Children Seen But Not Heard: When Parents Compromise Children's Online Privacy

Tehila Minkus
New York University
tehila@nyu.edu

Kelvin Liu
NYU Shanghai
kelvin.liu@nyu.edu

Keith W. Ross
NYU and NYU Shanghai
keithwross@nyu.edu

ABSTRACT

Children's online privacy has garnered much attention in media, legislation, and industry. Adults are concerned that children may not adequately protect themselves online. However, relatively little discussion has focused on the privacy breaches that may occur to children at the hands of others, namely, their parents and relatives. When adults post information online, they may reveal personal information about their children to other people, online services, data brokers, or surveillant authorities. This information can be gathered in an automated fashion and then linked with other online and offline sources, creating detailed profiles which can be continually enhanced throughout the children's lives.

In this paper, we conduct a study to see how widespread these behaviors are among adults on Facebook and Instagram. We use a number of methods. Firstly, we automate a process to examine 2,383 adult users on Facebook for evidence of children in their public photo albums. Using the associated comments in combination with publicly available voter registration records, we are able to infer children's names, faces, birth dates, and addresses. Secondly, in order to understand what additional information is available to Facebook and the users' friends, we survey 357 adult Facebook users about their behaviors and attitudes with regard to posting their children's information online. Thirdly, we analyze 1,089 users on Instagram to infer facts about their children.

Finally, we make recommendations for privacy-conscious parents and suggest an interface change through which Facebook can nudge parents towards better stewardship of their children's privacy.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: Models and Principles—*User/Machine Systems, human factors*; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

WWW 2015, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3469-3/15/05.
http://dx.doi.org/10.1145/2736277.2741124.

Keywords

Online social networks; privacy; human factors; measurement

1. INTRODUCTION

Technological advances present the modern parent with novel concerns. How much exposure should a child have to technology? Can children be trusted to retain appropriate privacy in a networked world? However, few parents view their own social media usage as a threat to their children. But as a new generation of adults joins the ranks of parents, mentions and photos of children and babies are popping up on Facebook, Instagram, and other social media with increasing frequency [7]. Facebook has become a "modern day baby book" [22], with the number of parents who post pictures of their children falling in the range of 66% [33] to 98% [4]. Are parents inadvertently compromising their children's privacy?

In this paper, we measure adults' sharing of children's personally identifiable information in online social networks, namely, Facebook and Instagram. This matter deserves attention for two reasons. Firstly, online social networks are public areas – since children are vulnerable, their information should not be publicly visible and archivable. This is a concern recognized by many parents [2]. Secondly, when parents post their children's information on Facebook, Instagram, or another social network, even in a non-public manner, they are effectively supplying the service provider with detailed information about the children. This limits children's ability to hide their online presences should they later wish to do so.

Specifically, we consider to what extent babies and young children – who do not even have their own Facebook accounts – can have their privacy compromised due to their parents' online behavior, and to what extent these privacy violations can be carried out in an automated fashion. We first apply off-the-shelf age detection software to the adults' public Facebook photos to automatically discover photos containing children. We then attempt to identify names and birthdays for the children through automated textual analysis. We find that for a large number of parents, one can learn the names and faces of their children; for many children, one can learn their birthdates.

By linking this information with publicly available data, one can obtain even more vivid profiles of young children. We demonstrate this by analyzing a set of adults for whom we have obtained the corresponding voter registration records. After detecting children on the public profile pages of these

adults, we can further determine the addresses of the families, the parents' birthdays and the parents' political affiliations. Such a seed profile could then be continually enhanced throughout the child's life by data brokers, government surveillance agencies, or Facebook itself. We extend this approach to Instagram, and we find that many parents are sharing not only their child's image but his birthday and name as well.

The automated attack described above is restricted to using public Facebook posts. Friends of a parent, Facebook itself, and organizations with access to one's Internet traffic can often see much more. As we cannot access the friends-only content, we cannot directly assess the full extent of what parents share online about their children. In order to gain a more complete picture of parental behavior online, we conduct a survey of parents who use Facebook. We find that the majority of parents report sharing their children's faces and names on Facebook, and many also report posting their children's birthdates.

This paper makes the following contributions:

1. Measure the occurrence of images of children posted by adults on Facebook.

2. Demonstrate how an attacker could infer, in an automated manner, attributes about children based on the posts of adults on Facebook.

3. Further demonstrate how this information can be linked with public records to create more detailed profiles of children.

4. Conduct a survey of parents on Facebook to learn about their posting habits with regard to their children.

5. Examine Instagram to determine how widespread parental oversharing is in this increasingly popular online social network.

6. Recommend better practices to parents and Facebook to protect children's privacy.

The structure of this paper is as follows: in Section 2, we discuss the threats facing children whose parents are active in online social networks as well as the legal and ethical considerations of this project. In Section 3, we present a method to gather a database of children whose likenesses and other information have been posted on Facebook without their participation. In Section 4, we survey parents who use Facebook to learn about their posting behaviors and privacy attitudes with regard to their children. In Section 5, we conduct a similar analysis using the Instagram photo-based social network. In Section 6, we discuss the ramifications of our findings and make some recommendations to parents and Facebook to ensure better privacy for children. Section 7 presents related work. Finally, in Section 8, we conclude.

2. PRELIMINARIES

In this section, we outline the different threats posed to a child whose information is shared on Facebook or Instagram. We also discuss the legal and ethical considerations involved in conducting this research.

2.1 Threats to Children on Online Social Networks

When parents post photos of children to Facebook or Instagram, they likely have only positive intentions: for example, to share updates about their family and life with grandparents, or to chronicle their lives' events [6]. However, the children may bear collateral risk as a result. We describe four threats to a child whose information is posted on Facebook:

- **Stranger danger.** When parents share information publicly about their children, they allow strangers to learn important facts about their children. For example, a public photo of a child with the caption "Happy birthday, Olivia!" provides an observer with knowledge of the child's face, name, and birthday. This could be exploited by criminals or predators local to the child, or by an identity thief who wishes to infer the child's personally identifiable information.

