



Ungezählte Köpfe: Statistiken zur Bevölkerungsentwicklung hinken der Realität oft hinterher. Eine schnelle Möglichkeit, Veränderungen zu erfassen, bieten Handydaten – zum Beispiel aus sozialen Netzwerken.

VOLKSZÄHLUNG PER FACEBOOK

TEXT: JAN SCHWENKENBECHER

Mehr als ein Drittel der Weltbevölkerung nutzt Facebook mindestens einmal im Monat. Dabei sammelt das soziale Netzwerk viele Daten von den Nutzerinnen und Nutzern – eine wertvolle, aber kaum erschlossene Quelle für die Wissenschaft. Emilio Zagheni und sein Team am Max-Planck-Institut für demografische Forschung in Rostock nutzen die Daten zum Beispiel, um Migrationsströme nachzuvollziehen und Trends zu erkennen.

Drei Tage lang hatten die Behörden den Vorboten eines Wirbelsturms schon beobachtet, der sich da durch den Atlantik immer weiter gen Westen schob, wuchs und wuchs und am Abend des 16. September 2017 so groß geworden war, dass die Meteorologen ihn zum Hurrikan erklärten und ihm einen Namen gaben: Maria. Und Maria wuchs weiter, Tag um Tag. Die maximale Intensität erreichte der Wirbelsturm drei Tage später, um drei Uhr in der Nacht. 280 Stundenkilometer Windgeschwindigkeit und ein extrem niedriger Luftdruck von 908 Millibar. Der zehntstärkste je aufgezeichnete

tropische Hurrikan. Einige Stunden später traf Maria auf die Küste Puerto Ricos. Der Wirbelsturm entwurzelte Bäume, riss Dächer von Häusern, wirbelte Autos durch die Luft. Die Strom- und die Trinkwasserversorgung brachen zusammen. Nur elf von 69 Krankenhäusern konnten den Betrieb aufrechterhalten. Maria wurde zu einer der tödlichsten Naturkatastrophen in der Geschichte der Vereinigten Staaten: Die Behörden sprachen zunächst von 45 Todesopfern, korrigierten die Zahl jedoch immer wieder – etwa ein Jahr später wurde sie mit 2975 angegeben. Tausende Puerto Ricaner verließen den Inselstaat, ein Außengebiet der USA, und suchten Zuflucht auf dem US-amerikanischen Festland. Wie viele genau, das wusste monatelang niemand so recht. Erst über ein Jahr später veröffentlichte das United States Census Bureau seine routinemäßig erhobenen Migrationsdaten.

Einer, der die Zahlen schon früher kannte, dazu mit zusätzlichen Details, ist Emilio Zagheni, der Daten aus allen verfügbaren Quellen zusammengetragen hatte. „Die Angaben der Behörden sind gut, sie sind sehr genau. Aber es sind jährliche Einmalaufnahmen, und im Zeitraum zwischen zwei Bildern kann viel passieren“, erklärt Zagheni. „Was die Menschen dazwischen gemacht haben – diese Informationen gehen verloren. Manche Puerto Ricaner kamen nur kurz aufs Festland der USA und kehrten nach einigen Wochen oder Monaten wieder zurück. Andere sind innerhalb des Landes weitergereist.“ Zagheni ist Direktor am Max-Planck-Institut für demografische Forschung. Er hat Statistik und Demografie studiert, in seiner Forschung versucht er nun beides zu kombinieren. Der Arbeitsbereich, den er leitet, heißt „Digitale und computergestützte Demografie“.



Was er da tut? Demografische Fragen mit digitalen Daten beantworten. Welche digitalen Daten das sind? Alle, die Zagheni bekommen kann – hauptsächlich allerdings solche aus sozialen Netzwerken wie Facebook, Instagram, Twitter oder auch LinkedIn. Wie genau das funktioniert, zeigt das Beispiel Maria.

