Alexander Kolodin (SBN 030826)
Roger Strassburg (SBN 016314)
Veronica Lucero (SBN 030292)
Arno Naeckel (SBN 026158)
Davillier Law Group, LLC
4105 North 20th Street, Suite 110
Phoenix, AZ 85016
602-730-2985
Email: akolodin@davillierlawgroup.com
rstrassburg@davillierlawgroup.com
vlucero@davillierlawgroup.com
anaeckel@davillierlawgroup.com
phxadmin@davillierlawgroup.com (file copies)
Attorneys for Plaintiffs

Laurin Mills (*pro hac vice*)
Samek | Werther | Mills, LLC
2000 Duke Street
Suite 300
Alexandria, VA 22314
703-547-4693
Email: laurin@samek-law.com

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF ARIZONA

Dr. Michael P. Ward, D.O., *et al.*;
Plaintiffs,

v.

Bennie G. Thompson, *et al.*;
Defendants.

Case No. 3:22-cv-08015-DJH

PLAINTIFFS' RESPONSE TO SELECT COMMITTEE'S MOTION TO DISMISS
(Oral Argument Requested)
(Hon. Diane J. Humetewa)

Defendants ask the Court to let them snoop through the telephone records of a rival political party's state Chair so that they can identify every person who called or texted her during the three-month period surrounding the last presidential election. Defendants seek access to data that would allow them to determine which organizations and grass-roots activists have a direct line to a swing-state party Chair in a presidential election year. If the Court accepts Defendants' argument, it will chill both Arizona citizens and aligned interest groups from working with or joining the Arizona Republican Party in the future because every person Chair Ward was in touch with will be placed at risk for a call or visit from federal Government investigators who are sharing information with a parallel DOJ criminal investigation into the same subject matter.

The chill is compounded by the fact that some citizens who called Dr. Ward are her patients and not political partisans. Dr. Ward's failure to protect her patients' privacy rights would be a violation of HIPAA. Congressional investigators, however, take the

position that a law that Congress passed is no constraint on their investigation.

The laws do, however, apply to Congress. They prevent the subpoena from being enforced here for at least three reasons: (1) Courts have the power to adjudicate the lawfulness of a Congressional subpoena; (2) the Select Committee ignored LRCiv 12.1(c) – rendering this Court unable to hear or decide any of their 12(b)(6) arguments; and (3) Plaintiffs have stated plausible claims. Firstly, that enforcement of the subpoena would violate the First Amendment by striking at the heart of the ability of Chairwoman Ward and the AZGOP's members to associate freely with the Republican Party and ideologically aligned interest groups and actors. And secondly, that enforcement of the subpoena would violate HIPAA. For just as the Committee has ignored this Court's rules of procedure, it has ignored Congress's own laws regarding when, by whom, under what circumstances, and with what preconditions HIPAA protected patient health information ("PHI") can be obtained by third-parties. Defendants' Motion must be DENIED.

FACTS

On January 24, the Select Committee ("Committee") issued a subpoena for Plaintiffs' phone records. Compl. Ex. 1 (ECF Doc. 1-1). The subpoena indiscriminately sought production of all call data records from any line on Plaintiffs' account for the period November 1, 2020-January 31, 2021. *Id*. In other words, investigators wanted to know everyone who called or texted Plaintiffs and when and for how long they communicated.[1]

Dr. Kelli Ward is both the chairwoman of the Arizona Republican Party ("AZGOP") and a practicing physician. Compl. ¶¶ 7, 49 (ECF Doc. 1). Her practice is exclusively in the field of medical weight loss. *Id*. ¶ 21. During the applicable time-period she used her Mole Medical line to conduct telemedicine visits, converse with her patients, talk to her family and friends, and for calls of a political nature. *Id*. ¶¶ 23-24.

The Committee is chaired by Dr. Ward's political rivals. *Id*. ¶ 52. If the subpoena is complied with, the personal telephone numbers and other contact details of patients,

[1] The Committee has recently withdrawn those portions of its subpoena that related to Dr. Michael Ward's line as well as the Wards' childrens' phone line. Mot. n.8.

family members, friends, and the party members most in communication with the AZGOP chair during one of the most contentious periods of modern political history would end up in the hands of these rivals who are criminalizing political participation. Every contact of Dr. Ward's is at risk for a call from Committee investigators or even a visit from the FBI.

ARGUMENT

I. This Court has jurisdiction over this action.

a. This Court has inherent authority to quash a congressional subpoena.

It is settled law that the authority that Courts possess to quash or modify grand jury subpoenas extends to congressional subpoenas. *Trump v. Mazars USA*, LLP, 39 F.4th 774, 787 (D.C. Cir. 2022) (holding that court had jurisdiction to consider President Trump's challenge to congressional subpoena). Indeed, as established by the Supreme Court during the McCarthy era, congressional subpoenas are susceptible to challenge in federal court on several grounds. These include that the subpoena is being "used to inquire into private affairs unrelated to a valid legislative purpose", that the subpoena extends "to an area in which Congress is forbidden to legislate," that the subpoena has been issued for "law enforcement" purposes," and that the subpoena violates one of the "specific individual guarantees of the Bill of Rights[.]" *Quinn v. United States*, 349 U.S. 155, 161 (1955). "The court must quash or modify the [Congressional] subpoena if it determines that the subpoena 'requires disclosure of privileged or other protected matter.'" *Comm. on the Judiciary of the United States House of Representatives v. McGahn*, 449 U.S. App. D.C. 1, 18 (2020) (citing Fed. R. Civ. P. 45(d)(3)); *Accord Trump v. Mazars USA, LLP*, 140 S. Ct. 2019, 2032 (2020) ("*Mazars*") (explaining that these include constitutional privileges).

b. Alternatively, this Court has jurisdiction to quash or modify the subpoena because T-Mobile is a party.

In *Republican Nat'l Comm. v Pelosi*, a case that concerned a subpoena by the Committee to Salesforce (one of the RNC's email vendors), the District Court for the District of Columbia agreed with the Committee that individual members of Congress were generally immune from suit but resolved the quandary by "assum[ing] without deciding" that it could treat Salesforce "as a state actor" for the purposes of the dispute

over the legality of the subpoena, *Republican Nat'l Comm. v. Pelosi*, Civil Action No. 22-659 (TJK), 2022 U.S. Dist. LEXIS 78501, at *37 (D.D.C. May 1, 2022), and therefore hear and decide the underlying claims. Similarly, in *Mazars*, the Supreme Court had no compunction about asserting jurisdiction to decide the constitutionality of a congressional subpoena for President Trump's papers where the subpoena was directed against his accounting firm, Mazars USA, LLP, and Mazars was a party. *Mazars* at 2027-28; *see also Eastman v. Thompson*, No. 8:22-cv-00099-DOC-DFM, 2022 U.S. Dist. LEXIS 59283, at *43-44 (C.D. Cal. Mar. 28, 2022) (no jurisdictional issue with suit to quash subpoena by Committee where the entity holding the emails was also made a party).[2]

Here, Plaintiffs have likewise alleged that **both** "[t]he production of these documents by T-Mobile concerning the Phone Number, **and** the Subpoena upon which this production would be based" violate the First (and Fourteenth) Amendments, as well as the Physician/Patient privilege and other applicable laws such as HIPAA, *see, e.g.,* Compl. ¶ 4 (ECF Doc. 1), and have made T-Mobile a party.[3]

In any event, the *Pelosi* Court assumed correctly. Courts treat private parties as state actors, subject to the same claims that may be asserted against the government if it were a party, under four different sets of circumstances: (1) they are exercising a public function; (2) they are engaging in joint action with the government; (3) governmental compulsion or coercion is present; or (4) there is a governmental nexus. *Sutton v. Providence St. Joseph Med. Ctr.*, 192 F.3d 826, 835-36 (9th Cir. 1999); *see also Brown v. Philip Morris, Inc.*, 250 F.3d 789, 801 (3d Cir. 2001) (using same analysis to determine whether a private party is a "federal actor"). Governmental compulsion or coercion is

---

[2] Although the issue of jurisdiction was not addressed in *Mazars* or *Eastman*, courts are required to act *sua sponte* to dismiss suits if jurisdiction is lacking. *Crowley v. Bannister*, 734 F.3d 967, 974 (9th Cir. 2013).

[3] Defendants state their Second through Fourth causes of action directly against the Committee and not against T-Mobile. However, to the extent this is relevant, this is easily curable by amendment. It would also have been curable had the Committee simply followed LrCiv 12.1(c) and met and conferred with Plaintiffs prior to filing their Motion to Dismiss.

obviously present here, as is a governmental nexus. T-Mobile was served with a congressional subpoena commanding it to produce Plaintiffs' documents, which but for that action it would not be compelled to produce. Thus, the Committee's attempt to invite the Court to reach the unnecessary question of whether the federal courts have jurisdiction over the Committee is a red-herring. *See Mills v. Rogers*, 457 U.S. 291, 305 (1982) ("It is this Court's settled policy to avoid unnecessary decisions of constitutional issue.").

II. The Committee failed to comply with LRCiv 12.1(c).

For seven months, the Committee has been seeking extensions to respond to the Complaint, but not once in that time did the Committee seek to work with Defendants concerning the scope of the subpoena. LRCiv 12.1(c) provides that Rule 12(b)(6) motions may **not** "be considered or decided unless the moving party includes a certification that, before filing the motion, the movant notified the opposing party of the issues asserted in the motion and the parties were unable to agree that the pleading was curable in any part by a permissible amendment[.]" The Committee has repeatedly averred to this Court and the parties that it was "actively engaged in studying the various alternatives in this and other related litigation." *See e.g.*, (ECF Doc. 30) pg. 2:20-26, (ECF Doc. 32) pg. 2:24-3:6, (ECF Doc. 35) pg. 3:1-4. How then were Plaintiffs to know that they even planned to file a motion to dismiss, much less what the contents would be? Certainly not because the Committee met its conferral obligation - the Committee failed to attach the certification required by LRCiv 12.1(c) to its Motion. Thus, the only matter that can currently be heard or decided is the question of this Court's jurisdiction.

