Evolution of Privacy Loss in Wikipedia

March 16th, 2016

Marian-Andrei Rizoiu

WIKIMEDIA FOUNDATION

Motivation

- Social media and online privacy are two of today's hot topics.
- Given that we know that digital traces reveal more than users might think [Kosinski, PNAS 2013], we ask the next questions.

- Goals of this work:
  - Does online user's privacy degrade over time?
  - What factors contribute most to revealing private traits?
  - Can I stop leaking personal information if I stop posting online?

Content of this presentation

Presentation outline

- Case study: Wikipedia dataset
- Profiling of editing behavior
- Measuring predictability of personal traits
- Marginal utility of features over time
- Conclusion and the way ahead

Case study: Wikipedia

Why Wikipedia?

- 13 years long, public: ideal for longitudinal study;
- tens of thousands of editors, of different geographic locations, religious, educational and political backgrounds;
- *apparently harmless dataset*: a reservoir of knowledge, no focus on personal information.

Dataset dimensionality:

- 188,805,088 revisions
- 117,523 editors
- 8,679 editor badges
- 22,172,813 edited pages
- 430,410 page categories
- Time extent: January 2001 - July 2013.

Encoding editing behavior (1)

Editor activity profiles:

- *basic set*: #revisions over 6 predefined categories (Wiki namespaces);
- *extended set*: adding Wikipedia's 23 high level thematic categories (Math, Geography, History etc.)

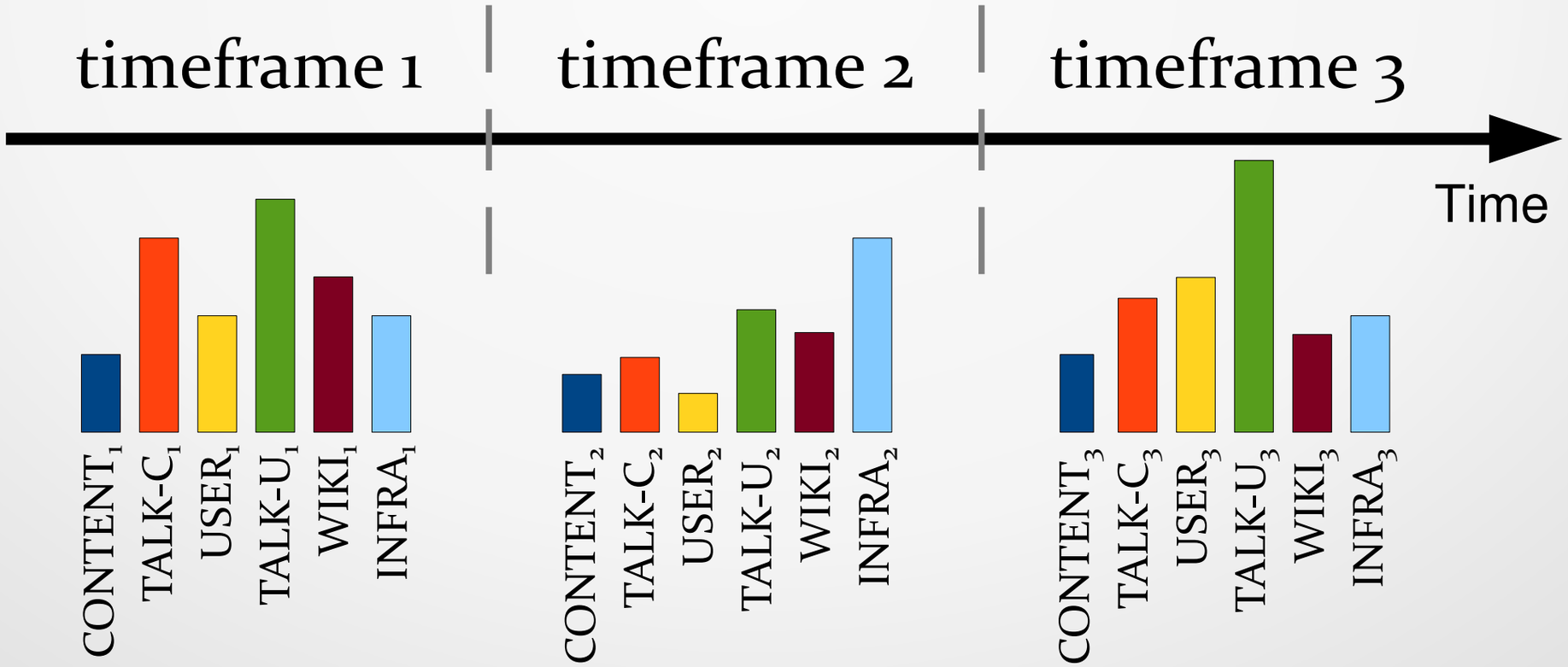
	Feature name	Namespaces
Basic feature set	CONTENT	0, 6
	TALK-C	1, 7
	USER	2
	TALK-U	3
	WIKI	4, 5
	INFRA	8, 9, 10, 11, 12, 13, 14, 15, 100, 101

Editor personal information:

- Extracted from the badges editors put on their editor pages;
- **Gender (6936 out of more than 117k)**, ethnic origin, **religious views (7685)**, **education (9224)**, sexual orientation *etc.*

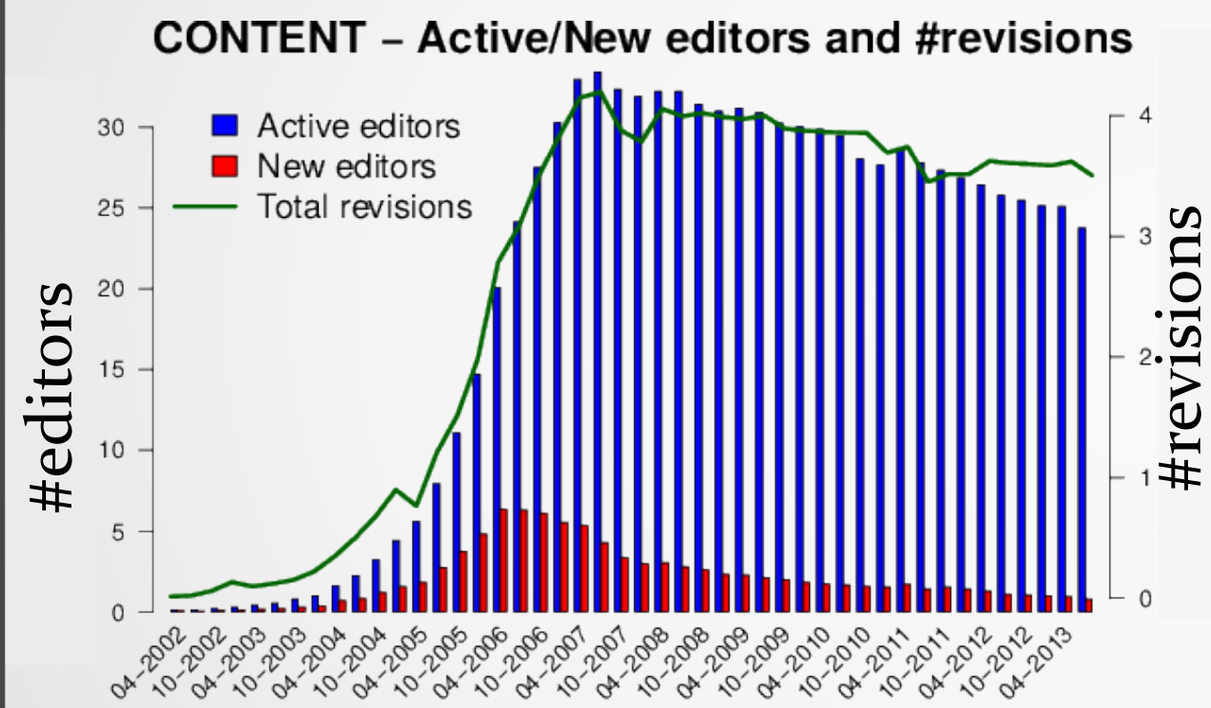
Encoding editing behavior (2)

