# Estimating Age Privacy Leakage in Online Social Networks


Ratan Dey
Polytechnic Institute of New York University
Brooklyn, NY, USA
Email: ratan@cis.poly.edu

Cong Tang
EECS, Peking University
China Academy of Electronics and Information Technology
Beijing, China
Email: congtang.cn@gmail.com

Keith Ross
Polytechnic Institute of New York University
Brooklyn, NY, USA
Email: ross@poly.edu

Nitesh Saxena
University of Alabama
Birmingham, USA
Email: saxena@cis.uab.edu



***Abstract*—We perform a large-scale study to quantify just how severe the privacy leakage problem is in Facebook. As a case study, we focus on estimating birth year, which is a fundamental human attribute and, for many people, a private one. Specifically, we attempt to estimate the birth year of over 1 million Facebook users in New York City. We examine the accuracy of estimation procedures for several classes of users: $(i)$ highly private users, who do not make their friend lists public; $(ii)$ users who hide their birth years but make their friend lists public.**

**To estimate Facebook users' ages, we exploit the underlying social network structure to design an iterative algorithm, which derives age estimates based on friends' ages, friends of friends' ages, and so on. We find that for most users, including highly private users who hide their friend lists, it is possible to estimate ages with an error of only a few years. We also make a specific suggestion to Facebook which, if implemented, would greatly reduce privacy leakages in its service.**


## I. Introduction

The current Online Social Networks (OSNs) allow users to control and customize what personal information is available to other users. For example, a Facebook user – let's call her Alice can configure her account so that her friends can see her photos and interests, but the general public can see only her name and profile picture. In particular, Alice has the option of hiding her attributes such as age, gender, relationship status, sexual preference, and political affiliation from the general public. Alice, of course, knows that the company providing the OSN service (let us say Facebook) has full access to any information she has placed on Facebook pages, including information that she provides only to her Facebook friends. However, Alice probably assumes that if she makes available only her name to the general public, third parties have access only to her name and nothing more. Unfortunately for Alice, by crawling OSNs and aggregating information provided by Alice's friends, third parties can potentially infer personal information – such as political affiliation, sexual orientation and gender – that Alice has not explicitly made public [1], [2], [3]. *To the extent this is possible, third parties not only can use the resulting information for online stalking and targeted advertising, but also can sell it to others with unknown nefarious intentions.* In this paper, we perform a large-scale study to quantify just how severe the privacy leakage problem is in Facebook.

As a case study, we focus on estimating birth year, which is a fundamental human attribute and, for many people, a private one. We have found that in our sample dataset of 1.47 million Facebook users from New York City, only 1.5% of them specify their age in their public profile, confirming that age is indeed a private attribute for most users. Motivated by this, we ask the question: *with what level of accuracy is it possible to estimate the age of the remaining users – i.e., those who aim to hide their ages – with a high accuracy?* We seek to answer this question using algorithms that are not Facebook specific, so that they can be applied to OSNs in general. For age estimation, we only use public profile and friendship information; we do not use image analysis or network/group information.

## II. Data Sets and Performance Measures

### A. Crawling NYC Users and Their Friends

In Facebook, when Bob visits Alice's profile page, the information that is displayed to him depends on his relationship with Alice (for example, whether she is a friend or not) and on Alice's privacy settings. Roughly speaking, when Alice is a Facebook friend of Bob, then he typically gets to see Alice's **full profile page**, which includes links to all of Alice's friends as well as all of the information and photos that Alice puts into Facebook; if Alice is not a friend, Bob only gets to see a **limited profile page**, which often includes no more than Alice's full name and her photo.

For the purpose of studying privacy leakages, we developed a multi-threaded crawler that visits Facebook user profile pages and stores the pages in a MySQL database. Using this crawler, in July 2009, we crawled all the users in NYC, obtaining their Facebook IDs and their *full profile pages*. We were able to do this because at that time $(i)$ users were, by default, assigned to regional networks; and $(ii)$ a user's full profile page was, by Facebook's default privacy setting, made public to all other users in the same network. By joining the NYC network, we obtained 1.67 million NYC user IDs and their corresponding full profiles. We refer to this dataset as the *July 2009 dataset*. Facebook fully deprecated regional networks as of late September 2009 [4]. A user's full profile is now, by default, only available to the user's friends.