III. Though not properly before the Court, Plaintiffs have stated plausible claims.

a. Standard on a motion to dismiss pursuant to 12(b)(6).

To survive a Rule 12(b)(6) motion, a complaint must contain factual allegations sufficient to "raise a right to relief above the speculative level." *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555 (2007). The task "is to evaluate whether the claims alleged [plausibly] can be asserted as a matter of law." *Adams v. Johnson*, 355 F.3d 1179, 1183 (9th Cir. 2004); *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009).

b. The Complaint states a plausible First Amendment claim.

Plaintiffs' Second Cause of Action alleges that the Subpoena violates their core First Amendment rights to associate with others for political purposes. An individual's First Amendment freedoms include a "correlative freedom to engage in group effort toward those ends." *Roberts v. United States Jaycees*, 468 U.S. 609, 622 (1984), *see also Perry v. Schwarzenegger*, 591 F.3d 1147, 1160 (9th Cir. 2010) (recognizing a First Amendment privilege). "Implicit in the right to engage in activities protected by the First Amendment [is] a corresponding right to associate with others in pursuit of a wide variety of political, social, economic, educational, religious, and cultural ends." *Brock v. Local 375, Plumbers Int'l Union*, 860 F.2d 346, 349 (9th Cir. 1988) (internal citations omitted). The "inviolability of privacy in group association" is in many circumstances "indispensable to [the] preservation of freedom of association, particularly where a group espouses dissident beliefs." *NAACP v. Alabama*, 357 U.S. 449, 462 (1958). This is one of those circumstances.

When such core political associational rights are at stake, the Court must apply what the Supreme Court calls "exacting scrutiny." Exacting scrutiny requires that there be "a substantial relation between the disclosure requirement and a sufficiently important governmental interest, and that the disclosure be narrowly tailored to the interest it promotes." *Ams. for Prosperity Found. v. Bonta*, 141 S. Ct. 2373, 2385 (2021). The exacting scrutiny standard applies when a litigant seeks to quash a congressional subpoena on First Amendment grounds. *See Pelosi* at *63 (citing *Bonta* at 2383, 2387). To satisfy this test, the subpoena must not cause an "unnecessary and unreasonable dissipation of precious constitutional freedoms" and "the investigative demand should not be substantially overbroad, meaning that a substantial portion of the information sought does not serve to advance the investigative goals." *Pelosi* at *62-63 (citing *Ward v. Rock Against Racism*, 491 U.S. 781, 799 (1989), *Watkins v. United States* 354 U.S. 178, 204

(1957), *Bonta* at 2386).[4] The inquiry must be "calibrated to fit the distinct issues raised in the context of each case." *Id.* (citing *Grutter v. Bollinger*, 539 U.S. 306, 333-34 (2003)).

The Committee argues that, since the Subpoena only seeks call detail records, it does not implicate any associational activities of Plaintiffs or their associates and exacting scrutiny does not apply. Br. 14 (ECF Doc. 46). It further contends that, even if the Subpoena were subject to the exacting scrutiny standard, Plaintiffs' First Amendment associational concerns are too speculative and the investigative needs of the Committee outweigh those concerns. *Id*. at 15. Neither of those arguments is well taken and, even were it otherwise, the Committee fails to address the "narrow tailoring" requirement.

i. The Subpoena plainly impacts Plaintiffs' associational rights.

The Committee's contention that a Subpoena that seeks the call detail records of the Chair of the AZGOP during the time of a contested presidential election does not implicate her First Amendment rights of political association is not a serious argument. "That telephone metadata do not directly reveal the content of telephone calls . . . does not vitiate the privacy concerns" arising out of the collection. *ACLU v. Clapper*, 785 F.3d 787, 794 (2d Cir. 2015). Telephone "[m]etadata can reveal civil, political, or religious affiliations . . . an individual's social status, or whether and when he or she is involved in intimate relationships." *Id*. Accordingly, "[w]hen the government collects [a party's telephone] metadata [the party and their associates'] interests in keeping their associations and contacts private are implicated, and any potential 'chilling effect' is created at that point." *Id*. 802-03 (going on to note that such collection confers standing to assert First Amendment claims). Indeed, Chairman Thompson, in a press release, has made plain that its investigation directly implicates associational activities, saying: "The inquiry includes examination of how various individuals and entities coordinated their activities leading up

[4] Internal quotations omitted from all citations in this paragraph.

to the events of January 6, 2021."[5] Another press release by Chairman Thompson explains that the Committee's stated goals are "Accountability under the law. Accountability to the American people. Accountability at every level: [down to] the local precincts in many states where Donald Trump and his allies attacked election workers for just doing their jobs."[6] *See Hutchinson v. Proxmire*, 443 U.S. 111, 133 (1979) ("newsletters and press releases" are "not entitled to the protection of the Speech or Debate Clause[.]"). In other words, the Committee seeks to expose and target not just the AZGOP's Chair, but its rank-and-file members and the scope of the inquiry goes far beyond the events in Washington on January 6. Not only are Plaintiffs' First Amendment rights implicated, the political associational rights of the entire AZGOP are at stake.

The *Pelosi* Court, in its analysis, pointed out the difference between the documents sought by the Committee's subpoena in that case and those sought in *Bonta*, saying "the Court there considered a challenge to a California regulation requiring tax-exempt organizations to disclose to the state **the names and addresses of certain donors**—information, unlike that here, that could directly chill individual associational rights." *Pelosi* at *58 (citing *Bonta* at 2380, 2387-88) (emphasis supplied). As set forth in section III(c) below, the records that the Committee seeks contain information that can easily be used to find the names and addresses of everyone who called Dr. Ward during a months'-long period of time. Thus, the information that the Committee seeks is exactly the sort of information that the *Bonta* court held presents the greatest threat of associational chilling.

Though *Pelosi* ultimately rejected the RNC's attempt to quash the Committee's subpoena on First Amendment grounds, it noted that "the RNC identified important First Amendment interests implicated by the subpoena **that would have presented a much**

---

[5] Press release available at: https://january6th.house.gov/news/press-releases/select-committee-subpoenas-organizers-rallies-and-events-preceding-january-6th (last accessed Aug. 22, 2022).

[6] Press release available at: https://january6th.house.gov/news/press-releases/thompson-cheney-luria-kinzinger-opening-statements-select-committee-hearing (last accessed Aug. 18, 2022).

different question for the Court had the materials at issue not been narrowed after discussions between the Select Committee and Salesforce." *Id* at *20 (emphasis supplied). *Pelosi* placed great weight on the fact that after negotiations, the Committee agreed not to seek "any disaggregated information about any of the RNC's donors, volunteers, or email recipients, including any person's personally identifiable information." *Id*. *25-26. Here, the Subpoena seeks disaggregated and personally identifiable information about everyone who called or texted Dr. Ward during a span of many months.

Public participation in politics is the life blood of our democracy. The criminalization of political activity, if not carefully constrained by the courts, will force legitimate political actors from the field. If the Committee gains access to Plaintiffs' call data records, virtually everyone Chair Ward talked to during the relevant time period is at risk to be contacted by Committee (or FBI) investigators and they will become implicated in the largest criminal investigation in U.S. history solely by virtue of the fact that they were in contact with the party Chair.[7] The chilling effect of that precedent on public participation in politics is palpable. The fact that the Committee is controlled by members of the rival political party, along with the existence of a parallel criminal investigation, also raises the legitimate concern that the Committee will use any information it obtains to harass or persecute political rivals by inquiring into their dealings with the party Chair.

In *NAACP v. Alabama*, Alabama sought the compelled disclosure of the NAACP's membership list. That compelled disclosure was significantly less intrusive than what the Committee seeks here. If the Committee prevails, it will get a list of who, when, and for how long the Chair of the AZGOP was in contact with party members at a sensitive time. This is exactly the sort of thing that may "induce members to withdraw" from the AZGOP "and dissuade others from joining it because of fear of exposure of their beliefs shown through their associations and of the consequences of this exposure." *NAACP* at 463

---

[7] Though the Committee might argue that whether such individuals will be contacted is speculative, how else is the Committee to know which individuals in the call records are political actors, which family and friends, and which patients?

(finding compelled disclosure under these circumstances to violate the First Amendment). Those in control of Alabama during the Jim Crow era would have drooled over the possibility of accessing such a trove of information about those they sought to persecute.

ii. Plaintiffs' First Amendment associational rights override the Committee's investigative needs for the information sought.

Dr. Kelli Ward is the Chair of the AZGOP. It is her job to contact and coordinate with members of her party and associated interest groups about elections. Her responsibilities are especially acute when there is public controversy concerning the outcome of a presidential election. Such a controversy was raging in Arizona (and nationally) during the time-period covered by the Subpoena.

Here, there is no dispute that Plaintiffs perceived that there were "issues" with the 2020 presidential election results in Arizona and elsewhere and they acted to send an alternate slate of electors to Washington in the event that the legal challenges to the Arizona results succeeded. Their beliefs and the actions they took were no secret. Their vote was posted on YouTube and Dr. Kelli Ward wrote a book about it. https://www.amazon.com/Justified-Americas-Dr-Kelli-Ward/dp/195725503X (last accessed August 14, 2022). The connection between that action, which took place on December 14, 2020, and the riot at the Capitol on January 6, is far from obvious. Nevertheless, the Committee has swept any actor who had concerns about the 2020 presidential election into its causative narrative about January 6.[8]

The Subpoena seeks to discover with whom Plaintiffs communicated about 2020 presidential election concerns. That will inevitably lead to the questioning of, and further subpoenas issued to, the thousands of Republicans in contact with Plaintiffs. If the Subpoena is not quashed, members of the AZGOP will be made to feel that every time they communicate with party leadership, they risk having those communications disclosed to law enforcement followed by a knock on the door (or worse) from federal investigators. A stronger risk of associational chilling can scarcely be imagined.