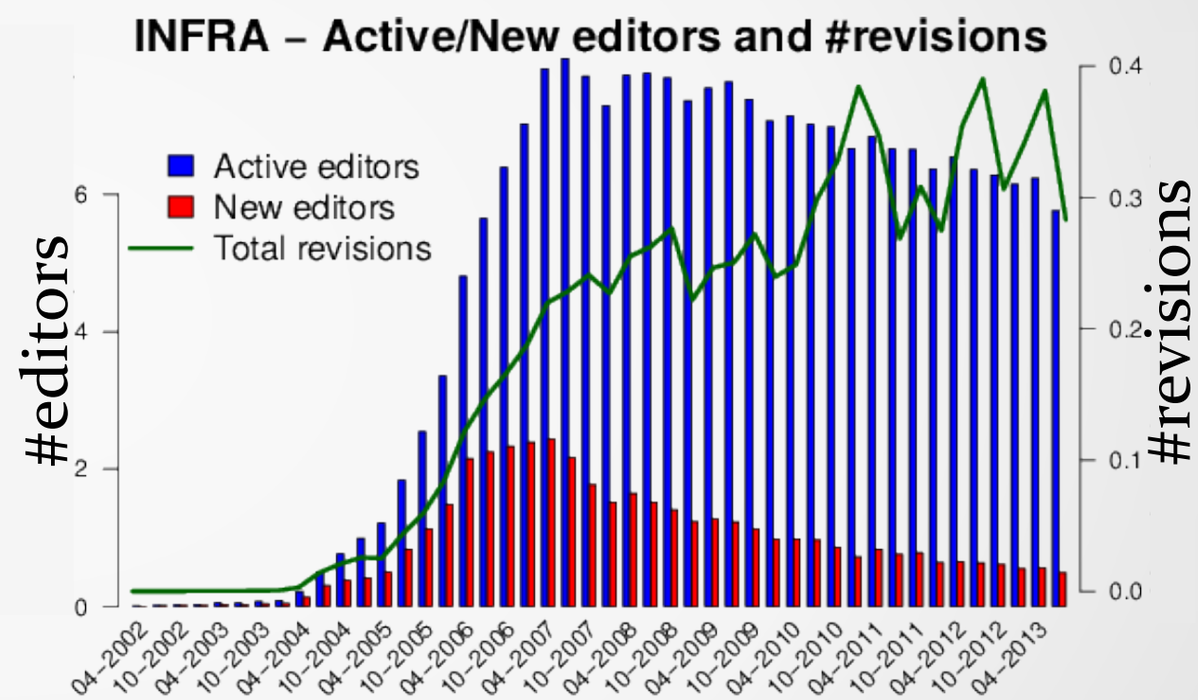
- 3-month timeframes
- description for each editor per timeframe, each feature counts revision over categories
- feature set temporally embeds increasing amounts of information



Profiling of editing behavior (1)