- **Overexposure to acquaintances.** Though media accounts often focus on children's abduction at the hand of strangers, it is far more common that children's kidnappers are from their family's social circles; a 1997 study by the FBI found that 76% of kidnappings and 90% of all violent crimes against juveniles were perpetrated by relatives or acquaintances [14]. As such, sharing personal information about one's child is not necessarily safe even when a parent has friends-only or friends-of-friends settings on their Facebook posts. Additionally, since many adult users of Facebook have 200 or more Facebook friends [31], these posts are less private than they realize. The same applies to a parent who has elected to use the privacy settings on Instagram.

- **Data Brokers.** Data brokers build profiles about people and sell them to advertisers, spammers, malware distributors, employment agencies, and college admission offices. Because the babies' and children's merchandise market is in the hundreds of billion dollars in the US alone, it is not surprising that data brokers are already seeking to compile dossiers on children [32] [30]. Using the information that parents post about their children, data brokers can create mini-profiles that can be continually enhanced throughout an individual's lifetime.

- **Surveillance.** In addition to the threats posed by other users, information posted to online social networks is subject to the threat of surveillance. By sharing a child's likeness and identifying information, a parent exposes his child to surveillance by the service provider and other parties, such as the NSA. This this can be problematic if children later wish to minimize or erase their digital footprints.

To measure the level of these risks, we undertake two methodologies. In the first methodology, we automate the analysis of public Facebook and Instagram pages to search for photos of children and accompanying personal information. This captures two of the four risks enumerated above, by demonstrating what information an unaffiliated viewer (e.g. a stranger or data broker) can glean about a child based on his parent's Facebook page. However, due to the

Facebook privacy settings, it does not provide a full view of the parent's posting behavior. In the second methodology, we conduct a survey of parents to learn about their self-reported habits with regard to their children on Facebook. Since parents are reporting their own behavior, this approach is not subject to the privacy-setting limitations of the first methodology. This allows us to assess what a friend, a surveillant authority, or Facebook itself might be able to infer about the children of the Facebook user.

2.2 Ethical and Legal Considerations

To conduct this research, we programmed crawlers that visited public pages on Facebook and Instagram and downloaded their contents. We then automated content extraction to detect faces, names and other information in the public comments.

Performing real-life research in online privacy can be ethically sensitive. Two stakeholders must be considered: the online service provider and the user. While crawling data from online service providers imposes a load upon their servers, we attempted to minimize the load by using a single process to sequentially download pages.

We emphasize that this research benefits Facebook and Instagram users by bringing to light an important aspect of children's online privacy. Any inferences we made were based on publicly available data. We intentionally limited the number of profiles analyzed in order to minimize the risk to any individual user. For the same reason, we also limited our analysis to a user's most recent posts.

3. AUTOMATED FACEBOOK ANALYSIS

According to Facebook's help pages, users must be at least 13 years old to join the social network[1]. Nevertheless, many children have an indirect social presence on Facebook through the photos and comments posted about them by adults. In this section, we seek to quantify the extent of this phenomenon by crawling the public Facebook posts of adults to see if they have posted photos and information about their children.

3.1 Methodology

Here, we describe the methodology we followed to discover posts about children on Facebook. See Figure 1 for a diagram of the process.

As basis for our exploration, we begin with a list of 2,383 Facebook users in a suburban city on the East Coast in the USA. This city has 20-30,000 households and a median household income in the $70-100,000 range, and its population is about 70% white, 10% African American, and 15% Asian. As we describe below, each one of these users has been matched with high certainty to a particular registered voter in a voter registration list [10]. Therefore, each user on the list is 18 or older.

For each user on the list, we collect the 20 most recent photos posted to the user's account. (This is the number of photos initially shown on a user's Facebook photos page.) We then analyze the photos using Face++[2], an online API which provides an age estimate for the faces detected in the photos. If a face's age estimate is beneath a certain threshold, we flag this photo as containing a child.

[1] https://www.facebook.com/help/210644045634222

[2] http://www.faceplusplus.com/

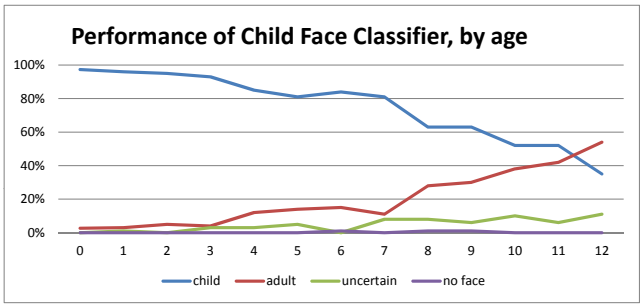

Figure 2: The performance of the Face++ age classification tool at each age, as judged by a human. Among the images that Face++ labeled as including children, some included only adults, people of uncertain ages, or no faces at all. We show the results for each age; the child-detection accuracy dips below 80% for ages eight and older.

Though face recognition and detection has become increasingly accurate, age estimation is still a hard problem; even humans may have a hard time guessing precise ages from a photo. In order to eliminate false positives among our flagged images, we examine only the images that contain a person whose age was estimated as seven or younger. We found through experimentation that this number helped limit false positives (i.e. young-looking teenagers or adults) while still returning most of the actual children in the sample. We established this threshold by labeling 100 random samples from each tagged age bucket and then calculating the proportion of false positives returned by the API. We used four labels: "child", "adult", "uncertain", and "no face pictured" (since some of the photos were incorrectly labeled as containing faces). In order to retain accuracy of at least 80% (without including faces of uncertain age), we opted to use only the photos that had been tagged with an age in the range of zero to seven, since these were less likely to be false positives. Figure 2 displays the results from each age group.

3.2 Results

From the 2,383 Facebook users, we collected 26,602 total photos from the photo pages of the accounts, for an average of 10.5 photos per account. The overall results are displayed in Table 1.

Of these 26,602 photos, Face++ estimated that 2,251 (8.5%) contained a child between zero and seven years old. 575 of these had public comments from which we could deduce a name for the child using the Stanford NER tool[3] [13]. In addition, 60 of the photos included the word "birthday" in the captions or comments and thus revealed the child's date of birth.