---

[8] All assertions made in this paragraph are for the purposes of the instant motion only.

The idea that there were legitimate concerns about the 2020 election results is considered abhorrent by many. That fact, however, has no bearing on whether Plaintiffs are entitled to the protections of the First Amendment. "The hallmark of the protection of free speech is to allow 'free trade in ideas'—even ideas that the overwhelming majority of people might find distasteful or discomforting." *Virginia v. Black*, 538 U.S. 343, 358 (2003) (quoting *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting)). Similarly, whether Plaintiffs' views are correct has no bearing on their entitlement to First Amendment protection. *Stanley v. Georgia*, 394 U.S. 557, 564 (1969) (citations omitted) ("It is now well established that the Constitution protects the right to receive information and ideas.... This right to receive information and ideas, regardless of their social worth, is fundamental to our free society.").

Neither the controversial nature of Plaintiffs' views nor their merit gives the Committee the right to disregard the core First Amendment rights of those who believe there was election fraud in 2020 and who engaged in political actions and discussions based on that belief. Yet, that is exactly what the Subpoena is designed to do. Allowing the Subpoena will forever chill the ability of partisan political actors to discuss and/or coordinate political activities in the wake of a close election. The attempt to criminalize partisan political activities via an intrusive investigation of political actors poses a greater threat to our democracy than the Capitol Riot.

The Committee advances no argument as to why any of the subpoenaed information is particularly important to its investigation (much the less why it requires **every single person** who called or texted Dr. Ward during a prolonged period of time to be identified). It fails to explain why it cannot obtain the material it needs from other sources, including publicly available sources, such as Dr. Ward's book. The Committee's argument is that the investigation itself is "important." Therefore, anything the Committee seeks is justified. This hardly satisfies exacting scrutiny and provides no basis for dismissing Plaintiffs' First Amendment claim.

c. The Subpoena violates HIPAA.

Congress is bound by its own laws. *See Wilkinson v. Legal Servs. Corp.*, 27 F. Supp. 2d 32, 48 (D.D.C. 1998) ("[F]ounding principle of this Republic" and requirement of the Due Process clause that all government officials are bound by the law). "Congress has spoken on the privacy of medical records through HIPAA." *Nat'l Abortion Fed'n v. Ashcroft*, 2004 U.S. Dist. LEXIS 4530, at *19-20 (S.D.N.Y. Mar. 18, 2004). The Committee does not dispute that, if HIPAA applies, then the Subpoena cannot be enforced. Instead, it argues that HIPAA does not apply at all for two reasons.

First, the Committee argues that HIPAA does not apply because the information sought in the Subpoena does not constitute PHI within the meaning of HIPAA. Br. 17:9-14. Second, it argues that "HIPAA's disclosure restrictions do not apply to this subpoena because neither the entity from which the records were sought – T-Mobile . . . nor the Select Committee . . . fit within HIPAA's definition of 'covered entity.'" *Id*. 16:24-28.

Those assertions are wrong. The telephone numbers of Dr. Ward's patients can easily be used to identify them. For this reason, HIPAA presumes patient telephone numbers to be PHI in the absence of an expert opinion otherwise. And HIPAA governs the disclosure, not because Defendants are covered entities but because Plaintiffs are.

i. The Subpoena seeks PHI.

When a patient seeking treatment for medical weight loss calls their doctor, the last thing they expect to happen is for the record of that call to be reviewed by a congressional investigator. In the absence of a protective order, such a patient might face uncomfortable questions from friends, reporters, and the public about why they are listed. For example, such patients could face the uncomfortable choice of admitting that they were seeking treatment for medical weight loss or living with the implication that they might have been partially responsible for the Capitol Riot.

Fortunately, Plaintiffs' patients need not face such a choice because the patient telephone numbers in T-Mobile's possession constitute "individually identifiable health information" ("PHI") that HIPAA protects from disclosure. PHI is defined as follows:

> *(6) INDIVIDUALLY IDENTIFIABLE HEALTH INFORMATION.— The term 'individually identifiable health information' means any information,*

> *including demographic information collected from an individual, that—*
> *(A) is created or received by a health care provider, health plan, employer, or health care clearinghouse; and*
> *(B) relates to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual, and—*
> *(i) identifies the individual; or*
> *(ii) with respect to which there is a reasonable basis to believe that the information can be used to identify the individual.*

42 U.S.C. § 1320d(6); *see also Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, HHS.gov ("HHS Guidance")[9] (phone numbers that "indicat[e] that the individual was treated at a certain clinic" are PHI). Here, there is a "reasonable basis" to believe that the telephone numbers of patients in T-Mobile's "metadata" files can be used to identify individual patients and expose their personal, medical information for at least three reasons.

First, it is a trivial task to look up the name of a caller once someone knows their telephone number. For example, performing a Google search for undersigned counsel's cellphone number yields, on the first page of results, a bevy of information from FastPeopleSearch.com including his name, address, previous addresses, the name of his wireless carrier, email addresses, aliases, relatives, and known associates. **Exh. 1**. If users wish to learn more, they are invited to purchase a "Full Background Report". Indeed, telephone metadata implicates a variety of privacy concerns. As reported in Mayer, et al., "Evaluating the privacy properties of telephone metadata" (2016), computer science researchers at the Security Laboratory in the Department of Computer Science at Stanford University demonstrated that telephone metadata can be readily reidentified to reveal the identity of an individual, like a patient, and can be used to discover personal health information. **Exh. 2** (the "Mayer Study"). The Mayer Study found that, "[T]elephone

[9] Available at https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html (last accessed Aug. 18, 2022).

numbers are trivially reidentifiable." *Id*. at 5538. After selecting a sample size of 30,000 phone numbers, the researchers found that searches using free, public interfaces matched identities for 32% (9,576) of the numbers with individuals. Using both manual and automated searches, the researchers were able to match 82% of the numbers with individuals. *Id*. at 5540, Table 2. Moreover, the researchers were able to connect individuals with a number of health organizations whose specialties disclosed the nature of the individual's health. *Id*. at Table 5. The researchers concluded that "Telephone metadata is densely interconnected, easily reidentifiable, and trivially gives rise to location, relationship, and sensitive inferences." *Id*. Additionally, it seems blindingly obvious that the file name of an attachment sent to Dr. Ward by a patient could constitute PHI (*e.g.*, a file name like "photo of my belly fat").

Second, metadata identifying patient telephone numbers calling the Wards' medical practice identifies the nature of the patient's health condition because of Dr. Kelli Ward's single specialty of weight loss, no less than the telephone numbers of patients calling an oncologist would similarly disclose the nature of the patient's cancer condition. *See Clapper* at 794 (noting that the fact of a call itself to a single-specialty provider may reveal whether someone is "a victim of domestic violence or rape; a veteran; suffering from an addiction of one type or another; [or] contemplating suicide[.]").

Third, in the absence of an expert determination otherwise, HIPAA **presumes** that patient telephone numbers are PHI and requires that they be redacted prior to disclosure to comply with HIPAA's "safe harbor" requirements for de-identified medical information. *See* HHS Guidance (citing 45 CFR § 164.514). There are few restrictions on the use or disclosure of de-identified health information. *See* 45 C.F.R. §§ 164.502(d)(2), 164.514(a) and (b). The de-identification safe harbor rule requires removal of metadata of the kind sought by the Committee here: The safe harbor rule (45 C.F.R. § 164.514(b)(2)(i)(A)-(Q)) requires removal of certain items of information--items that would appear in the metadata in the possession of T-Mobile and sought by the Committee – including telephone numbers, fax numbers, names, addresses, zip codes, email addresses, and internet protocol

(IP) addresses. Since the T-Mobile metadata includes such information, it could not qualify for the "safe harbor" of de-identified information protected by HIPAA.

ii. HIPAA applies because the Committee has subpoenaed information from a covered entity.

"HIPAA provisions provide for disclosure of medical information in the course of a judicial proceeding, but certain requirements are placed on the provider **and the party seeking the information**." *Montoya v. Arizona*, No. CV 18-08025-PCT-DGC (ESW), 2019 U.S. Dist. LEXIS 172561, at *2-3 (D. Ariz. Oct. 4, 2019) (emphasis supplied); *Accord Pyankovska v. Abid*, No. 2:16-cv-02942-JCM-PAL, 2018 U.S. Dist. LEXIS 233418, at *17-18 (D. Nev. Oct. 16, 2018), *Crenshaw v. Mony Life Ins. Co.*, 318 F. Supp. 2d 1015, 1028-29 (S.D. Cal. 2004). Similarly, when the government subpoenas PHI from a covered entity and no state law privilege applies, "[t]he privacy provisions promulgated under HIPAA . . . control the enforceability of the subpoena." *Nat'l Abortion Fed'n* at *19-20; *see also Crenshaw v. Mony Life Ins. Co.*, 318 F. Supp. 2d 1015, 1028 (S.D. Cal. 2004). "HIPAA, through its implementing regulations, speaks directly to the privilege" in this context. *Nat'l Abortion Fed'n* at *20.

HIPAA requires that the party serving the subpoena to either (a) obtain patient consent or (b) seek a qualified protective order. *Nat'l Abortion Fed'n* at *21-22 (citing 45 C.F.R. § 164.512(e)); *Accord Henderson v. Cty. of Santa Cruz*, No. 14-cv-03544-WHO, 2020 U.S. Dist. LEXIS 60271, at *2-3 (N.D. Cal. Apr. 6, 2020), *Montoya* at *2-3, *Pyankovska* at *17-18, *Kelso v. Redding Police Dep't*, No. 2:11-cv-1960-KJM-CMK, 2012 U.S. Dist. LEXIS 178250, at *3-4 (E.D. Cal. Dec. 17, 2012), *Crenshaw* at 1028-29. A qualified protective order must (a) prohibit the parties from using or disclosing the protected health information from any purpose other than the litigation or proceeding for which such information was requested and (b) require the return to the covered entity or destruction of the protected health information (including all copies made) at the end of the litigation or proceeding. *Henderson* at *2-3, *Crenshaw* at 1028-29, *Pyankovska* at *17-18, *Kelso* at *3-4. Though the Committee was placed on notice of this at the outset of this case, Ps.' Mot. to Quash 9:12-24 (ECF Doc. 2), it has neither disputed that these are the

requirements if HIPAA applies nor moved to secure a qualified protective order.