Decline of editorship and rise of maintenance

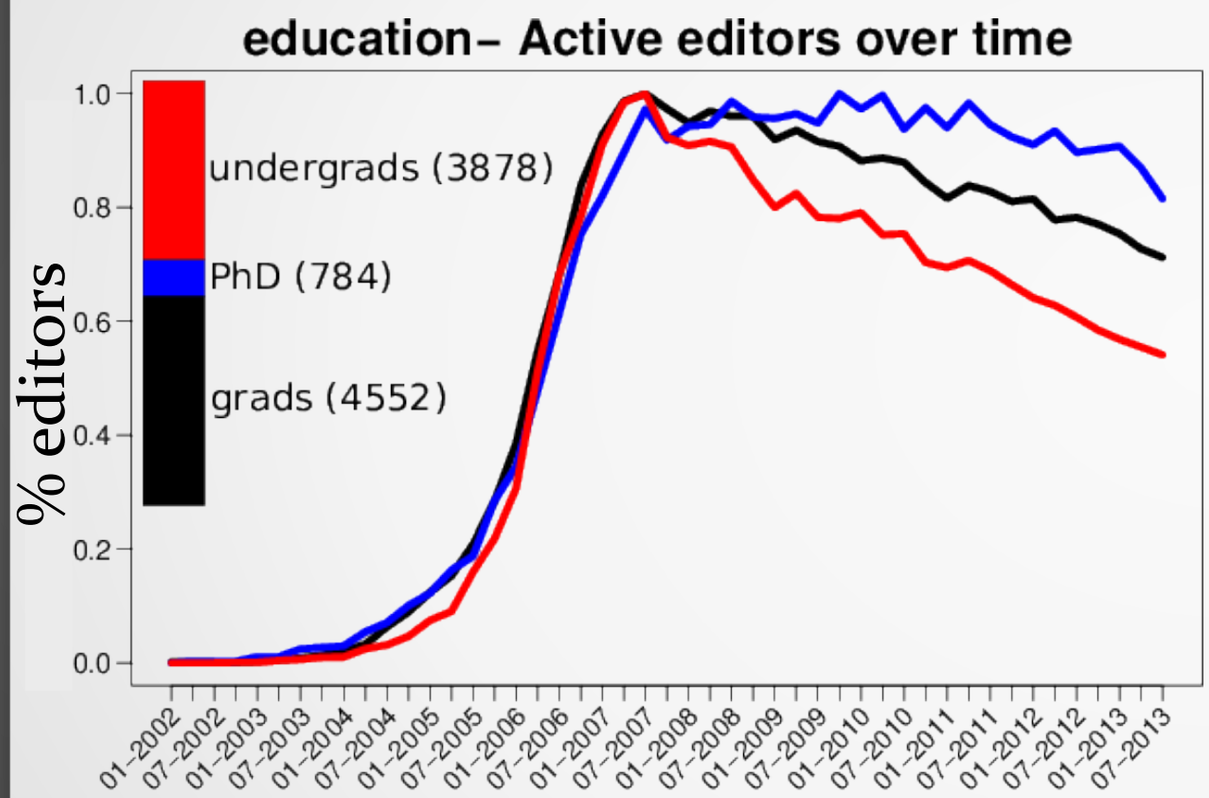


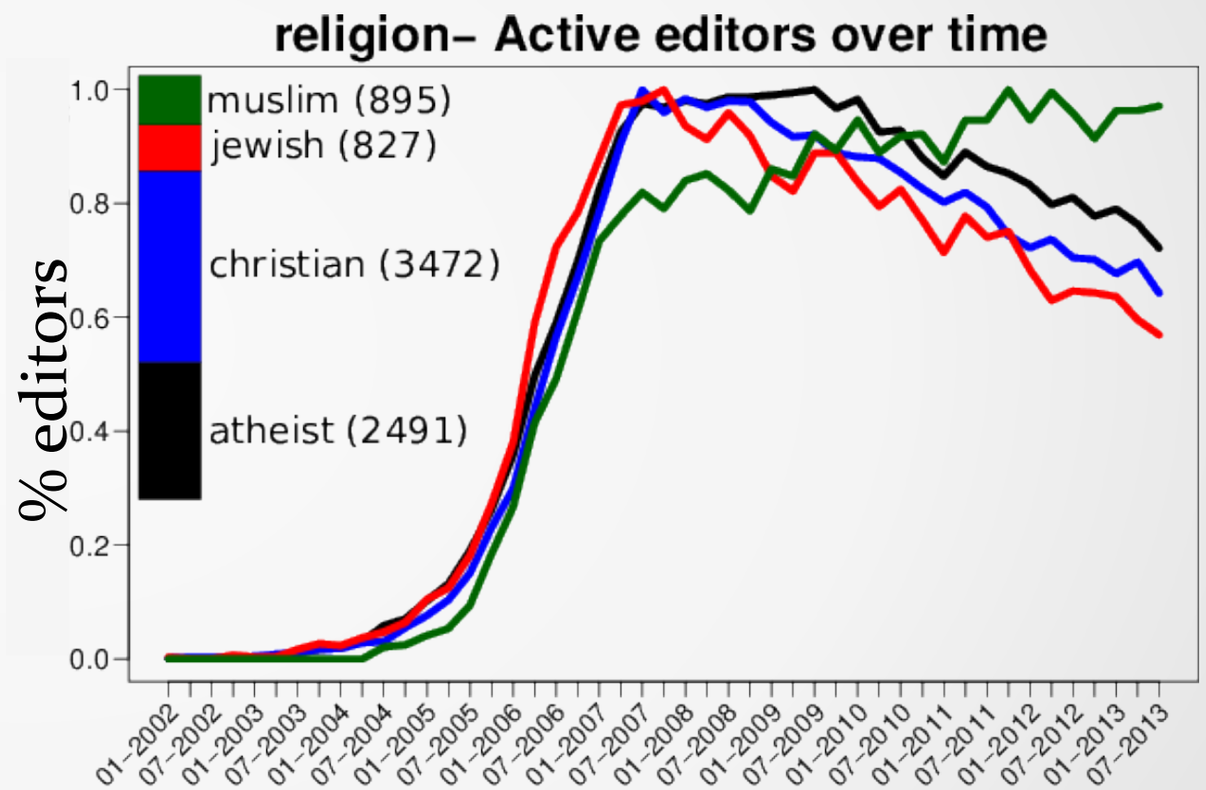


While the Wikipedia "slowdown" has been previously reported **[Suh '09, Halfaker '12]**, we break down this evolution per category and detect a *rise of maintenance effort*

Profiling of editing behavior (2)

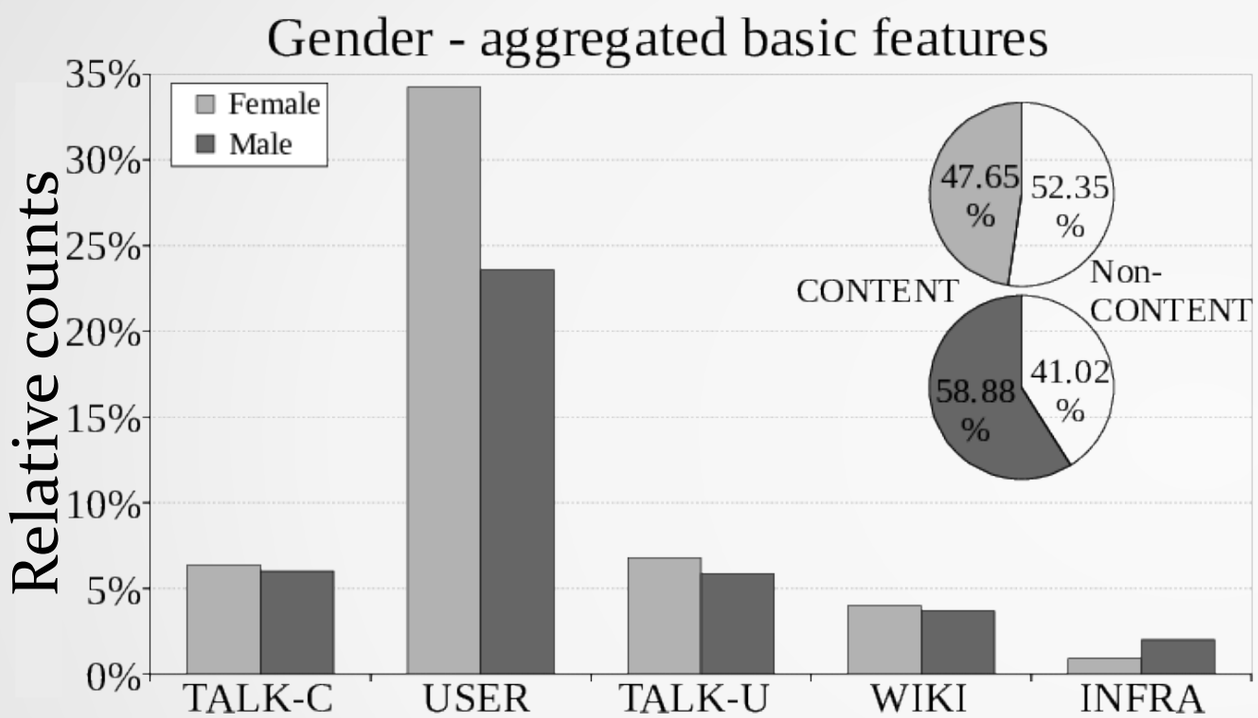
Different growth trends across editor demographics

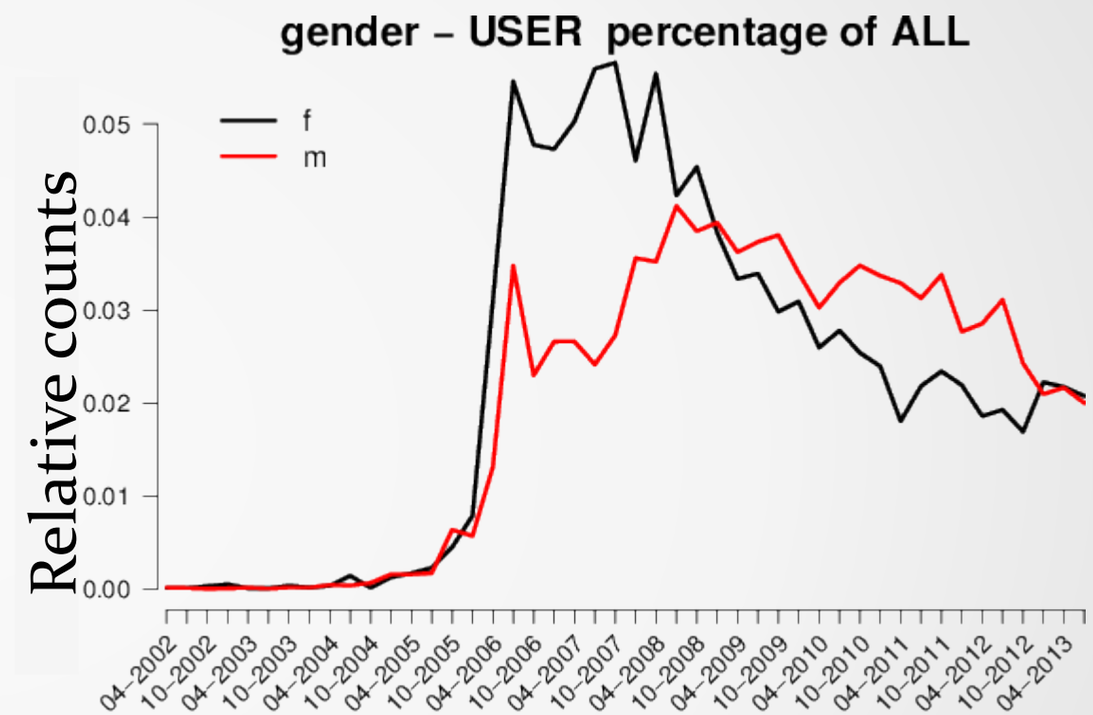




Evolution trends across editor categories are unequal, providing plausible explanations for the slowdown **[Gibbons '12]**, as well as *personal identification clues*.

