



PHOTO: ISTOCK/DMYTRO VARAVIN

Uncounted heads: statistics analyzing population development often lag behind reality. Cell phone data – from social networks, for example – offers a quick way of recording changes.

DIGITAL DEMOGRAPHY

TEXT: JAN SCHWENKENBECHER

More than one-third of the world's population uses Facebook at least once a month. In the process, Facebook collects a lot of data from its users, which represents a valuable but barely tapped source for scientific research. Emilio Zagheni and his team at the Max Planck Institute for Demographic Research in Rostock are using social media data in an attempt to track and predict international migration.

For three days, authorities observed the tropical wave that was making its way westward through the Atlantic as it gathered strength, and by the evening of September the 16th 2017 it had grown to such an extent that they declared it a hurricane and dubbed it "Maria". Day by day, Hurricane Maria continued to grow, finally reaching its maximum force three days later, at three o'clock in the morning. By that time, the storm was producing wind speeds of 280 kilometers per hour combined with an air pressure of 908

millibars. It was the tenth strongest tropical hurricane in recorded history. Maria slammed into the coast of Puerto Rico a few hours later, uprooting trees, ripping the roofs off houses, and flinging cars through the air. The electricity and fresh water supply collapsed and just eleven of 69 hospitals were still able to function. Maria became one of the most lethal natural disasters in U.S. history. The authorities initially claimed that it had caused 45 fatalities. However, they continued to correct this figure, and finally arrived at 2,975 victims about a year later. Thousands of Puerto Ricans – for several months, no one really knew exactly how many – fled the island nation, a U.S. territory, to seek refuge on the U.S. mainland. The U.S. Census Bureau only released the migration data it collects as a matter of routine over a year later.

Emilio Zagheni drew on all available sources and became aware of the underlying figures earlier on. He provided additional perspectives too. "Whilst the official data is good and extremely accurate," Zagheni explains, "it is based on annual snapshots and many things can happen between any two of these. Any information about what people have been doing in the meantime is lost. Some of them only stayed in continental U.S. for a few weeks or months. Others moved on to other states." Zagheni is a Director at the Max Planck Institute for Demographic Research. He graduated in statistics and demography, and tries to combine the two in his research. The working area he heads is known as "Digital and Computational Demography." What does he do there? In this area, digital data is studied and utilized to answer demo-

59



graphic questions. And which digital data in particular? Anything that he can get his hands on, but mainly data from Facebook, Instagram, Twitter, and LinkedIn. The case of Hurricane Maria shows how this approach works.

Better data, more targeted help

Zagheni and some of his colleagues had begun collecting aggregate and anonymous Facebook data for the U.S. in January 2017, some nine and a half months before Hurricane Maria made landfall. He would take a kind of snapshot every few months, which enabled him to see how many users were active in each of the 50 states and what their age and gender were. In the case of residents outside of continental U.S., the data also revealed from which countries or U.S. territories a particular user had come. Puerto Rico was one of these territories. Looking at the data, Zagheni and his colleagues found that the number of Puerto Ricans living on continental U.S. increased by 17 percent in the weeks after Hurricane Maria, which equates to around 185,200 people or 5.6 percent of the total population of Puerto Rico.

Most of them sought refuge in Florida, which is the closest U.S. state to Puerto Rico. But between 8,000 and 15,000 Puerto Ricans, most of them young men, also made their way to Pennsylvania, New York, Connecticut, and Massachusetts, which are far more distant. That information is important, because, as Zagheni explains, “To be able to help people, you first have to understand migration patterns and know who is going where and why. And so this kind of data is important, for example, to enable the authorities to send the right amount of resources and support to various places where it is needed.”

Hurricane Maria is just one of the many cases in which Zagheni has used Facebook data to explore human migration. He accesses the data via the ads

platform, which was specifically created for advertisers. Businesses that wish to place an advert via the app can specify which Facebook users should be able to see it, for example, men living in Berlin, aged between 30 and 40. Exactly how Facebook concludes that particular users have certain demographic characteristics and interests is not fully explained by the social media giant. However, Zagheni and his colleagues do not just have to have faith in the black box that is Facebook. As the old saying goes, “faith is good, but facts are better,” so Zagheni and his team checked how accurate the

data was for basic demographic data. In a recently published working paper, they demonstrated that Facebook’s data regarding gender, age, and place of residence is between 86 and 93 percent accurate. Facebook’s data offers researchers advantages not only in regard to Hurricane Maria because, for one thing, it contains much more granular details than traditional data from official sources. It can, for example, be downloaded on a monthly or even daily basis if required. It is also more readily available than government data and can be accessed much more rapidly. Another benefit is its

