Dear Differential Privacy: Put Up or Shut Up

Paul Francis

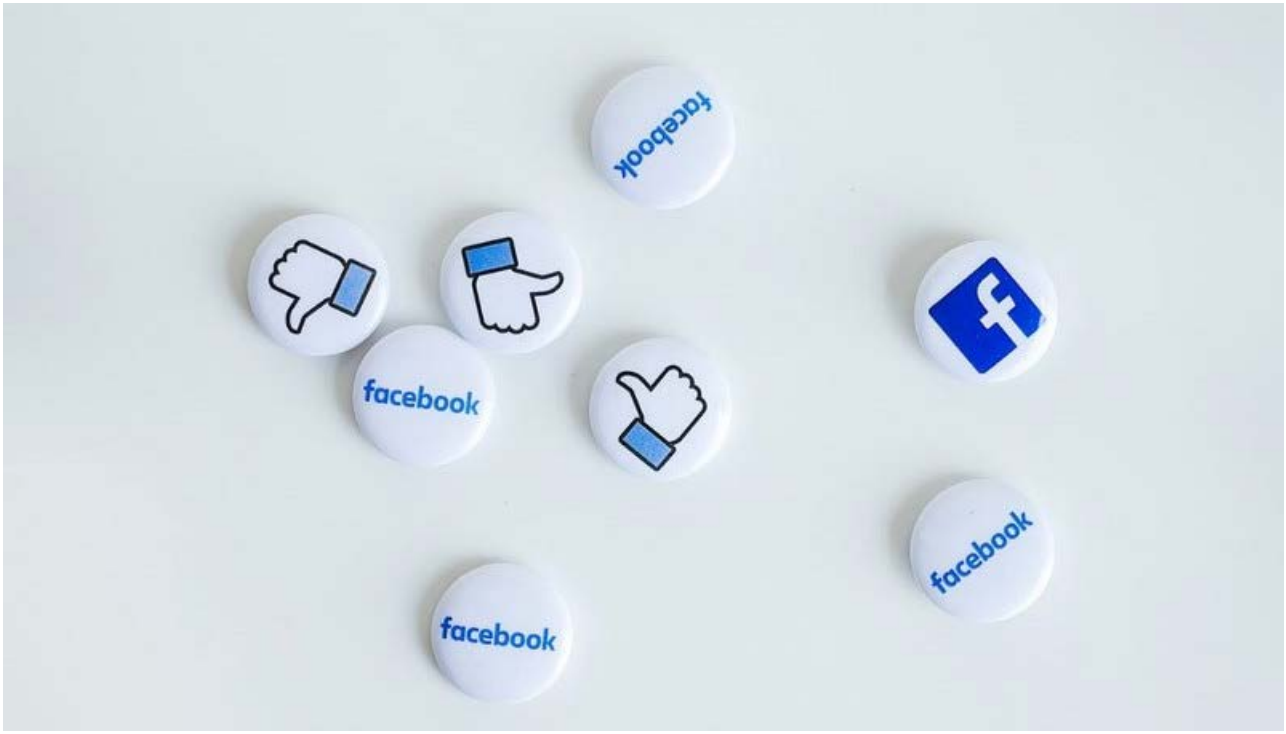
Technical Report MPI-SWS-2020-005

This technical report is to archive and make citable a blog post published on Jan. 9, 2020. It was published at the following URL:

https://medium.com/@francis_49362/dear-differential-privacy-put-up-or-shut-up-48ff255ec35

The pages following this one contain the blog.

Dear Differential Privacy, Put Up or Shut Up



Update: The original article was published in January 2020. In February 2020, Facebook released a new dataset using Differential Privacy. My thoughts on the new release is here.

Original Article

In July of 2018 I learned from this article that the Social Science One / Facebook partnership would be using Differential Privacy as the anonymization model for releasing Facebook data to researchers. I emailed one of the project leads, Gary King, and told him that this was unlikely to work because it is very hard to get decent analytics out of a Differential Privacy system with strong guarantees. I wasn't surprised, then, to read in Science Magazine and the New York Times over a year later that the project is in trouble because of poor data quality.