Bessere Daten, gezieltere Hilfe

Bereits im Januar 2017, neuneneinhalb Monate vor Maria, hatte Emilio Zagheni mit ein paar Kollegen begonnen, aggregierte und anonyme Facebook-Daten aus den USA zu sammeln. Alle paar Monate machte er eine Art Schnappschuss. Daraus konnte er ablesen, wie viele User in jedem der 50 Bundesstaaten aktiv waren und welches Alter und welches Geschlecht sie hatten. Außerdem konnte Zagheni dort sehen, aus welchen Regionen außerhalb des US-Festlands die Menschen stammten. Zu diesen Regionen gehört auch Puerto Rico. In ihren Daten sahen Zagheni und seine Kollegen, dass auf dem US-amerikanischen Festland die Zahl der Puerto Ricaner in den Wochen nach Maria um 17 Prozent anstieg. Das entspricht etwa 185 200 Personen und damit 5,6 Prozent der Gesamtbevölkerung Puerto Ricos.

Die meisten dieser Personen flüchteten nach Florida. Das liegt nahe, wortwörtlich. Aber auch in die entfernter gelegenen Bundesstaaten Pennsylvania, New York, Connecticut oder Massachusetts kamen zwischen 8000 und 15 000 Menschen aus Puerto Rico, die meisten von ihnen waren junge Männer. Wichtige Informationen, denn: „Um den Menschen helfen zu können, muss man Migration zunächst verstehen und wissen, wer wohin geht, um was zu tun“, sagt Zagheni. „Und dafür sind Zahlen wie diese wichtig, um zum Beispiel die richtige Menge an Ressourcen und Unterstützung an verschiedene Orte zu schicken.“

Hurrikan Maria ist dabei nur eines von zahlreichen Beispielen, in denen

Zagheni Facebook-Daten nutzt, um damit die Migrationsbewegungen zu erforschen. An die Daten gelangt er über die eigens für Werbekunden geschaffene Ads-Plattform. Dort können Unternehmen, die eine Werbeanzeige schalten möchten, angeben, welche Facebook-User die Werbung angezeigt bekommen sollen. Zum Beispiel: Männer, die in Berlin leben und zwischen 30 und 40 Jahre alt sind. Wie Facebook darauf kommt, dass eine bestimmte Person bestimmte demografische Merkmale und Interessen hat, das verrät das soziale Netzwerk nicht. Dennoch müssen Zagheni und sein Team der Black Box Facebook nicht blind vertrauen. Nach dem

Motto „Vertrauen ist gut, Kontrolle ist besser“ haben sie die Genauigkeit grundlegender demografischer Daten geprüft. In einem jüngst veröffentlichten Working Paper zeigen sie, dass die Facebook-Daten zu Geschlecht, Alter und Wohnort zu 86 bis 93 Prozent stimmen. Nicht nur im Fall von Hurrikan Maria bieten Facebook-Daten einige Vorteile. Zum einen, weil sie viel kleinteiligere Informationen enthalten als die klassischen, behördlichen Datenquellen. Etwa weil sie monats- oder bei Bedarf sogar tageweise abgefragt werden können. Dazu sind sie schneller verfügbar als Behördendaten. Viel schneller. Und ein weiterer Vorteil ist die Vergleich-



FOTO: SHUTTERSTOCK/ALESSANDRO PIETRI

Andauernde Katastrophe: Noch Wochen nach Hurrikan Maria waren in der puerto-ricanischen Hauptstadt San Juan manche Straßen überflutet.

Ausdauernde Menge: Kurz nach der Katastrophe bildeten sich am Flughafen lange Schlangen von Menschen, die teilweise tagelang auf einen Flug warteten.



FOTO: JOE RAEDLE/GETTY IMAGES

barkeit. Denn für klassische Datensets gilt oft: Verschiedene Länder erheben Daten mit unterschiedlichen Definitionen und mit unterschiedlicher Detailliertheit und Genauigkeit.

Frage man etwa in Polen und in Deutschland nach, wie viele Menschen im Jahr 2007 von Polen nach Deutschland ausgewandert sind, dann fallen die Angaben recht verschieden aus, sagt Zagheni. „Die Polen sprechen etwa von 14000, die Deutschen von 150000 Personen.“ Das liege mitunter daran, dass „die deutschen Behörden alle, die aus Polen kommen und sich beim Rathaus melden, als Migranten registrieren. Unabhängig davon, wie lange jemand bleibt“, sagt Zagheni. „Die polnischen Behörden hingegen zählen nur jene Menschen als Migranten, die die Absicht haben, dauerhaft in ein anderes Land zu ziehen.“ In den Facebook-Daten sind alle gleich – und damit auch vergleichbar. Doch gibt es auch einen großen Nachteil: Die Nut-

zer von Facebook sind für die Bevölkerung nicht repräsentativ. Um es kurz zu fassen: Viele über 60-Jährige und sehr viele unter 20-Jährige nutzen Facebook nicht. In anderen sozialen Netzwerken mag das anders aussehen, doch repräsentativ für die Bevölkerung sind sie eigentlich alle nicht. Für die Forschung heißt das: Absolute Aussagen zu treffen, wird schwieriger. Was dennoch gut geht: Trends beobachten und vergleichen. Denn konstante Verzerrungen können die Forscher herausrechnen.