In March 2010, we launched another extended crawl, during which we visited the 1.67 million NYC user IDs from the July 2009 dataset. Among the 1.67 million user IDs, we were able to re-visit 1.47 million of the users; the remaining accounts appear to have been deactivated or removed by Facebook between our two crawls. At the time of March 2010 crawl, we obtained the *limited profile pages* of the NYC users. As shown in Table I, only 82.73% of the limited profile pages publicize friend lists, and a mere 1.5% of them provide the users' ages.

During the March 2010 crawl, for each crawled user (say, Alice), in addition to obtaining Alice's limited profile page, we also collected the limited profile pages of her friends, whenever she made the friend list publicly available. By crawling the friends of the 1.47 million NYC users, we obtained an additional 47.79 million users, many of whom do not reside in NYC. Our *March 2010 dataset* has the limited profile pages of 49.26 million users, consisting of the 1.47 million NYC users and their friends. This data set contains approximately 306 million friendship links between NYC users and their friends. We emphasize that the data set does not include the friendship links between the 48 million non-NYC users collected, as that would have required significantly more computational and bandwidth resources than available at the time. The July 2009 dataset, containing full profile pages, is used for ground truth and evaluation of the methodology.

TABLE I
PROPERTIES OF THE MARCH 2010 DATASET (LIMITED PROFILES)

| **Property name** | **Value** |
|---|---|
| # users in NYC | 1,473,199 |
| # users in Total | 49.26M |
| % users who do not make friends public | 17.27% |
| % users who specified age | 1.5% |
| % users who make HS graduation year public | 21.6% |
| % users who provide work place network public | 3.7% |
| % users who provide grad/college info public | 19.0% |

### B. Reverse Friend Lookup

As shown in Table I, a significant fraction of users do not disclose their friend lists in their limited profiles. It is, however, possible to obtain partial friend lists for such users employing a novel reverse lookup mechanism. Specifically, if Alice hides her friend list, we can look at all other users who disclose their friend lists, and identify those who indicate they are friends with Alice. We applied reverse lookup to our dataset. Figure 1 shows the fraction of users among those hiding their friend lists for which reverse lookup can identify $x$ friends. For example, for 46.3% of these users we can find at least 15 (NYC) friends. Clearly, with a more extensive crawl, which would also obtain the friend lists of the non-NYC users, reverse lookup would yield a much more complete view of these otherwise hidden friend lists.

### C. Inactive Users

Although many Facebook users have hundreds of friends and 50% of users visit the site daily (as discussed in [5]),

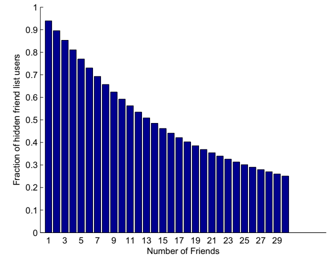


Fig. 1. Fraction of users for whom reverse lookup can identify $x$ friends

many accounts have few friends and no recent activity; we refer to such users as *inactive users*. In order to prevent these users from skewing the results of our study, we do not attempt to estimate the ages of users who satisfy all of the following conditions: $(i)$ the user has 10 or fewer friends; $(ii)$ the user does not provide his or her birth year. $(iii)$ the user does not provide high-school graduation year. After removing all users who satisfy all of the above three criteria, we have 1,191,758 NYC users, for whom we will attempt to estimate their ages.

### D. Estimation Performance Measures

In order to evaluate the performance of our age estimation procedures, we utilize two different measures: the Mean Absolute Error (MAE) and the Cumulative Score (CS). MAE is defined as the average of the absolute differences/errors between the estimated ages and "ground truth" ages, i.e., MAE $=\sum_{k=1}^{N} |x'_k - x_k|/N$, where $x_k$ is the ground truth age for the user $k$, $x'_k$ is the estimated age, and $N$ is the total number of test users. The MAE measure has previously been used in the context of age estimation based on facial images [6], [7], [8]. The cumulative score, on the other hand, is defined as $CS(j) = N_{e\leq j}/N \times 100\%$, where $N_{e\leq j}$ is the number of test users for which the age estimation procedure makes an absolute error no higher than $j$ years. For example, $CS(4)$ is the percentage of test users for which the absolute error is less than 4 years. This measure has previously been used in [6].

