

Digital discrimination: Political bias in Internet service provision across ethnic groups

Nils B. Weidmann,^{1*} Suso Benitez Baleato,^{1,2} Philipp Hunziker,³
Eduard Glatz,⁴ Xenofontas Dimitropoulos^{5,6}

The global expansion of the Internet is frequently associated with increased government transparency, political rights, and democracy. However, this assumption depends on marginalized groups getting access in the first place. Here we document a strong and persistent political bias in the allocation of Internet coverage across ethnic groups worldwide. Using estimates of Internet penetration obtained through network measurements, we show that politically excluded groups suffer from significantly lower Internet penetration rates compared with those in power, an effect that cannot be explained by economic or geographic factors. Our findings underline one of the central impediments to “liberation technology,” which is that governments still play a key role in the allocation of the Internet and can, intentionally or not, sabotage its liberating effects.

In the wake of the Arab Spring, the Internet has often been portrayed as a “liberation technology” (1). Specifically, it has been argued that the Internet fosters transparency and accountability of nondemocratic governments worldwide and can help opposition movements organize for collective action (2). This expectation, however, is based on the assumption that political activists have sufficient access to the Internet in the first place.

The socioeconomic background of individuals affects their access to the Internet (3, 4). Also, there is evidence of a global digital divide: Countries with democratic institutions and higher levels of development have higher Internet penetration rates (5). Still, we do not know how the provision of Internet services varies across societal groups in a country or how it is driven by politics. This information is key if we are to assess whether the

Internet can indeed empower politically marginalized populations.

In most developing countries, governments are the major, if not the only, provider of telecommunication services (6). At the same time, in many of these countries, politics operates along ethnic lines, so that one or more groups hold

¹Department of Politics and Public Administration, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany. ²Department of Political Science, University of Santiago de Compostela, Campus Vida, 15705 Compostela, Spain. ³International Conflict Research, ETH Zurich, Haldeneggsteig 4, 8092 Zurich, Switzerland. ⁴Computer Engineering and Networks Laboratory, ETH Zurich, Gloriastraße 35, 8092 Zurich, Switzerland. ⁵Foundation for Research and Technology Hellas, Nikolaou Plastira 100, 71110 Heraklion, Crete, Greece. ⁶Department of Computer Science, University of Crete, Voutes Campus, 70013 Heraklion, Crete, Greece.

*Corresponding author. Email: nils.weidmann@uni-konstanz.de

political power at the expense of other, marginalized ones (7). This allows Internet technology to be implemented in a way that benefits certain groups while neglecting others. Two mechanisms can account for this: First, ethnic groups in power can foster economic and technological development in their home regions (8), a phenomenon typically referred to as “ethnic favoritism.” Second, governments can attempt to strategically exclude certain groups from access to communication technology because they are afraid of facilitating political mobilization and unrest. Either (or both) of these mechanisms can lead to digital discrimination, in which politically marginalized groups suffer from reduced access to modern information and communications technology (ICT). Our aim was to test whether less politically favored ethnic groups are systematically deprived of this access by governments. The biggest challenge in doing so was to rule out alternative explanations (such as uneven economic status across groups) for the potential correlation between political status and Internet penetration.

We based our analysis on the Ethnic Power Relations (EPR) list of politically relevant groups (7). EPR distinguishes between politically “included” and “excluded” groups. The former are groups that have access to (executive) political power at the national level, such as by having representatives in the government of a country. This access can take a variety of forms, ranging from participation in ethnically shared governments to complete ethnic monopolies, in which the entire executive apparatus is effectively controlled by a single group. Conversely, excluded groups do not participate in the executive apparatus at the national level,

but they may have different levels of regional power or no power at all.

Typically, the main source of statistics about Internet penetration is the International Telecommunication Union’s World Telecommunications/ICT Indicators Database (9). This database provides more than 150 indicators of different aspects of ICT coverage, including access to and use of the Internet for 192 countries starting from 1992. However, it offers data only at the country level, making it useless for subnational analysis across ethnic groups. A source of subnational estimates could be the official statistics provided by the telecommunication ministries and regional operators, because they sometimes include data at the regional or local level. Unfortunately, this level of disaggregation is only available for a limited number of countries, mostly industrialized ones, which would severely limit the scope of our study. Yet another data source could be the large surveys that are periodically conducted by international agencies to describe trends in development, health, and demographics. For example, the Demographic and Health Surveys include questions about technological access, such as whether the Internet is available in a household, and about the use of certain Internet services (10). Some of these surveys also contain the (self reported) ethnicities of the individuals interviewed. However, although these surveys are generally representative at the country level, this is not necessarily the case at the level of ethnic groups. Moreover, relying on surveys would restrict our sample considerably because they are available for a limited set of countries only.

