

# Introduction to Computational Social Science

Summer Institute in Computational Social Science @ CU Boulder  
August 13th, 2018



University of Colorado  
Boulder

**SICSS**

# Overview

- What is computational social science?
- Why the Summer Institute in Computational Social Science (SICSS)?
- Logistics: schedule, materials, topics, visiting speakers, feedback



University of Colorado  
Boulder

SICSS

# What is computational social science?

<https://github.com/compsocialscience/summer-institute/blob/master/2018/materials/day1-intro-ethics/02-intro-computational-social-science.pdf>



University of Colorado  
Boulder

**SICSS**

# What is computational social science?

*“Anything that is cool”* - Matt Salganik

<https://github.com/compsocialscience/summer-institute/blob/master/2018/materials/day1-intro-ethics/02-intro-computational-social-science.pdf>



University of Colorado  
Boulder

**SICSS**

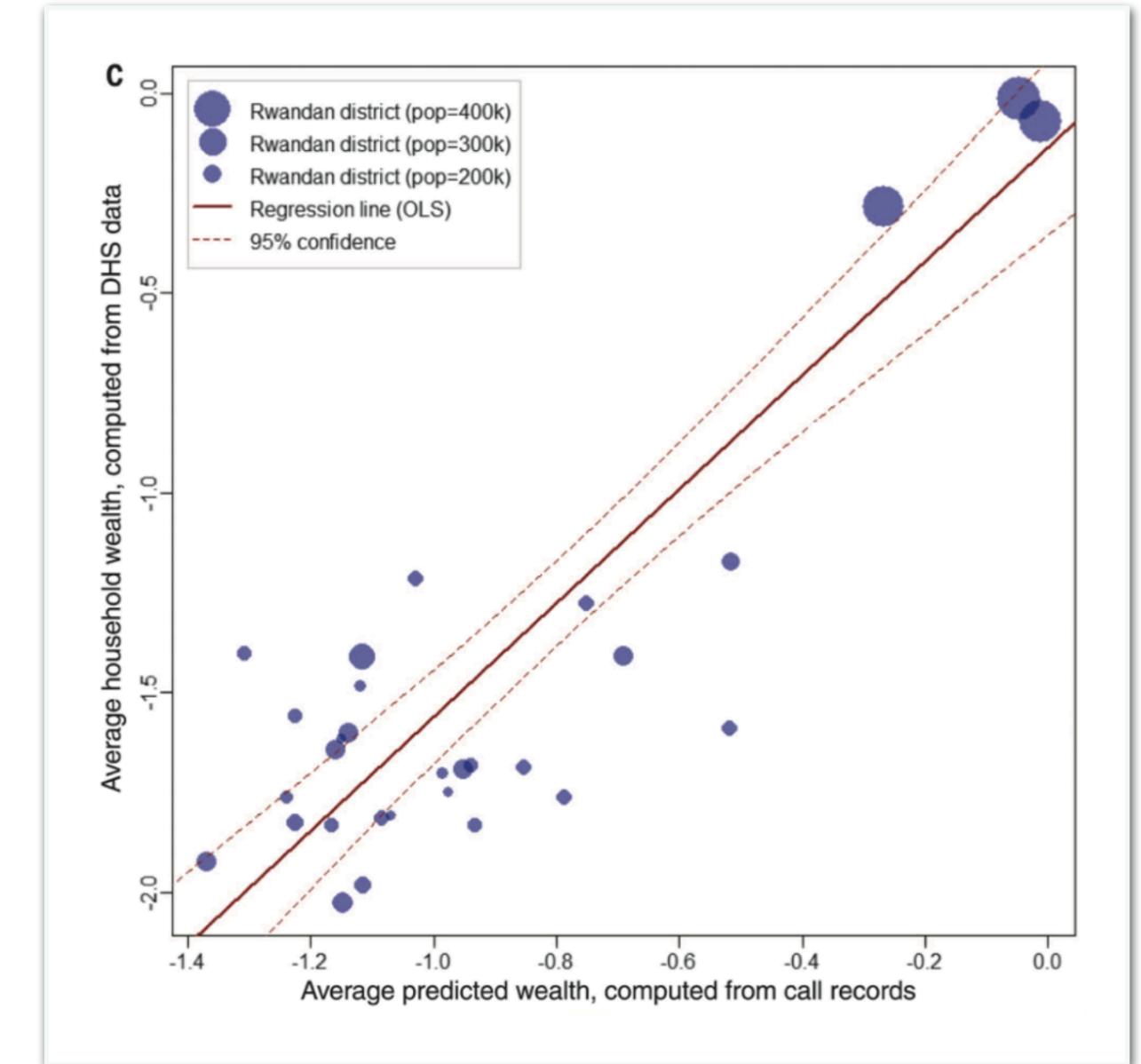
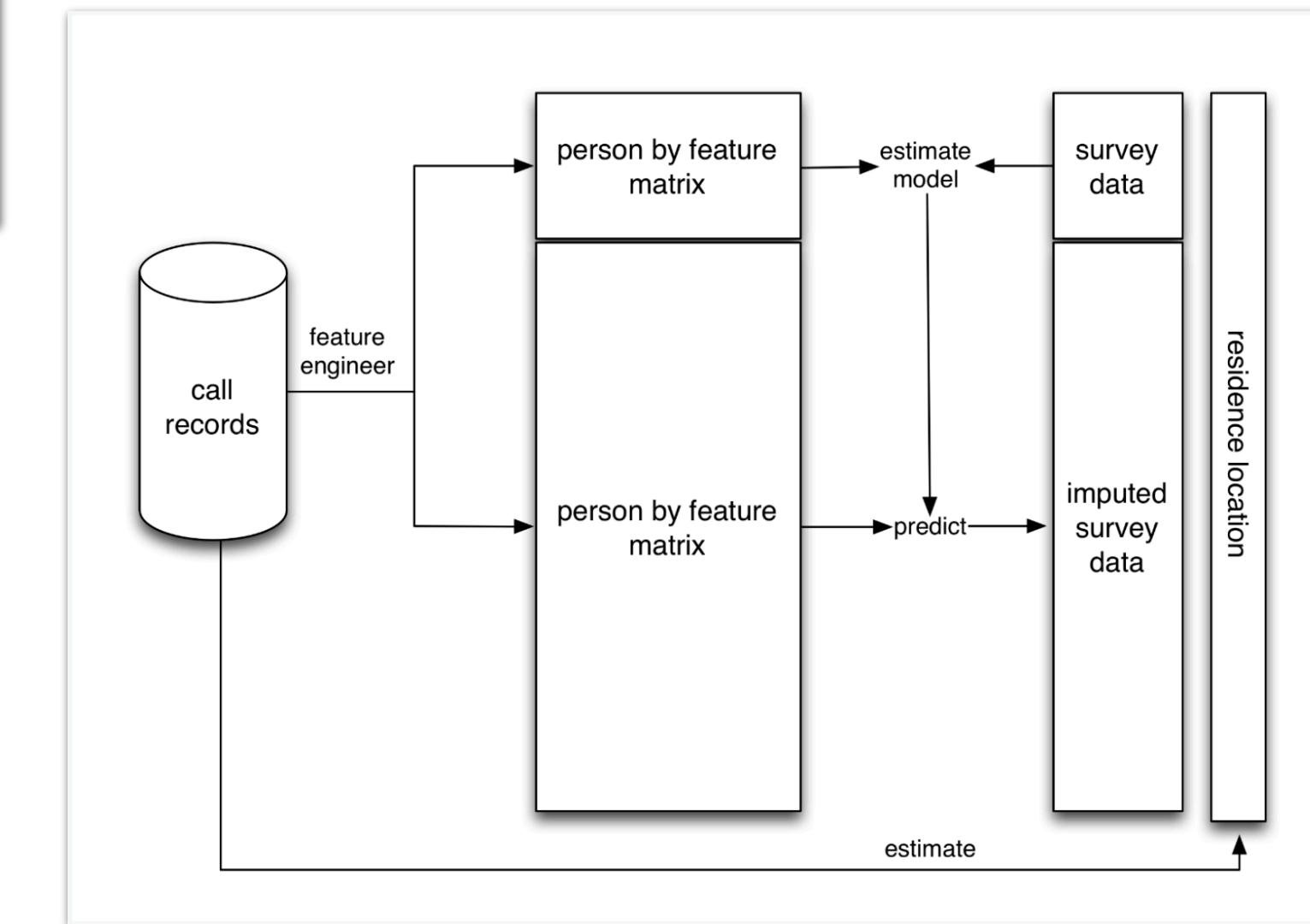
# Case Study #1