Edit behavior correlates with private traits

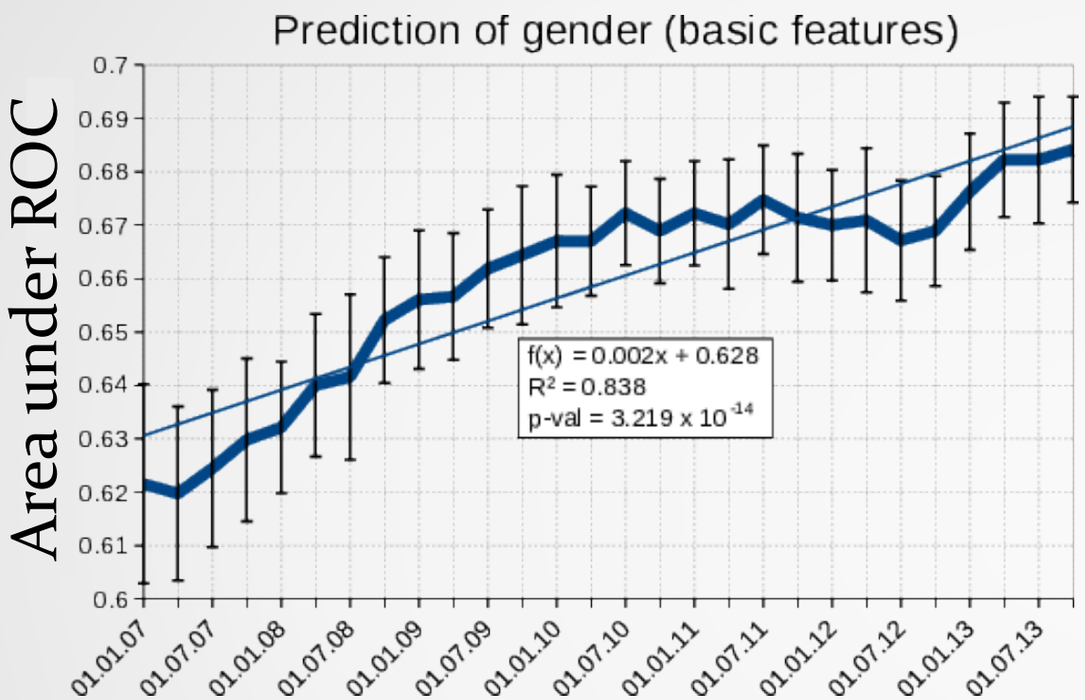


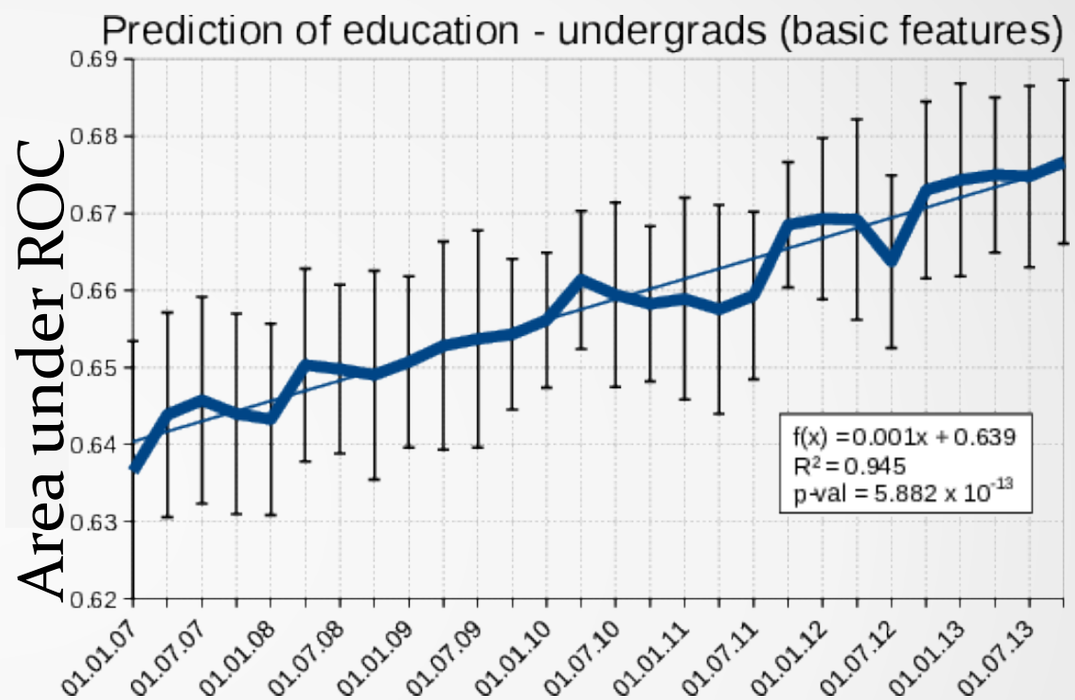


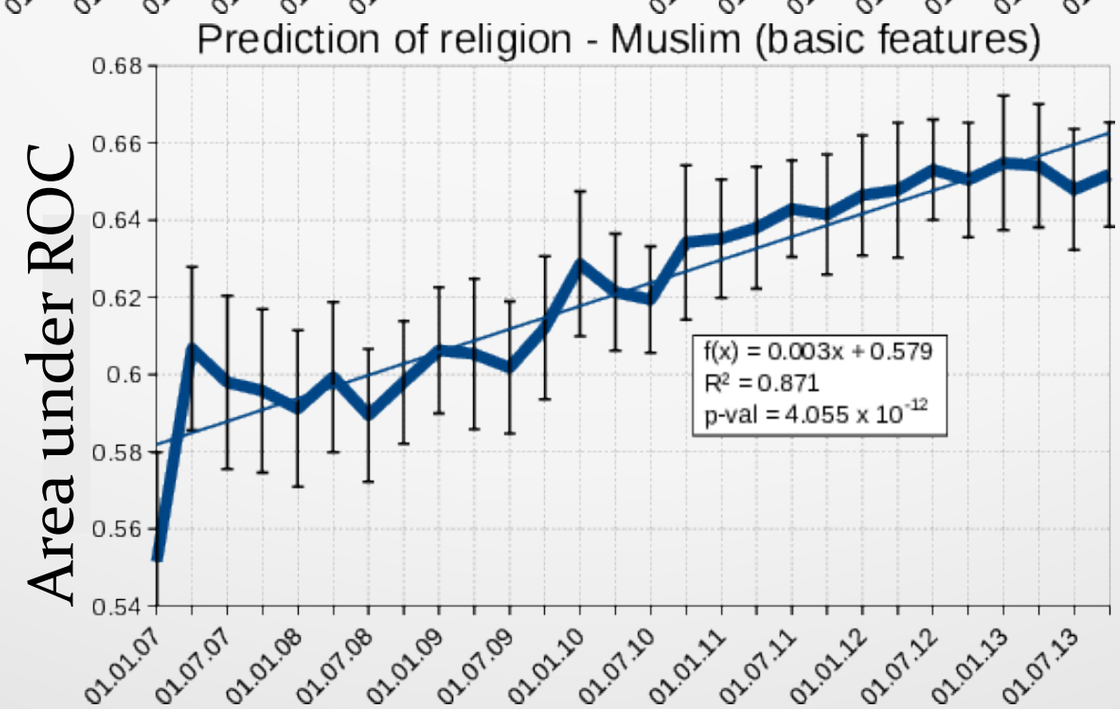
Mean editing behavior analysis shows regularities in the editing patterns for each sub-population.