Therefore, we estimated Internet penetration spatially using the method proposed in (11), which has shown that Internet penetration in countries and subnational administrative units (provinces or states) can be approximated by the number of active Internet subnetworks. We used subnetworks of size /24 (the slash notation refers to the number of bits used for the network address), which correspond to ~256 IP (Internet protocol) addresses each and are typically assigned to institutions or small providers.

Our estimation of Internet penetration from active subnetworks proceeded in three steps. First, we determined which were the active /24 subnetworks. This was necessary because large parts of the Internet address space are not used for digital communication. We did this in two different ways. One way was a passive measurement that relied on observed Internet traffic from a large Internet service provider in Switzerland. This data set included the outside IP addresses with which any host within the provider’s network communicated, from which we obtained a sample of 16 days for each year in our study period (2004–2012). The methodology for creating this data set is described in (12), and we refer to the subnetworks determined through this method as “observed” subnetworks. The other estimation approach used routing metadata (border gateway protocol data) collected by the Route Views project (13). This project provides daily snapshots of routing tables for the same annual samples as above. The IP addresses in these samples were collapsed to unique /24 subnetworks, and reserved addresses were removed. We retained those subnetworks that appeared on every day of the 16 day

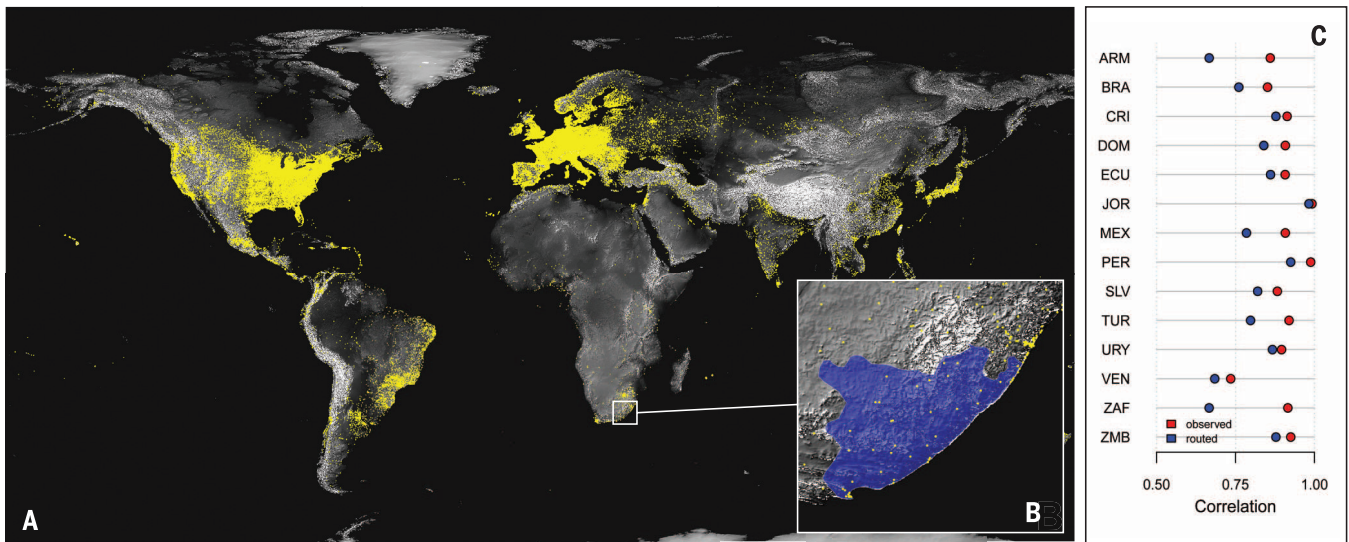


Fig. 1. Estimating Internet penetration from active Internet subnetworks. (A) Global map of active Internet subnetworks for the year 2012, where each yellow dot represents the subnetwork’s location. The areas with the highest Internet penetration are North America, Europe, and parts of Asia. (B) Active subnetworks were assigned to ethnic groups by identifying whether they were located within the settlement region of that group. As an example, the close-up view shows this for the Xhosa group (blue) in South Africa. (C) Results of the validation study of the Internet penetration measure. Active subnetworks