Mobilitätsdaten vom Karriereportal

Die Migration ist aber nicht das einzige Forschungsfeld, für das Zagheni und sein Team auf digitale Daten zurückgreifen. Ein zweiter Bereich ist die Alterung der Bevölkerung. „Wir untersuchen, inwieweit die Technologie für ältere Menschen von Nutzen sein

kann und wie sich der Zugang zur Technologie für verschiedene Gruppen verändert“, sagt Zagheni. Wie hoch ist der Grad der digitalen Kompetenz älterer Menschen, und wie könnte er verbessert werden? Und drittens untersuchen die Forscher die Auswirkungen des technologischen Wandels. Ein Beispiel: „In manchen Ländern haben Männer und Frauen möglicherweise unterschiedlichen Zugang zu digitaler Technologie“, so Zagheni. „Wenn wir diesen Zugang messen – indem wir schauen, wie viele Männer und wie viele Frauen soziale Netzwerke nutzen –, können wir womöglich Rückschlüsse auf den Grad der Selbstbestimmung von Frauen in einem bestimmten Land ziehen.“ Auch muss es nicht immer Facebook sein. Sarah Johnson zum Beispiel, Doktorandin in Zaghenis Forschungsgruppe, arbeitet gerade mit den Daten von LinkedIn. Dort können Nutzer ihre berufliche Expertise angeben, Firmen können rekrutieren. Fürs Anwerben von Mit-

arbeitenden gibt es auf LinkedIn eine spezielle Funktion – und die macht sich Sarah Johnson zunutze, um die Migration hochqualifizierter Fachkräfte zu untersuchen. „Über die Plattform bekommen wir wertvolle aggregierte Informationen über die Nutzer“, sagt Johnson. „Wir können zum Beispiel sehen, wie viele Personen bereit sind, von einem bestimmten Ort an einen anderen Ort zu ziehen.“ Das können LinkedIn-User nämlich in ihrem Profil angeben. Seit Sommer vergangenen Jahres sammelt Johnson in regelmäßigen Zeitabständen aggregierte Nutzerdaten wie diese. Die Auswertung läuft. „Wir wollen verstehen, welche Personen tatsächlich umziehen und was diese Gruppe kennzeichnet“, erläutert Johnson. „So können wir möglicherweise genauer erkennen, unter welchen Bedingungen jemand für einen Beruf wirklich umzieht und wann eher nicht.“

Findet Sarah Johnson Faktoren, die die Migration beeinflussen, so könnten diese in bestehende Migrationsmodelle eingearbeitet werden, um sie genauer zu machen. Dann könnten die Forscher Migrationsbewegungen genauer vorhersagen. Daran arbeitet auch Carolina Coimbra Vieira, eine

weitere Doktorandin in Zaghenis Team. „Oft geht man einfach davon aus, dass Migranten ins Nachbarland ziehen, wenn sie ein Land verlassen“, sagt Coimbra Vieira. „Aber es gibt viel mehr Faktoren, die eine Rolle spielen, als nur die Nähe. Die Sprache zum Beispiel oder auch der Wohlstand im Zielland.“ Sie untersucht einen weiteren möglichen Aspekt: die kulturelle Nähe zweier Länder. Nur: Wie kann man kulturelle Nähe messen? Die Forscherin sagt: „Mit Essen.“

