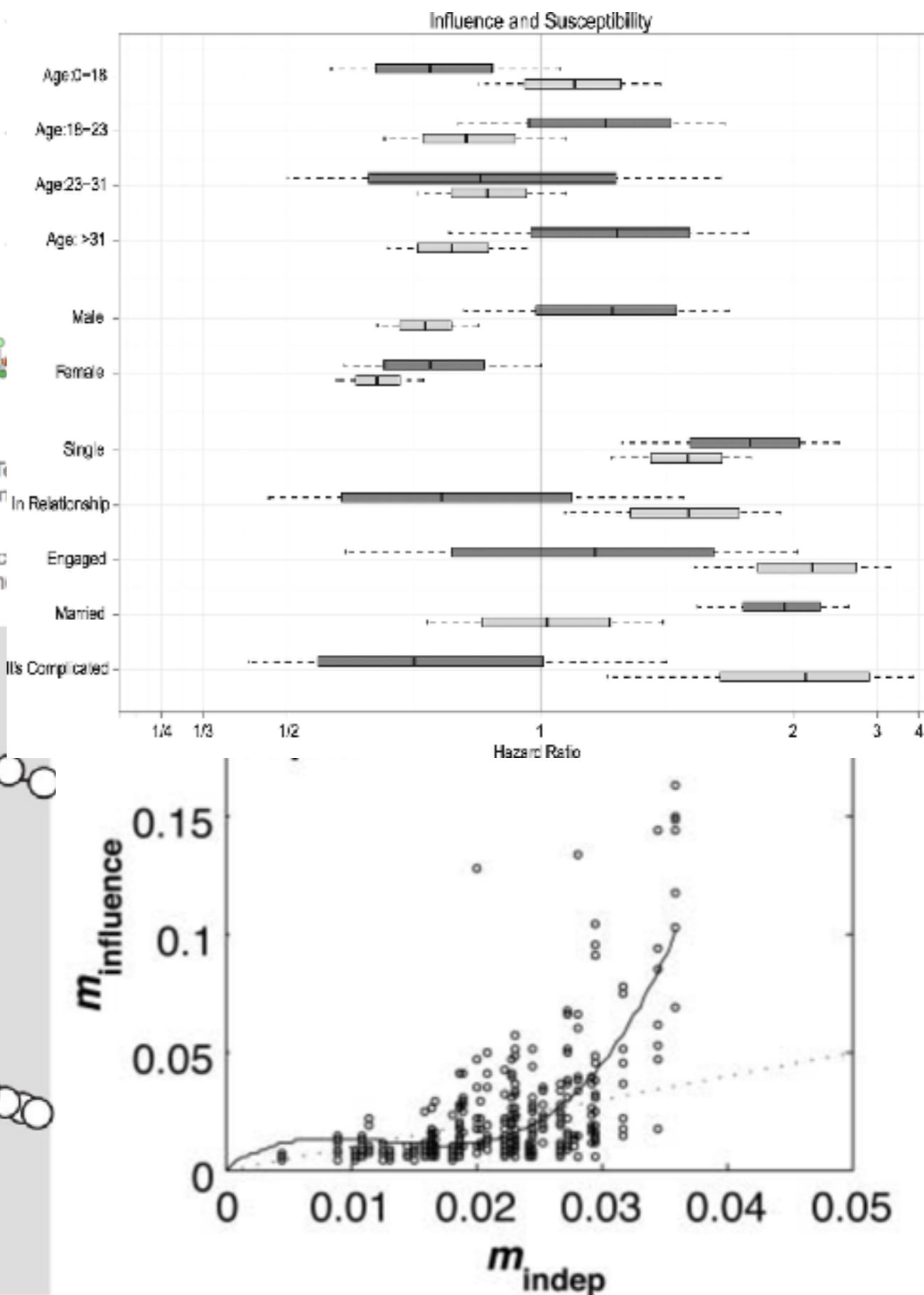
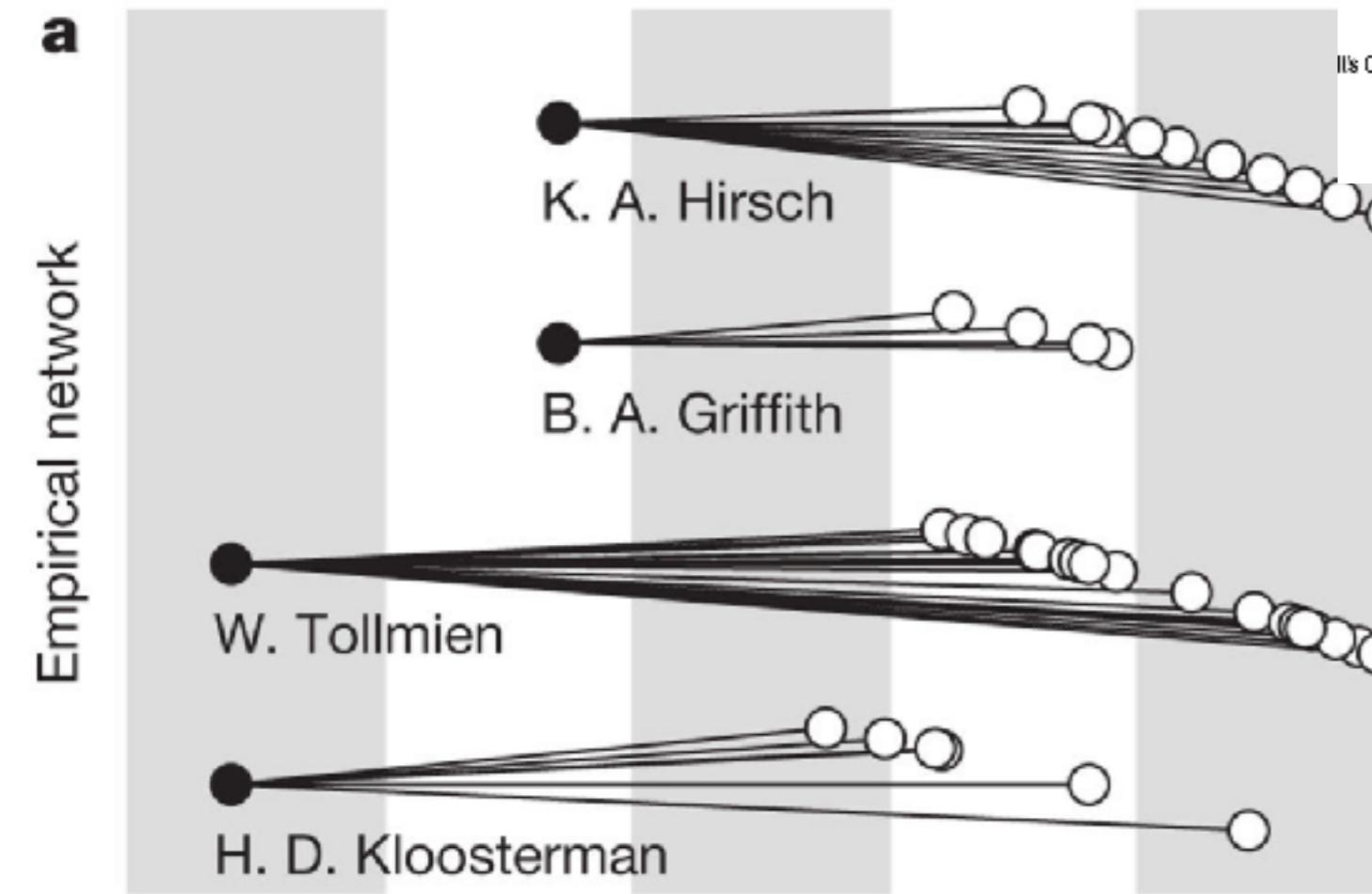
