Limitations of Technical Mechanisms as Proxies for Context

Catherine Dwyer

Seidenberg School of CS & IS
Pace University
NY, NY

What problem are you trying to solve?

- The challenge of representing context within technical systems
- Context needs to be operationalized, turned into ones and zeros
- Information flow has been operationalized, it is happening
- How is context currently being operationalized in technical systems?
- What technical mechanisms are used to determine (justify) information flow?

Contextual Integrity and Computer Science

- Computer science research “conceptualized contexts as concrete and at least partly technical … We note that as far as CI is concerned. It is essential that contexts be understood as normative, as one important trait of contexts is that they have ends, purposes, and values.”
- Benthall, S., Gürses, S., & Nissenbaum, H. (2017). *Contextual integrity through the lens of computer science*: Now Publishers Incorporated.

Applying CI Decision Heuristic to Location Data

- Hildebrandt, M. (2014). Location Data, Purpose Binding and Contextual Integrity: What's the Message? In L. Floridi (Ed.), *Protection of Information and the Right to Privacy-A New Equilibrium?* (Vol. 17, pp. 31-62). Switzerland: Springer.
- CI Decision Heuristic is a methodology that can identify "when and why some of these [privacy] alternatives provoke legitimate anxiety, protest, and resistance," Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*: Stanford University Press.

WEF and Personal Data

- Hildebrandt uses a framework for personal data introduced by the World Economic Forum (WEF)
- The WEF started a program entitled “Rethinking Personal Data.” The idea behind this project is that data, especially personal data, is a valuable asset that needs to be developed and optimized.
- Three categories of personal data: volunteered, observed, and inferred
  - WEF. (2011). *Personal data: The emergence of a new asset class*. Retrieved from https://www.weforum.org/reports/personal-data-emergence-new-asset-class
  - WEF. (2012). *Rethinking personal data: Strengthening Trust*. Retrieved from http://manuscritdepot.com/documentspdf/WEF_IT_RethinkingPersonalData_Report_2012.pdf

Personal data – a definition

For this report personal data is defined as data (and metadata) created by and about people, encompassing:

- Volunteered data – created and explicitly shared by individuals, e.g., social network profiles.
- Observed data – captured by recording the actions of individuals, e.g., location data when using cell phones.
- Inferred data – data about individuals based on analysis of volunteered or observed information, e.g., credit scores.

Source: World Economic Forum, June 2010.

From
WEF. (2011). *Personal data: The emergence of a new asset class*.

Volunteered, Observed, and Inferred Data

- Volunteered data are the data people deliberately provide, for example mailing address for an online purchase
- Observed data are the measurable behaviors of users that can somehow be "datafied," turned into machine readable data
  - Examples include click stream data (how long you look at a web page), proximity data (are you physically located near a point of interest), transportation behavior from toll collection devices
- Inferred data are the product of algorithm or other automated process that combines volunteered and observed data, and associates patterns of activity with particular trait or quality of interest
  - Inferred data are calculated values often expressed as a probability value, for example 90% likely to be female, 65% likely to be age 18-35