For calculating MAE and CS, we use the birth year data from the July 2009 dataset as ground truth. As described earlier, while crawling Facebook in July 2009, by default, we were able to obtain the full profile pages of the users in NYC. In the July 2009 data set, 515,000 users provide their birth years. In the second crawl (March 2010), we found that 486,686 of these user accounts were still active. However, some users blatantly lie about their ages, reporting ages over 80 when they are actually much younger. We therefore remove from our ground-truth data set any user who reports a birth year prior to 1931 (This step removes a small number of users who are actually over 80) and who is identified as inactive user as discussed in section II-C. At this stage, we have 419,395 users' birth years which will be used as ground truth to determine the accuracies of the age estimation methods.

## III. BIRTH YEAR ESTIMATION: BASIC METHODS

In this and the following section, we present our age estimation methodology. The methodology is based on fundamental attributes of OSNs, i.e., limited profile information and social

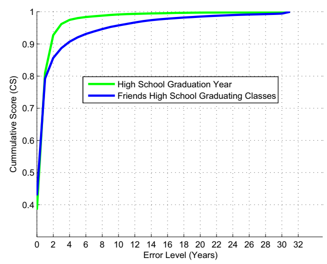


Fig. 2. Estimating birth year using high school graduation year and using friends' high school graduating class

links, and does not use features that are highly application (Facebook) specific. Let $\mathcal{G}$ be the set of all $1,191,758$ NYC Facebook users for which we will attempt to estimate the birth year. Our approach is to first find a subset $\mathcal{G}_0$ for which we can estimate the birth year with a high accuracy. Then, we find another disjoint subset $\mathcal{G}_1$ for which we can estimate the birth year with somewhat lesser accuracy. Iterating in this manner, we create a partition $\{\mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_N\}$ of $\mathcal{G}$ with a different estimation procedure and estimation confidence for each disjoint subset. The set $\mathcal{G}_0$ consists of $15,975$ users or $1.34\%$.

### A. Using High School Graduation Year

There are many users who do not make their birth years publicly available in their limited profiles, but nevertheless make their high school graduation year publicly available. Because most people complete high school between the ages of 17 and 19 years, the high school graduation year is clearly correlated with the birth year of an individual. To take advantage of this correlation, we build a training set for identifying the relationship between high school graduation year and birth year.

From our March 2010 dataset (including NYC users and their non-NYC friends), we found that $255,012$ users made both their birth year (BY) and high school graduation year (HSY) public. We fed these $255,012$ data points into Weka's [9] default linear regression method and obtained the following regression line with correlation coefficient $0.96$, MAE $1.01$ and root mean squared error $2.67$. We assume homoscedasticity across the distribution.

$$BY = 0.9368 \times HSY + 108.2107 \tag{1}$$

Let $\mathcal{G}_1$ be the set of NYC users who do not make their birth year publicly available, but make their high-school graduation year publicly available. In $\mathcal{G}_1$, there are $215,846$ users representing $18.11\%$ of users in $\mathcal{G}$. Using the Equation 1, we assign birth years for these $215,846$ users. We refer to this as Step 1. Of these $215,846$ users, $98,653$ belong to our ground truth data set, yielding an MAE of $1.11$. Figure 2 depicts the cumulative score for this linear regression estimation procedure; note that for $94\%$ of the users, the linear regression results in an error of $2$ years or less.

### B. Using Friends' High School Graduating Classes

A user may not publicize her birth year or her high school graduation year, but she may have many friends from her high school graduating class, from which we may be able to infer her high school graduating year. To create the subset $\mathcal{G}_2$, we use a grouping methodology that takes into account the high school name and graduation year of a user's friends. The methodology is as follows. For each user $u$ not in $\mathcal{G}_0 \cup \mathcal{G}_1$, among $u$'s friends we find the most frequently occurring high school graduating class (i.e., high school name and graduation year). If $u$ has $T$ or more friends in this high school graduating class, we put $u$ in $\mathcal{G}_2$ and assume that $u$ is also from this high school graduating class. Let $y_u$ be the corresponding graduation year. To estimate user $u$'s age, we then use $y_u$ as HSY in the regression Equation 1. We call this procedure Step 2. There are $919,680$ users in $\mathcal{G} - (\mathcal{G}_0 \cup \mathcal{G}_1)$. Using $T = 6$, we find $453,596$ users in $\mathcal{G}_2$. Using $T = 6$ gives us moderate coverage and accuracy (low MAE); if we choose a smaller value for $T$, coverage improves but accuracy degrades and if we set the value of T greater than 6, accuracy will increase but coverage will decrease. Of these $453,596$ users, $141,216$ are found in the ground truth verification set. For these $141,216$, the MAE for our estimation procedure is $1.86$. Figure 2 shows the corresponding cumulative score. We define $\mathcal{H} = (\mathcal{G}_0 \cup \mathcal{G}_1 \cup \mathcal{G}_2)$.