Zuckerrohrschnaps und Käsegebäck

Coimbra Vieira ist Brasilianerin, also testete sie die Idee am Beispiel Brasiliens. Zunächst suchte sie die 20 typischsten brasilianischen Spezialitäten heraus – vom Zuckerrohrschnaps Cachaça über den Bohneneintopf Feijoada und das Grillgericht Churrasco bis zu Pão de queijo, einem Käsegebäck. Dann schaute sie, wie viele Facebook-Nutzer sich in Brasilien und in 29 weiteren Ländern für brasilianisches Essen interessierten. Die Annahme: Je mehr Leute das tun, desto näher sind sich die Kulturen. „Einerseits sahen wir, dass die kulturelle Nähe zu Nachbarstaaten recht hoch ist, zum Beispiel zu Paraguay oder Bolivien“, sagt Coimbra Vieira. „Aber es war nicht nur die geografische Lage, welche die kulturelle Nähe bestimmte. Auch in Portugal oder Angola interessierten sich besonders viele Facebook-Nutzer für brasilianisches Essen.“ Beides Länder, in denen die Menschen Portugiesisch sprechen und die auch historisch eng mit Brasilien verknüpft sind. Aktuell testen Coimbra Vieira und ihre Kollegen das Modell über Brasilien hinaus für insgesamt 16 weitere Länder. Die ersten Ergebnisse sehen gut aus: „Auch da zeigt sich, dass kulturelle Ähnlichkeit oder kulturelle Distanz ein Prädiktor für Migration ist“, sagt Coimbra Vieira.

Bei allen Arbeiten von Emilio Zaghenis Forschungsgruppe ist Datenschutz ein wichtiges Anliegen. „Wir nutzen stark aggregierte Daten. Keine der Gruppen ist kleiner als 1000 Perso-



GRAFIK: GCO NACH 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

nen“, sagt Zagheni. „Da ist es nicht möglich, Rückschlüsse auf Einzelpersonen zu ziehen.“ Doch das sei nicht der einzige Aspekt. „Wir müssen auch auf den gruppenbezogenen Datenschutz achten.“ Damit meint der Forscher: Wenn man etwa in den Daten sehe, dass eine bestimmte Gruppe von Menschen von einer Region in eine andere zieht, könne man mit einer Publikation diese Gruppe in Gefahr bringen. Zum Beispiel, wenn es sich um Kriegsflüchtlinge handelt, die von einer Regierung verfolgt werden.

AUF DEN PUNKT GEBRACHT

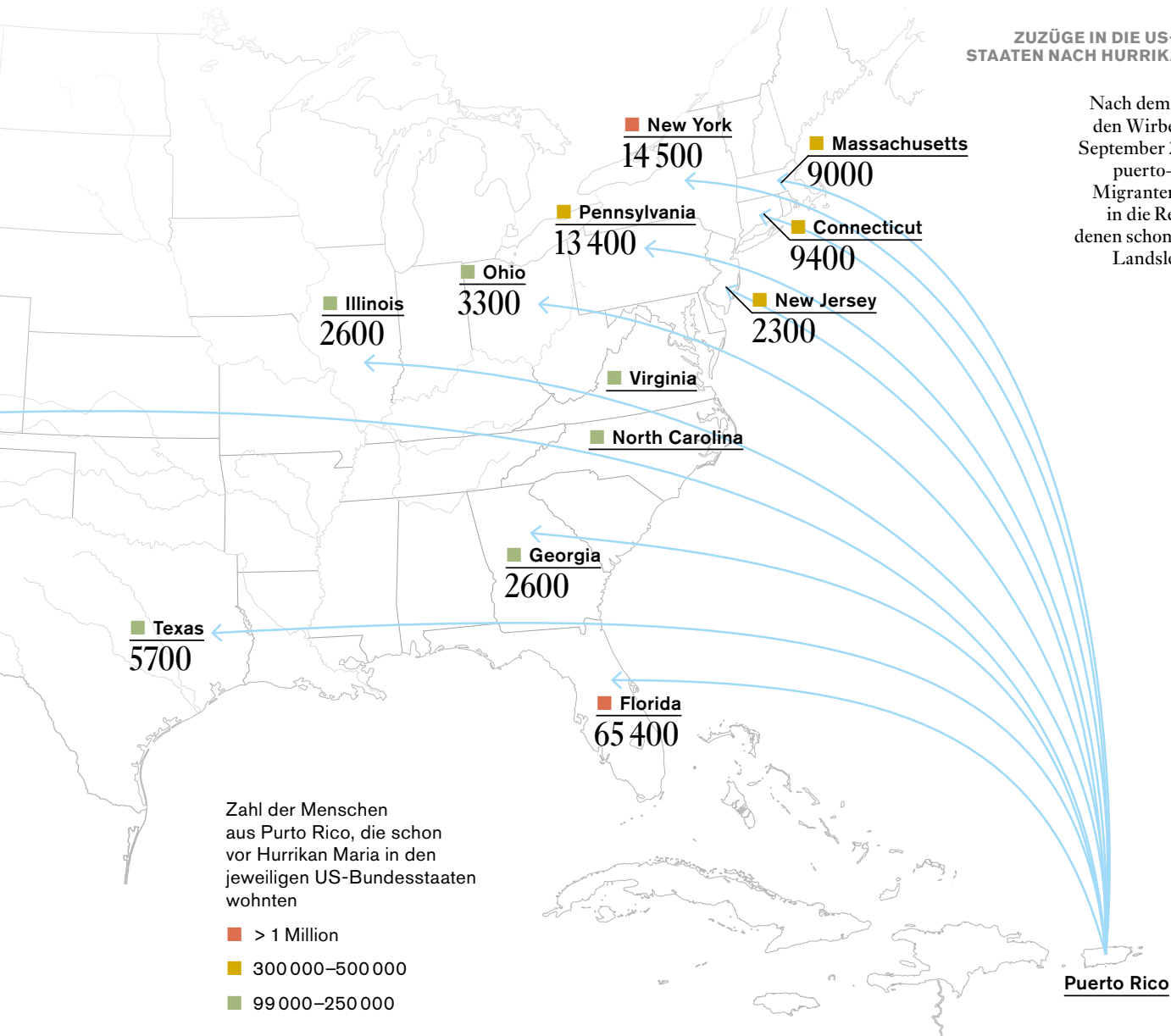
Daten aus sozialen Netzwerken sind schneller verfügbar und detaillierter als Behördendaten, um zu erfassen, wohin sich Menschen zum Beispiel nach einer Katastrophe flüchten.