In terms of accounts, 807 of the overall 2,383 accounts (34.8%) contained at least one photo of a child. Since children usually share their parent's last name, we are able to infer the last name for all the children in the photos. Of these 807 accounts, 45.2% (365 accounts) had posted or received a comment mentioning the child's first name, and 6.2% (50 accounts) had also revealed the child's date of birth. For 45 of the accounts, all three pieces of identifying information

[3] http://nlp.stanford.edu/software/CRF-NER.shtml

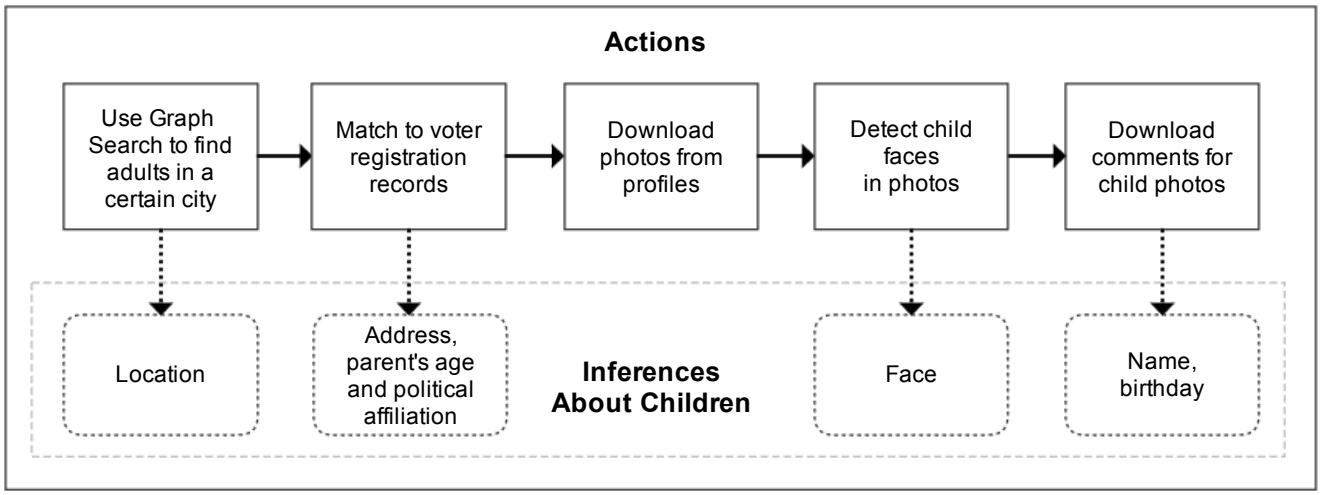

Figure 1: The process for downloading and inferring traits about children whose photos are posted on Facebook.

	Facebook	Instagram
Accounts		
Total collected	2,383	1,089
Sharing child photo	807	1,089
Sharing child name	365	689
Sharing child birthday	50	292
Photos		
Total collected	26,602	21,379
Child in photo	2,251	6,134
Name in comments	575	988
Birthday in comments	60	411

Table 1: The number of accounts and photos for each category examined in both Facebook and Instagram.

regarding a child – photo, name, and date of birth – were available in the parent's public photo albums.

Additionally, by examining the information in the parent's public Facebook pages, we can extend the profiles of the children by profiling their families. Additional information that can potentially be obtained for a child (who does not have a Facebook account) include the names and Facebook pages of both parents, siblings, and grandparents. These can be obtained by accessing the friend list of the parent. Moreover, one can infer the parents' religious and political affiliations, which are often adopted by their children, by the content of their status updates. The attacker can also augment his knowledge of the child by using the profiles of extended family members, if they have posted facts that were not included in the parents' public profiles.

3.3 Linking with Public Records

It is often possible to link Facebook accounts with other sources of offline and online information. For example, as described by Dey et al.[10], one can identify many of the Facebook users in a target town by using a combination of the Facebook graph search API, Facebook friends lists, and the voter registration list for the city. For the target suburban city considered in this paper, this technique was applied to obtain approximately 25,000 Facebook users who reside in the target city. Some of these Facebook users have the same name, and some these users' names match to multiple people in the voter list.

The 2,383 users studied here are a subset of the 25,000 likely-residents, with the following additional properties: (i) each of the 2,383 users has a unique name; (ii) each user recorded on Facebook that his hometown or current city is the target city; (iii) each user has at least five friends in the set of 25,000 likely-residents; and (iv) each user's name is an exact match with one name in the voter registration list of the target city. Owing to these properties, we believe that most (if not all) of the 2,383 Facebook users have been correctly linked to people in the voter registration lists.

Each record in a voter registration list corresponds to a person and contains the person's name, birth date, gender, political affiliation, and *address*. Thus, by linking a parent's Facebook page with his voter registration records, the attacker can further obtain the address of the child, which has clear potential for dangerous outcomes. Moreover, the attacker can obtain the political affiliations and birth dates of the child's parents; this would be informative to a data broker or surveillant authority.

In summary, when a parent posts photos of his or her children to Facebook, and the parent can be matched to a voter registration record, then the attacker can minimally obtain the child's face, last name, address, parents' names, parents' birthdays and parents' political affiliations. Additionally, the attacker may be able to determine the child's first name, birth date, and any additional information made publicly available in the parents' Facebook pages, such as parents' religion or employment. And as mentioned earlier, a Facebook friend of the parent, and Facebook itself, can significantly enhance such a profile with the information the parent shares only with Facebook friends.

3.4 Analysis of Users Posting Child Photos

Among the users who shared a child's photo, each user shared an average of 2.8 child photos within the user's 20 most recent photos.

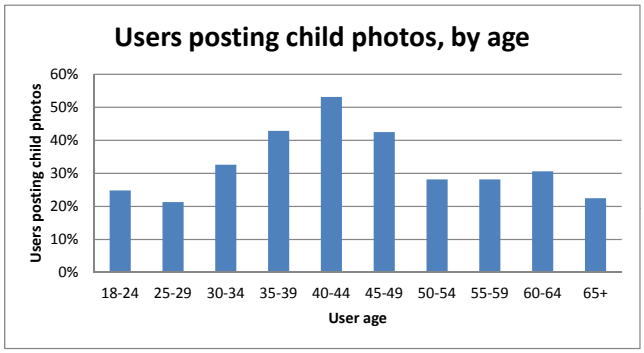

Figure 3: The percentage of users in each age group who shared photos of children on Facebook.

