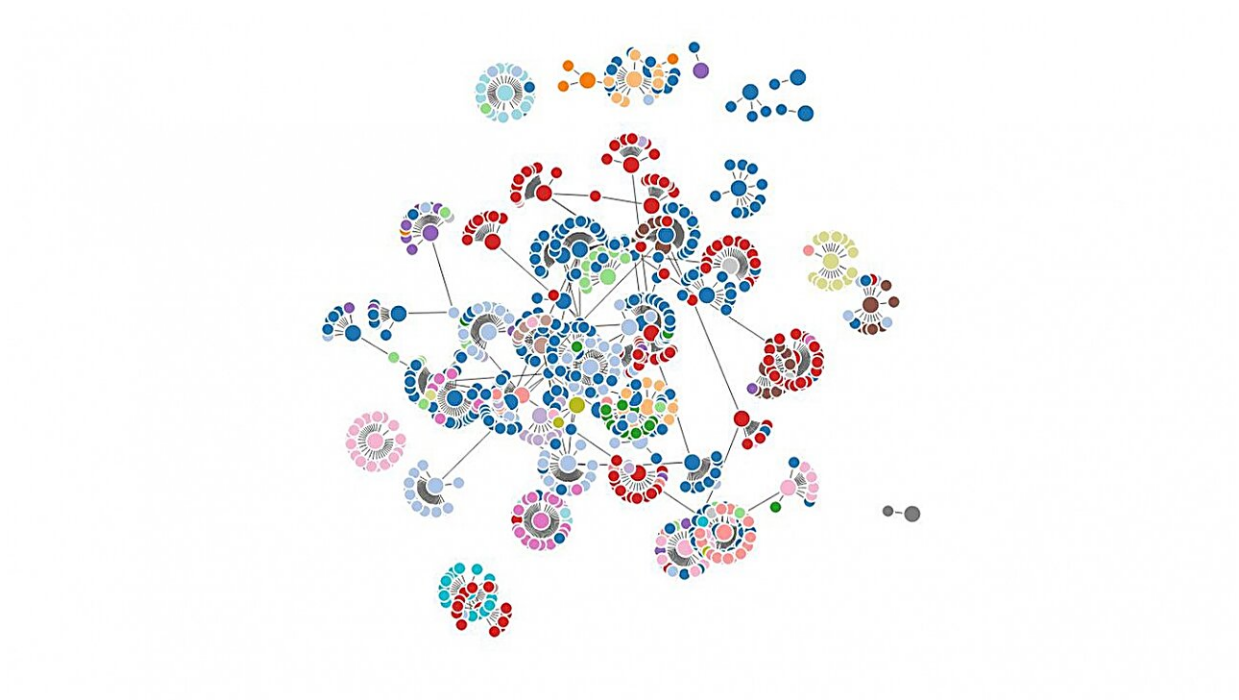
Investigating shadow profiles: The data of others

September 22 2023, by Jürgen Graf



Analysis of information linkage in a social network. Credit: David Garcia

Shadow profiles in social networks contain information about people who are not members. At the moment, shadow profiles are almost impossible to prevent using technical means, pose a collective problem for society and are a matter that has gone virtually unregulated so far. This is where the Center for Human | Data | Society at the University of Konstanz brings its perspective to the table: "Individual solutions will not

fully protect our privacy."

You don't even have to be a member of one of the many social networks or messenger services—the networks most likely have information about you anyway. They may even have created an invisible profile of you, a kind of virtual file, even if you have never been a member and never agreed to the corresponding terms and conditions. It is enough that a sufficient number of your contacts has an account there.

Through information and contact addresses that your friends share in the network, enough information can be puzzled together to draw conclusions about you. Put simply: if the network knows that most of your contacts play handball, live in Konstanz and are interested in migration policy, then the chances are good that the same applies to you.

How shadow profiles are created

It all begins with a quick click sharing your address book with the messenger service: The service now has access to contact data and can assign information to the telephone numbers in addition to establishing connections between them. The information puzzle is gradually pieced together using messages and photos that your circle of friends posts on the network, as well as any group memberships, comments and likes.

All this happens without any evil intent and completely unintentionally, even if no one means to share information about you or infringe upon your privacy. The messages your contacts post do not even need to state your name. However, a basic picture of you is derived from the sum of all the tiny bits of information provided by your social environment: your interests and political beliefs, ethnicity, place of residence, marital status, and even what you are likely to buy.

Such unofficial profiles, known as "shadow profiles," first came to light

in 2012 when a Facebook data leak revealed that the network possessed information it should not have had.

David Garcia is a researcher at Konstanz's Center for Human | Data | Society (CHDS). One of the center's overarching research topics focuses on shadow profiles and how to protect people from them. The strength of the CHDS is that its truly multidisciplinary team explores the phenomenon in all its complexity. The team of researchers from the fields of computer science, law, social and cultural sciences and psychology tackle the problem holistically.

Garcia is also professor of social and behavioral data science at the University of Konstanz and has been a member of the extended directorate of the Center for Human | Data | Society since October 2022. The computer scientist is an expert in the field of computational social science. Among other things, he studies the influence of digital media on people and society as well as collective emotions in online communities.

The CHDS employs this approach, because shadow profiles are not merely a technical problem—they also pose political, legal and cultural challenges for society. How is the right to privacy defined in the virtual world? And how far does this right extend? If publishing my data always reveals information about others at the same time, how much is my individual freedom to act and right to informational self-determination actually worth?

What types of shadow profiles are there?