Diese Daten sind zwar nicht repräsentativ, aber eine gute Quelle für die Beobachtung und die Vorhersage von Migrationstrends.

Die Forschenden achten dabei sorgfältig auf die Sicherheit gefährdeter Personengruppen.

ZUZÜGE IN DIE US-BUNDESSTAATEN NACH HURRIKAN MARIA

Nach dem verheerenden Wirbelsturm im September 2017 zogen puerto-ricanische Migranten vor allem in die Regionen, in denen schon viele ihrer Landsleute lebten.



Ein weiterer Punkt: „Je nach Personen-
gruppe müssen wir das Thema Da-
tenschutz verschieden behandeln. Die
Menschen sind unterschiedlich gut
informiert. Nicht alle sind sich be-
wusst, wie ihre Daten verwendet wer-
den können, und der Grad dieses Be-
wusstseins variiert zwischen den de-
mografischen Gruppen: jung oder alt,
niedrigere oder höhere Bildung,
Migranten oder Einheimische.“ In
manchen Fällen ziehen die Forschenden
eine rote Linie und gehen einer Frage
nicht weiter nach, obwohl es
technisch und rechtlich möglich wäre.

Bei Whatsapp zum Beispiel. Da gibt
es öffentliche Gruppen mit Tausenden
Mitgliedern. Den Chatverlauf
könnte jedes Mitglied problemlos
runterladen und zur Forschung nutzen.
Aber weil die meisten Menschen das gar
nicht wissen, lassen Zagheni
und sein Team es lieber bleiben.

Die Arbeitsgruppe hat ohnehin genü-
gend offene Forschungsfragen, für deren
Beantwortung sie neue Methoden
finden möchte. Zum Beispiel wollen
Zagheni und sein Team zukünftig
mehrere Datenquellen kombinieren,

etwa Facebook-Datensätze mit Tabel-
len des US-amerikanischen
Census-Büros. Oder in Zukunft ei-
gene Daten erheben. Wenn sie von den
30- bis 40-jährigen Berlinern, die sich
für brasilianisches Essen interessieren,
etwa auch noch wissen wollen, welche
Bücher sie lesen und welche Musik sie
hören, könnten die Forschenden tat-
sächlich eine Werbeanzeige für diese
Zielgruppe auf Facebook schalten und
sie zu einer Umfrage einladen. So ent-
stehen auf einmal ganz neue digitale
Datenquellen, die noch detailliertere
Erkenntnisse ermöglichen.

