When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning

Steven M. Bellovin

(Joint work with Renée Hutchins, Tony Jebara, Sebastian Zimmeck)

PATTERNS AND PREDICTIONS

- Machine learning can find all sorts of patterns
- Some uses of big data are fairly obvious, once we know how to do it
- Some aren't—like shaping legal doctrine
- For example: should the police need a search warrant to track someone's location?

SHOULD POLICE NEED A WARRANT FOR GPS TRACKING?

- No: movements are public
  - Police could just follow someone
  - You have no “reasonable expectation of privacy” in public activities
- No: in the 1982 *Knotts* case, the Supreme Court said that putting a beeper on a chemical shipment for three days is ok

SHOULD POLICE NEED A WARRANT FOR UPS TRACKING?

- Yes: One check on police abuse of their power is economic: they can't afford to trail very many people for a very long time

  ➔ GPS tracking is *much* cheaper

- Yes: *Patterns* of movement are very revealing

THE FOURTH AMENDMENT

"The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

Searches do not always require a warrant, but they have to be *reasonable*

MOSAIC THEORY

- *Mosaic Theory*: a large-enough collection of data points is very, very revealing, and violates "reasonable expectation of privacy"
- It is the *total pattern* of movements that is revealing
  - Law enforcement cannot afford to track (most) people for a month
- But—where do you draw the line? What is "large enough"?

US V. JONES (2012)

- Police attached a GPS tracker to Jones' car for 28 days
- The warrant had expired
- The Supreme Court overturned the conviction 9-0, but on classical Fourth Amendment grounds: a physical intrusion on his car

SOME JUDICIAL SUPPORT FOR MOSAIC THEORY

"Disclosed in [GPS] data . . . will be trips the indisputably private nature of which takes little imagination to conjure: trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, the criminal defense attorney, the by-the-hour motel, the union meeting, the mosque, synagogue or church, the gay bar and on and on."

Justice Sotomayor's concurrence in *Jones*

MORE SUPPORT

"We need not identify with precision the point at which the tracking of this vehicle became a search, for the line was surely crossed before the 4-week mark."

Justice Alito's concurrence in *Jones*,

joined by three other justices

BUT…

"[I]t remains unexplained why a 4-week investigation is 'surely' too long"

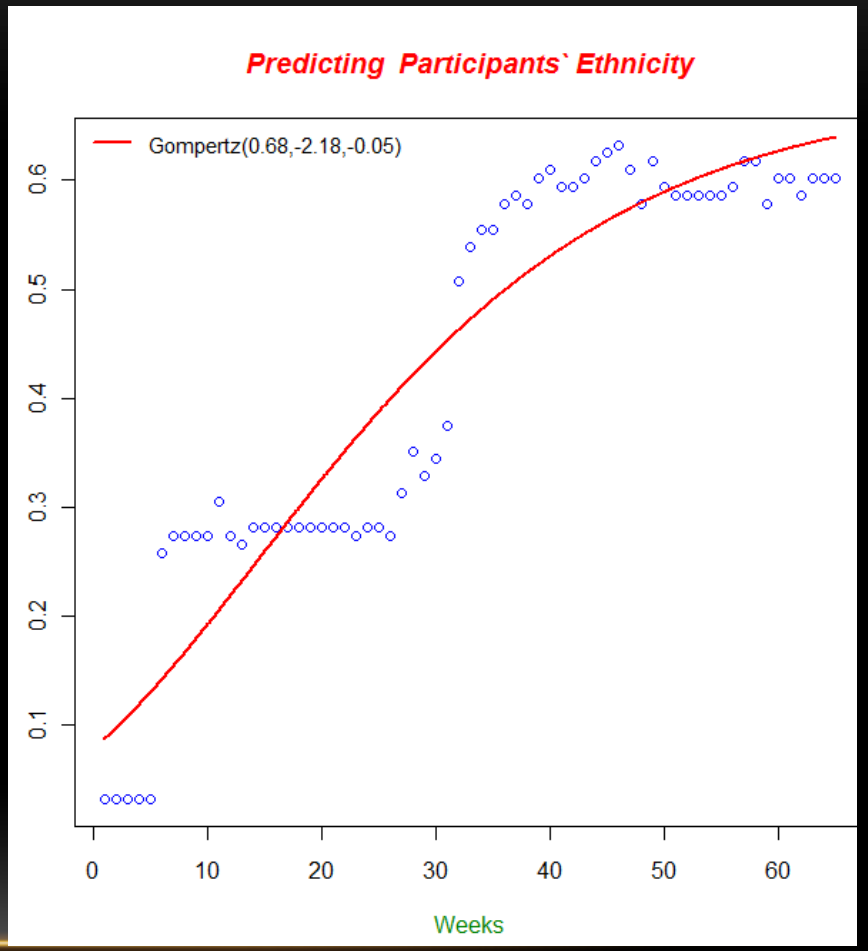
Opinion of the Court (by Justice Scalia) in *Jones*

MOSAIC THEORY AND MACHINE LEARNING: A HYPOTHESIS

- Use machine learning to make predictions based on location data
- When predictions are accurate enough, a mosaic exists
- In other words, use computer science to answer Justice Scalia's objection!

MACHINE LEARNING AND MOSAIC THEORY

- The technical literature supports the basic premise: with enough points, the whole *is* greater than the sum of its parts
- Note the jump in accuracy at 5 weeks and 28 weeks



(Graph from Altshuler et al.)

ONE WEEK IS THE LIMIT

- Experiments show that week-to-week movements are very predictable (Sadilek & Krumm)
- Weekend movements are *more* predictable, though of course different than weekday movement
- With seven days of observation, you have a very good picture of someone's life

THE FOURTH AMENDMENT

- Does Mosaic Theory make tracking "unreasonable"?
- Do people have a "reasonable expectation of privacy" in their location *and the inferences that can be made from it*?
- Is it "one that society is prepared to recognize as 'reasonable'"?

CONCLUSIONS

- From a technical perspective, Mosaic Theory is correct: you really can build a very full picture of someone with enough data points
- (The Massachusetts Supreme Court has set a limit of two weeks, though without giving a reason for that limit)
- Fundamentally, though, this is a legal question, not technical one
- Paper: http://lawandlibertyblog.com/s/Hutchins.pdf