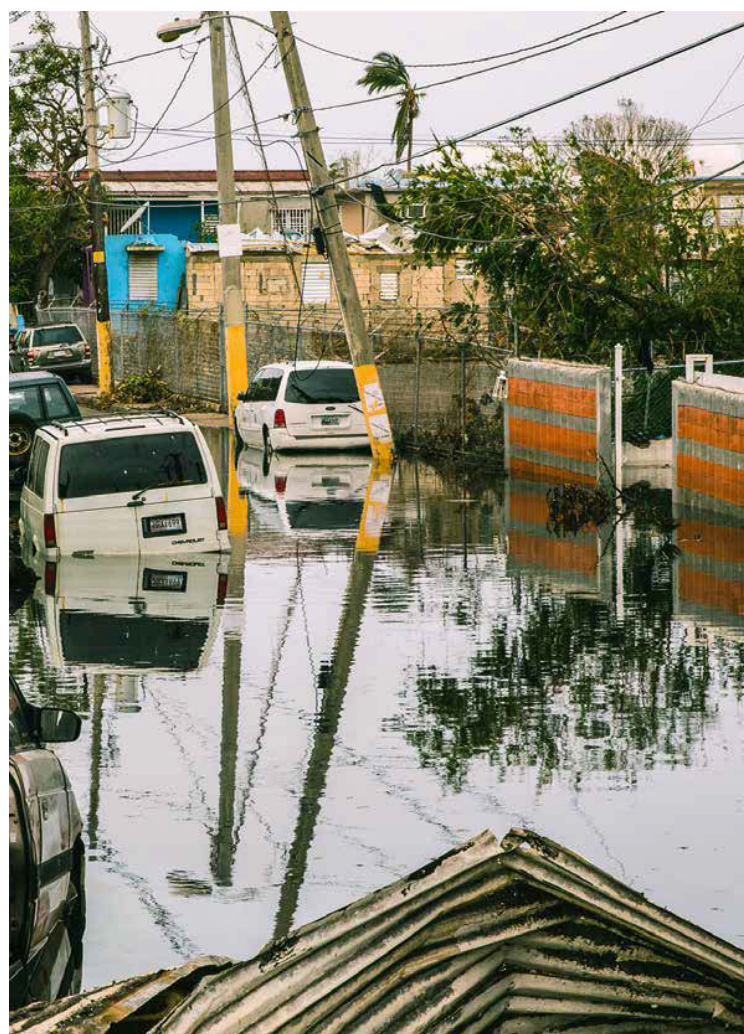


PHOTO: SHUTTERSTOCK/ALESSANDRO PIETRI

Ongoing disaster: some streets in San Juan, the capital of Puerto Rico, were still flooded weeks after Hurricane Maria had passed over.

A tenacious throng: long lines formed at the airport shortly after the disaster, where some people waited for days for a flight.



PHOTO: JOE RAEDLE/GETTY IMAGES

comparability because when it comes to traditional data sets it is often the case that different countries collect data using different definitions and at varying levels of detail and accuracy.

If, says Zagheni, one were to make inquiries in Poland and in Germany about how many people emigrated from Poland to Germany in 2007, the figures would be very different. “The Poles would say it was about 14,000, whereas the Germans cite 150,000,” one of the reasons for the discrepancy being because “the German authorities classify anyone who comes from Poland and registers at the town hall as an immigrant, regardless of how long he or she stays in Germany. The Polish authorities, on the other hand, only classify people as emigrants if they intend to relocate permanently to another country.” In terms of Facebook data, everyone is the same and therefore comparable. But there is also a big disadvantage: Facebook users are not representative of the entire

population. In a nutshell, a lot of over-60s and a huge number of under-20s do not even use Facebook. While the figures for other social networks may be different, none of them is actually representative of the population as a whole so, from a research perspective, it is becoming increasingly difficult to make absolute statements. Observing trends, on the other hand, enables researchers to filter out inherent distortions that remain constant.

Mobility data from the career portal

But migration is not the only field of research for which Zagheni and his team rely on digital data; another aspect concerns the aging of the population. “We’re currently analyzing the extent to which technology could be of use to senior citizens and how access to technology is changing for different groups,” Zagheni explains.