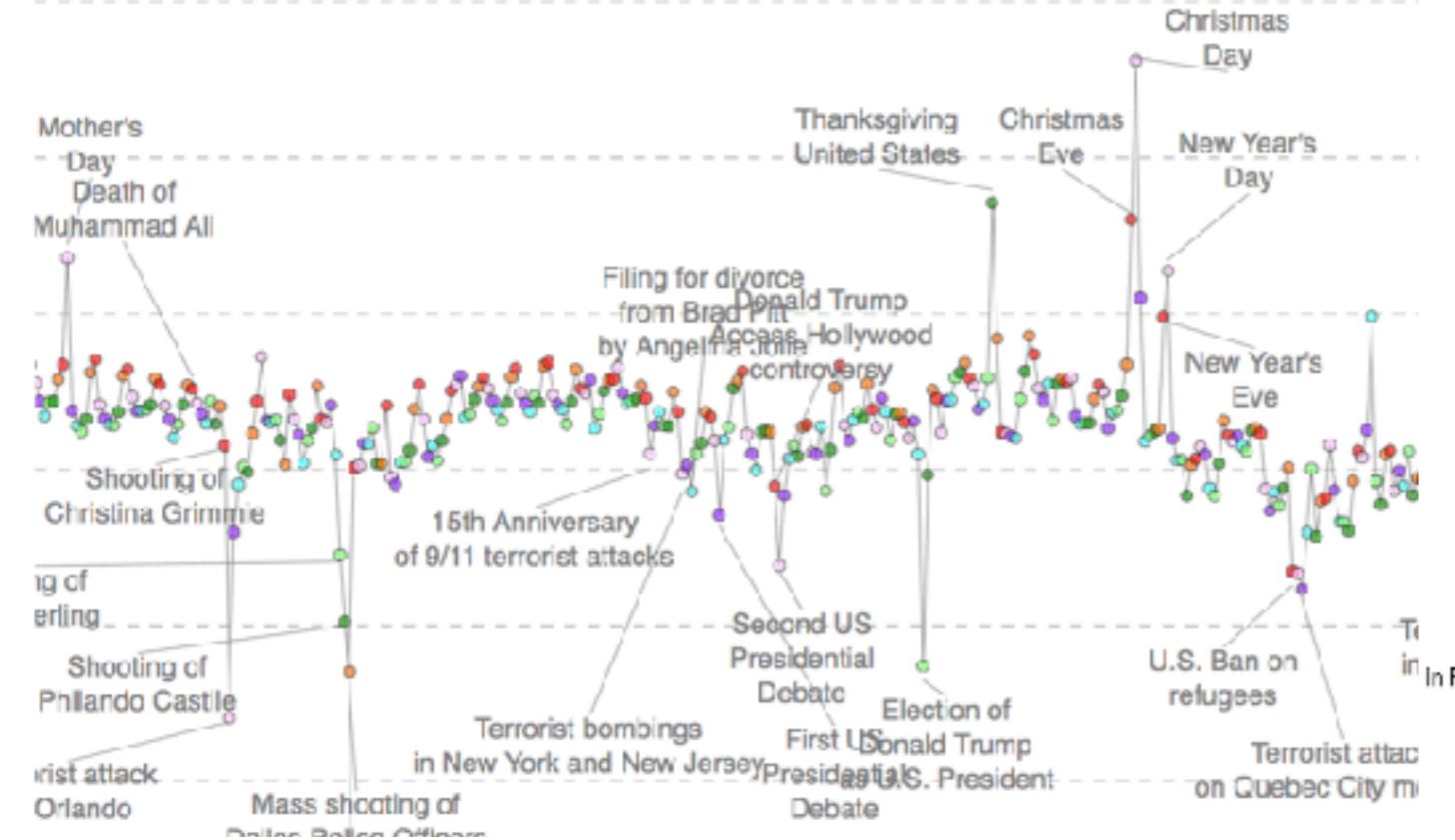