Table II summarizes the results from Steps $0$, $1$, and $2$. From these three steps, we have been able to estimate the ages of $57.51\%$ of the NYC users with a high-level of accuracy of MAE $1.5$. However, there still remains $506,341$ ($42.49$ %) NYC users outside of $\mathcal{H}$ for which we need to use more advanced procedures to estimate ages.

## IV. Iterative Method

### A. Method Description

The method in Section (III-B) makes use of the age distributions of a user's friends; however, it does not take advantage of the underlying network structure in the social network, which provides information about friends of friends, friends of friends' friends, and so on. To exploit this underlying network structure, we develop an iterative algorithm.

In our algorithm, at each iteration $i$, we have age estimates for a set of users, denoted $E(i)$. For each user $u \in E(i)$, let $x_u(i)$ be our estimate of $u$'s age at the $i$-th iteration. Also let $F_u$ be the set of $u$'s friends, and $F_u(i)$ be the set of $u$'s friends for which we have age estimates, that is, $F_u(i) = F_u \cap E(i)$.

In the iteration scheme, for any user $u \in \mathcal{H}$, we set $x_u(i) = a_u$, where $a_u$ is the age determined in the previous section. For a user $u \notin \mathcal{H}$ which has at least one friend with an age estimate (i.e., $F_u(i) \neq \phi$) we use iterations:

$$x_u(i+1) = \alpha x_u(i) + (1-\alpha)\Phi[x_v(i),\ v \in F_u(i)], \tag{2}$$

where $\Phi[\cdot]$ could be a simple algebraic expression or a more sophisticated clustering algorithm. We will soon provide some examples for $\Phi[\cdot]$. To initialize the iterations, we set $E(0) = \mathcal{H}$. We also set $E(i+1) = E(i) \cup \{u : F_u(i) \neq \phi\}$. Notice that this algorithm takes into account not only Bob's friends but also Bob's friends of friends when estimating his age. In first iteration, it takes into account only his friends and some of his friends may not be assigned ages at that time. After completion

TABLE II
SUMMARY OF RESULTS FROM STEPS 0,1,2

| Set | # NYC users | % of NYC users | # Ground Truth users | % of Ground Truth users | MAE on Ground Truth users | CS(4) on Ground Truth users |
|---|---|---|---|---|---|---|
| $\mathcal{G}_0$ | $15,975$ | 1.34% | 8339 | 1.99% | 0 | 100% |
| $\mathcal{G}_1$ | $215,846$ | 18.11% | $98,653$ | 23.52% | 1.11 | 96% |
| $\mathcal{G}_2$ | $453,596$ | 38.06% | $141,216$ | 33.68% | 1.86 | 91% |
| $\mathcal{H}$ | $685,417$ | 57.51% | $248,208$ | 59.18% | 1.5 | 95% |

of that iteration, some of his friends will be assigned ages and hence Bob's age may be changed in next iteration. In this way, user age depends not only on his friends but also his friends of friends. We will stop the iterative method when no additional users from the set $\mathcal{G} - \mathcal{H}$ will be assigned ages in two subsequent iterations.

Since the function $\Phi[\cdot]$ must be calculated for millions of users at each iteration, it is critical to choose a function that not only provides good estimates but is also computationally efficient. We examine two computationally-efficient approaches in this paper: linear regression and percentiles.

For the linear regression approach, we choose a linear function of the mean, median, and standard deviation of the user's friends; specifically, a function of the form

$$\begin{aligned}\Phi[x_v(i)),\ v \in F_u(i)] &= a_1 \times MEAN_u(i) \\ &+ a_2 \times MEDIAN_u(i) \\ &+ a_3 \times STD_u(i) + a_4 \qquad (3)\end{aligned}$$

where $MEAN_u(i)$ (respectively, $MEDIAN_u(i)$ and $STD_u(i)$) is the mean (respectively, the median and standard deviation) of the values in $F_u(i)$. This linear equation is efficient to calculate, but how should we choose the values for $a_1$, $a_2$, $a_3$, and $a_4$?