The Committee instead simply argues that HIPAA does not apply here because neither it nor T-Mobile "fit within HIPAA's definition of a covered entity." Mot. 16:24-28. However, as set forth in Section I(b), above, T-Mobile is a state actor in this context. Whether **it** is a "covered entity" is irrelevant. A state actor is simply treated as if it were the state itself. *Jackson v. Metro. Edison Co.*, 419 U.S. 345, 351 (1974). And, even were T-Mobile not a state actor, litigants may still assert their privileges in situations where subpoenas have been served to non-parties. *Orthoflex, Inc. v. ThermoTek, Inc.*, No. 12-MC-00013-PHX-JAT, 2012 U.S. Dist. LEXIS 42417, at *4 (D. Ariz. Mar. 28, 2012).

In harmony with these concepts, *Eastman* rejected the "sweeping proposition" that using a third-party communications provider that complies with subpoenas constitutes a waiver of applicable privileges. *Eastman* at *43-44. Thus, while the Committee correctly identifies itself as the party seeking the information, it fails to properly identify Plaintiffs as the true parties from whom the information is sought. Because Plaintiffs **are** a covered entity, HIPAA applies.

HIPAA's application, on its own, simply requires the Committee to secure a qualified protective order. Such a protective order would not impede the Committee's investigation. Rather, it would prevent the Committee from using the information received for anything **other** than its investigation (and prohibit it from leaking the PHI to the public or other government entities). To be sure, it would preclude the Committee from sharing PHI with the Justice Department, but "Congress may not issue a subpoena for the purpose of law enforcement" anyway "because those powers are assigned under our Constitution to the Executive and the Judiciary." *Mazars* at 2032 (quotations omitted).[10]

d. The Committee's other 12(b)(6) arguments fail.

The Committee, relying on *Trump v. Thompson* (and oral argument transcripts from other cases) argues that "four other courts" have rejected arguments that the subpoenas

[10] If the DOJ requires access to this information, HIPAA contains specific previsions by which law enforcement may subpoena PHI under limited conditions.

issued by the Committee lacked a legitimate legislative purpose. Br. 6. But Thompson did not overrule *Watkins's* clear command that, where challenged, Congressional subpoenas must be shown, with "undisputable clarity," to relate to an authorized and lawful purpose of the Committee's legislative investigation. *See Watkins* at 214-15 (going on to state "[t]o be meaningful, the explanation must describe what the topic under inquiry is and the connective reasoning whereby the precise questions asked relate to it."). To the contrary, the *Thompson* Court relied heavily on *Watkins*. *Thompson*, 20 F.4th at 24-25. *Thompson's* reason for finding the valid legislative purpose test was satisfied was that the Committee provided "detailed and substantial evidence" in that case of its "specific need" for the records it was seeking. *Id.* at 42. What's more, in *Thompson*, the Committee made a far more specific request, seeking only presidential records pertaining to the events of January 6th, the former President's claims of election fraud, and other, related, items. *Id*. at 16.

Here, by contrast, the Committee has made, at best, vague assertions that all Plaintiffs' phone records for a three-month period of time are required for its investigation. This falls far short of "substantial evidence" of a "specific need". The Committee was, indeed, so scattershot that, originally, it mistakenly subpoenaed the records for Dr. Ward's husband and children as well.

Additionally, Plaintiffs continue to maintain that the Committee was not constituted and does not operate in conformity with the House rules for the reasons stated in the Complaint and Motion to Quash and the Committee has advanced no reason for overlooking the plain text of the rules. As set forth above, neither this, nor any other 12(b)(6) issue is yet properly before the Court. Given the limitations of space, Plaintiffs will address these, and the Committee's other arguments, at the appropriate time and after any necessary refinement subsequent to conferral.

CONCLUSION

For the forgoing reasons, the Motion to Dismiss should be DENIED or, alternatively, leave to amend subsequent to conferral should be granted.

Respectfully submitted this 22nd day of August 2022

/s/ *Alexander Kolodin*
Alexander Kolodin
Roger Strassburg
Veronica Lucero
Arno T. Naeckel
Davillier Law Group, LLC
4105 North 20th Street
Suite 110
Phoenix, AZ 85016

/s/ *Laurin Mills*
Laurin Mills
SAMEK | WERTHER | MILLS, LLC
2000 Duke Street
Suite 300
Alexandria, VA 22314
(*Pro hac vice*)

CERTIFICATE OF SERVICE

I certify that on August 22, 2022 I electronically transmitted the attached document to the Clerk's Office using the CM/ECF System for filing, which electronically sends a copy to be served on all registered parties.

/s/ Alexander Kolodin

Exhibit 1

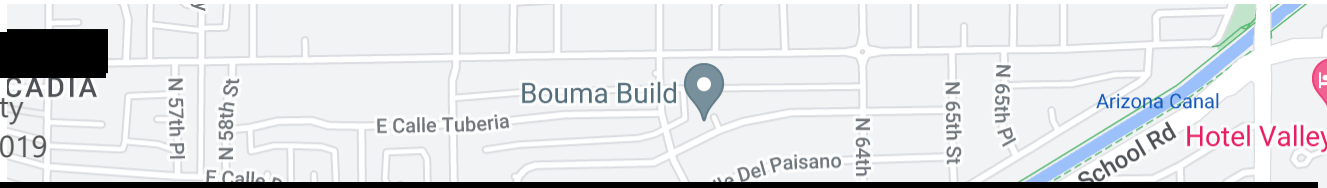
Alexander Kolodin in Scottsdale, AZ (Arizona)

Age 36

Current Address

[redacted]

Maricopa County
Since January 2019



[redacted]

Full Name:
Alexander Kolodin

Phone Numbers
for **Alexander Kolodin** in **Scottsdale, AZ**

[redacted] [redacted]
Wireless
[redacted]
First reported August 2015

[redacted]
Landline
[redacted]
First reported June 2013

[redacted]
Wireless
[redacted]
First reported November 2009

[redacted]
Landline
[redacted]
First reported July 2017

[redacted]
Landline
[redacted]
First reported October 2019

[redacted]
Landline
[redacted]
First reported October 2019

@ Email Addresses
for **Alexander Kolodin** in **Scottsdale, AZ**

[redacted]

Also Known As

Alexander M Kolodin

Alexander M Kollodin

Michael Del Rey Kolodin Alexander

Alex Kolodin

Previous Addresses
used by **Alexander Kolodin**

[redacted]
Maricopa County
Recorded February 2018
[redacted]

[redacted]

[redacted]
Maricopa County
Recorded July 2010
[redacted]

[redacted]
Maricopa County
Recorded December 2013
[redacted]

[redacted]
Maricopa County
Recorded November 2009
[redacted]

[redacted]
Maricopa County
Recorded November 2009
[redacted]

[redacted]
Maricopa County
Recorded December 2003

[redacted]
Maricopa County
Recorded November 2009
[redacted]

[redacted]
Maricopa County
Recorded May 2005

[redacted]
District Of Columbia County
Recorded September 2005

[redacted]

Is Alexander Kolodin Currently Married?

We can not find any public records stating that Alexander Kolodin is currently Married. It is not likely.

FREE Background Report for Alexander Kolodin in Scottsdale, AZ (Arizona)

Alexander Kolodin is 36 years old. Currently Alexander lives at the address [redacted]
Alexander has lived at this **Scottsdale, AZ** address for about 3 years, after moving in around January of 2019. Alexander previously lived at [redacted] for 2 years, starting in February of 2018. Going further back, starting in June of 2017, Alexander lived at [redacted].

Public records do not indicate that Alexander Kolodin is currently married. The following people are relatives or close associates of Alexander: [redacted]

Alexander's current phone number is [redacted] This Wireless number was issued by [redacted] first reported in public records on August of 2015. Past phone numbers for Alexander include [redacted] The primary email address for Alexander is [redacted]
Alexander has also used the following email accounts: and [redacted]

Alexander Kolodin's current address is [redacted]. Alexander has lived there for about 3 years, since January of 2019.

Who is related to Alexander Kolodin?

Alexander Kolodin is likely related to the following people: [redacted]

What is the best phone number for Alexander Kolodin?

Alexander Kolodin's latest phone number is a wireless number at [redacted]

What is the best email for Alexander Kolodin?

[redacted] is the most current email on record for Alexander Kolodin.

Is Alexander Kolodin alive today?

Yes! Alexander Kolodin is living today.

Does Alexander Kolodin go by any other names or aliases?

Alexander Kolodin may also go by the following names or aliases: Alexander M Kolodin, Alexander M Kollodin, Michael Del Rey Kolodin Alexander, Alex Kolodin

Who does Alexander Kolodin associate with?

The following people are friends, co-workers, partners, roomates, or otherwise associated with Alexander Kolodin: [redacted]

Where did Alexander Kolodin live previously?

Alexander Kolodin was registered, and likely lived at the following addresses in the past: | [redacted]

What email addresses have been used by Alexander Kolodin?