were assigned to subnational administrative regions for which the Internet penetration was known. The plots show the correlations between the (log-transformed) number of subnetworks per 1000 capita (red, observed; blue, routed) and the Internet penetration across the subnational administrative divisions for 14 countries. ARM, Armenia; BRA, Brazil; CRI, Costa Rica; DOM, Dominican Republic; ECU, Ecuador; JOR, Jordan; MEX, Mexico; PER, Peru; SLV, El Salvador; TUR, Turkey; URY, Uruguay; VEN, Venezuela; ZAF, South Africa; ZMB, Zimbabwe.

period to filter out subnetwork leaks due to misconfigurations (12). We call these the “routed” subnetworks. This approach, however, does not take into account whether the routed subnetworks actually transmit any data and may thus overestimate the number of active ones. However, because it is based on publicly available data, it has the advantage of being applicable

to other projects and in other contexts. In our analysis, we used these two methods as alternative measurement approaches for subnational levels of Internet penetration.

Second, we used a geolocation database to find the geographic location of the observed subnetworks (14) (Fig. 1A). This database translates IP addresses into geographic coordinates that best

approximate where on the globe a network is located. We used the MaxMind database, which is one of the most accurate databases with cover age going back as far as 2004 (15).

Third, we aggregated the active subnetworks (observed or routed) to the level of subnational geographic units (see Fig. 1B for an example). For our main analysis, we combined ethnic group

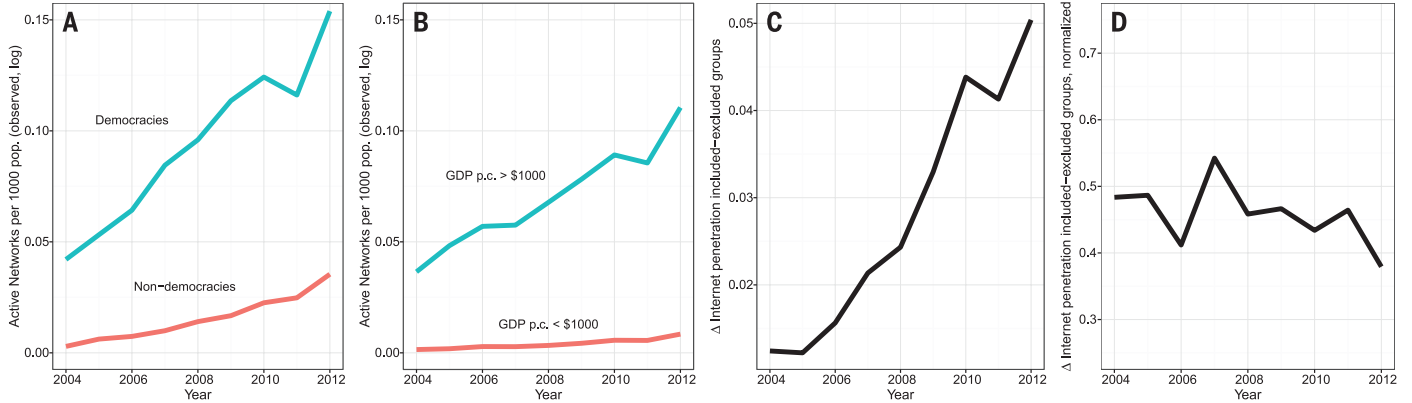


Fig. 2. Trends in Internet penetration for a global sample of ethnic groups. (A) Penetration rates in democracies and nondemocracies, the latter defined as having a democracy score of less than 6 according to (25). (B) Penetration rates in developed and less developed countries [gross domestic product (GDP) per capita (p.c.) data from (26)]. In (A) and (B), the log-transformed number of active subnetworks is shown per 1000 capita. (C) Yearly averages of the differences between included and excluded groups across all countries in the sample. (D) Differences between included and excluded groups, normalized by the country’s average level of Internet penetration.