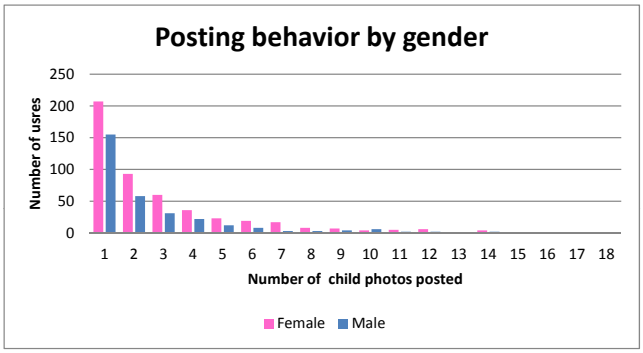

Figure 4: The number of women and men who shared photos of children on Facebook.

Which users account for the majority of the photo sharing? We analyzed several traits in our dataset to find what types of users were sharing photos of children. We note our relevant findings here:

Age.
In Figure 3, we show the percentage of users in each age group who shared child photos, using the set of 2,383 users. As we might expect, photo sharing is highest among users aged 30 to 50, as most parents in this community fall into this age bracket. The median age of the users who shared child photos is 41. Note that a significant fraction of users over 60 are sharing photos of children. We conjecture that the oldest users are sharing photos of their grandchildren.

Gender.
In our dataset of 2,383 users, more women shared child photos than men. 46% of the women in our sample shared a child's photo, as opposed to 23% of men. They also tended to share more photos per user. Among people who had shared child photos, women shared 2.9 photos on average while men shared 2.6. Figure 4 shows the distribution of photo sharing among men and women, respectively.

Politics.
We also examined users' political affiliations, as recorded in their voter registrations. The sample is dominated by political independents and Democrats, since the town profiled leans Democratic. As such, there are relatively few Republicans in the sample. Political affiliation did not make a large difference in whether users shared at least one photo of a child. In our sample, 34% of Democrats, 30% of Republicans, and 33% of independents had shared at least one photo of a child.

3.5 Examples of Oversharing Parents

We found several cogent examples of oversharing parents in our sample. In this section, we showcase some anecdotes to demonstrate more clearly the power that parents wield over their children's privacy. For their protection, we redact any identifying details and merely elaborate the categories known, using false names.

"Laura", age 7: full name and birthdate known. Her family consists of her mother (name and birthdate known), father (name and birthdate known), and an older sister. Her mother works in fashion retailing, and the family's street address is known. Her father is a registered Republican, and her mother is a political independent.

"Jerry", age 0: full name and birthdate known. His family consists of his father (name and birthdate known), mother (name known), and an older sister (name and birthdate known). The father's past and current occupations are known, and he is a football fan and political independent.

"Rebecca", age 2: full name and birthdate known. Her family consists of her father (name and birthdate known), mother (name and birthdate known), two older sisters, and an older brother. Her father is a lawyer and a Republican, and her mother is a political independent.

The data gleaned from a parent's Facebook profile can be rather personal; when conjoined with offline data sources, the information one learns about a child can be highly sensitive. Our work makes clear that this information could be valuable to data brokers, surveillant authorities, or unsavory adults. Parents may unwittingly do their children a disservice when they share too much information.

3.6 Limitations

The automated approach described in this paper for determining the presence of children on a Facebook page is easily scaled. However, it suffers from one major drawback: it can only detect publicly available posts. As a result, it fails to accurately model the threat of malicious acquaintances (who may be able to view friends-only posts) or of surveillant authorities (who can view a user's full posts, either through network traffic analysis, server backdoors, or data requests). To better explore these risks, we conduct a survey in Section 4 where parents report their overall Facebook usage with regard to their children's information.

Additionally, this approach assumes that adults only post photos of their own children on Facebook. In reality, though, adults may post photos of their nieces and nephews, students, friends or even child celebrities. This could be remedied by some heuristics; for example, by using a measure similar to TF-IDF for each face's unique ID, we may more safely determine if the child is a famous person or a family member specific to the account owner. However, we were

unable to implement this or other similar heuristics due to lack of groundtruth data.

A similar point can be made regarding name detection. A conversation in the comments of a photo may mention names other than that of the child. We did not filter for this effect in our experiment due to a lack of groundtruth data. However, an attacker with greater resources and access to more data, such as the social network provider or the NSA, would be able to employ more sophisticated unsupervised learning techniques to guess the correct name with higher accuracy.

Finally, we emphasize that this approach provides only a lower bound on how many parents may be exposing their children's personal information online. As we analyzed only the most recent public posts, it is likely that many more parents are actually sharing child-related information than those we detected. Our intention in the experiment was not to actually expose the children, but rather to prove that it could be accomplished. Notwithstanding the limitations of our approach, our research brings to light a new aspect of children's privacy which has not yet been measured at a large scale in the literature. By demonstrating the information that an attacker (or service provider) can gain about a child through adults' online activities, we hope to bring attention to the impact that a parent has on his child's online privacy.

4. SURVEY OF PARENTS ON FACEBOOK

In order to gain deeper insight into parents' sharing behaviors on Facebook, we conducted an online survey of parents who use Facebook. We used Amazon Mechanical Turk, a crowdsourcing platform, to recruit subjects and directed them to a survey hosted on the Qualtrics platform. We restricted our survey to respondents in the United States. Through a series of demographic questions, we narrowed down our respondent pool to Facebook users who are parents of at least one child under age 13 (the age at which Facebook allows teenagers to create their own accounts).

In order to ensure accurate reporting, we included attention-measuring questions wherein the respondent was directed to select a specific response. Respondents who did not follow the directions were assumed to be inattentive and were excluded from the final analysis.