Predictability improves over time

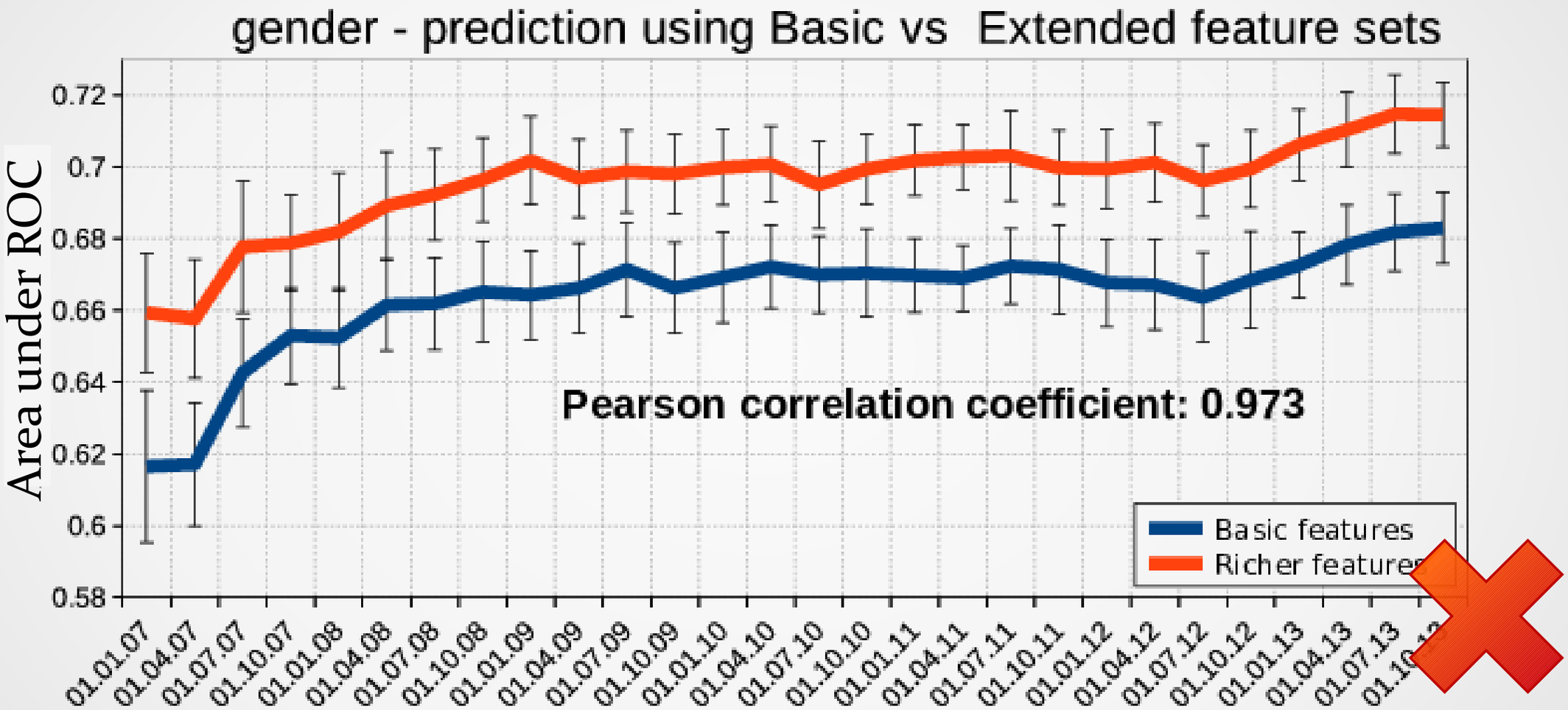
Privacy Loss as a *prediction problem*





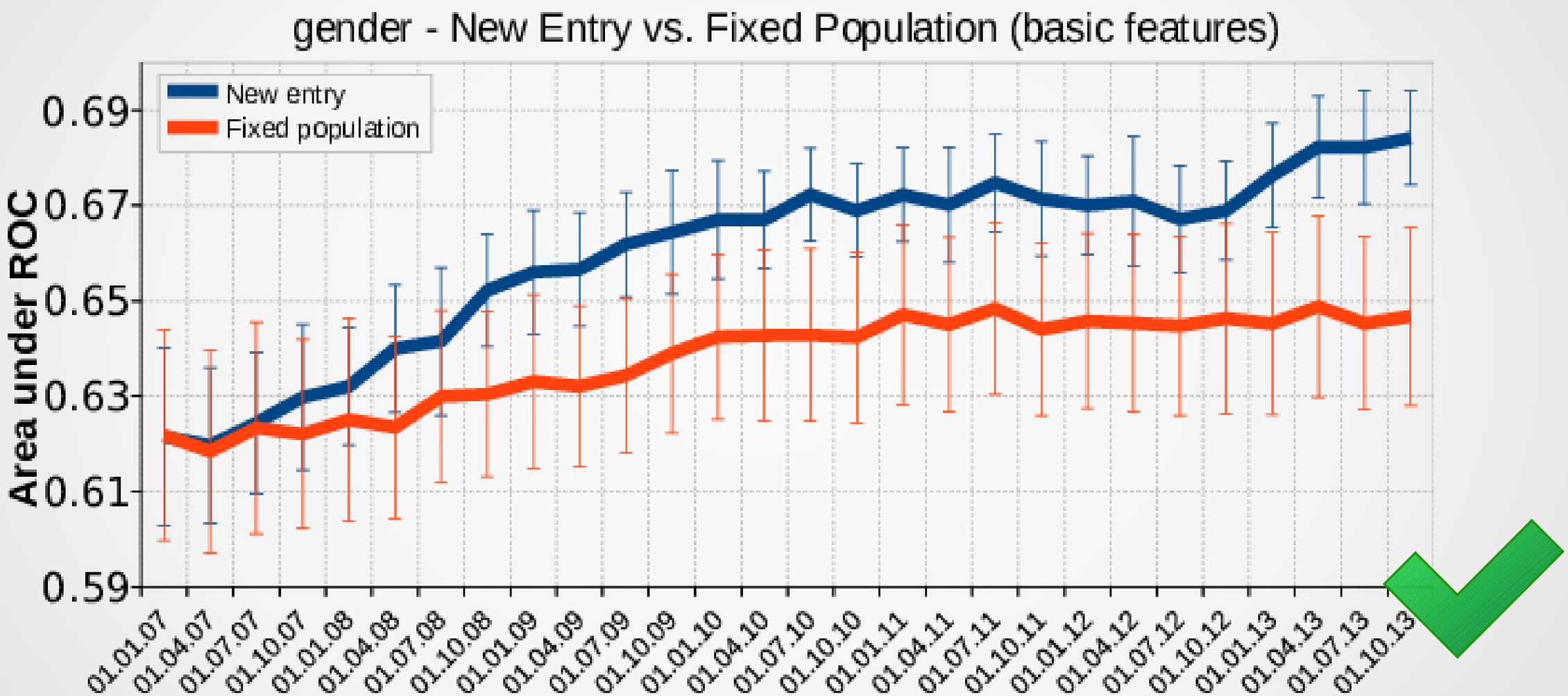


Sources of Privacy Loss (1)



Richer features: improve prediction,
but not *Privacy Loss*

Sources of Privacy Loss (2)



Newcomers: information from *newcomers* hurts privacy

Marginal utility of features over time

Information theory measures:

