

PRIVACY EVALUATION

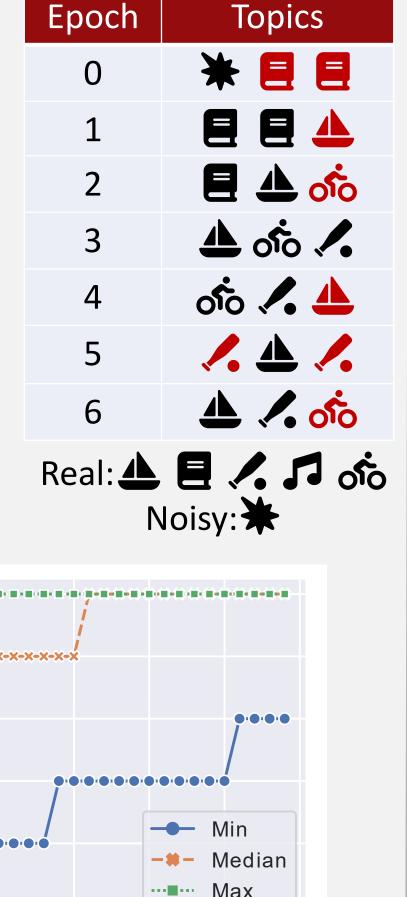
Identification of Noisy and Real Topics

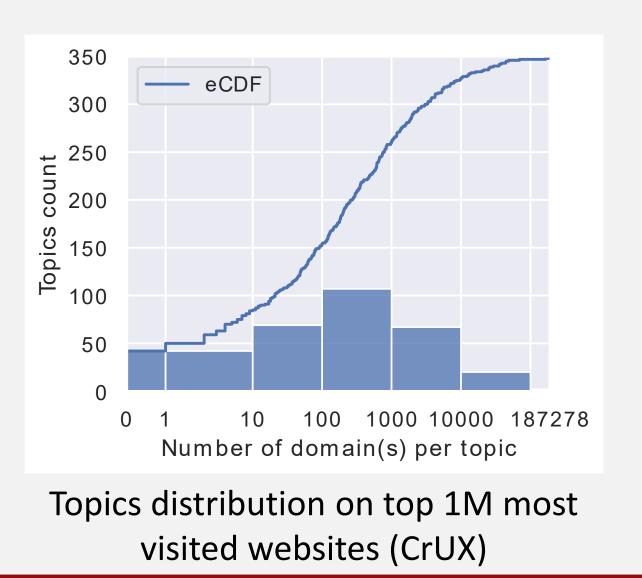
Asymmetric topics distribution on the web: our classifier considers every topic that does not appear at least on 10 websites among the top 1M as noisy.

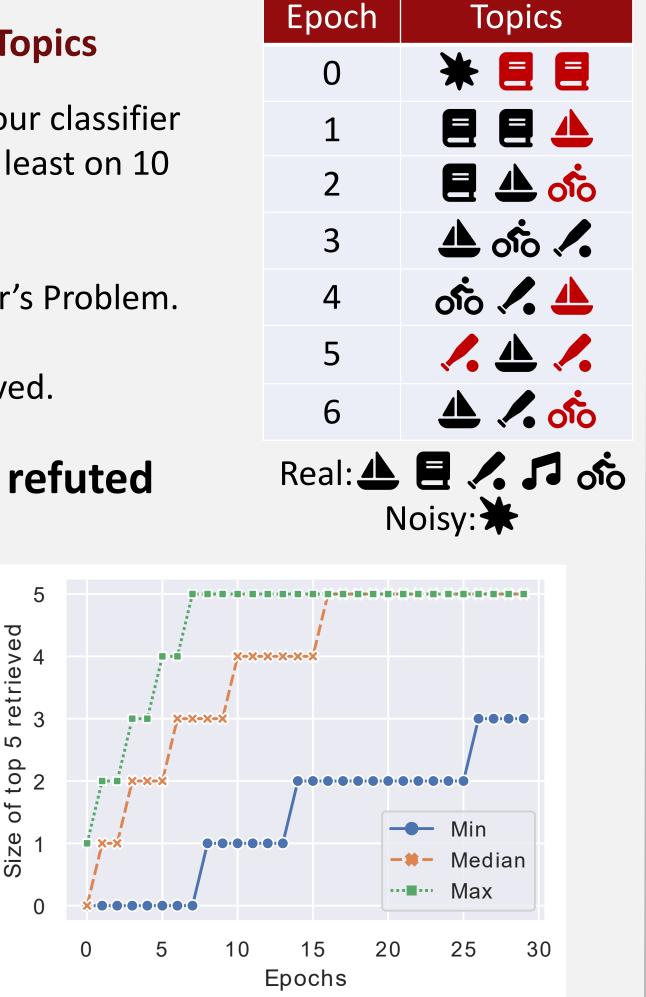
Repetitions leak real topics: Coupon Collector's Problem.

