Big Data and Big Cities

Feiyang Sun & Jan Whittington

NABE Tech Economics Conference
November 4, 2019
Seattle, Washington

UNIVERSITY of WASHINGTON

Cities and Data

Cities as Laboratories

Cities as sites of experimentation in the public interest

+ Detailed data can answer public policy questions
+ Concentrated market interest in cities
+ Representatives close to the voting populace
+ State policy interest is rising; Carpenter Case

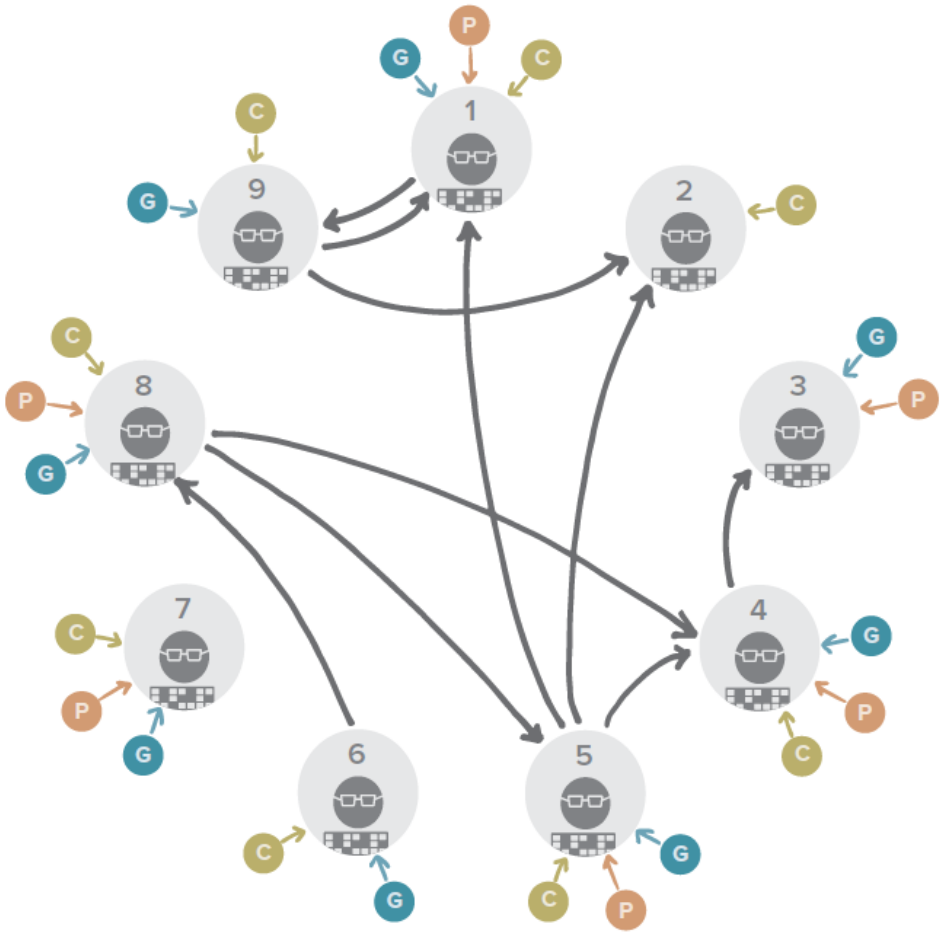
Municipal Open Data

+ Hopes of accountability, transparency, market benefit
+ Fears: Are subjects re-identifiable?

Public Records Acts

+ Requests are perceived to be in the public interest
+ Conflict: Are subjects re-identifiable?

Municipalities and Data



US FTC. *Data Brokers: A Call For Transparency and Accountability*. Washington DC, May 2014. Exhibit 2.

Cities **PUSH** data to open data portals

People **PULL** data out (e.g., public disclosure requests)

Cities **SPILL** data (e.g., external or internal breaches)

Whittington et al. *Push Pull and Spill*. BTLJ, 2015.

Hopes

"Having the data be open is an incredible source of accountability. It is a key to democracy."

"It could be useful for commercial benefit."

"If a public agency has, as you said, lots and lots of data … entrepreneurs can use it to make money. One purpose of most governments is to help the citizens thrive, which means to put money in your pocket."

Whittington et al. *Push Pull and Spill*. BTLJ, 2015.