Shadow profiles exist in various forms:

- A partial shadow profile is generated when someone has an account with a social network, but does not share certain information on the network (e.g., personal information or a

phone number), and the network fills in the "missing information" with data from the person's contacts.
- A full shadow profile is constructed when someone does not have an account with a social network and has never agreed to its terms and conditions, but the company still creates a profile of the person based on information from their contacts.
- A shadow profile can also be created when a person deletes their account with a social network. In this case, the network deletes all of the person's data from their account, but the company can partially recover the profile through indirect information provided by the person's contacts.

Since social networks have yet to publish a single shadow profile, it is difficult to tell how precise they actually are. In a study with publicly available data, David Garcia was able to show that it was possible to determine a person's city of residence within a radius of less than 50 km based on indirect contact information, and this was true for anywhere in the world with relatively sparse data. Social networks are likely to have much more detailed information at their disposal and thus have far more precise shadow profiles. It also took relatively little effort for Garcia to track information such as marital status or sexual identity in his data-driven simulation of a shadow profile.

Are shadow profiles illegal?

Our instinct tells us that shadow profiles should be illegal. In fact, however, there is a legal loophole here, as Liane Wörner points out. Wörner is the director of the CHDS and professor of criminal law, criminal procedural law, comparative criminal law, medical criminal law and legal theory at the University of Konstanz.

In Germany, laws addressing the illegal collection of data mainly do not apply to shadow profiles, but mainly address "digital data alteration" and

"data espionage."

- The offense of "digital data alteration"—the virtual equivalent of property damage—is defined as the act of altering, suppressing, or rendering data unusable as per § 303a of the German Criminal Code (Strafgesetzbuch StGB). None of this applies to shadow profiles, as the data remains intact.
- According to § 202a StGB, "data espionage" is when someone gains unauthorized "access to data not intended for them and, where this data is specially secured against unauthorized access, by overcoming the security measures in place." In addition to this, according to § 202a I StGB, the victim of the illegal access is the user themselves and not their contacts affected by the profile's existence.
- Here, too, the problem is that, without any bad intentions, the data was voluntarily published by the users in the network, and the data is usually visible to everyone. There are no access restrictions; anyone can enter (or leave) the social network, and the company that hosts it already has access to the data. It is also difficult to argue for whom publicly communicated messages are intended and for whom they are not.
- At the moment, a lot of data privacy issues are still open: Which data is protected, and which is not? When is the sharing of data permitted, and when is it not? In case of violations, those responsible for social networks should expect to be fined in accordance with § 83 paras. 4–6 General Data Protection Regulation (GDPR). At the same time, there are doubts about whether the provisions in § 42 Federal Data Protection Act (Bundesdatenschutzgesetz BDSG) can be applied to this matter. While the BDSG prohibits the dissemination of personal information that is not publicly available, it is unclear whether this can be applied to information provided by users of social networks.

"This matter is hardly ever prosecuted," says Liane Wörner, "and only if someone takes issue with it. But that rarely happens in the case of shadow profiles." After all, very few people even know that a secret profile of them exists. Indeed, (criminal) law does not prohibit anyone from engaging in the structured collection of publicly available data. In many cases, it can actually make sense to collect this data, which is often the idea behind services such as Wikipedia, because it provides a large amount of useful information. Such data is used, for example, to forecast weather or traffic jams on the Autobahn—or even on the bicycle bridge in Konstanz.

'My data is always simultaneously the data of others'

"When regulating digital networks, we very often lean towards an individualized solution: giving individuals control over what they share on the platform. But that only helps to a limited degree," explains David Garcia. "If we think that privacy is just an individual choice, we miss the bigger picture. Privacy is not an individual phenomenon. Privacy is a collective responsibility."

The computer scientist warns, "Individual solutions will not fully protect our privacy." He recommends using regulations that work on a collective level to prevent shadow profiling. "One approach would be to prevent centralized data collection, so that no one person or institution holds all the data," Garcia suggests. Furthermore, in his view, companies should be required to comply with standards that prevent shadow profiling and ensure greater transparency.

Garcia himself is studying technical approaches to protecting people from shadow profiles. One idea is to create "information noise"—in other words, to protect the real data by feeding networks "background noise" consisting of automated false data. Shadow profiles would then be worthless because they were based on false assumptions. This would

make it more difficult or even impossible to puzzle together profiles in a structured way. "However, there would still be the risk of creating a false history in the process," Liane Wörner adds, "If we do nothing to address the issue, then we will have a new problem to solve." Another question the Center for Human | Data | Society would like to investigate is how such noise would affect the precision of shadow profiles as well as the user experience.

The solution must be multidisciplinary

Garcia also wants to develop a model that can be used to identify a "red line" where a network has too much data and shadow profiles become too accurate. However, he always stresses that a solution cannot come from the technical or legal side alone, but must always come from a multidisciplinary perspective that factors in individual people as well as the cultural implications for society.

Liane Wörner agrees with her colleague. For the legal scholar, it is not just about creating a regulatory mechanism. In her view, the central question is: What kind of data-driven world do we really want to live in? And how can we shape it accordingly?

"Law is far too often reduced to the role of a regulator that only comes in after the harm has already been done. But a central task of laws is to shape our interpersonal relationships," Wörner emphasizes. She sees great opportunity in conscientiously and wisely shaping the digitalization and datafication of society. The prerequisite for this, however, is a multidisciplinary approach in which the fields of law, computer science and cultural studies work hand in hand. The Center for Human | Data | Society will do pioneering work in this area.

"The goal of our work in Konstanz is to create joint concepts for data sharing and push towards a meaningful world of datafied environments,

that we are all a part of," Wörner concludes. "We already have a fast internet. But a good internet? No, we don't have that. We want to have good data. And what exactly that means is something we have to discuss," says Liane Wörner.

Provided by University of Konstanz