Table 1. Regression results for Internet penetration as the dependent variable. Coefficients are shown with uncertainties (standard errors) in parentheses. Models 1 to 3 capture the absolute level of Internet penetration; models 4 to 6 capture Internet penetration relative to the country average. Models 1 and 4 include only the main independent variable (exclusion); models 2 and 5 use the group GDP indicator based on the G Econ data set (17) and the other control variables. Models 3 and 6 use the nighttime lights based indicator of development rather than the one based on the G Econ data set. These results are for observed subnetworks; results for routed subnetworks are shown in table S1. R^2 , coefficient of determination.

Dependent variable	Internet penetration, absolute			Internet penetration, relative		
	1	2	3	4	5	6
Excluded	0.027*** (0.003)	0.021*** (0.003)	0.019*** (0.003)	0.725*** (0.094)	0.481*** (0.094)	0.539 (0.090)
Group GDP p.c.		0.025*** (0.005)			0.749*** (0.155)	
Nighttime lights p.c.			0.023*** (0.004)			1.665*** (0.145)
Ruggedness		0.001*** (0.0001)	0.0004 (0.0001)		0.023*** (0.003)	0.011*** (0.003)
Distance to capital		0.006 (0.005)	0.001 (0.004)		0.942*** (0.133)	0.703*** (0.130)
Road density		0.110*** (0.033)	0.077** (0.032)		4.068*** (0.833)	3.108 (0.809)
Urbanization		0.525*** (0.026)	0.479*** (0.026)		2.782*** (0.748)	3.285*** (0.695)
Intercept	0.028 (0.054)	0.098* (0.052)	0.001 (0.050)	1.418 (1.628)	5.562*** (1.593)	0.65 (1.542)
Country × year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4663	4371	4581	2925	2811	2904
R^2	0.708	0.764	0.768	0.272	0.36	0.4
Adjusted R^2	0.623	0.694	0.698	0.119	0.017	0.071

* $P < 0.1$; ** $P < 0.05$; *** $P < 0.01$.

settlement regions from the GeoEPR data set (16) with the location of the subnetworks from the previous step. Using a simple GIS (geographic information system) overlay operation, we calculated the number of (observed or routed) subnetworks per group and year.

In a validation study, we confirmed that our indicator is able to capture subnational variation in Internet penetration. We aggregated the active subnetworks to the level of administrative units (provinces or districts) for which the level of Internet penetration is known (details are provided in section 1 of the supplementary materials). This analysis reveals generally high correlations (Fig. 1C). Results in sections 1.2 to 1.15 of the supplementary materials show that the log transformed number of observed subnetworks per population of 1000 achieves the best results, which is why we used this transformation for our main analysis (we provide additional results calculated using routed subnetworks as a robustness test in sections 3 to 6 of the supplementary materials). Because these tests were successful, we can assume that our method also works for ethnic group settlement regions, which in many cases are much larger than the administrative units in our validation analysis.

Our data illustrate that Internet penetration has increased for ethnic groups worldwide over time, but at higher rates in more democratic countries (Fig. 2A) and more developed ones (Fig. 2B). The “digital gap” between included and excluded groups within the same country widened over time in absolute terms (Fig. 2C). However, this increase was driven by the general increase in connectivity: The normalized difference between included and excluded groups (computed as the absolute difference as a proportion of the country’s average level of Internet penetration) actually decreased slightly (Fig. 2D). This suggests that excluded groups are catching up, if only slowly.

To determine the effect of political exclusion on Internet penetration at the level of groups, we used a regression analysis with the (log transformed) number per 1000 capita of active subnetworks per group to measure the absolute level of Internet penetration. In addition, we computed a relative indicator of Internet penetration that captures the Internet coverage of a group in relation to the other groups in a country (see section 2.1 of the supplementary materials for details). We ran our models with a separate intercept (fixed effect) for each country year to net out differences between countries that are due to different economic or political conditions or to country specific temporal trends in Internet adoption. To exclude the possibility that a group’s level of Internet penetration was simply a result of its level of development, geographic location, or settlement pattern (urban versus rural), we included a number of GIS derived control variables. Most importantly, we controlled for a group’s level of development by using either the disaggregated G Econ data set (17) or the group’s nighttime light emissions (18). These emissions have recently been proposed as an indicator

of economic performance (18, 19) and have been shown to strongly predict wealth at the local level (20). We also controlled for whether the group is located in a remote and inaccessible area by including an indicator of terrain ruggedness as well as the group’s distance from their country’s capital. Similarly, we included a control variable for the group’s level of urbanization (details are given in sections 2.2 to 2.7 of the supplementary materials).