Fears

"I fear that it would be used to lower public property values, redline insurance, et cetera, in neighborhoods with high crime."

"I'm worried that data about precincts where people don't vote much could lead politicians to write them off."

"It doesn't feel safe to me at all. There is a Seattle Public Utility LBGT group… there was a guy requesting all the members' name and information. He was just anti-gay."

Whittington et al. *Push Pull and Spill*. BTLJ, 2015.

Table 1: Attributes from Four Open Data Sets on a Continuum of Personalization.[113]

	Datasets				
Fields	Property Value	Tech. User Survey	Business License	Building Permits	Potential Privacy Concern
Name	●		●	●	
Address/Location	●			●	
Phone Number			●		
Age		●			
Gender		●			
Income		●			
Home Value	●				
Zip Code		●			
Sexual Orientation		●			
Race		●			
Level of Education		●			
Language		●			
Number in Household		●			
Employment		●			
Unpermitted Activity				●	
Internet Use		●			
Uses of Cable		●			
Incident Type/Descrip.			●		
Permitted Activity				●	
Value of Alteration				●	
Permit Type				●	

Persons

Groups

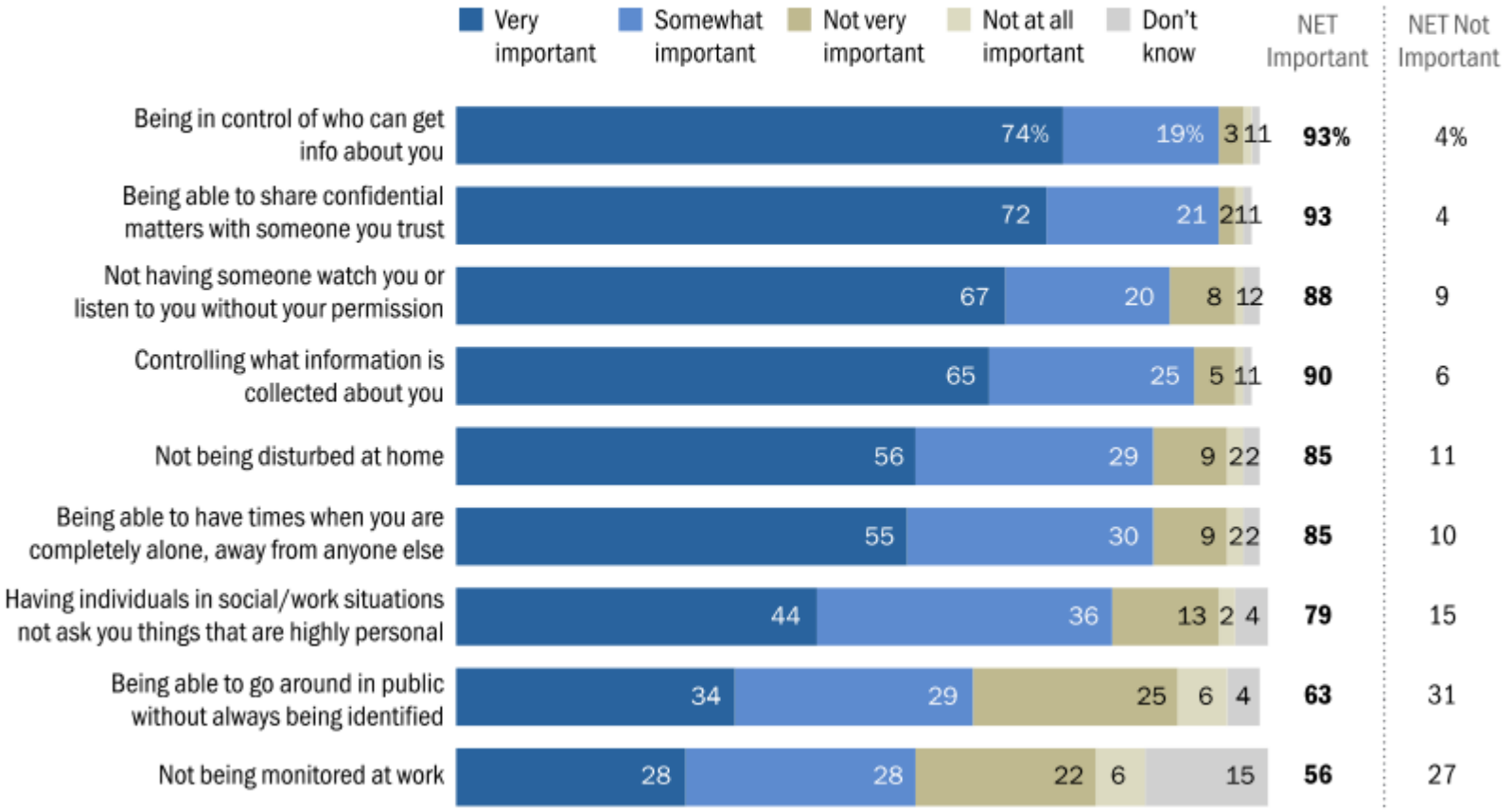
Unknown

Whittington et al. *Push Pull and Spill*. BTLJ, 2015.

Americans Hold Strong Views About Privacy in Everyday Life

In response to the following question: "Privacy means different things to different people today. In thinking about all of your daily interactions – both online and offline – please tell me how important each of the following are to you . . ."

% of adults who say ...



Source: Pew Research Center's Privacy Panel Survey #4, Jan. 27, 2015-Feb. 16, 2015 (N=461). Refused responses not shown.

PEW RESEARCH CENTER

Table 1: Attributes from Four Open Data Sets on a Continuum of Personalization.[113]

	Datasets				
Fields	Property Value	Tech. User Survey	Business License	Building Permits	Potential Privacy Concern
Name	●		●	●	
Address/Location	●			●	
Phone Number			●		
Age		●			
Gender		●			
