# Moving Beyond Set-It-And-Forget-It
Privacy Settings on Social Media

<table>
<thead>
<tr>
<th>Authors</th>
<th>Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainack Mondal†*</td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Mohammad Taha Khan§</td>
<td>§ University of Illinois at Chicago</td>
</tr>
<tr>
<td>Chris Kanisch§</td>
<td></td>
</tr>
<tr>
<td>Günce Su Yılmaz†</td>
<td>Noah Hirsch†</td>
</tr>
<tr>
<td>Michael Tang†</td>
<td>Christopher Tran§</td>
</tr>
<tr>
<td>Blase Ur†</td>
<td>Elena Zheleva§</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>†University of Chicago</td>
<td>§ University of Illinois at Chicago</td>
</tr>
<tr>
<td>*IIT Kharagpur</td>
<td></td>
</tr>
</tbody>
</table>
Need to revisit old privacy settings

2009

Privacy setting: “all friends”
Need to revisit old privacy settings

2009

Privacy setting: “all friends”

Undergraduate friends
Need to revisit old privacy settings

2019

Privacy setting: “all friends” !!
Need to revisit old privacy settings

2019

Privacy setting: “all friends”!!
Need to revisit old privacy settings

2019

Privacy setting: “all friends” !!
Need to revisit old privacy settings

2019

Privacy setting: "all friends"!!
Need to revisit old privacy settings

2019

Students

Undergraduate friends

Graduate school friends

Professional colleagues

Privacy setting: “all friends” !!
Issue: Users take a “set-it-and-forget-it” approach to privacy settings for social media posts

Need: Retrospectively manage privacy
Retrospective privacy management is difficult

Undergraduate friends
Retrospective privacy management is difficult
Retrospective privacy management is difficult
State of the art

No proposal for a predictive model or mechanism

[Bauer et al. 2013]
[Ayalon et al. 2013]
State of the art

No proposal for a predictive model or mechanism [Bauer et al. 2013] [Ayalon et al. 2013]

Focus of our study

Measure privacy activity and preferences
Predictive models for retrospective privacy management
Outline

- Our data collection approach
- Privacy settings and friend network over time
- Preferences for changing privacy settings
- Automated classifiers
Outline

Our data collection approach

Privacy settings and friend network over time

Preferences for changing privacy settings

Automated classifiers
Study overview

Privacy-preserving data-collection infrastructure → 78 Facebook users → Two surveys
Generic survey

- Overall Facebook usage over time
- Use of Facebook’s privacy features
- Participant demographics
Consent process

Consent form

Highlights
Data collection process
Data collection process

Programmatic
No humans ever view raw HTML

Hash names and IDs;
No images collected

Never access friends’ profiles
Facebook Timeline data

Chose not to store images!
Facebook Activity Log data

ALL Facebook activities by user (friendship, likes, comments,...)
Post-specific survey

1. Desired privacy settings for 5 random posts per user
Post-specific survey

1. Desired privacy settings for 5 random posts per user

Click here to see Post 1. Current privacy setting: Public
Post-specific survey

1. Desired privacy settings for 5 random posts per user

Click [here to see Post 1]. Current privacy setting: Public

- Keep same setting
- Change setting to:
  - [Options]
  - [Options]
  - [Options]
  - [Options]

- Delete
Post-specific survey

1. Desired privacy settings for 5 random posts per user

Click [here](#) to see Post 1. Current privacy setting: Public

Change setting to:

- [ ] Keep same setting
- [ ] Delete

Why?
Post-specific survey

1. Desired privacy settings for 5 random posts per user
2. Desired privacy settings for 6 specific friends per post
Post-specific survey

1. Desired privacy settings for 5 random posts per user
2. Desired privacy settings for 6 specific friends per post

This question concerns Post 1 and one of your Facebook friends: Blase Ur
You can visit Blase Ur's profile by clicking his picture:
Post-specific survey

1. Desired privacy settings for 5 random posts per user
2. Desired privacy settings for 6 specific friends per post
Post-specific survey