# **What is Computational Social Science?**

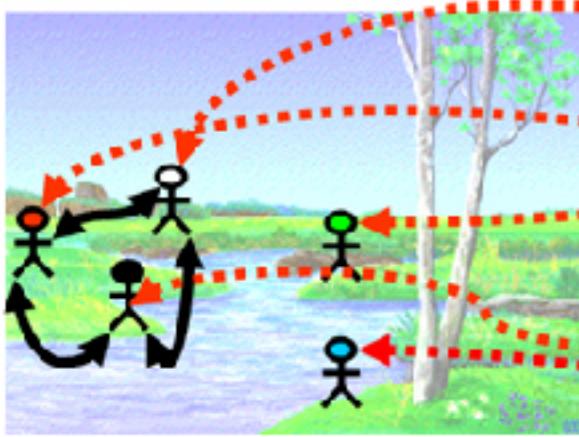
**Adam Pah  
MORS & NICO**

# Computational Social Science has enabled significant advances

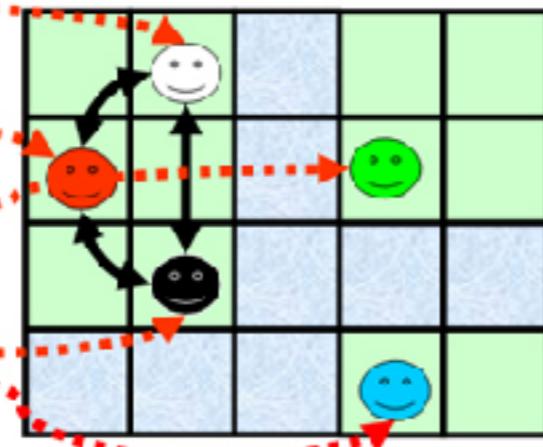