Analysis of These Categories Within Technical Systems

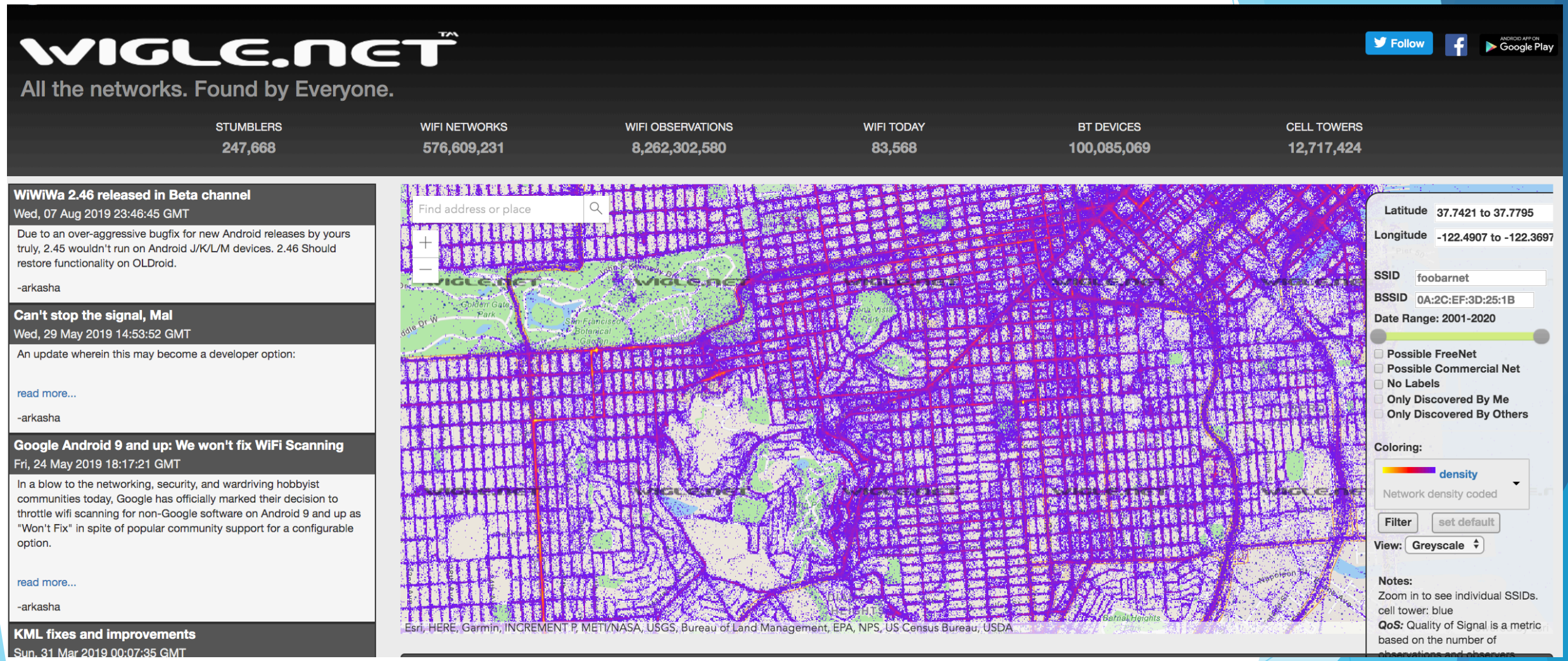
- Can these categories – volunteered, observed, and inferred provide some insight into how information flow is built within technical systems? And how that flow is managed (justified)?
- Two cases: location data and social media data

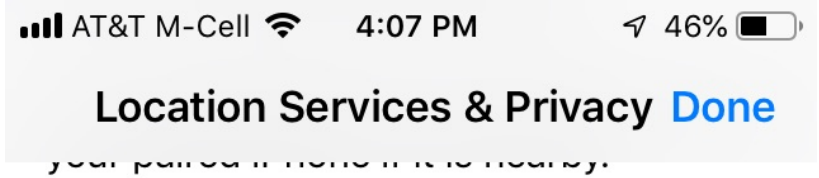
Location data and smartphones

- Location is represented as two floating point numbers, one that represents latitude, one for longitude (~16 bytes)
- Since smartphones come with gigabytes of memory, the storage and transmission demands for location data are miniscule
- Location data are the perfect type of data when considering cost versus benefit, it uses bare minimum of computing resources and delivers significant added value to advertisers

Alice and Bob and location

- **Volunteered**: Alice wants to share her location with Bob, so she explicitly requests than an app such as Google Maps transmit her location data to Bob's phone
- **Observed**: Bob opens the app Uber Eats to find some food. When Bob installed UberEats, he gave it permission to view location. UberEats accesses (observes) Bob's location, and sends it to mobile advertising partner, AdMob. In most cases involving advertising, location is associated with a pseudo-anonymous identifier rather than Bob's explicit identify
- **Inferred**: Location can also be inferred from proximity to wireless access points. As Alice travels around her phone picks up nearby wireless signals. Google GeoLocation API will estimate location from two or more nearby wireless access points