1. Desired privacy settings for 5 random posts per user
2. Desired privacy settings for 6 specific friends per post

This question concerns Post 1 and one of your Facebook friends: Blase Ur
You can visit Blase Ur’s profile by clicking his picture:

Keep sharing post 1 with Blase Ur

Stop sharing post 1 with Blase Ur

Why?
Demographics

- AMT workers from US
- 69% identified as female
- 46% reported age 25-34
- 18% reported CS background
### Facebook usage

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account age (Years)</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>#Friends</td>
<td>-</td>
<td>224</td>
</tr>
<tr>
<td>#Timeline posts</td>
<td>253,122</td>
<td>1,840</td>
</tr>
<tr>
<td>#Activity log entries</td>
<td>1,738,303</td>
<td>20,263</td>
</tr>
</tbody>
</table>
Facebook usage

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account age (Years)</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>#Friends</td>
<td>-</td>
<td>224</td>
</tr>
<tr>
<td>#Timeline posts</td>
<td>253,122</td>
<td>1,840</td>
</tr>
<tr>
<td>#Activity log entries</td>
<td>1,738,303</td>
<td>20,263</td>
</tr>
</tbody>
</table>

Active users with old accounts and lots of posts
Facebook usage

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account age (Years)</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>#Friends</td>
<td>-</td>
<td>224</td>
</tr>
<tr>
<td>#Timeline posts</td>
<td>253,122</td>
<td>1,840</td>
</tr>
<tr>
<td>#Activity log entries</td>
<td>1,738,303</td>
<td>20,263</td>
</tr>
</tbody>
</table>

Active users with old accounts and lots of posts

67% reported reduced Facebook usage over time
Outline

Our data collection approach

Privacy settings and friend network over time

Preferences for changing privacy settings

Automated classifiers
Privacy settings over time

% of Posts

Year

Privacy settings over time

Majority of old posts are shared with all “friends”
Change in number of friends
Change in number of friends

- Year
  - 2009
  - 2010
  - 2011
  - 2012
  - 2013
  - 2014
  - 2015
  - 2016
  - 2017
  - 2018

- % of Current Friends
  - 0
  - 20
  - 40
  - 60
  - 80
  - 100

- Change in number of friends

- % of current friends

- Participants
Change in number of friends

- % of Current Friends: 0, 20, 40, 60, 80, 100

- Change in number of friends: 25%
Change in number of friends

Year

% of Current Friends


0 20 40 60 80 100

25%

50%

Change in number of friends

Substantial change in the meaning of “friends” privacy setting
Outline

Our data collection approach

Privacy settings and friend network over time

Preferences for changing privacy settings

Automated classifiers
Desired privacy setting for old posts

Post-specific survey: Desired privacy setting for 390 random posts
### Desired privacy setting for old posts

Post-specific survey: **Desired privacy setting for 390 random posts**

<table>
<thead>
<tr>
<th>Current setting</th>
<th>Public</th>
<th>Friends+</th>
<th>Friends</th>
<th>Custom</th>
<th>Only Me</th>
<th>Custom (Decreased)</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>58</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Friends+</td>
<td>3</td>
<td>27</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td>21</td>
<td>4</td>
<td>177</td>
<td>3</td>
<td>5</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>Custom</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>19</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Only Me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

Total: 389
Desired privacy setting for old posts

Post-specific survey: **Desired privacy setting for 390 random posts**

<table>
<thead>
<tr>
<th>Current setting</th>
<th>Public</th>
<th>Friends+</th>
<th>Friends</th>
<th>Custom</th>
<th>Only Me</th>
<th>Custom (Decreased)</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>58</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Friends+</td>
<td>3</td>
<td>27</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Friends</td>
<td>21</td>
<td>4</td>
<td>177</td>
<td>3</td>
<td>5</td>
<td>-</td>
<td>31</td>
</tr>
<tr>
<td>Custom</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>19</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Only Me</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9</td>
<td>-</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8

Participants **desire to change audience for 25% of old posts!**
Post-specific survey: **Desired privacy setting for 390 random posts**

