celebrity dataset).