How digitally literate are senior citizens and how could the situation be improved? Researchers are also studying the impact of technological change, for example: “Access to technology might differ between men and women in some countries,” as Zagheni explains. “Gauging this differential access – by looking at how many men and how many women use social media – would enable us to draw conclusions about the degree of women’s empowerment in a particular country.” Nor does the relevant data always have to come from Facebook. Sarah Johnson, a doctoral researcher in Zagheni’s research group, for example, is currently using data from LinkedIn, an online site where users can cite their professional expertise and companies can recruit new personnel. LinkedIn includes a special function for attracting employees, which Johnson is using to study the migration of highly educated professionals. “The platform provides us with highly valuable aggregate-level

information,” she says: “For example, we can see how many people are willing to move from one place to another.” Every LinkedIn user has the option of specifying this feature in their profile. Johnson has been collecting this type of aggregate-level data at regular intervals since last summer and is currently evaluating it. “We want to understand which types of users actually relocate and what characterizes that particular group,” she says: “That might enable us to make a more accurate assessment of the conditions under which someone would really relocate for a job and when they are less likely to do so.”

If Johnson succeeds in identifying specific factors that influence migration, they could be incorporated into existing migration models to make them more accurate. Researchers would then be able to make more accurate predictions of migration movements. Another doctoral researcher in Zagheni’s team, Carolina Coimbra Vieira, is also working on this question. “People often simply assume that migrants relocate to a neighboring country whenever they move abroad,” says Coimbra Vieira. “But mere proximity is just one of many factors that play a role, such as the language spoken or

the economic conditions in the destination country.” She is also researching another potential factor, namely the cultural proximity between the two countries. Yet, how can one measure cultural proximity? “Using food,” says Coimbra Vieira.

Sugar cane schnapps and cheese crackers

Being a Brazilian, Vieira tested the idea using Brazil as her representative sample: she first identified the ten most typical Brazilian dishes, from churrasco to cachaça and pão de queijo to feijoada. She then analyzed how many Facebook users in Brazil and in 29 other countries were interested in Brazilian food on the assumption that the more people who share the interest, the closer their respective cultures are. “On the one hand, what we observed was that the cultural proximity to neighboring countries, such as Paraguay and Bolivia, is fairly high,” says Coimbra Vieira. “Yet geographical proximity was not the only factor that determined cultural proximity. An especially large number of Facebook users in both Portugal and Angola were also interested in Brazilian food.” Both are Lusophone countries and also have close historic ties to Brazil. Coimbra Vieira and her colleagues are currently testing the model for 16 other countries beyond Brazil and the initial results are promising. “Again, what we see is that the degree of cultural proximity or distance is a predictor of migration,” says Coimbra Vieira.

Data protection is a key issue in all of the research being conducted by Zagheni’s working group. “We use highly aggregated data sets in which no group contains less than 1,000 people,” Zagheni explains. “That makes it impossible to draw conclusions about specific individuals.” But that is not the only aspect of data protection. “We also need to think about

GRAPHIC: GCO BASED ON M. ALEXANDER, K. POLIMIS, E. ZAGHENI, THE IMPACT OF HURRICANE MARIA ON OUT-MIGRATION FROM PUERTO RICO: EVIDENCE FROM FACEBOOK DATA, POPULATION AND DEVELOPMENT REVIEW 45, 3 (2019); WIKIPEDIA



group privacy,” he says. If, for example, the data shows that a certain group of people is migrating from one region to another, you could endanger the group by publishing the data. For example, it could be a group of war refugees who are fleeing from government persecution.

There is another aspect to data privacy: “You have to take a different approach to data protection depending on the group of people involved. People vary