Income		●			
Home Value	●				
Zip Code		●			
Sexual Orientation		●			
Race		●			
Level of Education		●			
Language		●			
Number in Household		●			
Employment		●			
Unpermitted Activity				●	
Internet Use		●			
Uses of Cable		●			
Incident Type/Descrip.			●		
Permitted Activity				●	
Value of Alteration				●	
Permit Type				●	

Persons

Groups

Unknown

Whittington et al. *Push Pull and Spill*. BTLJ, 2015.

Geospatial re-identifiability

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
APPLIED PHYSICS
APPLIED MATHEMATICS
STATISTICS
COMPUTATIONAL SCIENCE

Received
1 October 2012

Accepted
4 February 2013

Published
25 March 2013

Correspondence and requests for materials should be addressed to Y.-A. de M. (yva@mit.edu)

Unique in the Crowd: The privacy bounds of human mobility

Yves-Alexandre de Montjoye[1,2], César A. Hidalgo[1,3,4], Michel Verleysen[2] & Vincent D. Blondel[2,5]

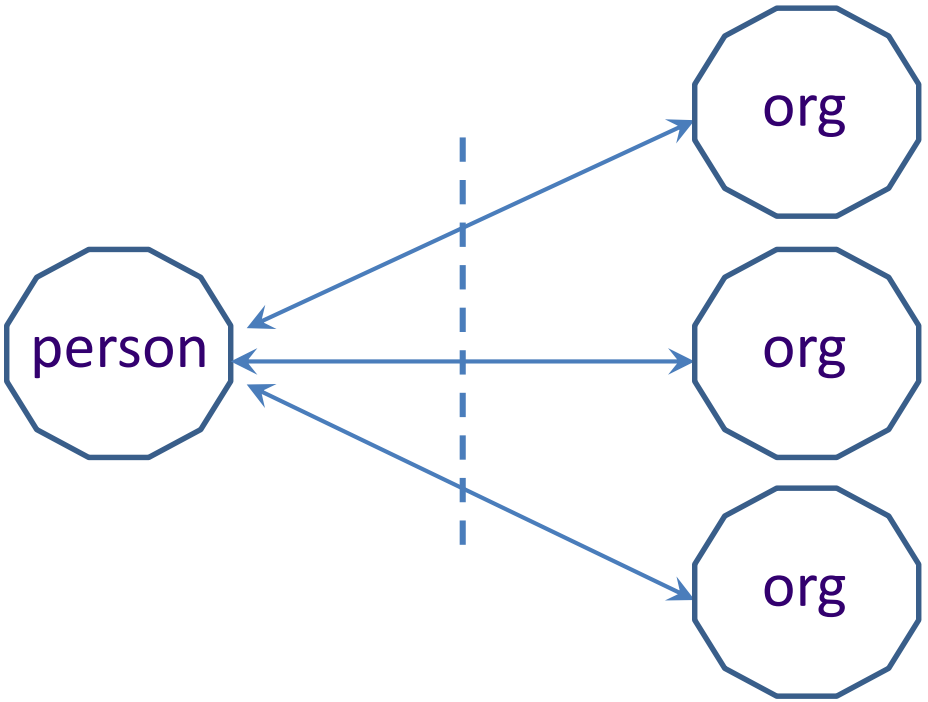
[1]Massachusetts Institute of Technology, Media Lab, 20 Ames Street, Cambridge, MA 02139 USA, [2]Université catholique de Louvain, Institute for Information and Communication Technologies, Electronics and Applied Mathematics, Avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium, [3]Harvard University, Center for International Development, 79 JFK Street, Cambridge, MA 02138, USA, [4]Instituto de Sistemas Complejos de Valparaíso, Paseo 21 de Mayo, Valparaíso, Chile, [5]Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals. We coarsen the data spatially and temporally to find a formula for the uniqueness of human mobility traces given their resolution and the available outside information. This formula shows that the uniqueness of mobility traces decays approximately as the 1/10 power of their resolution. Hence, even coarse datasets provide little anonymity. These findings represent fundamental constraints to an individual's privacy and have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals.