<table>
<thead>
<tr>
<th>Current setting</th>
<th>Public</th>
<th>Friends+</th>
<th>Friends</th>
<th>Custom</th>
<th>Only Me</th>
<th>Custom (Decreased)</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>58</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Friends+</td>
<td>3</td>
<td>27</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td>21</td>
<td>4</td>
<td>177</td>
<td>3</td>
<td>5</td>
<td>-</td>
<td>31</td>
</tr>
<tr>
<td>Custom</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>19</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Only Me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

Desire to limit audience: 54 posts

Desire to increase audience: 45 posts

Participants **desire to change audience for 25% of old posts!**
Effectiveness of Facebook's privacy tools

Limit The Audience for Old Posts on Your Timeline

If you choose to limit your past posts, posts on your timeline that you've shared with Friends of friends, and Public posts, will now be shared only with Friends. Anyone tagged in these posts, and their friends, may also still see these posts.

If you want to change who can see a specific post, you can go to that post and choose a different audience. Learn about changing old posts

Privacy Checkup

Hi Charlie!

We have a new tool that helps you quickly review a few of your privacy settings to make sure they’re set up the way you want.

It should take a minute or two to use. Do you want to check it out?

No Thanks  Let’s Do It!

Found no significant correlation between usage of these tools and the desire to change posts’ privacy settings
Outline

Our data collection approach

Privacy settings and friend network over time

Preferences for changing privacy settings

Automated classifiers
A human-in-the-loop design
A human-in-the-loop design

**Inspiration**
- **People You May Know**
  - **John Doe**
    - Johnny Doe and 14 other mutual friends
  - **Jenny Doe**
    - Jane Doe and 6 other mutual friends
  - **Matt Doe**
    - Robert Doe and 41 other mutual friends
  - **Julia Doe**
    - Robert Doe and 40 other mutual friends
  - **Judith Doe**
    - Robert Doe and 35 other mutual friends

**Our vision**
- Stop sharing with
  - **Malayak Mondal**
    - August 15, 2014
    - is procrastinating
    - Like
    - Comment
    - Share

- Stop sharing with
  - **Malayak Mondal**
    - March 12, 2015
    - is feeling dumb.
    - Like
    - Comment
    - Share

- Stop sharing with
  - **Malayak Mondal**
    - November 16, 2017
    - I have lost the will to live.
    - Like
    - Comment
    - Share

- Stop sharing with
  - ???
Prediction task

**Prediction task**

Predict if a user wants to “stop sharing” a given post with a given friend

**Output**

List of friend-post pairs ordered by probability

**Ground truth**

Privacy decisions for 78 participants x 5 posts X 6 friends = 2,340 pairs
## Features for prediction

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-specific</td>
<td>#friends, age of the account, life change, Facebook privacy tool usage, user age, CS-background</td>
</tr>
<tr>
<td>Post metadata</td>
<td>Age of the post, #likes, #comments, previous change in privacy setting, type of post, tagged friend</td>
</tr>
<tr>
<td>Post content</td>
<td>Word2vec embeddings, Google content-classification categories, sentiment</td>
</tr>
<tr>
<td>Friend-specific</td>
<td>Days since first and last communication, #wall words exchanged, #likes from friend to user</td>
</tr>
</tbody>
</table>
Prediction algorithms

Supervised learning algorithms with cross validation

**Random Forests, XGBoost**, Decision Trees, Logistic Regression, Support Vector Machines, Deep Neural Networks

Baselines

Random: Randomly predicts “stop sharing” for a pair
Interaction: Low interaction level $\rightarrow$ “stop sharing”
Are our models better than the baselines?
Are our models better than the baselines?

The fraction of posts the user actually wanted to stop sharing that were predicted
Are our models better than the baselines?

- Precision
- Recall

The fraction of predictions for which the user actually wanted to stop sharing

The fraction of posts the user actually wanted to stop sharing that were predicted
Are our models better than the baselines?
Are our models better than the baselines?
Are our models better than the baselines?
Are our models better than the baselines?
Are our models better than the baselines?