Since this survey uses self-reported data, it provides a more comprehensive picture of parents' posting behaviors on Facebook. Unlike the scraping approach, it is not limited to public posts; rather, parents reported their overall Facebook usage patterns.

Demographics and Family Makeup.

After filtering for attentiveness, we received 357 responses. 48% of the respondents were male, and 52% were female. A majority (52%) of respondents reported that they had one child; 31% reported 2 children; and 17% of respondents reported 3 or more children.

Behavior on Facebook.

Respondents were directed to check their setting for posts to Facebook. 13% of the users had their posts set to public, and another 77% had chosen friends. See Figure 5 for a full breakdown of the choices.

Though choosing "friends" may appear to preserve a large amount of privacy by limiting one's audience, the parents

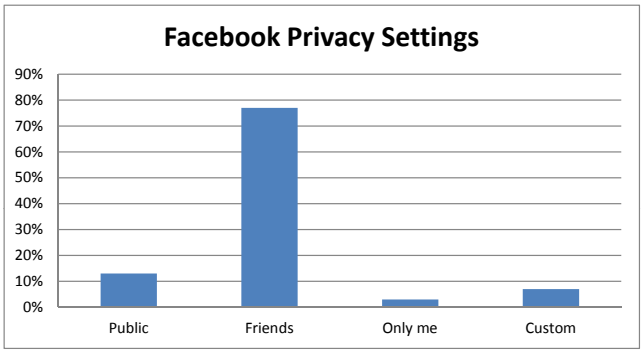

Figure 5: The self-reported Facebook privacy settings chosen by parents of children younger than 13.

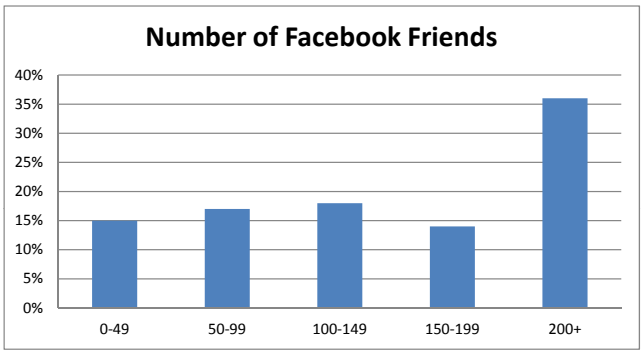

Figure 6: The self-reported Facebook friend count for parents of children younger than 13.

in our survey reported high numbers of friends consistent with findings of other research [31]. The plurality of parents (36%) reported having 200 or more friends, with fully half of respondents reporting a number of friends in the range of 150 and higher. (See Figure 6 for details.) This indicates that even parents who are posting on a friends-only basis are still sharing their photos and comments with large numbers of people.

How much information do parents on Facebook report sharing about their kids? 82% of respondents said they had posted a picture of their child at least once. 77% of the parents said they had mentioned their child's name in a post on Facebook, and 54% of parents said they had mentioned their child's birthday or date of birth. A summary of their responses can be seen in Figure 7.

36 parents - 10% overall - admitted to posting all three of these pieces of information: a photo, name, and birthday. These pieces of personally identifiable information combine to create strong identifiers for their children. (Notably, some parents might not realize that a "Happy birthday" post reveals their child's birthday. Therefore, this number may be underreported.)

Privacy Attitudes.

We included questions in our survey to deduce whether parents were concerned about their children's privacy. We found that parents trended towards moderate regard for privacy, both for themselves and for their children. On a Likert

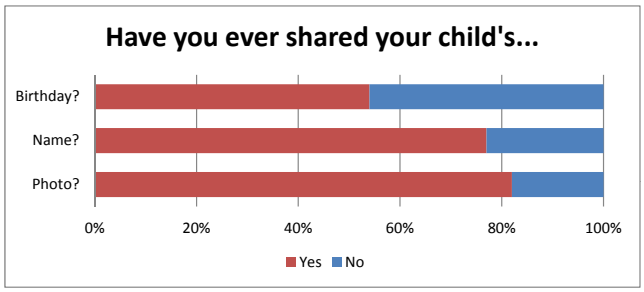

Figure 7: The responses of parents who were asked about their Facebook sharing behaviors.

scale from 1 to 5, parents rated their personal privacy concerns as 3.75, and their privacy concerns for their children as 3.8. Contrary to our expectations, parents were not significantly more concerned for their children's privacy than for their own, as shown by a paired t-test of the scores.

We also asked parents if they believed that they had posted something about their children which could be embarrassing. 11% of parents answered yes, 35% answered no, and 54% were unsure.

Discussion of Findings.

Recall from Section 3 that 35% of the 2,383 Facebook users in our sample publicly shared at least one photo of a child. In particular, among the users in the parenting age group of 30 to 49, 43% shared a photo of a child on their public pages. Although these percentages are substantial, we find that when we ask users to self-report their overall Facebook behavior (not only their public postings), the sharing rates becomes even higher. 82% of the survey respondents said that they share photos of their children. The respondents also indicate significantly more sharing of their children's names and date of birth than what we observed from the public pages of the 2,383 adults. We can therefore conclude that although a substantial percentage of parents are compromising the privacy of their children in their public Facebook pages, significantly more are doing so among Facebook friends. As we note in Section 1, these friends-only photos can still pose a privacy threat to their children.

5. AUTOMATED INSTAGRAM ANALYSIS

As the fastest-growing social site [26], Instagram is rapidly becoming the go-to service for sharing images and photos. As of November 2014, Instagram has more than 200 million active users, and an average of 60 million pictures are uploaded daily [1].

Unlike Facebook, Instagram profiles and posts are fully public by default, and other users can follow the account without approval. Instagram follows a broadcast model (similar to Twitter) unless users specifically change their settings. Instagram's terms of service also state that users must be at least 13 years old[4].

In this section, we describe our analysis of more than 1,000 Instagram accounts. We examine Instagram users who are likely to be parents to find photos and information about their children and predict what an outside viewer may be able to infer about the children.