Derived from the Latin Privatus, meaning "withdraw from public life," the notion of privacy has been foundational to the development of our diverse societies, forming the basis for individuals' rights such as

Privacy as a TCE problem

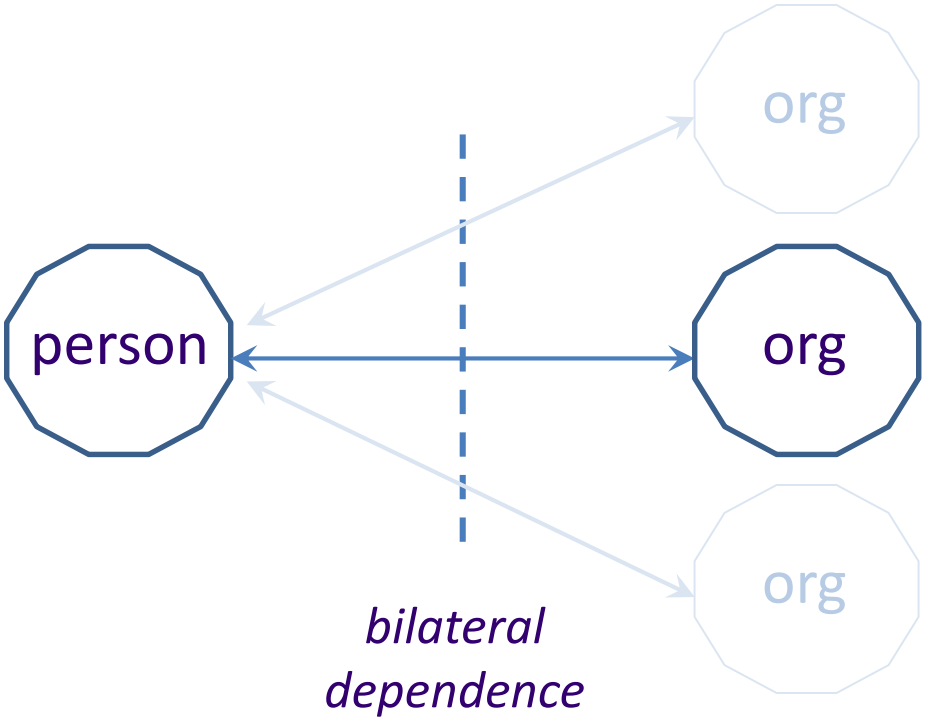
Presumption of competition for consumers in data markets



Hoofnagle and Whittington. Free. UCLA, 2014.
Whittington and Hoofnagle. *Unpacking Privacy's Price.* NCLR, 2012.