Alexander Kolodin has used the following email addresses: [redacted]

Name Directory:

A B C D E F G H I J K L M N O P Q R S T U V W

X Y Z

Phone Directory: 1 2 3 4 5 6 7 8 9

FREE People Search | FREE Reverse Phone Lookup | FREE Address Lookup |

Fast People Search API

Terms | Privacy | Contact

Exhibit 2

Evaluating the privacy properties of telephone metadata

Jonathan Mayer[a,b,1], Patrick Mutchler[a], and John C. Mitchell[a]

[a]Security Laboratory, Department of Computer Science, Stanford University, Stanford, CA 94305; and [b]Stanford Law School, Stanford University, Stanford, CA 94305

Edited by Cynthia Dwork, Microsoft Research Silicon Valley, Mountain View, CA, and approved March 1, 2016 (received for review April 27, 2015)

Since 2013, a stream of disclosures has prompted reconsideration of surveillance law and policy. One of the most controversial principles, both in the United States and abroad, is that communications metadata receives substantially less protection than communications content. Several nations currently collect telephone metadata in bulk, including on their own citizens. In this paper, we attempt to shed light on the privacy properties of telephone metadata. Using a crowdsourcing methodology, we demonstrate that telephone metadata is densely interconnected, can trivially be reidentified, and can be used to draw sensitive inferences.

surveillance | privacy | telephone | metadata | social network

Communications privacy law, in the United States and many other nations, draws a distinction between "content" and "metadata" (1). The former category reflects the substance of an electronic communication; the latter includes all other information about the communication, such as parties, time, and duration (2).*

When a government agency compels disclosure of content, the agency must usually comply with extensive substantive and procedural safeguards. Demands for metadata, by contrast, are often left to the near-total discretion of authorities. In the United States, for instance, a law enforcement officer can request telephone calling records with merely a subpoena—essentially a formal letter from the investigating agency (3). An intelligence program by the National Security Agency (NSA) has drawn particular criticism; under the business records provision of the USA PATRIOT Act (4), the agency acquired a substantial share of all domestic telephone metadata (5).†

In this paper, we empirically investigate factual assumptions that undergird policies of differential treatment for content and metadata. Using crowdsourced telephone logs and social networking information, we find that telephone metadata is densely interconnected, susceptible to reidentification, and enables highly sensitive inferences.‡

The balance of the paper is organized into three parts. First, we discuss our data collection methodology and properties of our participant population. We next present our results. Finally, we discuss implications for policy and future quantitative social science research. Additional methodological detail and figures are available in the *Supporting Information*.

Methods

We collected the data in this study through an Android smartphone application (Fig. 1).§ Potential participants could discover the project through academic websites, the Google Play store, and references in media coverage. The application automatically retrieved historical call and text message [Short Message Service (SMS)] metadata from device logs.¶ In addition, the application retrieved information from a participant's Facebook account, to be used as ground truth for potential inferences.# Participants were provided an opportunity to view individualized features of their phone metadata, and then they were invited to uninstall the application. In total, 823 participants volunteered their metadata, which included 251,788 calls and 1,234,231 text messages. The *Supporting Information* provides additional detail on data sources and dataset properties (Figs. S1–S5 and *1. Dataset Methodology*, *1.1. Data Collection*, *1.2. Participants*, *1.3. Logs*, and *1.4. Sampling Bias*).

Ethical Considerations. Given the quantity and sensitivity of the data associated with this project, we instituted several informed consent mechanisms. Participants received extensive disclosure notices, both in the application and on the study website. In addition, the Facebook software library notified participants of the categories of social network information that the application was requesting. Each screen of the application, until information upload was complete, provided participants with an opportunity to withdraw. Furthermore, participants were furnished contact information for research staff such that they could request deletion of their information after using the application. The university institutional review board suggested helpful methodological refinements, and we began collecting data only after receiving the board's approval.

We also took a number of security precautions to safeguard participant information. Our application transmitted information to a cloud storage service only over an encrypted and authenticated connection [transport layer security (TLS)], and we retrieved information only over TLS. Credentials for accessing the data were restricted to the research team, and once the data were retrieved, the data were stored on encrypted devices at academic facilities.

Dataset. We provide a detailed treatment of our dataset in the *Supporting Information*. We note here, importantly, that our crowdsourced dataset is not a

Significance

Privacy protections against government surveillance are often scoped to communications content and exclude communications metadata. In the United States, the National Security Agency operated a particularly controversial program, collecting bulk telephone metadata nationwide. We investigate the privacy properties of telephone metadata to assess the impact of policies that distinguish between content and metadata. We find that telephone metadata is densely interconnected, can trivially be reidentified, enables automated location and relationship inferences, and can be used to determine highly sensitive traits.

Author contributions: J.M., P.M., and J.C.M. designed research; J.M. and P.M. performed research; J.M. and P.M. contributed new analytic tools; J.M. and P.M. analyzed data; and J.M., P.M., and J.C.M. wrote the paper.

Conflict of interest statement: C.D. was jointly funded with J.C.M. on a 3-year Sloan Foundation grant in 2011, on a no-cost extension through 2015.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 5467.

[1]To whom correspondence should be addressed. Email: jmayer@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1508081113/-/DCSupplemental.

*The contours of the content–metadata distinction are well established for telephony and messaging, but are far more elusive for newer forms of communication.

†While this article was in submission, Congress enacted the USA FREEDOM Act (6). Provisions codify the two-hop limit voluntarily imposed by the executive branch, as well as the proposed 18-month-duration limit. Data that are not associated with a query result will also remain with telecommunications services. These changes took effect on November 29, 2015.

‡In the interest of providing timely input on matters of public controversy, we presented our preliminary results in a series of online postings (webpolicy.org/2013/11/27/metaphone-seeing-someone/, webpolicy.org/2013/12/09/metaphone-the-nsa-three-hop/, webpolicy.org/2013/12/23/metaphone-the-nsas-got-your-number/, and webpolicy.org/2014/03/12/metaphone-the-sensitivity-of-telephone-metadata/).

§We initially approached several telecommunications providers about collaboration. All declined.

¶Metadata included the time of the call or SMS, whether the call or SMS was incoming or outgoing, the other phone number participating in the call or SMS, and the length (in seconds) of the call or the length (in characters) of the SMS.

#Facebook information included age, gender, relationship status, political leanings, religious affiliation, occupation, location, and interests.

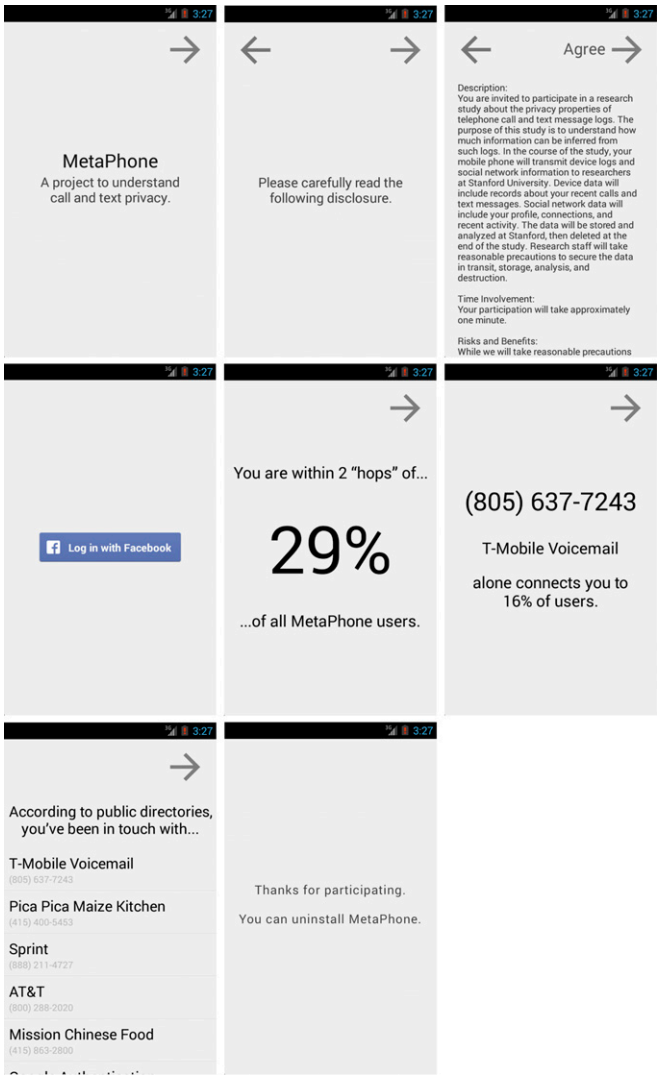


Fig. 1. Example user experience flow in MetaPhone, the Android application that we developed to crowdsource a telephone metadata dataset.

random, representative sample of the US population. Participant requirements and recruiting introduced biases, reflected in skewed demographics. In addition, participant Facebook profiles did not include a uniform set of fields. Our results are, however, strongly suggestive of properties in the larger population. The privacy phenomena that we observe are not subtle, and their causes are generalizable.

Results

In the following sections, we pose open questions about the privacy properties of telephone metadata. We then use our crowdsourced data to provide approximate empirical answers.

Graph Structure. Certain metadata surveillance programs impose a "hop" constraint, most notably the NSA's domestic telephone program (7, 8).[‖] After accessing metadata on a suspected ("seed") telephone number, an analyst can retrieve records for numbers one or more edges ("hops") distant in a connectivity graph.[**] These restrictions are intended to constrain the volume of metadata that an agency can access. Although the NSA program initially allowed three hops, executive officials scaled it back to two hops following criticism (9).

Durational limits are another form of constraint on metadata surveillance. In the NSA's program, analysts can retrieve metadata for 5 years prior. A revision to the program, proposed by the White House, would shorten the accessible history to 18 months—the current retention period under federal communications regulation (10).

Our dataset enables us to quantify the impact of these surveillance limitations. We begin with a discussion of the structure of the telephone connectivity graph, then describe how we accounted for longitudinal considerations, and finally quantitatively assess the efficacy of these constraints.

Prior work on telephone graphs has emphasized a small-world network topology (11), largely treating the graphs as diffuse social networks. The literature emphasizes monotonic, heavy-tailed degree distributions, and especially power law distributions (12–19).