[4]`http://instagram.com/about/legal/terms/`

5.1 Methods

To find Instagram accounts that were likely to be parents, we used the Instagram API to search for parenting-related hashtags, such as `#mybigboy`, `#mybiggirl`, `#growingtoofast`, `#stopgrowing`, `#sleepingbaby`, `#mylittleprince`, `#mylittleprincess`, `#toddlerbirthday`, and `#firstbirthday`.

This returned a list of photos corresponding to the keywords, along with their associated accounts. A manual analysis of the accounts revealed that not all were relevant; to account for this, we excluded accounts that were associated with less than two of the parenting keywords. We assume that accounts using two or more parenting-related keywords are likely to belong to parents.

After filtering, we downloaded the most recent posts for each of the remaining 1,089 accounts. We then queried the Face++ API for estimates of the ages of the photo subjects. Again, we consider all photos with an estimated age of seven or younger to belong to children. We then proceeded to infer personally identifiable information from the associated comments. If posts contained the words "birthday" or "born," they revealed the child's date of birth. We once again used the Stanford NER tool to extract proper names from the comments after employing some simple sanitation techniques to the text.

5.2 Results

Overall, we considered 1,089 Instagram accounts. We downloaded a total of 21,379 photos, approximately 20 photos per account (as the Instagram API returns the 20 most recent photos by default). The overall results can be viewed in Table 1.

Of the 21,379 photos, 6,134 (28%) were labeled as containing the face of a child. 6,070 (99%) of the child photos included comments or tags. In 988 of these photos, we were able to detect a proper name. Thus, we inferred a name for 16% of the child photos analyzed.

Among the child photos, 317 (5%) mentioned a birthday in the comments. Another 94 photos (2%) mentioned the word "born", which can be used to infer a date of birth. As such, we were able to infer a date of birth for 7% of the child photos.

With respect to accounts, all of the accounts had at least one photo with a child. This indicates that our filtering method (described above in Section 5.1) to locate parents was very accurate. Of the 1082 accounts, 689 (63%) mentioned a child's name in at least one photo. 292 accounts (27% overall) referenced a birth date. 19% of the accounts overall (or 209 individual accounts) referenced both a child's name and date of birth.

Since we do not have auxiliary information about Instagram users, we present a more basic analysis of users' posting behaviors. Among users who posted child photos, the average number of child posts detected in their 20 most recent posts was 5.6, and the median was 5. A distribution of how many photos each user posted can be seen in Figure 8.

5.3 Limitations

Like the Facebook profiling attack, this approach scales quite easily. However, in the process it also suffers from some uncertainty; for example, we cannot verify that the names detected in comments actually belong to the child pictured. Nonetheless, as we state in Section 3.6, a more

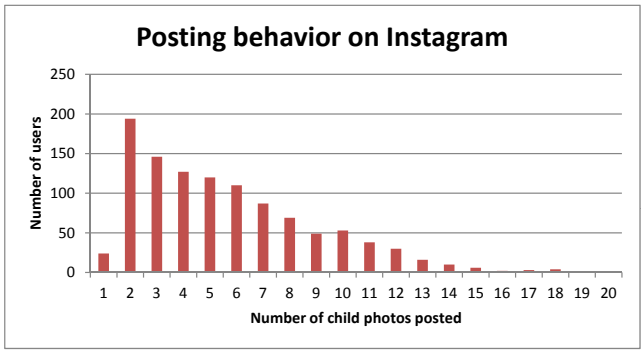

Figure 8: A histogram of the number of child photos shared by users on Instagram.

determined and informed attacker could employ measures to estimate the probability of a specific name belonging to the child. As such, the amount of data shared on Instagram profiles is a matter of concern.

5.4 Comparison of Facebook and Instagram

We find significant variance between the results for inferring children's information on Facebook and Instagram. On Instagram, posts are fully public by default, and the search function for hashtags facilitates finding accounts which belong to parents. Due to these features, it is easier to directly discover children and their data.

Among Facebook users who publicly post a child's photo, less than 50% share the child's name and less than 10% share the child's date of birth. But as shown in Table 1, these numbers are significantly higher for Instagram, with 63% and 27% of the users sharing a child's name and birthday, respectively. We believe that this is largely due to Facebook's more private default sharing settings, whereas Instagram sharing is fully public by default. Indeed, given the Instagram results and the survey results in Section 4, it appears that parents are quite casual about sharing their children's photos, names, and dates of birth in both Facebook and Instagram. However, owing to the difference in the default privacy settings, the information is more accessible to the public when posted on Instagram.

On the other hand, since Facebook encourages users to post many personal attributes on their profiles, more datapoints can be inferred from a Facebook profile. Additionally, the use of Facebook's Graph Search allows an attacker to target a specific geographical area, which in turn enables the profile to be matched more easily to offline data sources such as voter records. Finally, with Instagram, it is not always as easy to infer the last name of a child, as Instagram users often register with pseudonyms.

6. DISCUSSION

In this section, we discuss takeaways from our findings. We also recommend more private behaviors to parents who use these online social tools. Finally, we suggest a Facebook design modification to better protect the privacy of children who are posted on the social network.

6.1 Giving Kids a Chance at Privacy

For children nowadays, navigating the boundaries between public and private is tougher than ever before. While some scholars claim that as "digital natives", adolescents have shed any concern for privacy, teens and children still do care about privacy, as shown by Boyd [5]. Rather, their non-private behaviors are often symptoms of immaturity or ignorance of the specific technologies that can help maintain their desired levels of privacy.

However, we are seeing a move towards more private behavior online, even among children. Applications such as Snapchat, which circumvent the permanence of most digital communications, are very popular among adolescents and teens, since they allow users to share intimate moments without the drama or long-term consequences of persistent messaging applications [5]. Moreover, privacy tools are beginning to become more usable; in particular, Facebook's Privacy Checkup tool urges users to review and update their privacy settings [15].

Currently, adult users of Facebook and Instagram have provided their data by choice, presumably having decided that any potential loss of privacy is worth the utility of a convenient and well-populated social tool. However, the children of these adults have provided no such consent. When a parent shares a child's information online, the child is exposed to non-negligible privacy risk without receiving the attendant benefits of social networking. This is problematic inherently, and it also can reduce a child's privacy agency later in life when the main online service providers are already aware of his presence, personal information, and familial ties.