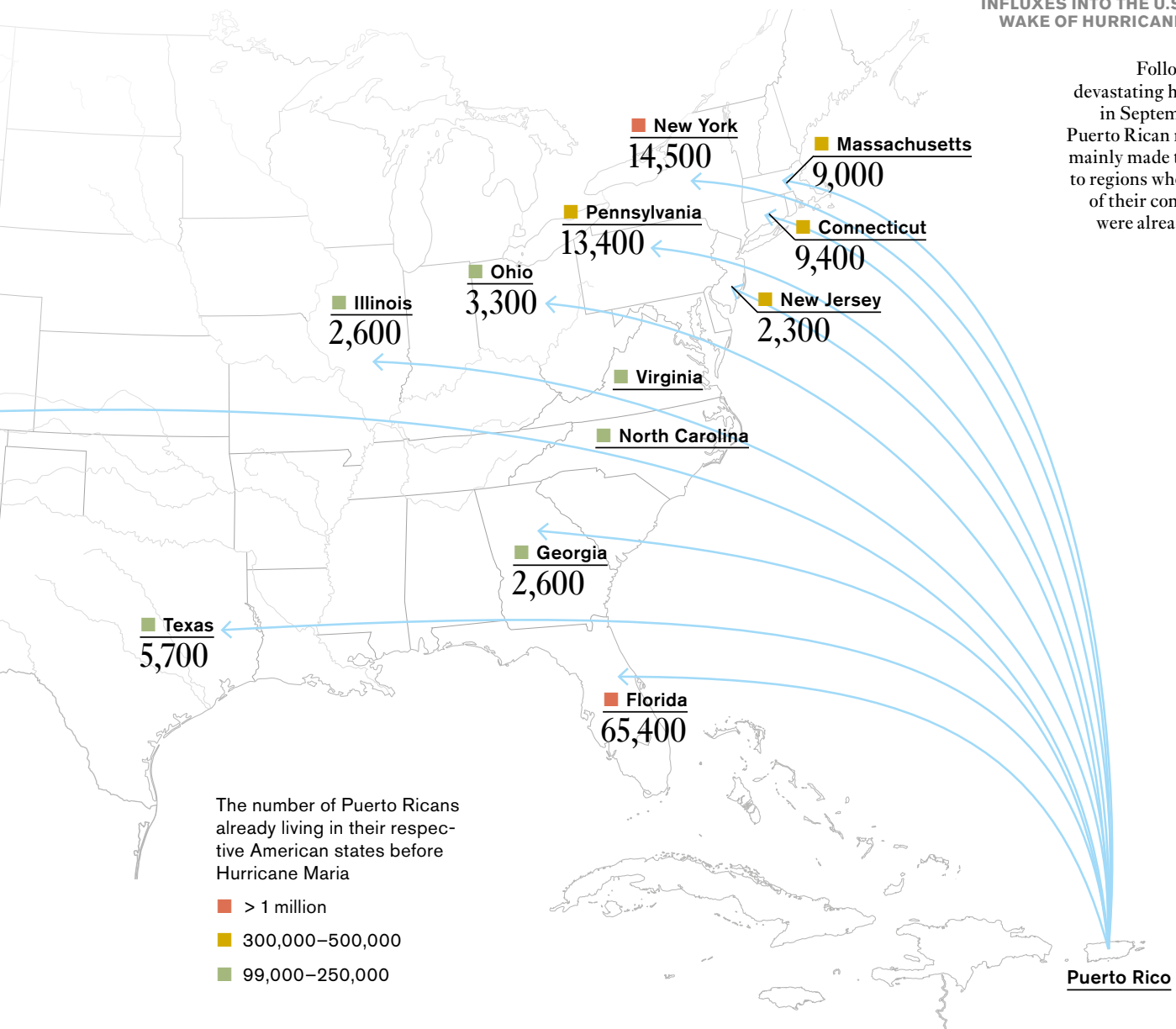
SUMMARY

When it comes to such things as tracking people’s movements after a natural disaster, social network data is more quickly accessible and more detailed than government data and, although it is not representative, it does provide a good source for monitoring and predicting migration trends.

In doing so, the researchers pay careful attention to the safety of vulnerable groups.

INFLUXES INTO THE U.S. IN THE WAKE OF HURRICANE MARIA

Following the devastating hurricane in September 2017, Puerto Rican migrants mainly made their way to regions where many of their compatriots were already living.



63

in terms of how well informed they are. Not everyone understands how their data could be exploited, and that level of awareness differs between various demographic groups: young or elderly, a lower or higher level of education, migrants or nationals.” There are also cases in which researchers draw a line and refuse to pursue a particular question although it would be technically and legally feasible to do so. There are public groups with thousands of members on social

media sites such as WhatsApp, whose chat history could easily be downloaded by any member of the group and used for research purposes. But because most of people don’t know about this possibility, Zagheni and his colleagues choose not to do so.

Zagheni’s research group already has enough unanswered research questions for which they will need to find new approaches, perhaps by combining multiple data sources such as

Facebook records with data tables from the U.S. Census Bureau, or by collecting their own data in the future. If, for example, there was an interest in which books or music those 30- to 40-year-old Berliners who are interested in Brazilian food read and listen to, then Zagheni and his colleagues could simply place an advert aimed at this target group on Facebook and invite them to take part in a survey. In this way, they suddenly have access to entirely new digital data sources.