- Uncertainty about private information → **entropy of target variable Y**

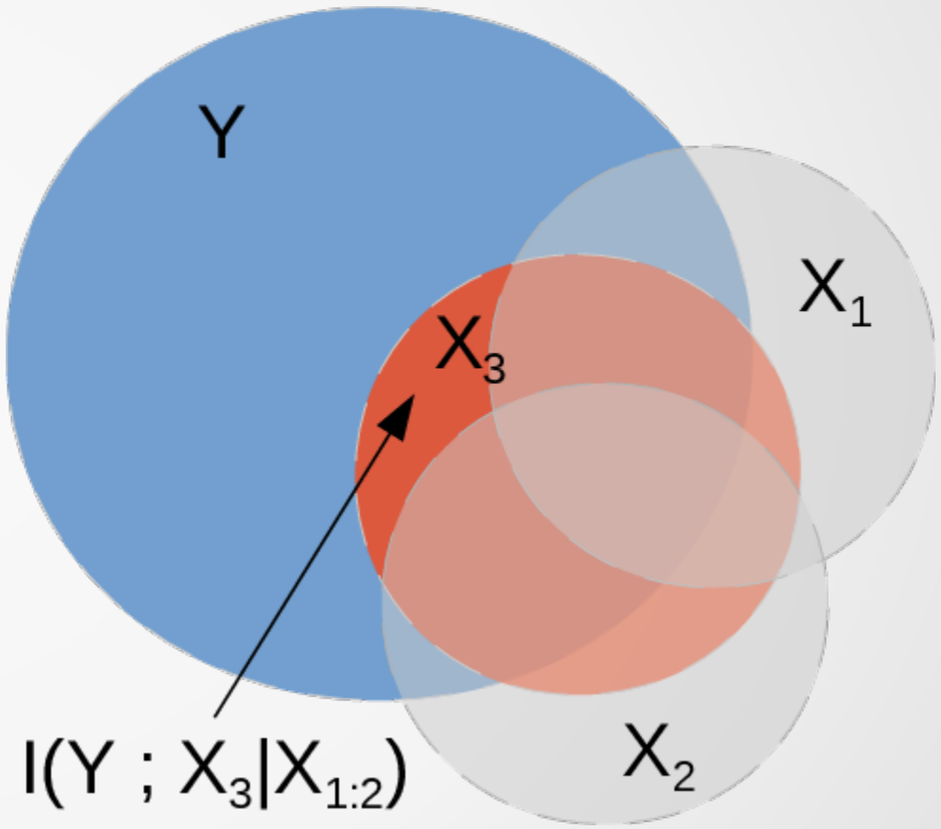
$$\mathbf{H(Y)}$$

- Amount of information disclosed by a feature X about Y → **mutual information of X and Y**

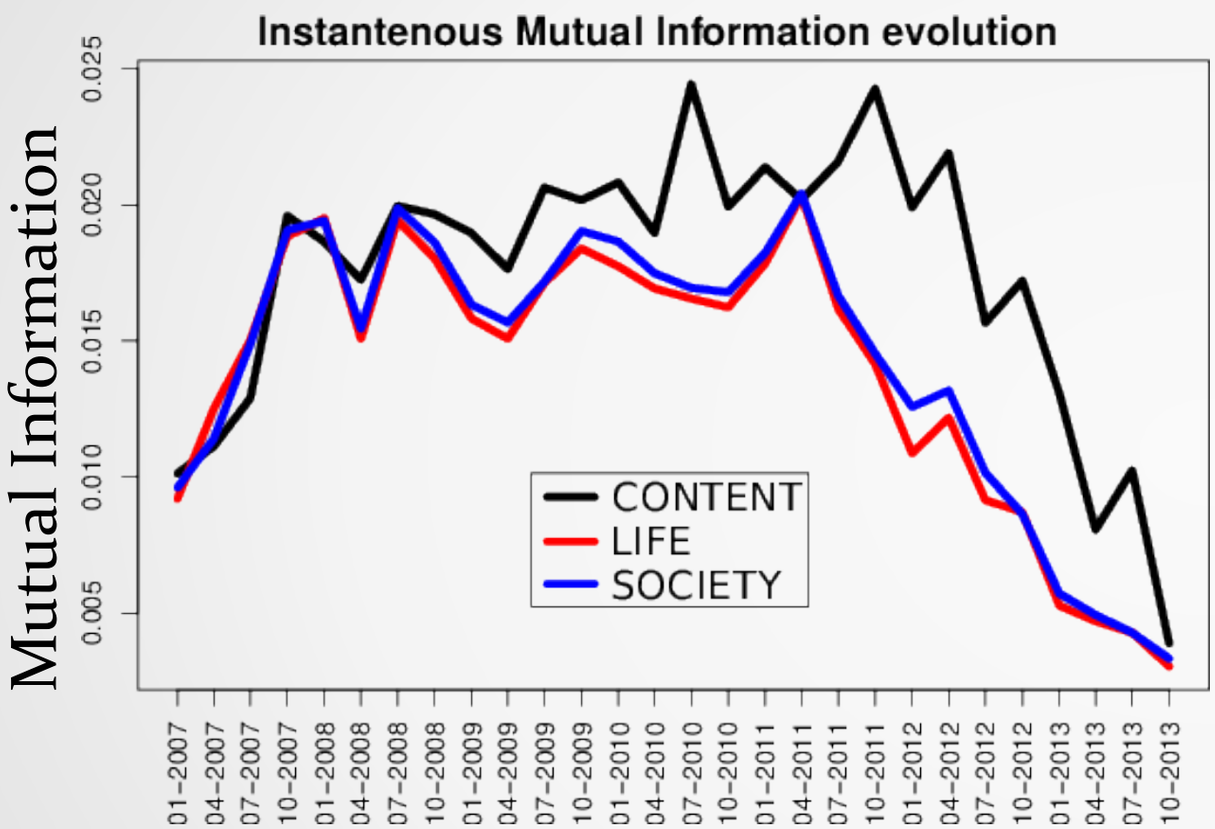
$$\mathbf{I(Y; X)}$$

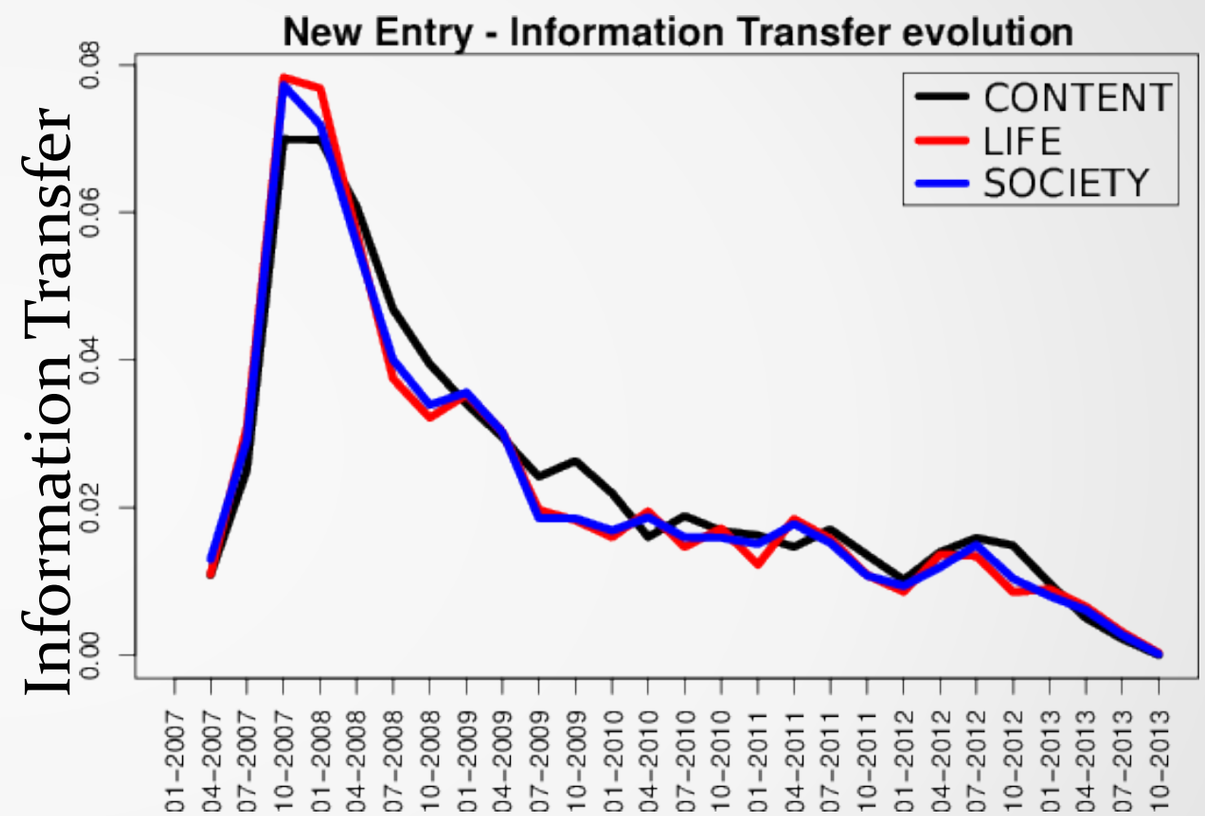
- Amount of *new information* disclosed by a feature at time t X_t → **Information Transfer**