6.2 Recommendations to Parents

We make the following recommendations to parents who want to preserve their children's online privacy while continuing to use online social networks:

- **Check your Facebook privacy settings.** By using more private settings, parents can limit the audience of potential viewers. Though the service provider (namely, Facebook and any ancillary applications) will still host the data, this can protect children from stranger danger or unsavory acquaintances.
- **Make your Instagram account private.** When a user makes his Instagram private, other users must be approved before viewing the photos on the account. This whitelisting method would allow a parent to share photos with grandparents and other relatives while protecting his child from stranger danger, though again it would not hide the data from the service provider.
- **Think before you share.** A parent can serve as an advocate for his child's privacy by imagining himself in her shoes. How would the parent feel if someone else had shared embarrassing incidents or personal information from his youth in a permanent and semi-public forum?
- **Avoid sharing personally identifiable information whenever possible.** To reduce the likelihood of an adversary learning the child's full identity, we recommend that parents avoid sharing personal information about their children whenever possible. For example, parents should not post children's cell numbers, full names, or birthdates.
- **Encrypt uploaded photos.** Tools such as Cryptagram [34] help users to encrypt any photos uploaded

Figure 9: A mockup interface that Facebook could implement to nudge users towards more private sharing with regard to children's photos.

to Facebook. The photo's owner can then share the key with any user he chooses, allowing them to view it. This hides the photo from unwanted viewers and from surveillant authorities. However, like many applications using cryptography, we recognize that this may not be the most intuitive tool for the average user since it also requires that their friends and family adopt its usage.

Realistically, using any free online service will entail some trade-offs. However, it is important that parents consider the risks before engaging in online sharing about their children.

6.3 Recommendations to Facebook

How can Facebook better protect the privacy of the children who are posted on Facebook by parents or other adults? Similar to the work of Wang et al. [35], we suggest a privacy-preserving mechanism that nudges users to consider more private sharing behaviors with regard to children. If a child's face is detected in a photo, a message can be displayed to encourage the user to select more private settings for the post; see Figure 9 for a graphical example.

Alternatively, Facebook could implement a policy to automatically restrict photos containing children to a more private sharing setting. This would be similar to a past policy regarding teens, whose posts could only be shown to friends-only audiences [21].

7. RELATED WORK

Families and Facebook.

As human interactions move increasingly to the digital realm, research has explored how this affects family dynamics. Burke et al. [6] examine family conversations to find that family member's roles extend into their Facebook conversations; for example, parents of adult children are likely to ask them to call or inquire how the grandchildren are faring. Morris [28] found that new mothers exhibit specific behavioral patterns on Facebook and discusses how these findings can be leveraged to better support women at this critical transition. Kumar and Schoenebeck [22] interviewed 22 mothers of young children and found that mothers often encountered social pressure to share photos of their children. Jomhari et al. [20] described the interactions of mothers in online blogs and social networks as using the new media to tell stories about their children. In a 2012 survey, Bartholomew et al. [4] found that 95% of new mothers and 89% of new fathers had shared images of their babies online.

The medical community has also conducted research on how social media usage can affect children. O'Keeffe et al. [29] point to benefits of social media, such as socialization and enhanced learning opportunities, but they also indicate several risks that apprehend youths on social media. For example, youths may experience cyberbullying, privacy risks, advertising influences, and "Facebook depression," a phenomenon where teens and preteens develop symptoms of depression after excessive social media usage.

Third-party risks to privacy.

Considerable research has demonstrated that a person's privacy can be weakened by the actions of others. In a famous paper, Jernigan and Mistree [19] found that people's sexual orientation could be accurately predicted by the sexual orientation of their friends on Facebook. Similar results were found for age [11], gender [37], and political associations [23], where users' private traits could be predicted by their friends' information.

Children's online privacy.

As access to Internet-connected devices grows, there has been a growing conversation about keeping children safe online. This was formalized with the passage of COPPA, the Children's Online Privacy Protection Act, in the USA. COPPA limits the amount of information that websites may collect about users under age 13 [36]. However, Hargittai et al. [16] found that in many cases, this motivated children to lie about their age with parental consent in order to gain access to more services or features. Dey et al. [9] showed that by lying about their ages, children inadvertently reduced the privacy of their friends who had honestly entered their age. Additionally, Livingstone and Helsper [25] found the surprising result that parental attempts to monitor and limit children's online behavior were not associated with a reduction in overall online risk to the children.

Cranor et al. [8] examined parents' attitudes about their teens' online privacy and found that overall, parents did not take their teens' claims to privacy as seriously as the teens did. However, both Ahern et al. [2] and Kumar and Schoenbeck [22] found that parents practiced some self-censorship, choosing not to share naked or negative photos of their children on Facebook.

Instagram.

As Instagram becomes ever more popular, the research community has started to examine it more closely. Ferrara et al. [12] analyzed its community structure and popular topics, and Manikonda et al. [27] explored user locations and activities. Hu et al. [18] manually coded the content of 50 user profiles to determine what types of content are posted by users. Bakshi et al. [3] found that photos with faces accrued more likes on Instagram. Another work, by Hosseinmardi et al. [17], looked at cyber-bullying on Instagram. While Litt and Hargittai studied users' privacy preferences with regard to online photo sharing [24], we are aware of no work exploring the technical aspects of privacy

on Instagram or the role of children whose photos are posted on Instagram.

Our Contribution.

Past research about children's privacy has focused on two main threats: children's carelessness, or malicious third parties. In this paper, we show that even well-meaning parents can unwittingly compromise their child's privacy by sharing seemingly innocuous updates on Facebook and Instagram. We measure this though two large-scale crawl-based experiment as well as a survey of Facebook users with children under age 13 and determine that the practice of parental oversharing on Facebook can have serious implications.