This is both unfortunate and unnecessary. Unfortunate because important research on social media's impact on elections and democracy is delayed or may even not happen. Unnecessary because traditional anonymization mechanisms would have been perfectly adequate.

How is it that one of the largest and most sophisticated IT companies in the world could not manage to safely provide access to data for twelve groups of academics while census bureaus the world over safely release data to the public, and medical data is routinely safely released privately to researchers?

The fault for this lies with whoever made the decision to use Differential Privacy. Contrary to the hype surrounding Differential Privacy, it is not suitable for the vast majority of use cases, this one included, and the project's technical advisers should have known better.

There are only two reasons that Facebook (FB) could have decided to use Differential Privacy (DP). First, they believed that the resulting anonymized data would be adequate for the proposed research. Second, they believed that any alternative to DP would be regarded as too weak, opening them to criticism or even legal action.

Both of these reasons stem from the two anonymity beliefs that are perpetrated by privacy academics in general, and DP researchers in particular:

Vulnerability Belief: No anonymization technique other than DP can protect privacy, either because of known attacks or future attacks we haven't thought of.

Utility Belief: DP is already a usable technology, and only becoming

more so because of the vast amounts of research being done.

I revisit these beliefs later in the article. Before that, let's look at what happened.

Following is a timeline of events associated with the data sharing initiative:

2018

- April 9: Facebook (FB) announces initiative to share data through the independently managed and transparent SS1/FB partnership.
- July 11, 2018: SS1 is publicly launched.
- July 11: Facebook releases description of the planned anonymized URL dataset (*not based on DP*). Research proposals are solicited based on assumption of this data or data of similar fidelity.
- July 12: The Electronic Privacy Information Center (EPIC) sends a letter urging immediate suspension of the project pending a thorough and independent investigation of the privacy protections for Facebook users.

2019

- Aug. 27: After more than a year FB has not made the data available. Funders threaten to cancel funding if dataset not made available.
- Sept. 18: SS1/FB makes available the data, as described by the URL "light" dataset description (protected by DP). The data quality is far below that of the originally described data and is not suitable for many if not all of the studies. (I reached out to the grant recipients. Four of twelve responded, and of those four all of them were essentially unable to do the proposed research.)

2020

- I've been told by SS1 that FB will release a new dataset early in 2020.

How poor is the Differentially Private data quality?

The **original dataset proposed by FB** at the beginning of the project was not protected by DP. Rather, FB engineers proposed traditional

anonymization techniques. Specifically, FB used two of the most effective and common traditional mechanisms; removal of individual identifiers and aggregation.

This dataset contained:

- Information about web pages identified by **URLs**, including for instance whether they were fact-checked or considered hard news,
- **Events** associated with the URLs, for instance when and how often they were clicked, viewed, shared, and liked, and
- The demographics of the associated **users** including location, gender, age, and FB's computed political ideological score.

Event times were aggregated into 7-day buckets. User age's were aggregated into 5-year buckets, and locations into states or roughly state-sized regions internationally. The counts of events within these aggregates is exact: *no noise was added*.

The URLs themselves were scrubbed of potentially private information, and were included only if publicly available and shared by at least 20 different users.

Compared to the original proposed dataset, the **released dataset protected by Differential Privacy** has *removed all demographic information* about users: age, gender, location, and political ideology. It has *removed all information about time* except when URLs were first posted and fact-checked. It has *removed all location information* with the exception that it indicates which country had the most shares per URL. Finally, all event counts have Gaussian noise with sigma=200 added: counts can be off by easily plus or minus 500 or more.

Compared to the original proposed data, the released data eliminates the ability to study the demographics of users, as well as the ability to observe how URLs were shared over time and over geography.

> Of the four questions that could have been studied about the original data: **what** was shared, **who** shared it, **where** was it shared, and **when** was it shared, only the first question remains.

It is not surprising then that the proposed research cannot be carried out.

How insecure was the original data?