Our results are broadly consistent, with two refinements. We find that at the low end of node degree, among participants, probability density includes a peak and a one-sided heavy tail (Fig. 2, Fig. S6, and *2. Graph Structure and Analysis Methodology*, *2.1. Individual Participant Structure*). The intuitive explanation is that a small proportion of telephone subscribers makes essentially no telephone use, and another small proportion makes unusually heavy use. In future work, a nonmonotonic distribution—such as a variant of a log-normal distribution—would better approximate individual telephone use behavior (see ref. 18).

More importantly, we find that at the high end of node degree, there are hubs that connect meaningful proportions of the entire participant population (Fig. S7 and *2. Graph Structure and Analysis Methodology*, *2.2. Hub Structure*). These widely shared telephone numbers include customer service lines, Voice over Internet Protocol (VoIP) bridges, two-factor authentication services, and telemarketers (Fig. S8). Critically, for purposes of surveillance regulation, these high-degree nodes establish two-hop paths between large volumes of individual telephone subscribers (Fig. S9).

Because participants varied in the duration of telephone logs that they provided, and because some surveillance programs (including the NSA's) extend beyond the time window of our dataset, we are compelled to extrapolate a longitudinal distribution of participant degree. We accomplished this by fitting curves to longitudinal degree data (Fig. S10 and *2. Graph Structure and Analysis Methodology*, *2.3. Estimating the Effects of Surveillance Regulation*).

With these preliminaries, we are able to quantitatively estimate the reach of a telephone metadata surveillance program under particular hop and duration limits. Fig. 3*A* depicts expected reach

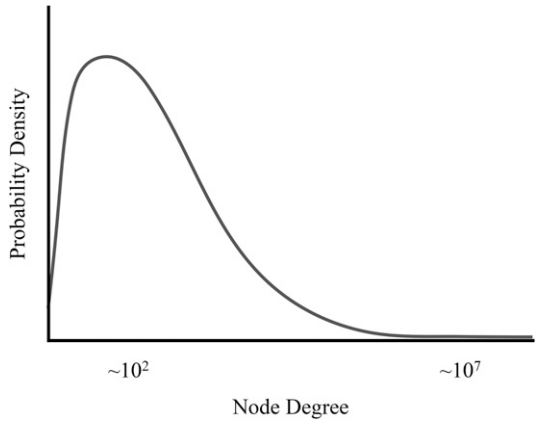


Fig. 2. Notional distribution of node degree in the telephone call and text message graphs, over approximately 1 year.

[‖]Our description emphasizes telephone metadata because that component of NSA bulk surveillance has been declassified. Officials have neither confirmed nor denied bulk surveillance of text messages.

[**]Our understanding of the NSA program is that, at each hop, an analyst can retrieve the subscriber's communications records. Disclosures have not been entirely clear on this point.

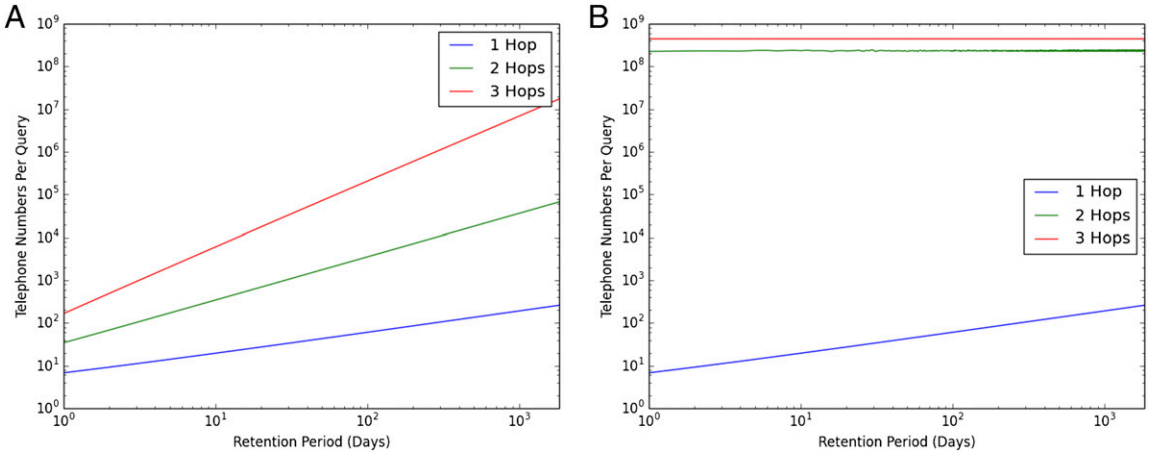


Fig. 3. Approximation of expected surveillance authority reach with one seed, in a combined US call and text message graph. (*A*) Naïve approach, assuming solely individual subscribers. (*B*) Bootstrapping approach, incorporating national hubs.

against time and hop count, assuming the call graph only includes individual subscribers. Fig. 3*B* uses a bootstrapping algorithm to incorporate the effects of high-degree hubs (Algorithm S1 and *2. Graph Structure and Analysis Methodology*, *2.3. Estimating the Effects of Surveillance Regulation*). An additional 3D surface visualization is provided in Fig. S11 and *2. Graph Structure and Analysis Methodology*, *2.3. Estimating the Effects of Surveillance Regulation*.

Applied to the NSA's program, our results strongly suggest that until 2013, analysts had legal authority to access telephone records for the majority of the entire US population. Under the more recent two-hop rule, the proposed 18-month-retention period, and an assumption that national and local hub numbers are removed from the call graph,[††] an analyst could in expectation access records for ∼25,000 subscribers with a single seed.

Reidentification. One of the chief defenses of metadata surveillance programs, including the NSA's, has been that the information is not identified. By relying on data that are not "personally identifiable information" (PII), the argument goes, metadata programs have a lesser privacy impact.[‡‡]

Prior work has demonstrated that the policy distinction between PII and non-PII is not based on sound science. Researchers have demonstrated "reidentification" risks in a number of applications, including health records (21, 22), location histories (23–25), web search queries (26), web browsing activity (27–29), movie reviews (30), and social network graphs (31, 32).

We contribute to this literature with an unsurprising result: telephone numbers are trivially reidentifiable. We conducted both automated and manual attempts at reidentification, and we found that both approaches were highly successful.

To quantify the feasibility of automated telephone number reidentification, we leveraged existing directory, search, and social network application programming interfaces (APIs). We randomly selected 30,000 numbers from our dataset and queried free, public interfaces hosted by Yelp, Google Places, and Facebook using these numbers. This approach matched identities for 9,576 (32%) of the numbers (Table 1). Matches included both businesses (from Yelp and Google Places) as well as individuals (from Facebook). These results are necessarily conservative; with access to commercial databases, a business or government agency would be able to achieve substantially higher match rates.

To assess the efficacy of manual reidentification, we randomly selected 250 phone numbers from our dataset and used two separate strategies for manual reidentification. First, we used a manual query interface for an inexpensive commercial database (Intelius). Second, we performed manual Google web searches and examined the results for identifying information. In total, we spent $19.95 for a month subscription to Intelius and 70-min running web searches. With these limited resources—far below those available to a large business or intelligence agency—we were still able to identify the overwhelming majority of the numbers (Table 2).

Location Inferences. The policy and law surrounding telephone metadata has conventionally distinguished call and text records from mobile location records. We used our dataset to investigate the extent to which location could be inferred from calls and text messages.

Prior work on mobile phone location has relied upon precise and dense Global Positioning System (GPS), wireless network, and cell tower measurements, using them to predict personal locations and movement patterns between those locations (33–35). In comparison, we show that home locations can often be predicted using imprecise and sparse telephone metadata. We accomplish this in two steps: (*i*) locating the businesses in a participant's phone logs using the reidentification techniques described above; and (*ii*) using those business locations to predict home locations.

Both Yelp and Google Places provide street addresses for reidentified businesses. We determined the latitude and longitude of these addresses using the Google Geocoding API. Following the intuition that most of the businesses an individual calls are clustered around their home, we used the DBSCAN algorithm (36) to find the largest cluster of calls based on business location information. We then predicted home location at the median latitude and longitude of the cluster.

Table 1. Performance of telephone number reidentification (automated approaches)

Look-up source	Matched, %
Google Places	16.6
Yelp	10.5
Facebook	13.7
All Automated Sources	31.9

[††]The Foreign Intelligence Surveillance Court has authorized the NSA to identify high-degree nodes (e.g., ref. 5). It is not apparent whether the NSA elects to eliminate these nodes when marking portions of the call graph as eligible for analysis, or whether the NSA merely eliminates these nodes when conducting subsequent analysis.

[‡‡]Definitions of PII vary. Some authorities do consider telephone numbers to be PII (e.g., ref. 20).

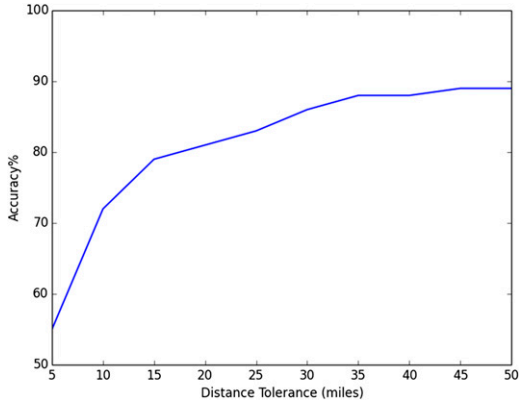


Fig. 4. Performance of automated home location prediction.

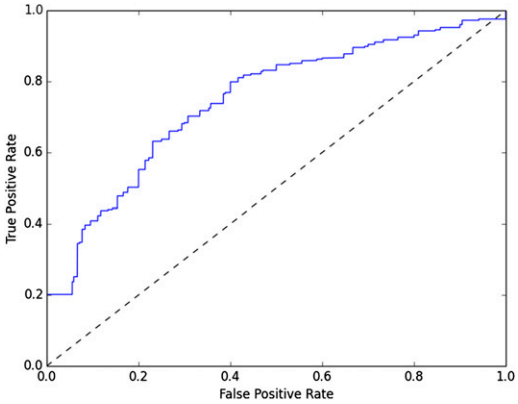


Fig. 5. Average performance of automated personal relationship prediction.