ECONOMICS

## Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,<sup>1\*</sup> Gabriel Cadamuro,<sup>2</sup> Robert On<sup>3</sup>

Science, 350(6264), 1073-1076, 2015



<https://github.com/compsocialscience/summer-institute/blob/master/2018/materials/day1-intro-ethics/02-intro-computational-social-science.pdf>



University of Colorado  
Boulder

SICSS

# Case Study #2

## Forecasting elections with non-representative polls

Wei Wang<sup>a,\*</sup>, David Rothschild<sup>b</sup>, Sharad Goel<sup>b</sup>, Andrew Gelman<sup>a,c</sup>

<sup>a</sup> Department of Statistics, Columbia University, New York, NY, USA

<sup>b</sup> Microsoft Research, New York, NY, USA

<sup>c</sup> Department of Political Science, Columbia University, New York, NY, USA

*International Journal of Forecasting*, 31(3), 980-991, 2015

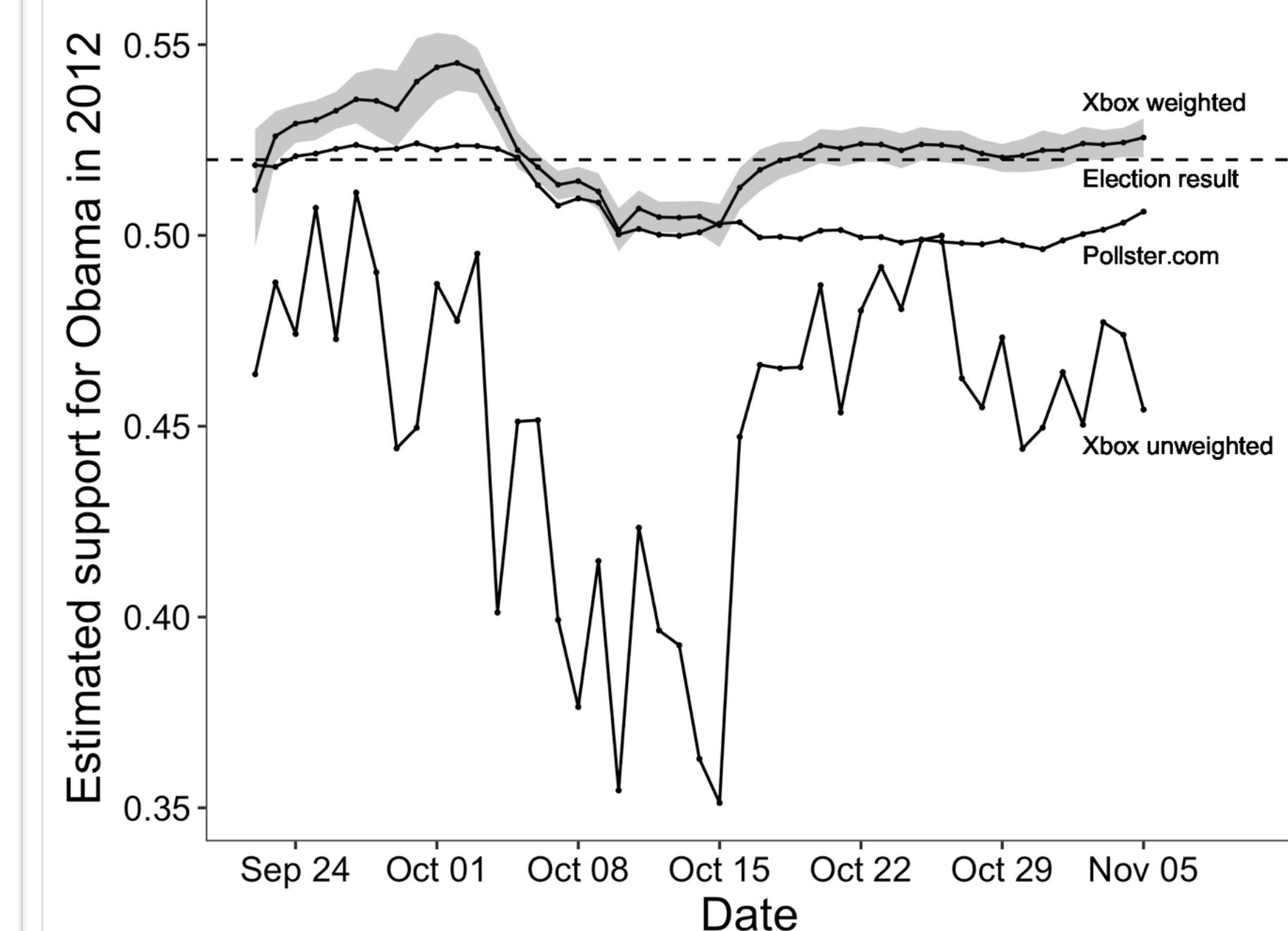


Figure 3.8: Estimates from W. Wang et al. (2015). Unadjusted XBox sample produced inaccurate estimates. But, the weighted XBox sample produced estimates that were more accurate than an average of probability-based telephone surveys. Adapted from W. Wang et al. (2015), figures 2 and 3.

Salganik 2018, p. 106



University of Colorado  
Boulder

SICSS

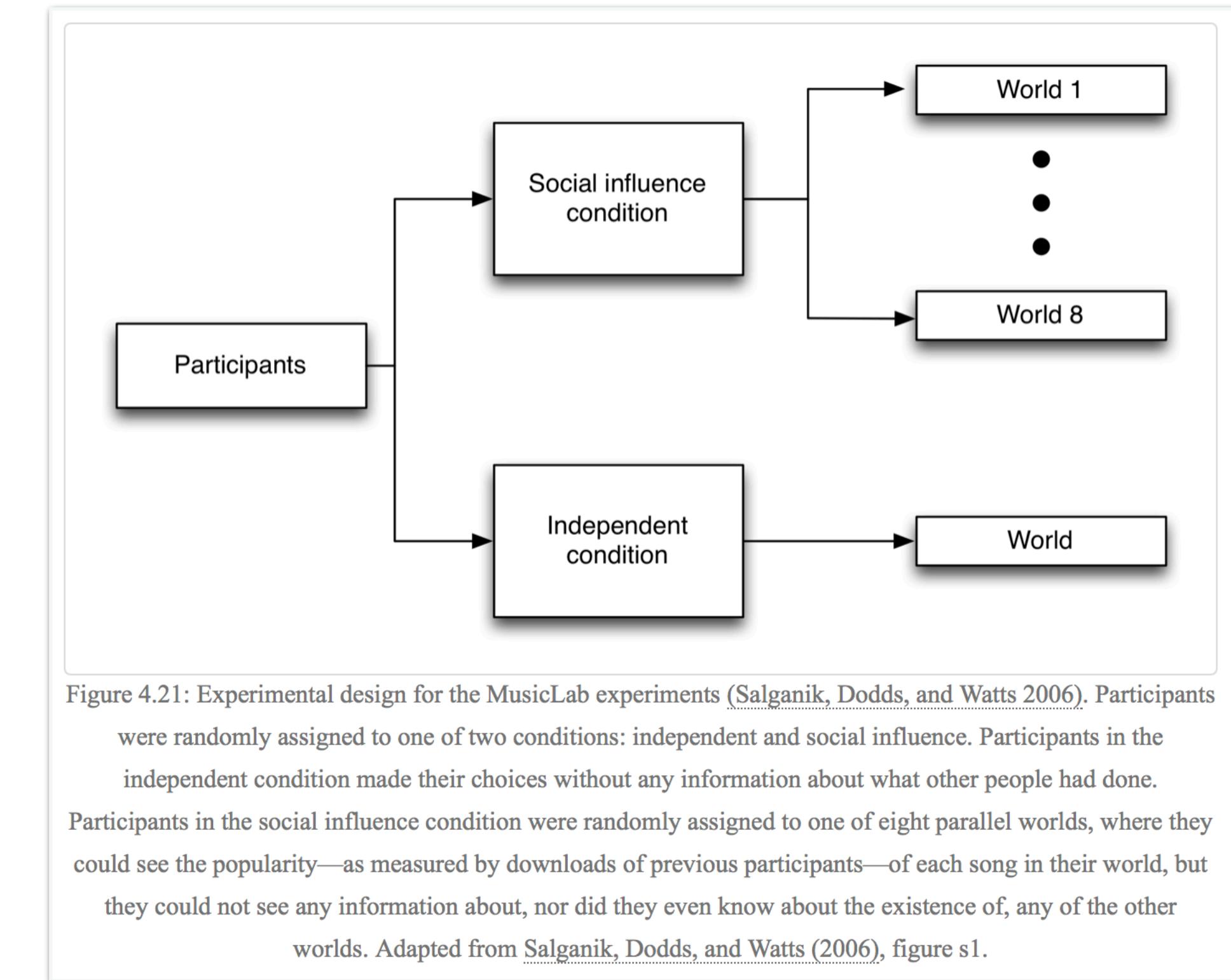
# Case Study #3

REPORTS

## Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market

Matthew J. Salganik,<sup>1,2,\*</sup> Peter Sheridan Dodds,<sup>2\*</sup> Duncan J. Watts<sup>1,2,3\*</sup>