WIGLE.NET
All the networks. Found by Everyone.
Follow
Google Play
STUMBLERS
247,668
WIFI NETWORKS
576,609,231
WIFI OBSERVATIONS
8,262,302,580
WIFI TODAY
83,568
BT DEVICES
100,085,069
CELL TOWERS
12,717,424
WiWiWa 2.46 released in Beta channel
Wed, 07 Aug 2019 23:46:45 GMT
Due to an over-aggressive bugfix for new Android releases by yours truly, 2.45 wouldn't run on Android J/K/L/M devices. 2.46 Should restore functionality on OLDroid.
-arkasha
Can't stop the signal, Mal
Wed, 29 May 2019 14:53:52 GMT
An update wherein this may become a developer option:
read more...
-arkasha
Google Android 9 and up: We won't fix WiFi Scanning
Fri, 24 May 2019 18:17:21 GMT
In a blow to the networking, security, and wardriving hobbyist communities today, Google has officially marked their decision to throttle wifi scanning for non-Google software on Android 9 and up as "Won't Fix" in spite of popular community support for a configurable option.
read more...
-arkasha
KML fixes and improvements
Sun, 31 Mar 2019 00:07:35 GMT
Find address or place
Esri, HERE, Garmin, INCREMENT P, METI/NASA, USGS, Bureau of Land Management, EPA, NPS, US Census Bureau, USDA
Latitude 37.7421 to 37.7795
Longitude -122.4907 to -122.3697
SSID foobarnet
BSSID 0A:2C:EF:3D:25:1B
Date Range: 2001-2020
Possible FreeNet
Possible Commercial Net
No Labels
Only Discovered By Me
Only Discovered By Others
Coloring:
density
Network density coded
Filter
set default
View: Greyscale
Notes:
Zoom in to see individual SSIDs.
cell tower: blue
QoS: Quality of Signal is a metric based on the number of



your paired iPhone if it is nearby.

If Location Services is on, your iPhone will periodically send the geo-tagged locations of nearby Wi-Fi hotspots and cell towers in an anonymous and encrypted form to Apple, to be used for augmenting this crowd-sourced database of Wi-Fi hotspot and cell tower locations.

By enabling Location Services, location-based system services such as these will also be enabled:

Traffic — If you are physically moving (for example, traveling in a car), your iPhone will periodically send GPS locations and travel speed information in an anonymous and encrypted form to Apple, to be used for augmenting a crowd-sourced road traffic database.

Popular Near Me — Your iPhone will periodically send locations of where, and when, you have purchased or used Apps in an anonymous and encrypted form to

"Your iPhone will periodically send the geo-tagged locations of nearby Wi-Fi hotspots and cell towers in encrypted form …. to augment this crowd-sourced database of Wi-Fi hotspot and cell tower locations."

Google GeoLocation Request and Response

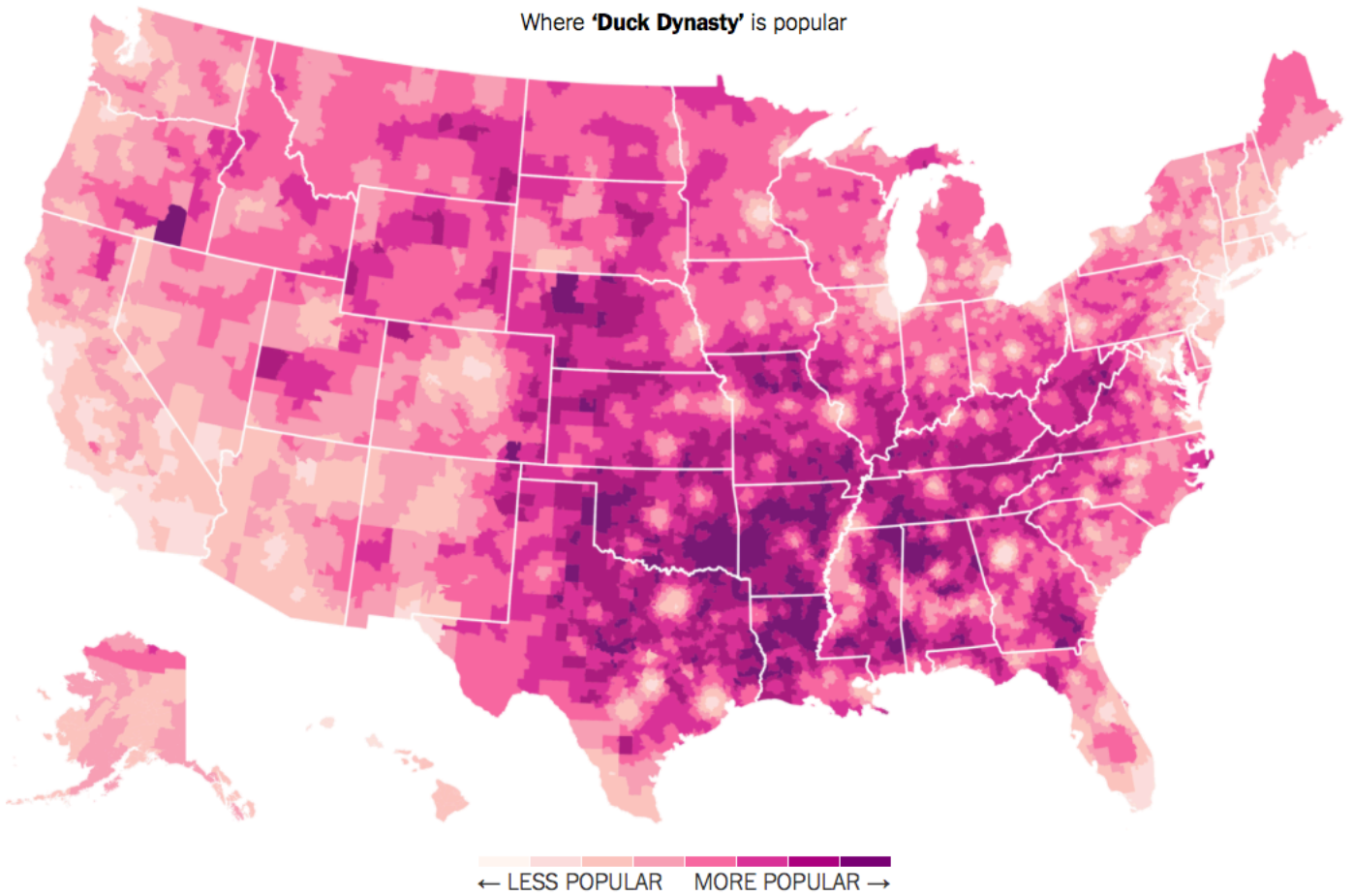
```
#GeoLocation Request:
{
  "considerIp": "false",
  "wifiAccessPoints": [
    {
        "macAddress": "00:25:9c:cf:1c:ac",
        "signalStrength": -43,
        "signalToNoiseRatio": 0
    },
    {
        "macAddress": "00:25:9c:cf:1c:ad",
        "signalStrength": -55,
        "signalToNoiseRatio": 0
    }
  ]
}
```

```
#GeoLocation Response:

{
  "location": {
    "lat": 33.3632256,
    "lng": -117.0874871
  },
  "accuracy": 20
}
```