Among participants in our study, 418 listed a current city on Facebook.[§§] Of these participants, 241 (60%) had at least 10 calls to reidentified businesses. We were able to correctly predict the Facebook current city of 130 (57%) participants using the method described above. Fig. 4 presents the prediction accuracy at varying distance tolerances, measured from the center of a participant's current city.

Relationship Inferences. Another policy concern surrounding telephone metadata is that the metadata could be used to infer categories of interpersonal relationships. To understand the feasibility of drawing such inferences with an automated methodology, at scale, we focused on studying romantic relationships.

Prior work has applied supervised learning to a small sample of smartphone sensor and usage data and achieved good performance at predicting marital status (37, 53).[¶¶,##,||||] Related research has also demonstrated the feasibility of inferring relationship status from an online social network graph (38).

We built a classifier for whether a person was in a relationship, based on his or her call and text records. We began by selecting participants who were, according to their Facebook profile, single ($N = 148$) or in a relationship ($N = 309$). We then generated a range of features from telephone metadata and trained a support vector machine (*3. Relationship Inference Methodology*). Fig. 5 depicts the receiver operating characteristic for the resulting classifier.

Once a participant was labeled as in a relationship, we found that identifying the participant's partner was trivial. We tested several heuristics against the subset of participants with an identified relationship partner ($N = 211$) and achieved good performance (Table 3).

In sum, it appears feasible—with further refinement—to draw Facebook-quality relationship inferences from telephone metadata.

Sensitive Trait Inferences. Perhaps the greatest policy concern surrounding telephone metadata has been the possibility of drawing sensitive inferences. The issue is neatly encapsulated in a pair of December 2013 federal court opinions. One judge invalidated the NSA program, noting that "metadata from each person's phone 'reflects a wealth of detail about her familial, political, professional, religious, and sexual associations'" (39, 40). Less than 2 weeks later, another judge sustained the NSA program, dismissing sensitive inferences as merely a "parade of horribles" (41).

Data privacy researchers have not been so divided. In academic publications (e.g., ref. 42), court filings (e.g., ref. 40), and opinion pieces (e.g., ref. 43), scholars have persuasively argued that telephone metadata is highly sensitive. These claims have, however, been largely based on hypothetical scenarios and experiential intuition—not empirical results.

The closest related work has attempted inferences from a range of mobile phone features, including communications records, location estimates, and (in some papers) sensor and application logs. Prior results have suggested the feasibility of inferring age, gender, employment, and personality from these mobile phone data sources (refs. 37, 44–48 and ¶¶, ##, and ||||). Our study, motivated by the NSA's program and the state of American law, examines only call and text records.[***] We also attempt to draw particularly precise and particularly sensitive inferences about the participants in our study.[†††]

Using our dataset of reidentified phone numbers, we estimated the feasibility of drawing sensitive inferences from phone metadata. As with the reidentification task, we include results from both automated and manual approaches.

Automated inferences can be made directly from the results of Google Places and Yelp queries, which include business category information in their results. By labeling certain categories as sensitive, we identified the portion of participants that made a call or text to a potentially sensitive organization. Table 4 shows the portion of participants that made calls or texts to organizations matching sensitive categories.

Health Services was the most common category of sensitive organization. We further labeled medical specialist subsets of this category using more precise labels obtained from Google and Yelp queries. Table 5 shows the specialist categories that appear in at least 1% of participants' call logs.

Calls to religion-affiliated numbers provided an opportunity to validate the accuracy of automated sensitive inferences. A subset of participants both placed a call to a religious group and provided a religion on Facebook ($N = 18$). Among these, the most-called religious group overwhelmingly matched the Facebook religion ($N = 14$).

Our results suggest that, even without human review, a business or agency could draw sensitive inferences from a significant share of telephone records.

To simulate the inferences that might be drawn from manual telephone record analysis, we focused on participants who held a high proportion of their phone conversations with sensitive numbers. We then applied our automated and manual reidentification

[§§]This information is self-reported and may be out of date; future work would benefit from a more reliable source of location data.

[¶¶]Brdar S, Ćulibrk D, Crnojević V, Demographic attributes prediction on the real-world mobile data, Mobile Data Challenge 2012 (by Nokia) Workshop, June 18–19, 2012, Newcastle, UK.

[##]Mohrehkesh S, Ji S, Nadeem T, Weigle MC, Demographic prediction of mobile user from phone usage, Mobile Data Challenge 2012 (by Nokia) Workshop, June 18–19, 2012, Newcastle, UK.

[||||]Ying JJC, Chang YJ, Huang CM, Tseng VS, Demographic prediction based on user's mobile behaviors, Mobile Data Challenge 2012 (by Nokia) Workshop, June 18–19, 2012, Newcastle, UK.

[***]The majority view in American courts is that the Fourth Amendment does not protect mobile phone location records. As a matter of statute, a court order is still required to obtain those records.

[†††]As a natural consequence of attempting such specific inferences, we examine fewer types of inference and draw inferences with lesser reliability than in prior work.

Table 2. Performance of telephone number reidentification (manual and combined approaches)

Look-up source	Matched, %
Intelius	65
Google search	58
All automated sources	26
All sources	82

Table 3. Performance of relationship partner identification heuristics

Heuristic, maximum	Accuracy, %
Calls	81
Call duration	45
Days with a call	77
Texts	76
Text length	68
Days with a text	76

approaches, attempting to identify as many of each participant's contacts as possible.[‡‡‡]

The following vignettes are reflective of the types of inferences we were able to draw.

i) Participant A held conversations with a pharmacy specializing in chronic care, a patient service that coordinates management for serious conditions, several local neurology practices, and a pharmaceutical hotline for a prescription drug used solely to manage the symptoms and progression of relapsing-remitting multiple sclerosis.

ii) Participant B received a long phone call from the cardiology group at a regional medical center, talked briefly with a medical laboratory, answered several short calls from a local drugstore, and made brief calls to a self-reporting hotline for a cardiac arrhythmia monitoring device.

iii) Participant C placed frequent calls to a local firearm dealer that prominently advertises a specialty in the AR semiautomatic rifle platform. He also placed lengthy calls to the customer support hotline for a major firearm manufacturer; the manufacturer produces a popular AR line of rifles.

iv) Participant D placed calls to a hardware outlet, locksmiths, a hydroponics store, and a head shop in under 3 weeks.

v) Participant E made a lengthy phone call to her sister early one morning. Then, 2 days later, she called a nearby Planned Parenthood clinic several times. Two weeks later, she placed brief additional calls to Planned Parenthood, and she placed another short call 1 month after.

Using public sources, we were able to confirm that participant B had a cardiac arrhythmia and participant C owned an AR rifle. As for the remaining inferences, regardless of whether they were accurate, the mere appearance of possessing a highly sensitive trait assuredly constitutes a serious privacy impact.[§§§]

Our results lend strong support to the view that telephone metadata is extraordinarily sensitive, especially when paired with a broad array of readily available information. For a randomly selected telephone subscriber, over a short period, drawing these sorts of sensitive inferences may not be feasible. However, over a large sample of telephone subscribers, over a lengthy period, it is inevitable that some individuals will expose deeply sensitive information. It follows that large-scale metadata surveillance programs, like the NSA's, will necessarily expose highly confidential information about ordinary citizens.

Table 4. Participant interaction with sensitive organizations

Category	Participants with ≥1 calls, %
Health services	57
Financial services	40
Pharmacies	30
Veterinary services	18
Legal services	10
Recruiting and job placement	10
Religious organizations	8
Firearms sales and repair	7
Political officeholders and campaigns	4
Adult establishments	2
Marijuana dispensaries	0.4

Table 5. Participant interaction with health organizations

Category	Participants with ≥1 calls, %
Dentistry and oral health	18
Mental health and family services	8
Ophthalmology and optometry	6
Sexual and reproductive health	6
Pediatrics	5
Orthopedics	4
Chiropractic care	3
Rehabilitation and physical therapy	3
Medical laboratories	2
Emergency or urgent care	2
Hospitals	2
Cardiology	2
Dermatology	1
Ear, nose, and throat	1
Neurology	1
Oncology	1
Substance abuse	1
Cosmetic surgery	1

Discussion

The results of our study are unambiguous: there are significant privacy impacts associated with telephone metadata surveillance. Telephone metadata is densely interconnected, easily reidentifiable, and trivially gives rise to location, relationship, and sensitive inferences. In combination with independent reviews that have found bulk metadata surveillance to be an ineffective intelligence strategy (7, 8), our findings should give policymakers pause when authorizing such programs.

More broadly, this project emphasizes the need for scientifically rigorous surveillance regulation. Much of the law and policy that we explored in this research was informed by assumption and conventional wisdom, not quantitative analysis. To strike an appropriate balance between national security and civil liberties, future policymaking must be informed by input from the relevant sciences.

Our results also bear on commercial data practices. It is routine practice for telecommunications firms to collect, retain, and transfer subscriber telephone records, often dubbed "Customer Proprietary Network Information" (49, 50). Telecommunications regulation should also incorporate a scientifically rigorous understanding of the privacy properties of these data.

There remains much future work to be done in this space. To conduct this study, we were compelled to rely on a small and

[‡‡‡]Although several of these participants consented to being identified in this publication, out of recognition for the associated privacy risks, we use only pseudonyms.

[§§§]More generally, a probabilistic sensitive inference—even with less than even likelihood—could constitute a significant privacy risk.

unrepresentative dataset. Future efforts would benefit from population-scale data; the challenges are in sourcing the data, not computing on them. Future work could also pair telephone records with more comprehensive ground truth than the Facebook data we accessed. Subscriber records and cell site location information, for instance, would better enable testing for inferences. Another potential direction is testing more advanced approaches to automated inferences; the machine-learning techniques we applied in this study were effective, although relatively rudimentary.