$$\mathbf{I(Y; X_t \mid X_{1:t-1})}$$



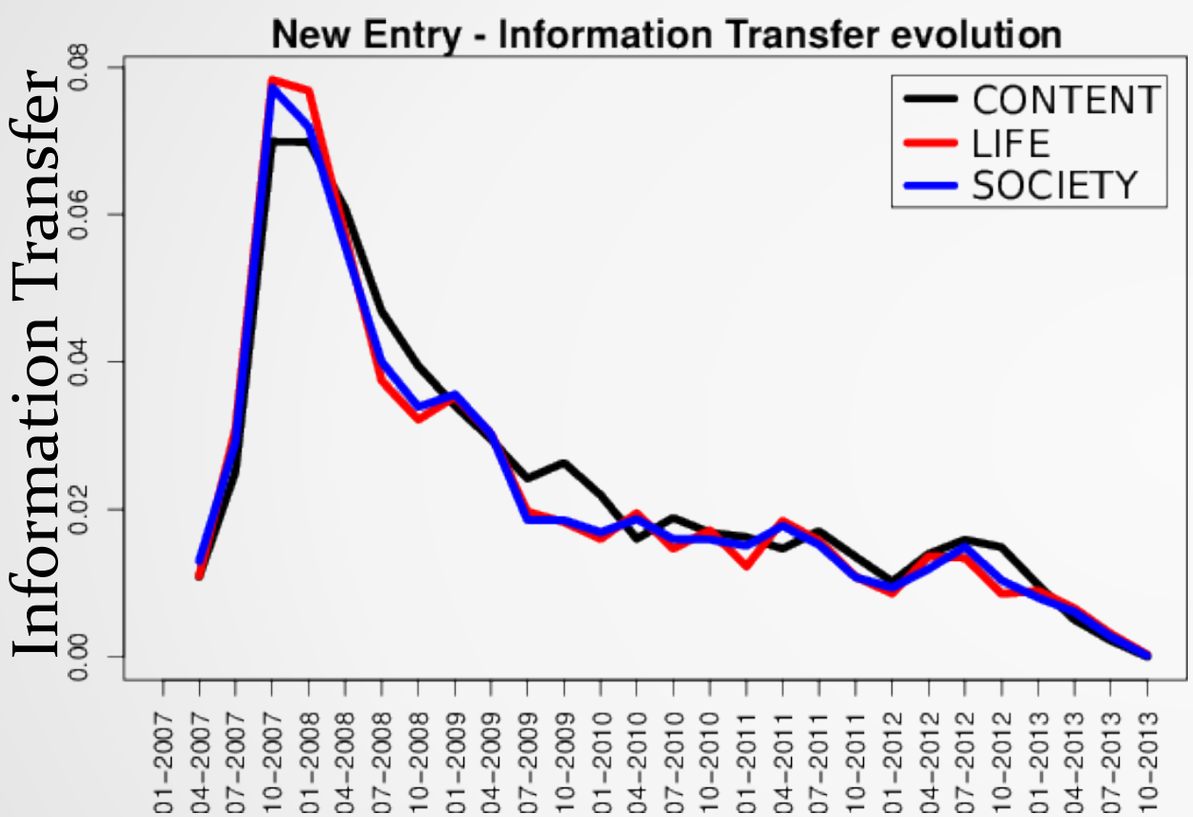
Effect of *online breadcrumbs*

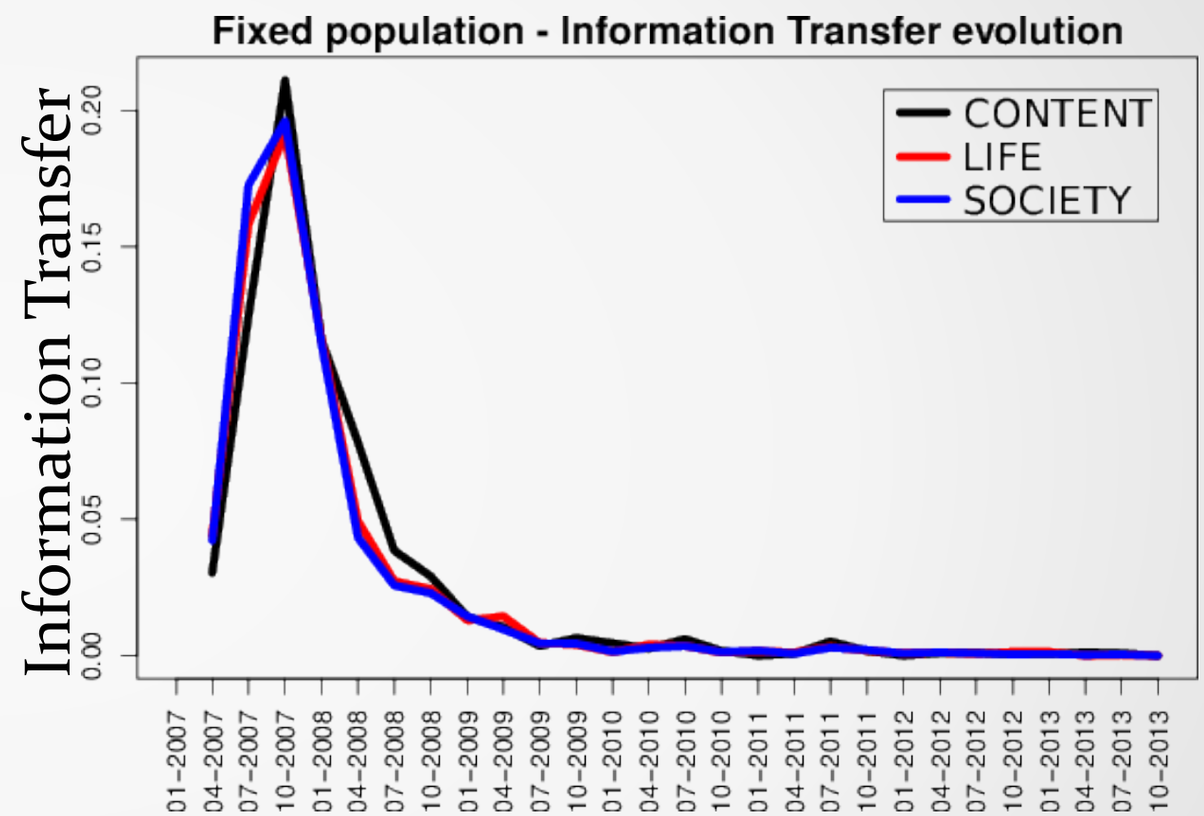




While later edits contain just as much information about a user's privacy as the earlier edits, they tend to be less harmful since most of the information they bring has already been learned.