The regression results in Table 1 demonstrate that excluded groups’ political status leads to significantly lower Internet penetration rates compared with included groups in the same country. This result is not driven by the groups’ level of development, their geographic location and quality of infrastructure, or their urban versus rural settlement pattern. The coefficients indicate that on average, excluded groups have between 0.019 and 0.021 subnetworks fewer per 1000 capita than politically included groups in the same country (models 2 and 3). To put this result into context, models 5 and 6 indicate that exclusion leads to a reduction in relative Internet penetration by a factor of about 0.6 ($e^{-0.481}$ and $e^{-0.539}$, respectively). This means that, all other factors being equal, an included group with an average level of Internet penetration for its country would receive only ~60% of that level if it were an excluded group. In additional analyses, we confirmed that this result is driven not only by lower penetration rates associated with exclusion (when comparing only groups that have some level of penetration) but also by excluded groups’ higher probability of having no Internet coverage at all (tables S2 and S3). Table S4 shows that, using the approach of (21), the inclusion of our control variables appears to be sufficient to remove bias from unobserved confounding variables. To study variation in the effect of exclusion across countries, we also used multilevel regression models (supplementary materials, section 6). These models fail to provide evidence that democracy alleviates the negative effect of exclusion on Internet coverage. Rather, they suggest that if democracies exclude groups politically, their level of digital discrimination is comparable to that of nondemocracies. Overall, however, because democratic countries have much lower percentages of excluded populations [on average 6%, compared with 21% in nondemocracies, according to the EPR data (7)], digital discrimination is a much more severe issue in nondemocratic countries.

What are the effects of reduced Internet coverage on excluded groups? Resource mobilization theories explain the emergence of social movements by their capability to mobilize a sufficient number of supporters (22), and the Internet should be a crucial technology to this end (23). If, however, politically excluded groups suffer from reduced levels of Internet access, as we have shown above, a group’s level of Internet penetration should be less important in determining whether it is able to mobilize collectively. We conducted additional regression analyses on the set of excluded groups to test whether a group’s level of Internet penetration helps to predict whether it engages in

collective violence against the government (supplementary materials, section 7). We found little evidence that it does. Overall, the effect of Internet penetration on the probability of collective mobilization was insignificant. When estimating the effect across different levels of development among ethnic groups (as measured by per capita light emissions), a positive and significant effect emerged for poor groups only, suggesting that the Internet can increase collective mobilization capacity if few other resources are available.

In sum, the politically motivated digital discrimination against ethnically marginalized groups that we identify in this analysis constitutes a challenge to proponents of liberation technology. In many countries, access to modern ICT, and in particular to the Internet, is determined by national governments. As we have shown, this can lead to selective provision of digital communication, with governments extending these services primarily to politically favored groups. Although the Internet clearly has the potential to foster collective organization and political change, governments can prevent this effect through their key role in the allocation and control of digital communication. This may be one explanation for the finding that Internet penetration is a weak predictor of collective mobilization at the level of ethnic groups.

This insight has important consequences for research and policy. First, students of the political effects of ICT will have to rethink if and under what circumstances the Internet can catalyze collective mobilization and political change. This scholarship needs to pay closer attention to the role of governments and the extent to which they can prevent these effects. Second, development policies aiming to promote peace and democratization through the Internet need to take into account the uneven provision of digital services within countries. Only if this digital inequality is alleviated can we expect these modern channels to empower people and societies in order to foster lasting political and economic development. Third, it is a frequent assumption that the uneven global distribution of digital technology can be mitigated by economic forces and incentives; for example, the recent *World Development Report 2016* suggests that more competition and better regulation of the ICT sector are necessary (24). Again, however, this suggestion needs to carefully consider the role of local political actors in shaping this process. If the allocation of digital services has become a means to reward loyalists and fend off challengers, improved connectivity for every one may not be in the interest of governments.