ACKNOWLEDGMENTS. We thank the participants in this study, who altruistically gave up their privacy, such that the public might better understand government surveillance. We also thank Dan Boneh and Edward Felten for invaluable suggestions. This project was supported by the National Science Foundation Team for Research in Ubiquitous Secure Technology Research Center.

1. Electronic Communications Privacy Act, 18 U.S. Code Sect. 2510(8) (2012).
2. United States v. Forrester, 512 F.3d 500, 9th Cir. (July 6, 2007).
3. Stored Communications Act, 18 U.S. Code Sect. 2703(c)(2) (2012).
4. USA PATRIOT Act, 50 U.S. Code Sect. 1861 (2012).
5. In re FBI Application, No. BR 13-109, FISA Ct. (August 29, 2013).
6. USA FREEDOM Act, Pub. L. No. 114-23, 129 Stat. 268 (June 2, 2015).
7. President's Review Group (December 12, 2013) *Liberty and Security in a Changing World* (President's Review Group on Intelligence and Communications Technologies, Washington, DC). Available at www.whitehouse.gov/sites/default/files/docs/2013-12-12_rg_final_report.pdf. Accessed April 2016.
8. Privacy and Civil Liberties Oversight Board (2014) *Report on the Telephone Records Program Conducted Under Section 215 of the USA PATRIOT Act and on the Operations of the FISC* (Privacy and Civil Liberties Oversight Board, Washington, DC).
9. Privacy and Civil Liberties Oversight Board (2015) *Recommendations Assessment Report* (Privacy and Civil Liberties Oversight Board, Washington, DC).
10. Retention of Telephone Toll Records, 47 C.F.R. Sect. 42.6 (2015).
11. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442.
12. Abello J, Pardalos PM, Resende MGC (1998) On maximum clique problems in very large graphs. *External Memory Algorithms. DIMACS Workshop: External Algorithms and Visualization, May 20–22 1998*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, eds Abello JM, Vitter JS (American Mathematical Society, Providence, RI), pp 119–130.
13. Abello J, Resende MG, Sudarsky S (2002) Massive quasi-clique detection. *LATIN 2002: Theoretical Informatics, 5th Latin American Symposium, Cancun, Mexico April 2002 Proceedings*, Lecture Notes in Computer Science, ed Rajsbaum S (Springer, Berlin), Vol 2286, pp 598–612.
14. Aiello W, Chung F, Lu L (2001) A random graph model for power law graphs. *Expo Math* 10(1):53–66.
15. Nanavati AA, et al. (2006) On the structural properties of massive telecom call graphs: findings and implications, Conference on Information and Knowledge Management, CIKM '06, November 5–11, 2006, Arlington, VA (Association for Computing Machinery, New York), pp 435–444.
16. Onnela JP, et al. (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA* 104(18):7332–7336.
17. Pandit V, et al. (2008) Extracting dense communities from telecom call graphs. *Third International Conference on Communication Systems Software and Middleware and Workshops, COMSWARE 2008* (IEEE, New York), pp 82–89.
18. Seshadri M, et al. (2008) Mobile call graphs: beyond power-law and lognormal distributions, The 14th ACM SIGKDD International Conference on Knowledge, KDD '08, August 24–27, 2008, Las Vegas (Association for Computing Machinery, New York), pp 596–604.
19. Wang P, González MC, Hidalgo CA, Barabási AL (2009) Understanding the spreading patterns of mobile phone viruses. *Science* 324(5930):1071–1076.
20. California Online Privacy Protection Act, Cal. Bus. and Prof. Code Sect. 22577 (2012).
21. Sweeney L (2002) k-anonymity: a model for protecting privacy. *Int J Uncertain Fuzziness Knowl Based Syst* 10(5):557–570.
22. Sweeney L, Abu A, Winn J (2013) *Identifying Participants in the Personal Genome Project by Name, Technical Report 1021-1* (Harvard Univ Data Privacy Lab, Cambridge, MA).
23. Golle P, Partridge K (2009) On the anonymity of home/work location pairs. *Proceedings of the 7th International Conference on Pervasive Computing* (Springer, Berlin), pp 390–397.
24. Zang H, Bolot J (2011) Anonymization of location data does not work: A large-scale measurement study. *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom '11* (Association for Computing Machinery, New York), pp 145–156.
25. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the Crowd: The privacy bounds of human mobility. *Sci Rep* 3:1376.
26. Ohm P (2010) Broken promises of privacy. *UCLA Law Rev* 57:1701–1777.
27. Krishnamurthy B, Naryshkin K, Wills C (2011) Privacy leakage vs. Protection measures: The growing disconnect. *IEEE Secur Priv* 11(3):14–20.
28. Mayer JR, Mitchell JC (2012) Third-party web tracking: Policy and technology. *2012 IEEE Symposium on Security and Privacy (SP), May, 20–23, 2012, San Francisco* (IEEE, New York), pp 413–427.
29. Englehardt S, et al. (2015) Cookies that give you away: The surveillance implications of Web tracking. *Proceedings of the 24th International Conference on World Wide Web, WWW '15* (Association for Computing Machinery, New York), pp 289–299.
30. Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparse datasets. *Proceedings of the 2008 IEEE Symposium on Security and Privacy, SP '08* (IEEE Computer Society, Washington, DC), pp 111–125.
31. Backstrom L, Dwork C, Kleinberg J (2007) Wherefore art thou R3579X? Anonymized social networks, hidden patterns, and structural steganography. *Proceedings of the 16th International Conference on World Wide Web, WWW '07* (Association for Computing Machinery, New York), pp 181–190.
32. Narayanan A, Shmatikov V (2009) De-anonymizing social networks. *30th IEEE Symposium on Security and Privacy* (IEEE, New York), 173–187.
33. Zheng Y, Li Q, Chen Y, Xie X, Ma WY (2008) Understanding mobility based on GPS data. *Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08* (Association for Computing Machinery, New York), pp 312–321.
34. Zheng Y, Zhang L, Xie X, Ma WY (2009) Mining interesting locations and travel sequences from GPS trajectories. *Proceedings of the 18th International Conference on World Wide Web, WWW '09* (Association for Computing Machinery, New York), pp 791–800.
35. Anagnostopoulos T, Anagnostopoulos C, Hadjiefthymiades S (2012) Efficient location prediction in mobile cellular networks. *Int J Wirel Inf Networks* 19(2): 97–111.
36. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *Data Mining and Knowledge Discovery* 2(2):169–194.
37. Zhong E, Tan B, Mo K, Yang Q (2013) User demographics prediction based on mobile data. *Pervasive and Mobile Computing* 9(6):823–837.
38. Backstrom L, Kleinberg J (2014) Romantic partnerships and the dispersion of social ties. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '14* (Association for Computing Machinery, New York), pp 831–841.
39. Klayman v. Obama, 957 F. Supp. 2d 1, Dist. Ct. DC (December 16, 2013).
40. Declaration of Professor Edward W. Felten, ACLU v. Clapper, No. 13-cv-03994, Southern Dist. Ct. NY (August 26, 2013).
41. ACLU v. Clapper, 959 F. Supp. 2d 724, Southern Dist. Ct. NY (December 27, 2013).
42. Landau S (2013) Making sense from Snowden. *IEEE Secur Priv* 11(4):54–63.
43. Blaze M (June 19, 2013) Phew, NSA is just collecting metadata. *Wired*. Available at www.wired.com/2013/06/phew-it-was-just-metadata-not-think-again.
44. Chittaranjan G, Blom J, Gatica-Perez D (2013) Mining large-scale smartphone data for personality studies. *Pers Ubiquitous Comput* 17(3):433–450.
45. de Montjoye YA, Quoidbach J, Robic F, Pentland A (2013) Predicting Personality Using Novel Mobile Phone-Based Metrics. *Social Computing, Behavioral-Cultural Modeling and Prediction*, Lecture Notes in Computer Science (Springer, Berlin) Vol 7812, pp 48–55.
46. Arai A, et al. (2014) Understanding user attributes from calling behavior. *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia, MoMM'14* (Association for Computing Machinery, New York), pp 95–104.
47. Jahani E, et al. (2015) Predicting gender from mobile phone metadata, NetMob 2015, April 7–10, 2015, Cambridge, MA, eds Moro E, de Montjoye Y-A, Blondel V, Pentland A, pp 110–113.
48. Toole JL, et al. (2015) Tracking employment shocks using mobile phone data. *J R Soc Interface* 12(107):pii: 20150185.
49. The Telecommunications Act, 47 U.S. Code Sect. 222 (2012).
50. Customer Proprietary Network Information, 47 C.F.R. Sects. 64.2001–64.2011 (2015).
51. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci USA* 106(36):15274–15278.
52. Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland A (2014) Daily stress recognition from mobile phone data, weather conditions and individual traits. *Proceedings of the 22nd ACM International Conference on Multimedia, MM'14* (Association for Computing Machinery, New York), pp 477–486.
53. Laurila JK, et al. (2013) From big smartphone data to worldwide research: The mobile data challenge. *Pervasive and Mobile Computing* 9(6):752–771.
54. Apple (2015) *ResearchKit Technical Overview* (Apple, Cupertino, CA). Available at researchkit.org/docs/docs/Overview/GuideOverview.html. Accessed April 2015.
55. The White House Office of the Press Secretary (2014) *Presidential Policy Directive/PPD-28* (The White House, Washington, DC).
56. Dwork C (2006) Differential privacy. *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, Lecture Notes in Computer Science (Springer, Berlin), Vol 4052, pp 1–12.
57. Mayer J, Narayanan A (2013) Privacy substitutes. *Stanford Law Rev Online* 66:89–96.
58. Pew Research Center (2015) *The Smartphone Difference* (Pew Research Center, Washington, DC), Available at www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015. Accessed April 2015.
59. Federal Communications Commission (2014) *Local Telephone Competition: Status as of December 31, 2013* (Federal Communications Commission, Washington, DC).
60. Pedregosa F, et al. (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830.