We use linear regression to determine the coefficients $a_1$, $a_2$, $a_3$, and $a_4$. Specifically, for each of the $685,417$ users in $\mathcal{H}$, we determine the mean, medium, and standard deviation of the user's friends' ages. For each user in $\mathcal{H}$, we have a data point consisting of the user's age as well as the associated mean, median and standard deviation. We feed these $685,417$ data points into Weka's [9] default linear regression procedure to obtain the values of $a_1$, $a_2$,$a_3$, and $a_4$. The resulting regression equation becomes as follows with correlation coefficient $0.90$, MAE $2.10$ and root mean squared error $4.12$:

$$\begin{aligned}BY &= 0.3583 \times MEAN + 0.6654 \times MEDIAN \\ &- 0.3596 \times STD - 45.5534 \qquad (4)\end{aligned}$$

For the percentile approach, with a given value of $q$, $\Phi[\cdot]$ is simply the $q$ percentile of the ages in $F_u(i)$. For example, with $q = 70$, we take the age $x$ for which $70\%$ of the users in $F_u(i)$ are younger than $x$. Note that $q = 50$ is simply the median of the ages in $F_u(i)$. We experimented with using different percentiles such as 50th (median), 60th, 70th, 80th etc and found that 70th percentile provided the best estimation accuracy in terms of MAE and CS.

### B. Results for Iteration

We first applied the regression equation 4 for the function $\Phi[\cdot]$ and then applied the 70th percentile of friends ages for the function $\Phi[\cdot]$. If a user has more than 20 friends with known ages, we assign less weight ($\alpha$) to the new estimates; and if the user has at most 20 friends with known ages, we assign more weight to new estimates, with the hope that some of user's friends will be assigned ages in subsequent iterations. We have set the value $\alpha = 0.6$ for users who have at most 20 friends (with known ages) and $\alpha = 0.9$ for users who have more than 20 friends (with known ages) in both of the cases.

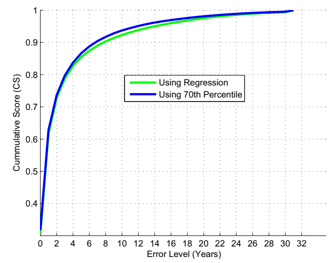


Fig. 3. Overall accuracy after combining basic and iterative methods

There are $506,341$ users in the set $\mathcal{G} - \mathcal{H}$. After running the iterative method for **5** iterations (as no additional users were assigned ages from 4th iteration to 5th), we were able to assign ages to 505 thousands additional users in both approaches. Of these 505 thousands users, $171,157$ belong to our ground truth data set. Over the set $\mathcal{G} - \mathcal{H}$, iterations with regression gave an MAE of 5.13 and CS(4) of 66.8%, whereas iterations with percentiles gave MAE of $4.48$ and CS(4) of $69.3\%$.

For the remaining few thousand users, we simply use mean birth year (i.e., $1980$), which we found to yield better results than the median. Figure 3 shows the accuracy of overall methodology (combining basic profile information, reverse friend lookup, and iterations with regression and percentiles). The overall method using iterations with 70th percentile, we obtain an MAE of $2.71$ and CS(4) of $83.8\%$. Our age inference approach is simple enough for naive attackers to develop and execute with effect. This implies that Facebook age privacy can be violated rather easily for most Facebook users.

## V. HIGHLY PRIVATE USERS

### A. Using Reverse Lookup

In Section I, we defined a user as *highly private* if the user hides his/her age, high-school graduation year, *and* friend list. Without reverse lookup, it would be difficult to accurately estimate the age of a private user, or determine any of his/her attributes such as political orientation or religious affiliation. We now investigate to what degree can age be accurately estimated using reverse lookup.

There are $235,377$ highly private users in our March 2010 data set. Using reverse lookup of friendship links, we can find at least 11 friends for each of the $128,641$ users among these

$235,377$ users. Let R be the set highly private users with at least $11$ known friends through reverse lookup. Now we apply our step-by-step age estimation methodologies for these $128,641$ users. Step $0$ and step $1$ are not applicable, since they require information directly from the users limited profile which, by definition of a highly private user, is not available. Let $R_2$ be the set of users whose ages can be estimated using their friends' high school graduating classes, that is, using step 2. We found $22,221$ users in $R_2$; among them $3,682$ users are in the ground truth data set, yielding an MAE of 2.8. Then we applied our iterative method to the remaining $106,420$ users. Using iterations with 70th percentile, we can assign ages to $105,315$ users, and remaining $1,105$ users are assigned mean birth year of March 2010 dataset, i.e., $1980$ as their birth year. Among these $106,420$ users, $13,170$ users can be found in the ground truth data set, yielding an MAE of 2.81.