REFERENCES AND NOTES

1. L. Diamond, *J. Democracy* **21**, 69–83 (2010).
2. M. M. Hussain, P. N. Howard, *Int. Stud. Rev.* **15**, 48–66 (2013).
3. W. T. Kilenthong, P. Odton, *Telecomm. Policy* **38**, 1146–1159 (2014).
4. S. Mendonça, N. Crespo, N. Simões, *Telecomm. Policy* **39**, 192 (2015).
5. H. V. Milner, *Comp. Polit. Stud.* **39**, 176–199 (2006).
6. T. C. Boas, *How Revolutionary Was the Digital Revolution? National Responses, Market Transitions, and Global Technology*,

- J. Zysman, A. Newman, Eds. (Stanford Univ. Press, 2006), pp. 361–378.
7. L. E. Cederman, A. Wimmer, B. Min, *World Polit.* **62**, 87–119 (2010).
 8. R. Hodler, P. A. Raschky, *Q. J. Econ.* **129**, 995–1033 (2014).
 9. International Telecommunications Union, World Telecommunication/ICT Indicators Database (2013); www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx.
 10. Demographic and Health Surveys Program, <http://dhsprogram.com/>.
 11. S. Benitez Baleato, N. B. Weidmann, P. Gigis, X. Dimitropoulos, E. Glatz, B. Trammell, in *Proceedings of the Passive and Active Measurement Conference*, J. Mirkovic, Y. Liu, Eds. (Springer, 2015), pp. 220–231.
 12. A. Dainotti et al., *Comput. Commun. Rev.* **44**, 99–100 (2014).
 13. University of Oregon Route Views Project, www.routeviews.org.
 14. MaxMind, GeolIP2 Databases; www.maxmind.com/en/geolip2-databases.
 15. B. Huffaker, M. Fomenkov, K. Claffy, “Geocompare: A comparison of public and commercial geolocation databases” (CAIDA Tech Report, Center of Applied Internet Data Analysis, 2011); www.caida.org/publications/papers/2011/geocompare-tr/geocompare-tr.pdf.
 16. J. Wucherpfennig, N. B. Weidmann, L. Girardin, L. E. Cederman, A. Wimmer, *Conflict Manage. Peace Sci.* **28**, 423–437 (2011).
 17. W. D. Nordhaus, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3510–3517 (2006).
 18. J. V. Henderson, A. Storeygard, D. N. Weil, *Am. Econ. Rev.* **101**, 194–199 (2011).
 19. X. Chen, W. D. Nordhaus, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8589–8594 (2011).
 20. N. B. Weidmann, S. Schutte, *J. Peace Res.* 10.1177/0022343316630359 (2016).
 21. J. G. Altonji, T. E. Elder, C. R. Taber, *J. Polit. Econ.* **113**, 151–184 (2005).
 22. S. Tarrow, *Power in Movement: Social Movements, Collective Action and Politics* (Cambridge Univ. Press, 1994).
 23. J. Earl, K. Kimport, *Digitally Enabled Social Change: Activism in the Internet Age* (MIT Press, 2011).
 24. The World Bank, *World Development Report 2016: Digital Dividends* (The World Bank, 2016); www.worldbank.org/en/publication/wdr2016.
 25. M. G. Marshall, T. R. Gurr, K. Jagers, “Polity IV Project: Political regime characteristics and transitions, 1800–2015” (2015); www.systemicpeace.org/polity/polity4.htm.
 26. The World Bank, World Development Indicators (2016); <http://data.worldbank.org/data-catalog/world-development-indicators>.

ACKNOWLEDGMENTS

We acknowledge financial support from the Alexander von Humboldt Foundation (Sofja Kovalevskaja Award to N.B.W.). We are grateful to M. Baum, M. Becher, L. E. Cederman, D. Lazer, A. Little, and P. Selb for comments. We do not have any real or apparent conflicts of interest. Data for replication are available at <http://dx.doi.org/10.7910/DVN/Y3VPIG>. There are two restrictions. First, we cannot share the active subnetworks data because these are computed from traffic traces (netflow) captured from an Internet service provider. These traces were made available to us under a strict nondisclosure agreement that prevents further sharing or public release. However, we share the complete list of routed subnetworks (annual observations) because they were derived from a public source. The second restriction concerns the spatial coordinates for each subnetwork. MaxMind does not permit the publication of any information contained in their database, so we share a random sample of 5% of the routed subnetworks with truncated geographic coordinates, but without IP addresses, along with code that executes the spatial aggregation to the group polygons in SQL (structured query language).