8. CONCLUSION

What role do parents play in their children's online privacy? In this paper, we show that parents and other adults can inadvertently compromise the privacy of children by oversharing on online social networks. We describe four threats and implement two experiments to quantify the extent of parental oversharing. Firstly, we run an automated analysis of public Facebook pages to discover evidence of children in adults' photo albums and comments. We show how, when correlated to offline data sources, the photos of children on Facebook can trigger a chain reaction of privacy violations. We also conduct a survey to examine parent's self-reported behaviors and attitudes about their children's data on Facebook. We find that many adults are sharing personally identifiable information regarding their children on Facebook, thus weakening their children's privacy with regard to strangers, acquaintances, and surveillant authorities. We then extend the automated analysis to Instagram. Finally, we propose better practices for parents and suggest that Facebook change its interface to encourage better privacy stewardship on the part of parents.

Acknowledgements

This work was supported in part by the NSF (under grants CNS-1318659 and DGE-0966187). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of any of the sponsors.

9. REFERENCES

[1] Instagram press. *http://instagram.com/press/*.

[2] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 357–366. ACM, 2007.

[3] S. Bakhshi, D. A. Shamma, and E. Gilbert. Faces engage us: photos with faces attract more likes and comments on instagram. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 965–974. ACM, 2014.

[4] M. K. Bartholomew, S. J. Schoppe-Sullivan, M. Glassman, C. M. Kamp Dush, and J. M. Sullivan. New parents' facebook use at the transition to parenthood. *Family relations*, 61(3):455–469, 2012.

[5] D. Boyd. *It's Complicated: the social lives of networked teens*. Yale University Press, 2014.

[6] M. Burke, L. A. Adamic, and K. Marciniak. Families on facebook. In *ICWSM*, 2013.

[7] A. Considine. Making facebook less infantile. *New York Times*, August 9 2012.

[8] L. F. Cranor, A. L. Durity, A. Marsh, and B. Ur. Parents' and teens' perspectives on privacy in a technology-filled world. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[9] R. Dey, Y. Ding, and K. W. Ross. Profiling high-school students with facebook: how online privacy laws can actually increase minors' risk. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 405–416. ACM, 2013.

[10] R. Dey, Y. Ding, and K. W. Ross. Profiling city residents using publicly available information. Technical report, New York University, Computer Science and Engineering, October 2014.

[11] R. Dey, C. Tang, K. Ross, and N. Saxena. Estimating age privacy leakage in online social networks. In *INFOCOM, 2012 Proceedings IEEE*, pages 2836–2840. IEEE, 2012.

[12] E. Ferrara, R. Interdonato, and A. Tagarelli. Online popularity and topical interests through the lens of instagram. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 24–34. ACM, 2014.

[13] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 363–370. Association for Computational Linguistics, 2005.

[14] D. Finkelhor and R. Ormrod. *Kidnaping of Juveniles: Patterns from NIBRS*. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, 2000.

[15] J. Guyn. Facebook rolling out privacy checkup for users. *USA Today*, September 4 2014.

[16] E. Hargittai, J. Schultz, J. Palfrey, et al. Why parents help their children lie to facebook about age: Unintended consequences of the 'children's online privacy protection act'. *First Monday*, 16(11), 2011.

[17] H. Hosseinmardi, S. Li, Z. Yang, Q. Lv, R. I. Rafiq, R. Han, and S. Mishra. A comparison of common users across instagram and ask.fm to better understand cyberbullying. *ArXiv Preprints*, 2014.

[18] Y. Hu, L. Manikonda, and S. Kambhampati. What we instagram: A first analysis of instagram photo content and user types. In *Proceedings of ICWSM*. AAAI, 2014.

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

[20] N. Jomhari, V. M. Gonzalez, and S. H. Kurniawan. See the apple of my eye: baby storytelling in social space. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, pages 238–243. British Computer Society, 2009.

[21] H. Kelly. Facebook changes privacy settings for teens. *CNN*, October 31, 2013.

[22] P. Kumar and S. Schoenebeck. The modern day baby book: Enacting good mothering and stewarding privacy on facebook. In *Computer Supported Cooperative Work and Social Computing (CSCW '15)*. ACM, 2015.

[23] J. Lindamood, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham. Inferring private information using social network data. In *Proceedings of the 18th international conference on World wide web*, pages 1145–1146. ACM, 2009.

[24] E. Litt and E. Hargittai. Smile, snap, and share? a nuanced approach to privacy and online photo-sharing. *Poetics*, 42:1–21, 2014.

[25] S. Livingstone and E. J. Helsper. Parental mediation of children's internet use. *Journal of broadcasting & electronic media*, 52(4):581–599, 2008.

[26] I. Lunden. Instagram is the fastest-growing social site globally, mobile devices rule over pcs for access. *TechCrunch*, January 21, 2014.

[27] L. Manikonda, Y. Hu, and S. Kambhampati. Analyzing user activities, demographics, social network structure and user-generated content on instagram. *ArXiv Preprints*, 2014.

[28] M. R. Morris. Social networking site use by mothers of young children. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1272–1282. ACM, 2014.

[29] G. S. O'Keeffe, K. Clarke-Pearson, et al. The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4):800–804, 2011.

[30] N. Singer. Senator opens investigation of data brokers. *New York Times*, October 10, 2012.

[31] A. Smith. 6 new facts about facebook. *Pew Research Center*, February 3, 2014.

[32] S. Stecklow. On the web, children face intensive tracking. *Wall Street Journal*, September 17, 2010.

[33] A. Sultan and J. Miller. Facebook parenting is destroying our children's privacy. *CNN*, May 25 2012.

[34] M. Tierney, I. Spiro, C. Bregler, and L. Subramanian. Cryptagram: photo privacy for online social media. In *Proceedings of the first ACM conference on Online social networks*, pages 75–88. ACM, 2013.

[35] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, and L. F. Cranor. Privacy nudges for social media: an exploratory facebook study. In *Privacy and Security in Online Social Media (PSOSM)*. ACM, 2013.

[36] J. Warmund. Can coppa work-an analysis of the parental consent measures in the children's online privacy protection act. *Fordham Intell. Prop. Media & Ent. LJ*, 11:189, 2000.

[37] 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 *Proceedings of the 18th international conference on World wide web*, pages 531–540. ACM, 2009.