From this analysis, we have shown that it is very hard for a user to avoid privacy leakages, even if the user takes maximal measures to do so.

### B. Recommendations for Reducing Privacy Leakage

We now briefly discuss what a Facebook user and the company Facebook can do to avoid age privacy attacks. The user can configure her privacy settings so that age, high-school graduation year, and friend lists are not available in her limited profile (that is, to non-friends). However, this alone will not fully protect the user, since an attacker can still perform reverse-friend lookup. With reverse friend lookup, the attacker may find a group of friends all from the same high-school graduation class, which – as we saw – can provide highly accurate estimates of age. The attacker can also apply iterations, as previously described, to obtain good estimates for age. Note that reverse lookup can also be potentially used to infer not only age, but also other attributes including religious and political preferences. *To prevent reverse friend lookup, when Alice chooses to hide her friends in her limited profile, Facebook could also automatically remove Alice from the friend lists in all her friends' limited profiles. We strongly recommend that Facebook adopt this policy.*

## VI. Related work

We now review the prior work that considers inference of one or more private attributes in OSNs. To the best of our knowledge, this is the first paper that examines in-depth age estimation in online social networks. Furthermore, our data set is at least one order of magnitude larger than all of those in the prior work on inference of private attributes (in the papers cited below).

Zheleva and Getoor [1] proposed techniques to predict the private attributes of users in four real-world datasets (including Facebook) using general relational classification and group-based classification. They looked at prediction of genders and political views, but not at age estimation. Other papers [10], [11], [12], [3], [13], [14] have also attempted to infer private information inside social networks, although none of these papers consider age estimation except Becker and Chen's [14]. To our knowledge, [14] is the only other existing study that considers age estimation. Age estimation is not a focus of their study, and their dataset size has only $49$ users. For this very limited study, their heuristics gave a success rate of 72.3%. Jernigan and Mistree [2] demonstrated a method for accurately predicting the sexual orientation of Facebook users by analysing friendship associations. Thomas *et al* examine scenarios where conflicting privacy settings between friends will reveal information that at least one user intended remain private [15].

## VII. Conclusion

To estimate Facebook user ages, we exploited the underlying social network structure to design an iterative algorithm, which derives age estimates based on friends' ages, friends of friends' ages, and so on. We found that for most users, including private users who hide their friend lists, it is possible to estimate ages within a few years. We also made a specific suggestion to Facebook which, if implemented, would greatly reduce privacy leakages in its service.

## References


[1] E. Zheleva and L. Getoor, "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles," in *WWW*, 2009.

[2] B. F. M. Carter Jernigan, "Gaydar: Facebook friendships expose sexual orientation," *First Monday*, vol. 14, no. 10, 2009.

[3] C. Tang, K. W. Ross, N. Saxena, and R. Chen, "A name-centric approach to gender inference in online social networks," in *SNSMW*, 2011.

[4] "Facebook to fully deprecate regional networks by september 30," available at: http://www.insidefacebook.com/2009/08/05/facebook-to-fully-deprecate-regional-networks-by-september-30.

[5] "Facebook statistics," available at: http://www.facebook.com/press/info.php?statistics.

[6] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *ACM Multimedia*, 2006, pp. 307–316.

[7] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

[8] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *CVPR*, 2009, pp. 112–119.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.

[10] R. Heatherly, M. Kantarcioglu, B. Thuraisingham, and J. Lindamood, "Preventing Private Information Inference Attacks on Social Networks," University of Texas at Dallas, Tech. Rep. UTDCS-03-09, 2009.

[11] W. Xu, X. Zhou, and L. Li, "Inferring Privacy Information via Social Relations," in *24th International Conference on Data Engineering Workshop*, 2008, pp. 154–165.

[12] J. He, W. W. Chu, and Z. Liu, "Inferring privacy information from social networks," in *ISI*, 2006, pp. 154–165.

[13] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *WSDM*, 2010.

[14] J. Becker and H. Chen, "Measuring privacy risk in online social networks," in *W2SP*, 2009.

[15] K. Thomas, C. Grier, and D. M. Nicol, "unfriendly: Multi-party privacy risks in social networks," in *Privacy Enhancing Technologies*, 2010, pp. 236–252.