The original proposed data was by any reasonable standard very strongly protected. With the exception of one easily fixable aspect, it is protected *better than census data*, which is publicly released and for which no privacy breaches have ever been reported, and *far better protected than medical data* anonymized by HIPAA standards.

Compared to census data, the original proposed data has far fewer user attributes, and larger aggregates; per state rather than per census block, or aggregates of millions of individuals versus a few thousand individuals. Compared to medical data, which is frequently shared privately for research purposes, not only are the aggregates larger, but user identifiers have been removed.

> The original proposed FB dataset was protected better than census data in most respects, and far better than HIPAA-standard medical data in all respects.

(Note that HIPAA-standard anonymized data is not typically strongly anonymous. However to minimize risk it is shared under contractual agreements. In any event I have been unable to find any reported incidences of malicious re-identification of medical data, so either the safeguards are effective or attacks are under-reported.)

Were *any* users in the original dataset at risk of re-identification? Maybe. But if so, these could have been handled by a one of two

additional commonly used mechanism; removing users that are not in adequately large aggregates, or placing such users in larger aggregates. So long as the number of such at-risk users is small (as I would expect in this case), removing them does not hurt the data quality very much. With these well-known techniques, the data would be better protected than publicly-released census data.

Remember too that the FB data is not being shared publicly but rather with a small number of researchers. Furthermore, the anonymized data is not be transmitted to the researchers. Rather the researchers are required to query the data on FB systems, thus providing yet another layer of protection.

Why is Differential Privacy hard to use?

DP is a mathematical measure of potential privacy loss. To use it, one designs a mechanism, proves that it follows the math, and then computes the measure. DP has two serious limitations.

First, if the mechanism is complex, it is very hard if not impossible to derive a meaningful DP measure. There are a number of useful anonymization mechanisms, including the ones used in the first FB dataset, that cannot easily be used in a DP mechanism. DP therefore limits the types of mechanisms that can be used, making it harder to build mechanisms that offer both strong privacy and good utility.

Second, DP measures *potential* privacy loss, not actual privacy loss. If the measure is very low, then one can be sure the mechanism is safe. A high measure, by contrast, does not mean that the mechanism is not safe. A high measure doesn't really say anything about whether the mechanism is safe or not. If one is going to work with a high measure,

one may as well not use a DP measure at all.

In short, DP is useful as a privacy loss measure only when the measure is low, and for most use cases nobody knows how to build a mechanism that achieves a low measure while allowing for useful analytics.

For more information, this article published by the Center for the Governance of Change expands on these points.

What about Google, Apple, and the US Census Bureau?

I don't know everything that Google and Apple use DP for, but based on this and this, it appears to be good for use cases where a lot of noise is acceptable, where it isn't necessary to understand correlations between different types of data, and where understanding the long tail of infrequent data is not necessary. For use cases where DP isn't adequate, Apple and Google use traditional mechanisms. Both Google Analytics and Apple App Analytics for instance provide public-facing anonymized user data, but neither uses DP. This is not to suggest that the data release isn't safe, only that it isn't DP.

An interesting case is the US Census Bureau, which announced in August 2018 plans to use DP for the 2020 Census. Like FB/SS1, the US Census Bureau is publicly committing to DP before it really understands how it will work. Unlike the FB/SS1 project, there are thousands of researchers and others who rely on US Census data. There are concerns that the use of DP will unnecessarily degrade the quality of the data. In an open letter to the US Census Bureau leadership, 4407 academics, planners, journalists, and researchers from the government, non-profits, and the private sector request clarifications and closer cooperation with

stake-holders.

I won't be surprised to read in a few years that the US Census Bureau has backed-off on its plans for DP.

Harm versus harm

All anonymized data releases require that one harm is weighed against another. One is the harm that comes to individuals if the data is maliciously re-identified. The other harm is the loss of knowledge when data can't be analyzed, or in obtaining incorrect knowledge when anonymization distorts the data too much.