- **One-shot:** 25% of noisy topics removed.
- Multi-shot: 49-94% (15-30 epochs) removed.

Result: **plausible deniability** can be **refuted**





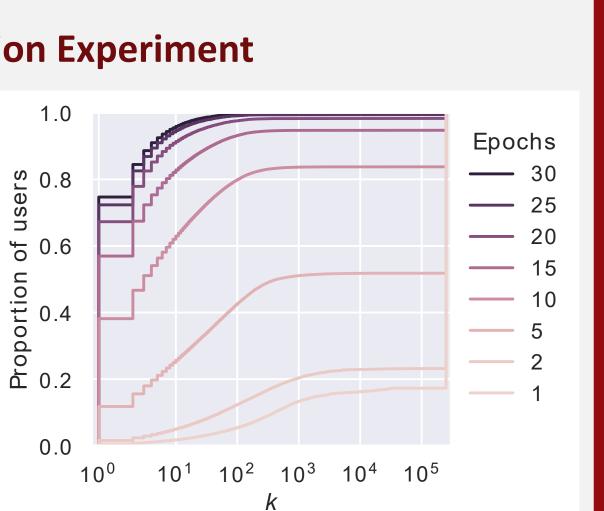


Simulation on 250k stable users

Re-identification Experiment

Simulation: quantification of the fingerprinting risk of Topics for an arbitrarily large population of users (250k) over time (30 epochs).

k-anonymity across time: How *"difficult"* is it to re-identify *"significant numbers of users* across sites"?



Result: users can be **fingerprinted by the Topics API**

Measurement on Real Browsing Histories

Real data: 1207 users from Germany over 5 weeks in October 2018. **Uniqueness:** 93% have unique topics profiles.

Stability: at least 47% have 3 or more stable topics.

Topics observation(s)	Users re- identified
1	28 (2%)
2	37 (3%)
3	50 (4%)

Result: third parties can **track** users across websites by observing their topics

TAKEAWAYS

Topics API can be used to Fingerprint Users

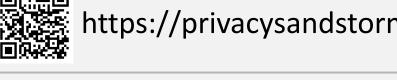
- Topics can not guarantee non re-identification across websites to all users.
- Users have stable and unique web behaviors that need to be considered.
- Google's non-reproducible analyses are disconnected from reality and lack systematization in their approach.

Some Utility Retained, but Classification can be Manipulated

- Topics returned are somewhat aligned with users' interests.
- Utility buckets introduced after advertisers' feedback is making topics profiles more unique (privacy-utility tradeoff).
- Unclear if Google's current mitigation (external attestation mechanism) will prevent further abuse.

Need for a (Research) Sandstorm through the Privacy Sandbox

- Call for reproducible analyses and release of tools and datasets.
- More evaluations are required to understand all potential impacts.
- Launch of a new research hub at https://privacysandstorm.com



https://yohan.beugin.org

🗶 🔳 yohhaan



yohan@beugin.org