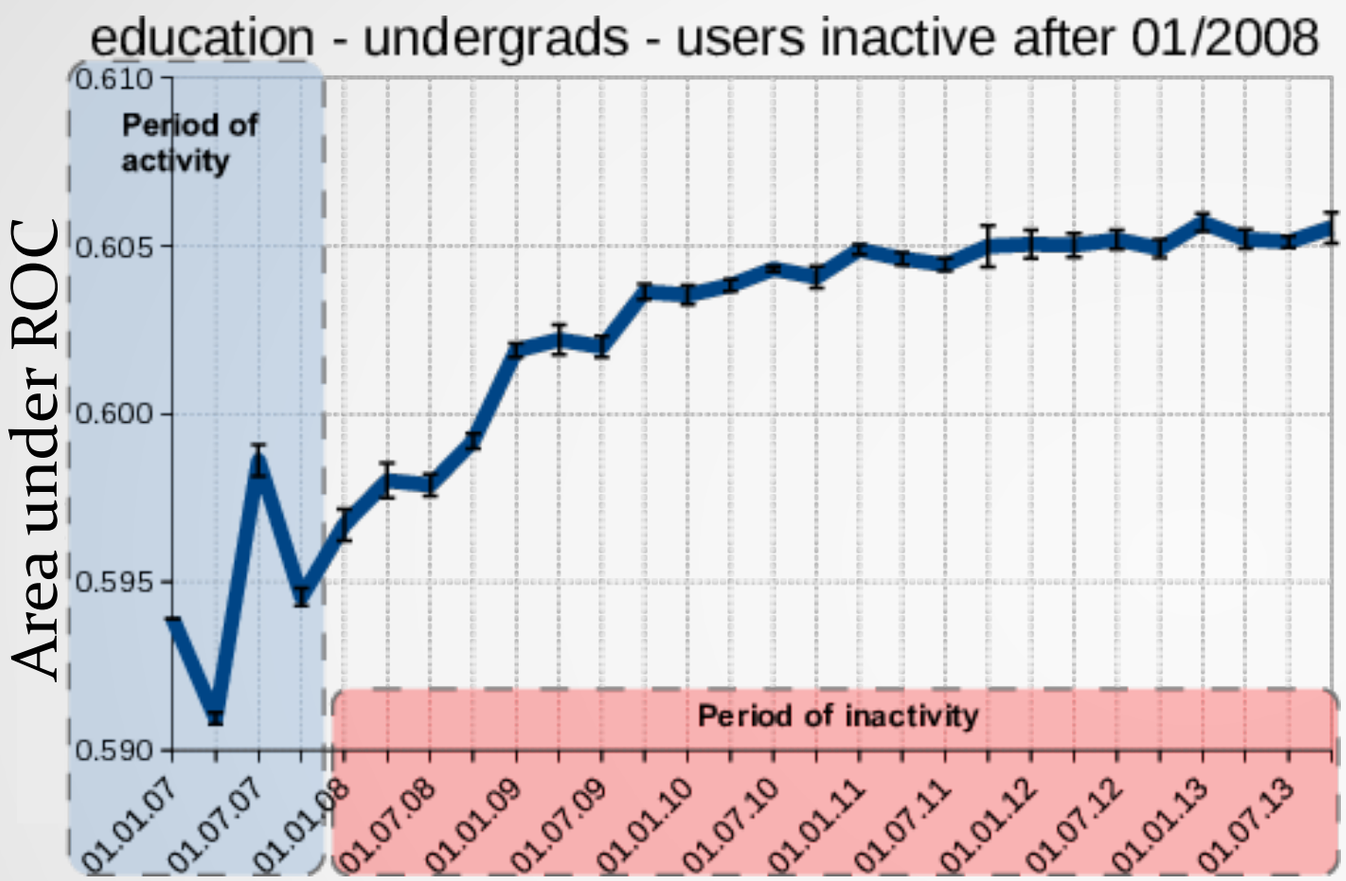
Effect of *newcomers*

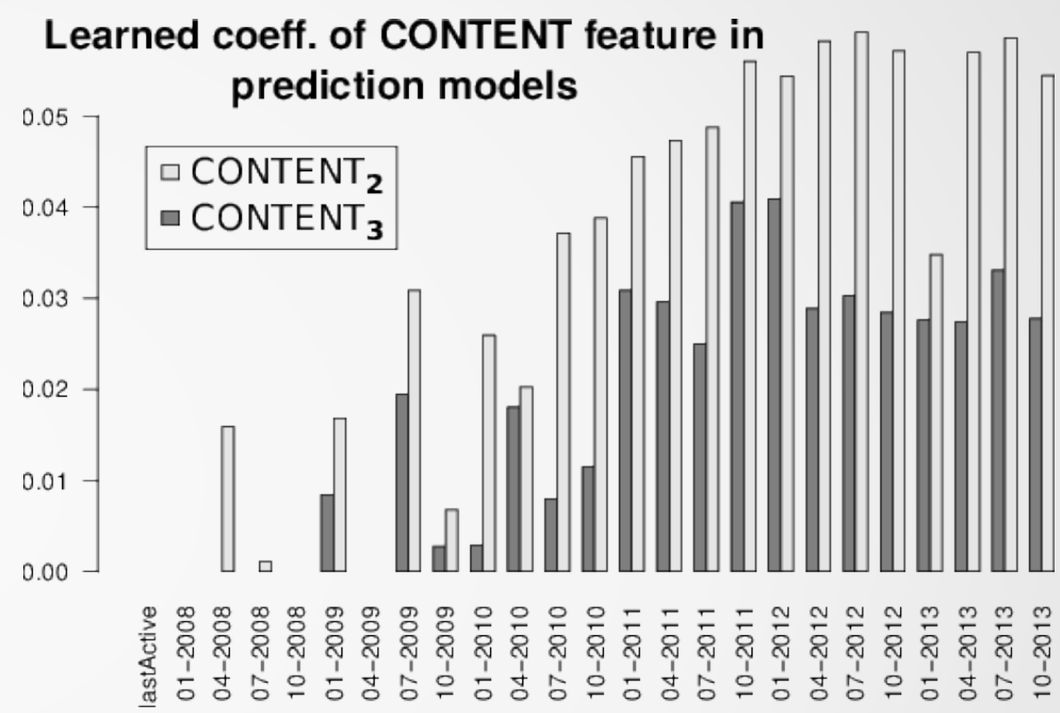




The information inferred from newcomers seems to be moderate, but consistent over time.

Privacy erodes even for *retired editors*





Plausible explanation: observed prediction improvement originates with currently active editors, whose activity overlaps with exited editors

Conclusion

3 main conclusions:

- Time has an adverse effect on privacy
- Factors influencing Privacy Loss:
  - *online breadcrumbs* (i.e. editor's own activity)
  - activity of other editors and newcomers
- Privacy erodes even for *retired* editors

Users don't have complete control over the consequences of the information they release

Thank you!

3 main conclusions:

- Time has an adverse effect on privacy
- Factors influencing Privacy Loss:
  - *online breadcrumbs* (i.e. editor's own activity)
  - activity of other editors and newcomers
- Privacy erodes even for *retired* editors

Users don't have complete control over the consequences of the information they release