DP is focused almost entirely on the harm of re-identification. It's useful operating range (low potential privacy loss measure) severely limits how much analytics is possible, either because of too much noise, too few queries, or limited types of queries. By committing to DP early, the FB/SS1 project effectively chose to place much more weight on protecting users than on getting useful research done. In many scenarios this might be a perfectly valid choice to make. In this case, it is the wrong choice.

How did we get into this mess?

Some readers might be skeptical about the claim that the original FB dataset was strongly protected. Perhaps readers have heard of the paper by Paul Ohm titled "*The Broken Promise of Privacy: Responding to the Surprising Failure of Anonymization,*" or have seen tweets like this one from Luc Rocher stating that "*Anonymizing data is not enough to protect privacy anymore.*" Maybe this has led readers to believe, like Aleksandra Korolova, that any form of anonymization other than DP is

susceptible to a "*failure of imagination*" on the part of the designers.

Data anonymization is a complex topic, and unfortunately the general understanding of anonymity as well as the quality of reporting on anonymization is very poor. Part of the problem is that the word "*anonymization*" is used to describe a range of techniques that differ vastly in strength. Just as a magnitude 7 earthquake is millions of times stronger than a magnitude 2 earthquake, the strongest anonymization techniques offer thousands of times better protection than the weakest techniques. Whereas reporting on earthquakes however always include a magnitude measure, this never happens with reporting on anonymity. The large majority of high-profile attacks on anonymization take place on the weaker forms of anonymization. The resulting declarations that anonymization doesn't work are almost as ridiculous as deducing that earthquakes are not dangerous from the observation of magnitude 2 and 3 earthquakes.

A counter-example is the attack that the US Census Bureau ran on it's own data which, in the words of the John Abowd, Chief Scientist of the U.S. Census Bureau, is one of a class of attacks that is the "*death knell for traditional data publication systems*." This attack is what led the U.S Census Bureau to commit to DP.

Taking into account the considerable risk to research based on census data that using DP entails, it is surprising to learn that the attack is by no means a home run. A good description of the attack can be found in the section "Database Reconstruction and Re-Identification" on page 11. The bottom line is that, if one tried to re-identify an individual based on this attack, the chances that that re-identification is correct is small.

The class of attack referred to by Abowd is the "Database Reconstruction

Theorem" (page 6), which Abowd refers to as a "powerful result" from a 2003 paper by Dinur and Nissim, which Abowd characterizes as:

> Too many statistics published too accurately from a confidential database exposes the entire database with near certainty.

This, by and large, is the basis of Vulnerability Belief. While the result is powerful, it is not necessarily general. The Dinur paper concerns a specific attack on a specific anonymization technique (adding random noise to answers) with specific query capabilities (ability to filter for individual users). Certainly the attack used by the U.S. Census Bureau came nowhere near exposing the entire database with near certainty.

It is easy to see, however, how the two beliefs would lead the U.S Census Bureau to adopt DP. The attack does indeed expose a new vulnerability, and even though the attack is not particularly effective, the Vulnerability Belief would suggest that this attack is just the tip of the iceberg, and that it is just a matter of time before more effective attacks are found. Indeed, since the data is made public and can't be taken back, it would certainly be a huge problem if an effective attack were found.

The Vulnerability Belief would discourage the U.S. Census Bureau from exploring any fixes other than DP, and the Utility Belief would give them the confidence that it could be done.

(In the interest of full disclosure, Nissim successfully executed a reconstruction attack on an earlier version of my own anonymization design, Diffix Birch. A subsequent version defends against that specific attack, but it is an open question as to whether another form of the attack exists, and if so whether additional defenses can be designed.)

The European General Data Protection Regulation (GDPR) made some

progress in distinguishing between weak and strong forms of anonymization by specifically labeling a commonly used weak form as *pseudonymization*. Under the GDPR, pseudonymized data is still regarded as personal data, while anonymized data is not.

Our general inability to talk meaningfully about anonymity is painfully apparent in the open letter that the Electronic Privacy Information Center (EPIC) wrote to SS1. The EPIC letter repeatedly refers to FB/SS1 releasing *personal* data to researchers when in fact even the initial proposed release was strongly anonymous, and would, with minor additional tweaks, **not be regarded as personal data** by the GDPR, at least in my experience. The letter bizarrely complains about FB/SS1 failing to implement pseudonymization when in fact FB/SS1 was clearly going far beyond mere pseudonymization.