Science, 311(5762), 854-856, 2006



Salganik 2018, p. 193



University of Colorado  
Boulder

SICSS

# What is computational social science?

- Often involves ethical/privacy questions that are now considered complex
- Combines *readymades* and *custommades* (Salganik 2017)
- Involves five key communities: social science, data science, business people, privacy advocates, policy makers

<https://github.com/compsocialscience/summer-institute/blob/master/2018/materials/day1-intro-ethics/02-intro-computational-social-science.pdf>



University of Colorado  
Boulder

SICSS

# What is computational social science to you?

**The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks**

Christopher A. Bail<sup>a</sup>

## Computational Social Science

David Lazer,<sup>1</sup> Alex Pentland,<sup>2</sup> Lada Adamic,<sup>3</sup> Sinan Aral,<sup>2,4</sup> Albert-László Barabási,<sup>5</sup> Devon Brewer,<sup>6</sup> Nicholas Christakis,<sup>1</sup> Noshir Contractor,<sup>7</sup> James Fowler,<sup>8</sup> Myron Gutmann,<sup>3</sup> Tony Jebara,<sup>9</sup> Gary King,<sup>1</sup> Michael Macy,<sup>10</sup> Deb Roy,<sup>2</sup> Marshall Van Alstyne<sup>2,11</sup>

Using Facebook ad data to track the global digital gender gap

Masoomali Fatehkia <sup>a</sup>, Ridhi Kashyap <sup>b,\*</sup>, Ingmar Weber <sup>c</sup>

Big Data,  
Digital Media,  
and  
Computational  
Social Science:  
Possibilities and  
Perils

... and many more!

**“Voters of the Year”: 19 Voters Who Were Unintentional Election Poll Sensors on Twitter**

William Hobbs, Lisa Friedland, Kenneth Joseph, Oren Tsur, Stefan Wojcik, David Lazer

**Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data**

Pablo Barberá

Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks

Sinan Aral, Dylan Walker

**Impact of heterogeneity and socioeconomic factors on individual behavior in decentralized sharing ecosystems**

Arnaud Gavaldà-Miralles<sup>a,b,c</sup>, David R. Choffnes<sup>d</sup>, John S. Otto<sup>e</sup>, Mario A. Sánchez<sup>e</sup>, Fabián E. Bustamante<sup>e</sup>, Luís A. N. Amaral<sup>c,f,g,h</sup>, Jordi Duch<sup>a,1</sup>, and Roger Guimerà<sup>b,i</sup>

**Women through the glass ceiling: gender asymmetries in Wikipedia**

Claudia Wagner<sup>1,2\*</sup> , Eduardo Graells-Garrido<sup>3</sup>, David Garcia<sup>4</sup> and Filippo Menczer<sup>5</sup>



University of Colorado  
Boulder

**SICSS**

# Social scientists ↔ Data / computer scientists

Computer science	Social science
Study anything	Study social things
Methods driven	Question driven
Large found data	Small designed data
Prediction	Explanation

*“[C]omputer scientists may be interested in finding the needle in the haystack—such as [...] the right Web page to display from a search—but social scientists are more commonly interested in characterizing the haystack.” (Hopkins & King 2010)*

<https://cacm.acm.org/magazines/2018/3/225484-computational-social-science-computer-science-social-data/fulltext>

[http://videolectures.net/icml2015\\_wallach\\_social\\_science/](http://videolectures.net/icml2015_wallach_social_science/)

<https://github.com/compsocialscience/summer-institute/blob/master/2018/materials/day1-intro-ethics/02-intro-computational-social-science.pdf>



University of Colorado  
Boulder

**SICSS**

# What is SICSS?

## History of the Summer Institute in Computational Social Science (SICSS):

- 2017: Single program led by Chris Bail & Matt Salganik at Princeton University
- 2018: Seven satellite programs organized by alumni, most running concurrently with the main program (mid-June)