Substantial improvement over baselines
Prediction task

Predict if a user wants to “stop sharing” a given post with a given friend

Output

List of friend-post pairs ordered by probability

Ground truth

Privacy decisions for 78 participants x 5 posts X 6 friends = 2,340 pairs
Recommendation accuracy of our models

![Graph showing precision@k vs K]
Recommendation accuracy of our models

![Graph showing precision at different cutoffs for Random Forest and XGBoost models. The x-axis represents the prediction cutoff K, ranging from 1 to 30. The y-axis represents precision@k, ranging from 0.0 to 1.0. The graph compares the performance of Random Forest (teal diamonds) and XGBoost (purple diamonds) models.]
Recommendation accuracy of our models

![Graph showing precision@k for Random Forest and XGBoost models.](image)
Recommendation accuracy of our models

30 recommendations with good precision!
Understanding inaccurate predictions

Qualitative data from survey: “Why” did desired setting change?
Understanding inaccurate predictions

Qualitative data from survey: “Why” did desired setting change?

“I no longer participate in these activities and don’t find them appropriate any longer.”
Understanding inaccurate predictions

Qualitative data from survey: “Why” did desired setting change?

“I no longer participate in these activities and don’t find them appropriate any longer.”

“Because the people I feel close to has changed in the years since that post.”
Understanding inaccurate predictions

 Qualitative data from survey: “Why” did desired setting change?

“I no longer participate in these activities and don’t find them appropriate any longer.”

“Because the people I feel close to has changed in the years since that post.”

“ it shows a time that I was upset and I would rather not relive that.”
Understanding inaccurate predictions

Qualitative data from survey: “Why” did desired setting change?

“I no longer participate in these activities and don’t find them appropriate any longer.”

“Because the people I feel close to has changed in the years since that post.”

“it shows a time that I was upset and i would rather not relive that.”

Coded this data to identify additional predictive features for future efforts
Future features to collect

<table>
<thead>
<tr>
<th>Features of posts</th>
<th>Features from external content (image/video)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classes of sensitive information (e.g., children)</td>
</tr>
<tr>
<td></td>
<td>Similarity of content with user’s current interest</td>
</tr>
</tbody>
</table>
## Future features to collect

<table>
<thead>
<tr>
<th>Features of posts</th>
<th>Features from external content (image/video)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classes of sensitive information (e.g., children)</td>
</tr>
<tr>
<td></td>
<td>Similarity of content with user’s current interest</td>
</tr>
<tr>
<td>Features of friends</td>
<td>Interests, likes and dislikes of specific friends</td>
</tr>
<tr>
<td></td>
<td>If particular friends are close family or related</td>
</tr>
<tr>
<td></td>
<td>Frequency of offline interaction</td>
</tr>
</tbody>
</table>
Summary

Median 1,840 #old posts, 400% increase in audience size within 10 years

Users want to change privacy for 25% of old posts

Predictive recommender system achieves > 80% precision identifying some friend-post pairs needing privacy change
Summary

Median 1,840 #old posts, 400% increase in audience size within 10 years

Users want to change privacy for 25% of old posts

Predictive recommender system achieves > 80% precision identifying some friend-post pairs needing privacy change

Moving Beyond Set-It-And-Forget-It Privacy Settings on Social Media
Mainack Mondal, Günce Su Yilmaz, Noah Hirsch, Mohammad Taha Khan, Michael Tang, Christopher Tran, Chris Kanich, Blase Ur, Elena Zheleva.
Extra slides!
Privacy setting change over time

Normalized by users

(#posts for 55 participants who created their account on or before 2009)
Worry about others checking old posts

86% expect other people check their old posts
55% reported checking others’ old posts
Purposes: checking past relationships, family history

Might partially motivate changes in privacy settings
Correlating interaction / friend-addition time

Post-specific survey: privacy setting change for friend-post pairs
Friends chosen based on interaction frequency and addition time

Hypothesis: Desired privacy changes correlate with interaction frequency
    Found only weak correlation
    Mostly offline communication with some close relations

Hypothesis: Desired privacy changes correlate with when friend was added
    Found no correlation