A key argument of the EPIC letter is that the SS1/FB program violates the FTC's 2011 Consent Order with FB. That order states that FB must obtain affirmative express consent before disclosing personal information to third parties. The whole argument falls apart if the data is not personal data.

Nevertheless, given the Cambridge Analytica debacle, it is understandable that FB would fall back on the reputation of DP for guaranteed privacy. In the face of the two beliefs, I can well imagine that it would be hard to convince organizations like EPIC that the original data was safe.

Next steps

It is easy to make the U.S. Census data perfectly safe. Just don't collect the data in the first place! What is hard is making data both useful and

private.

> A guarantee of privacy is absolutely meaningless if the data can't be used for the intended research.

The only way forward for this project is to stop using DP. I have been in contact with the SS1 co-chairs, both of whom agree with this. If FB cannot agree to this, then the project should be canceled. Nevertheless I have been told that FB is working on another DP data release that will have better utility.

SS1 is currently pursuing a "safe harbor" designation under GDPR for FB data made accessible to researchers for the social good. Such a designation would limit or eliminate FB's liability with respect to GDPR regulations so long as certain oversights are in place. These would include vetting researchers and proposals, vetting facilities where the research would be performed, recording and potentially auditing research activity, pre-publication reviewing of research results to ensure no privacy leakage, and even criminal penalties for researchers who engage in malfeasance.

SS1/FB may of course not be able to get safe harbor approved. Even if it is approved, this does not obviate the need to anonymize data so far as doing so doesn't impede the research. Towards this end, I suggest the following:

1. Expand the membership of the SS1 Privacy and Security Committee to include researchers that work with traditional anonymization techniques and professionals that understand risk assessments, ideally including people from national Data Protection Authorities.
2. Be transparent about the privacy evaluation process. Trust in FB is very low, and if the EPIC letter is an indication, so is trust in researchers that work with FB.
3. Consider running an anonymization bounty program. The startup Aircloak, for which I am a co-founder, did this for it's implementation of Diffix Birch. Besides helping us find weaknesses in the algorithm, it has been effective in building trust.
4. Consider using Diffix for future phases of the project. Anonymization for the first

dataset, which doesn't encode any user behavior, is relatively straight-forward. More complex datasets will be harder to anonymize, and using Diffix can speed up the process without compromising privacy.

A message to the academic community

The computer science academic community studying data anonymity needs to take this DP failure to heart. Real harm, in the form of blocked or delayed research, has been done because of the widespread acceptance of the two beliefs.

In academia there is currently a de facto freeze on any anonymity research that isn't DP. If the Utility Belief is wrong, then the academic community is missing an opportunity to explore alternative forms of anonymization that, even if not perfect, can add strong protection to a wide range of use cases, including this one.

This may be wishful thinking, but I encourage academics, especially DP researchers, to do the following:

1. **Stop overstating the failure of anonymity.** Most attacks are on weak mechanisms, and no attack that I'm familiar with is broadly applicable to different anonymization mechanisms.
2. **Stop implying that DP is usable.** Or maybe as a first step, stop believing that DP is usable. There is ample evidence that it rarely usable, and by not explicitly stating so, researchers imply that it is.
3. **Stop rejecting papers simply for not being DP.** Few researchers are going to work on alternative mechanisms, especially informal ones, if they know it will be hard to publish.

Deja vu all over again

Recently I learned from this article that Microsoft and Harvard are launching a project to develop an open source platform for sharing data privately. The article says that it will draw from Differential Privacy, and will "*show how differential privacy provides the strongest possible privacy protections available.*" I emailed one of the project leads, Gary

King, and told him that this was unlikely to work because it is very hard to get decent analytics out of a Differential privacy system with strong guarantees....

By Paul Francis on January 9, 2020 .

Canonical link

Exported from Medium on May 12, 2020.