SICSS @ CU Boulder is primarily based on the 2018 program (+ network analysis -  
(surveys in the digital age + mass collaboration))

Material is modeled after Matt Salganik's book *Bit by Bit*: [bitbybitbook.com](http://bitbybitbook.com)

[https://cbail.github.io/Why\\_SICSS.html#/](https://cbail.github.io/Why_SICSS.html#/)



University of Colorado  
Boulder

**SICSS**

# Why SICSS?

- Goal #1: Provide state of the art training
- Goal #2: Challenge disciplinary divides
- Goal #3: Reach a broad audience
- Goal #4: Open-source
- Goal #5: Teach the Teachers

[https://cbail.github.io/Why\\_SICSS.html#/](https://cbail.github.io/Why_SICSS.html#/)



University of Colorado  
Boulder

**SICSS**

# Schedule & Materials

- Online: <https://compsocialscience.github.io/summer-institute/2018/boulder/#schedule>
- Topics:
  - Ethics (Brian)
  - Collecting Digital Trace Data (Allie & Yotam)
  - Network Analysis (Aaron & Dan)
  - Automated Text Analysis (Allie & Brian)
  - Experiments & Causal Inference (Brian & Amanda)
- Organized around skill-based lectures, small group activities, and research talks



University of Colorado  
Boulder

**SICSS**

# Schedule & Materials

Discuss -sion	Lectures & Group Activities	Research Talks
Discuss -sion	Lectures & Group Activities	Research Talks
Discuss -sion	Lectures & Group Activities	Research Talks
Discuss -sion	Lectures & Group Activities	Research Talks
Discuss -sion	Lectures & Group Activities	Research Talks

Materials for the group activities and slides will be posted the day of:

[https://compsocialscience.github.io/  
summer-institute/2018/boulder/  
#schedule](https://compsocialscience.github.io/summer-institute/2018/boulder/#schedule)



University of Colorado  
Boulder

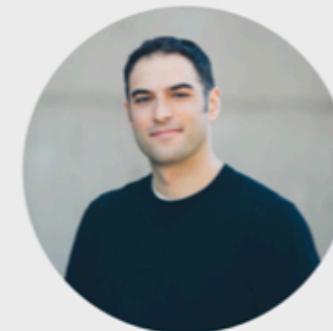
**SICSS**

# Speakers



## Aaron Clauset

Aaron Clauset is an Assistant Professor in the Department of Computer Science and the BioFrontiers Institute at the University of Colorado Boulder, and is External Faculty at the Santa Fe Institute. He received a PhD in Computer Science, with distinction, from the University of New Mexico, a BS in Physics, with honors, from Haverford College, and was an Omidyar Fellow at the prestigious Santa Fe Institute. In 2016, he was awarded the Erdos-Renyi Prize in Network Science. Clauset is an internationally recognized expert on network science, computational social science, and machine learning for complex systems. His work has appeared in many prestigious scientific venues, including Nature, Science, PNAS, JACM, WWW, ICWSM, STOC, SIAM Review, and Physical Review Letters. His work has also been covered in the popular press by the Wall Street Journal, The Economist, Discover Magazine, New Scientist, Wired, Miller-McCune, the Boston Globe and The Guardian.



## Jake Hofman

Jake Hofman is a Senior Researcher at Microsoft Research in New York City, where he works in the field of computational social science. Prior to joining Microsoft, he was a Research Scientist in the Microeconomics and Social Systems group at Yahoo! Research. He holds a B.S. in Electrical Engineering from Boston University and a Ph.D. in Physics from Columbia University. He is an Adjunct Assistant Professor of Applied Mathematics and Computer Science at Columbia University and runs Microsoft's Data Science Summer School to promote diversity in computer science. His work has been published in journals such as Science, Proceedings of the National Academy of Sciences, and Management Science, and has been featured in popular outlets including The New York Times, The Wall Street Journal, The Financial Times, and The Economist.



## Daniel Larremore

Daniel Larremore is an Assistant Professor in the Department of Computer Science and the BioFrontiers Institute at the University of Colorado at Boulder. His research develops statistical and inferential methods for analyzing large-scale network data, and uses those methods to solve applied problems in diverse domains, including public health and academic labor markets. In particular, his work focuses on generative models for networks, the ongoing evolution of the malaria parasite and the origins of social inequalities in academic hiring and careers. Prior to joining the University of Colorado faculty, he was an Omidyar Fellow at the Santa Fe Institute 2015-2017 and a post-doctoral fellow at the Harvard T.H. Chan School of Public Health 2012-2015. He obtained his Ph.D. in Applied Mathematics from the University of Colorado at Boulder in 2012, and holds an undergraduate degree from Washington University in St. Louis.