Privacy as a TCE problem

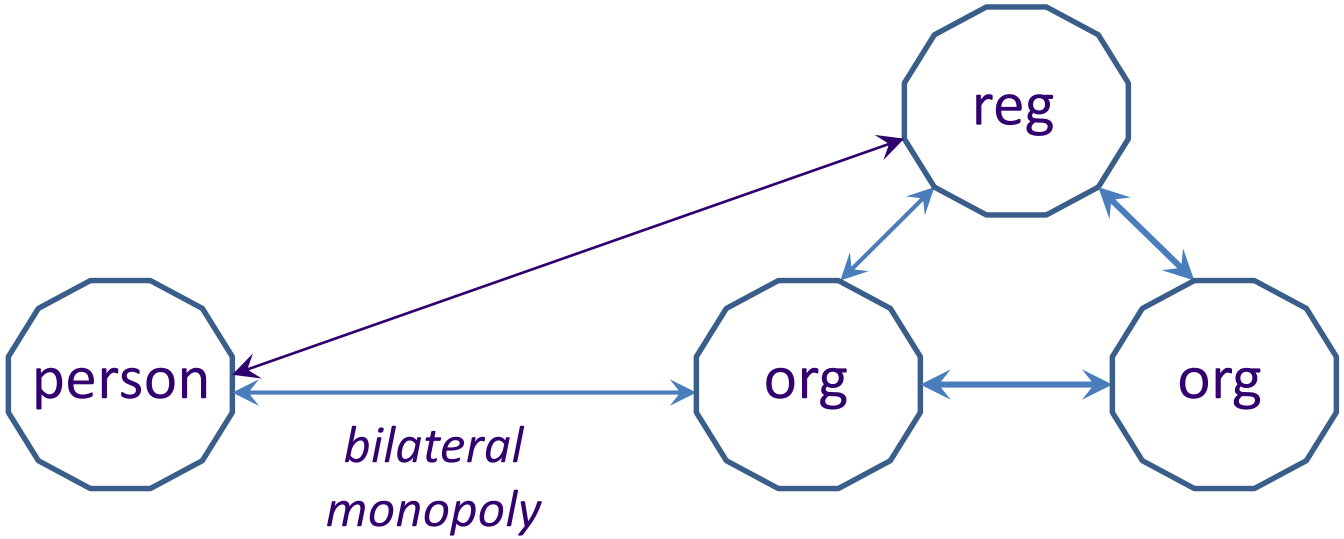
Identification of the consumer presents bilateral dependence



Hoofnagle and Whittington. Free. UCLA, 2014.
Whittington and Hoofnagle. *Unpacking Privacy's Price.* NCLR, 2012.

Consumer Protection

Global trend toward third party or regulatory governance



Young et al. *Beyond Open v. Closed*, ACM FAT*, 2019.
Hoofnagle and Whittington. Free. UCLA, 2014.
Whittington and Hoofnagle. *Unpacking Privacy's Price.* NCLR, 2012.

Directions in City Data

Protecting privacy in data

+ Low bar of re-identifiability (math)
+ Threat varies geographically (land use and density matter)
+ Need query-based solutions (narrow mosaic ‘attack surface’)

Mitigating bias in data

+ What are the populations represented in the data?
+ Algorithmic fairness, accountability and transparency [FAT*]
+ Purpose, from the firm, public, & subjects' POV

Directions for analytics and governance

+ Merge technical & legal protection
+ Develop new institutional rules of the game [PIAs, Surveillance Acts]
+ Recognize public need for evidence-based policy
+ Inform of trajectories for re-identifiability & harm

Referenced

Meg Young, Luke Rodriguez, Emily Keller, Feiyang Sun, Boyang Sa, Jan Whittington, Bill Howe. "**Beyond open vs. closed: Balancing individual privacy and public accountability in data sharing**" *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, 2019, Atlanta, GA: Association for Computing Machinery (ACM), pp. 191-200.

Jan Whittington, Ryan Calo, Mike Simon, Jesse Woo, Meg Young, Peter Schmeideskamp. "**Push, Pull, and Spill: a Transdisciplinary Case Study in Municipal Open Government**." *Berkeley Technology Law Journal,* vol. 30, no. 2, 2015, pp. 1899-1966.

Chris Jay Hoofnagle and Jan Whittington. "**Free: Accounting for the Costs of the Internet's Most Popular Price.**" *UCLA Law Review*, vol. 61, no. 3, 2014, pp. 606-670.

Jan Whittington and Chris Jay Hoofnagle. "**Unpacking Privacy's Price**." *North Carolina Law Review*, vol. 90, no. 5, 2012, pp. 1327-1370.

Contacts

Feiyang Sun
Interdisciplinary PhD in Planning
University of Washington
fs377@uw.edu

Jan Whittington
Associate Professor
University of Washington
janwhit@uw.edu

UW Urban Infrastructure Lab

UW Tech Policy Lab

UW TDC