Alice and Bob on Facebook

- Cambridge Analytica used Facebook data to infer political orientation
- Alice and Bob are friends on Facebook
- Alice signs up for personality quiz “This Is Your Digital Life,” and agrees (volunteers) to to share her list of friends with the app, including Bob
- Through Alice, the app accesses (observes) Bob’s public information, including name, home town, and public “likes.” The Facebook data of Alice and Bob is used to calculate personality profiles for Alice and Bob
- Alice is predicted (inferred) as a likely Trump voter, based in part on her “like” of Duck Dynasty
- By “liking” Empire, Bob is predicted (inferred) to be a likely Clinton voter, and receives political advertising designed to diminish enthusiasm for the Clinton candidacy (https://www.nytimes.com/interactive/2016/12/26/upshot/duck-dynasty-vs-modern-family-television-maps.html)



'Duck Dynasty' vs. 'Modern Family': 50 Maps of the U.S. Cultural Divide

Discussion

- As shown in these two cases, information flow can disclose sensitive information such as a person's location or political orientation
- Each incremental step in the information flow is explained by its status as volunteered, observed, or inferred
- But when the analysis of information flow focuses on the little steps that lead to sensitive data, there is no clear place to fix the flow. An app running on a smart phone has access to the identity of nearby wireless access points, that seems reasonable, Facebook "likes" are publicly available
- So instead of ethical norms informing information flow, the assumptions tied to the categories of volunteered, observed, and inferred are used to inform technology design decisions

Future Work

- It seems that these categories are used to manage information flow, the question is how?
- Develop an app for the Google Android OS that will calculate location through all three methods – volunteered, observed, and inferred. Once the app is operational, subjects will be recruited to use the app and record their reactions as their location is derived from different mechanisms
- Use of Facebook Onavo VPN app and other "beta testing" tools used to extract data from users https://techcrunch.com/2019/01/29/facebook-project-atlas/ and https://www.wsj.com/articles/facebooks-onavo-gives-social-media-firm-inside-peek-at-rivals-users-1502622003?mod=article_inline

Questions?

- Cathy Dwyer
- cdwyer@pace.edu
- Twitter: @ProfCDwyer