## Casey Fiesler

Assistant Professor Casey Fiesler is a social computing researcher who primarily studies governance in online communities, technology ethics, and fandom. She is a Senior Fellow in the Silicon Flatirons Institute for Law, Technology, and Entrepreneurship, an ATLAS fellow, and holds a courtesy appointment in Computer Science. Also a public scholar, she is a frequent commentator and speaker on topics of technology ethics and policy, as well as women in STEM (including consulting with Mattel on their computing-related Barbies). Her work is supported in part by a \$3 million collaborative National Science Foundation grant focused on empirical studies of research ethics. Fiesler holds a PhD from Georgia Tech in Human-Centered Computing and a JD from Vanderbilt University Law School.



## Yotam Shmargad

Yotam Shmargad is a computational social scientist with an interest in political networks and privacy. In his research, he runs experiments, links and analyzes large datasets, and uses natural experiments to study how digital media augment the patterns of connectivity between people – the size, density, and diversity of our social networks - and the implications that these bigger networks have for our social and political lives. Shmargad's recent projects look at how political candidates can overcome financial shortcomings with Twitter, and how the partisan composition of one's social network influences the information they choose to share online. Before joining the University of Arizona as an Assistant Professor, Shmargad received his PhD in Marketing from Northwestern University's Kellogg School of Management. He holds an MS in Operations Management from Columbia University and a BS in Mathematics from UCLA.



## Amanda Stevenson

Amanda Jean Stevenson is a sociologist trained in demographic and computer science methods. She studies the impacts of and responses to abortion and family planning policy. She is an Assistant Professor of Sociology at the University of Colorado Boulder. In her current research, she uses demographic methods to study the impacts of reproductive health policies, and computational and qualitative methods to study social responses to these policies. At Boulder she leads a team using massive administrative data at the Census Bureau to evaluate the life course consequences of access to (as opposed to use of) highly effective contraception. And she contributes to a variety of ongoing evaluations of reproductive health policies and develops new strategies for measuring fertility with administrative data. Another line of research examines the social responses to reproductive health policies. In a current project, she uses Twitter responses, website content, media coverage, and in-depth interviews to examine the social movement response to Texas' 2013 abortion restrictions.



## Chenhao Tan

Chenhao Tan is an assistant professor of computer science at University of Colorado Boulder. He obtained his PhD degree in the Department of Computer Science at Cornell University and bachelor's degrees in computer science and in economics from Tsinghua University. Prior to joining CU Boulder, he spent a year at University of Washington as a postdoc. His research interests include natural language processing and computational social science. He has published papers primarily at ACL and WWW, and also at KDD, WSDM, ICWSM, etc. His work has been covered by many news media outlets, such as the New York Times and the Washington Post. He also won a Facebook fellowship and a Yahoo! Key Scientific Challenges award.



## Michael Paul

Michael Paul is an Assistant Professor of Information Science at the University of Colorado Boulder. His research analyzes and develops methods in machine learning and natural language processing, and applies those methods to problems in public health and epidemiology. He is a pioneer of using novel data sources from social media for public health monitoring, summarized in his 2017 book, "Social Monitoring for Public Health." He received a PhD in Computer Science from Johns Hopkins University in 2015, and a BS in Computer Science from the University of Illinois at Urbana-Champaign in 2009.



University of Colorado  
Boulder

SICSS

# Feedback

- Daily surveys: Keep, stop, start format
- Overall survey at the completion of the workshop
- Reach out to Brian ([Brian.Keegan@Colorado.EDU](mailto:Brian.Keegan@Colorado.EDU)) or Allie ([Allison.Morgan@Colorado.EDU](mailto:Allison.Morgan@Colorado.EDU)) for more immediate feedback



University of Colorado  
Boulder

**SICSS**

# Your Responsibilities

Be open, patient, generous, and help foster togetherness.

Any questions?

Let's start! 

[https://cbail.github.io/Why\\_SICSS.html#/](https://cbail.github.io/Why_SICSS.html#/)



University of Colorado  
Boulder

**SICSS**