Social Science One set to release massive trove of Facebook data for research purposes

July 11 2018

A new AI tool created to help identify certain kinds of substance abuse based on a homeless youth's Facebook posts could provide homeless shelters with vital information to incorporate into each individual's case management plan. Credit: CC0 Public Domain

It seems Christmas is coming early this year for social scientists.

That's because, just months after Albert J. Weatherhead III University Professor Gary King wrote an academic paper about a system that would allow researchers to access the massive data troves held by Facebook and other private companies, it is set to become a reality.

Along with his collaborator Nathaniel Persily at Stanford, King created an organization called Social Science One which will lead the effort to identify data inside Facebook, prepare it for researchers, and fund numerous scholars to analyze the data.

The organization is today making the first of what King says will be many data sets available for research—a massive trove of more than half a trillion numbers that includes every link ever clicked on by Facebook users in the last year, including the types of people who clicked, what they clicked on, and indicators for whether links were judged to be intentionally false news stories.

"As social scientists, our goal is to understand and solve the greatest challenges that affect human society," King said. "Twenty years ago, almost all the data in the world to address these challenges was created by those of us in the academy, by governments and given to us, or by private companies and sold to us," he said. "But the problem is that even though we have more data than ever before, we have a smaller fraction of the data that the world is creating. Most of the data that would be useful for social science is now locked up inside private companies. Social Science One is an important mechanism for unlocking that data for social scientists."

And the amount of data they'll have access to is almost beyond imagining.

"The data we're going to be providing access to is extraordinary—in quantity it may rival the total amount of data that currently exists in the

social sciences."

"This commission has the potential to open a new chapter in social science research, and in the overall acquisition of knowledge, in which the organizations that possess critically important information about people and institutions, like social media platforms, and professional researchers will be able to more effectively collaborate to address some of the most difficult problems facing our society," said Matthew Baum, Marvin Kalb Professor of Global Communications at the Harvard Kennedy School, and member of the Social Science One commission.

Outlined by King and Persily in a working paper in April, the framework that underpins Social Science One consists of two parts.

The first, he said, is a commission of distinguished academics from across the globe who will work with Facebook officials to identify potential data sets which they will make available to researchers through a process in which study proposals are submitted and peer reviewed. Once study ideas are approved, researchers get access to the data as well as grants to support their work provided by seven charitable foundations. The foundations span the ideological gamut but their money is pooled, and all decisions will be made by academics, and so no one viewpoint can dominate. And the outside researchers will have complete academic freedom without having to give Facebook prepublication approval rights.

"The key part of the process is that the commission, as a trusted third party, can look at the proposals and decide that some not be funded—even if scientifically appropriate—for reasons not publicly known, such as if they would touch on litigation that has not been made public," he continued. "And if Facebook reneges on this agreement and does not make data available that Social Science One requests, we are obligated to report that to the public. So this system is incentive compatible for the public, for the company, and for the social scientific

community. We think of this as essentially a work of political [science](#), where we came up with a constitution that works for all parties."

Social Science One is being incubated at the Harvard's Institute for Quantitative Social Science, which King directs. Over the years, IQSS has taken on this type of activity many times. It has regularly incubated and spun off nonprofit research groups and for profit companies, as well as centers, programs, and research projects now housed at IQSS, elsewhere at Harvard, and at other institutions.

As exciting as it may be for researchers to have access to Facebook's data store, the use—and misuse—of Facebook data has made worldwide headlines in recent months, something King and colleagues have developed procedures to avoid. They built safeguards into their procedures, the first of which is simple—researchers won't be given Facebook data, they'll only be given access to it.

"No academic will be handed data, like before," King said. "Instead, we'll make data access available to academics so that individual privacy is always preserved."

In addition, the organization plans to make use of a mathematical concept known as "differential privacy," to ensure that the data that is made available can't be traced back to individual users. "We have some of the leading experts in the world studying this concept here at Harvard, including Cynthia Dwork, the Gordon McKay Professor of Computer Science in the Harvard John A. Paulson School of Engineering & Applied Sciences, and Salil Vadhan, the Vicky Joseph Professor of Computer Science and Applied Mathematics, both of whom are members of the commission," King said. "The idea is that you can take a data set and add special types of random noise to make it impossible to identify any single person, but when you aggregate it, it doesn't alter the overall patterns you want to examine."

But by far the strongest security measure, King said, is related to the system that allows academics to access the data. "When academics access the data, every character they type will be logged and audited," he said. "So if they type the letter 'k,' we will know they typed that letter. So there is no possibility of them copying or misusing the data. This means that we are switching from a model of individual responsibility, that has the researcher violating the rules as a single point of failure, to one of collective responsibility, where no one person can violate privacy without everyone knowing and being able to stop it."

Ultimately, King said, the goal of Social Science One is to develop ways for Facebook—and eventually other companies—to make their vast data stores available to researchers in the hope of finding solutions to the social problems that continue to plague humanity.

"Facebook has highly informative data on two billion people," King said. "That's an incredible privilege, and with the privilege comes considerable responsibility. It only makes sense that Facebook also use some of that information and power to help the public and contribute to social good."

It's an idea that's not without precedent, King said.

Over the decades, several large companies have built large research divisions—perhaps most notably with the creation of Bell Labs by AT&T and Microsoft Research at Microsoft—that allowed scientists the freedom to explore topics as varied as information theory to the development of lasers and the development of transistors.

With the release of the first data set today, King and colleagues hope to continue that tradition—but in a manner designed especially for the social science-related businesses.

"This is just our first data set—we have quite a lot of others that will be coming after this, and we have funding from seven generous foundations, and so we hope to begin getting researchers up and running fast," King said. "We also hope to extend this collaboration beyond Facebook and to partner with other companies as well."

"The discoveries we make using these data sets are not going to interrupt these companies' businesses, but they could help solve some of the challenges that affect human society," King said. "And if there's a way to do that, who wouldn't want to contribute to that mission?"

Provided by Harvard University

Citation: Social Science One set to release massive trove of Facebook data for research purposes (2018, July 11) retrieved 26 April 2024 from https://phys.org/news/2018-07-social-science-massive-trove-facebook.html