# But the definition of computational social science is not static

Target System

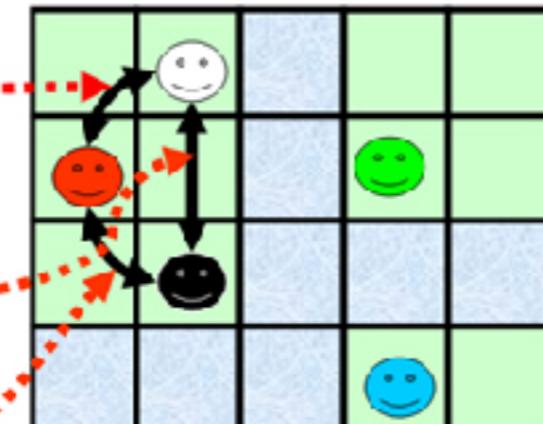
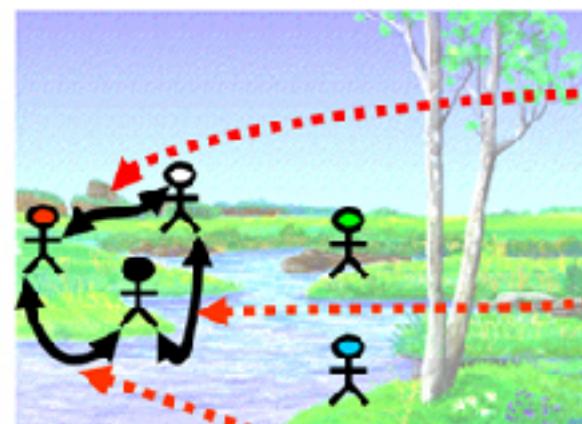


Agent based model



Entities

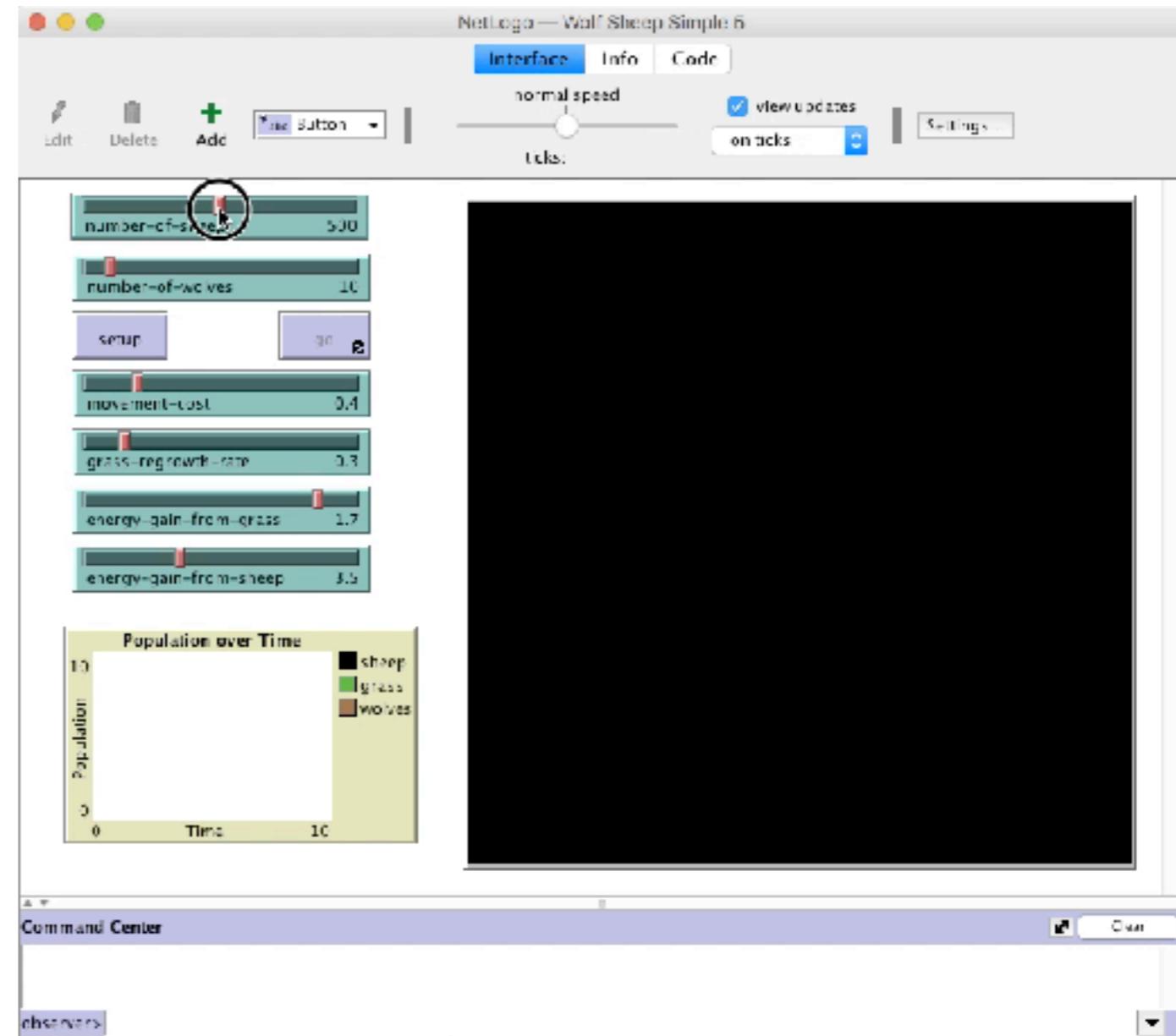
Agents



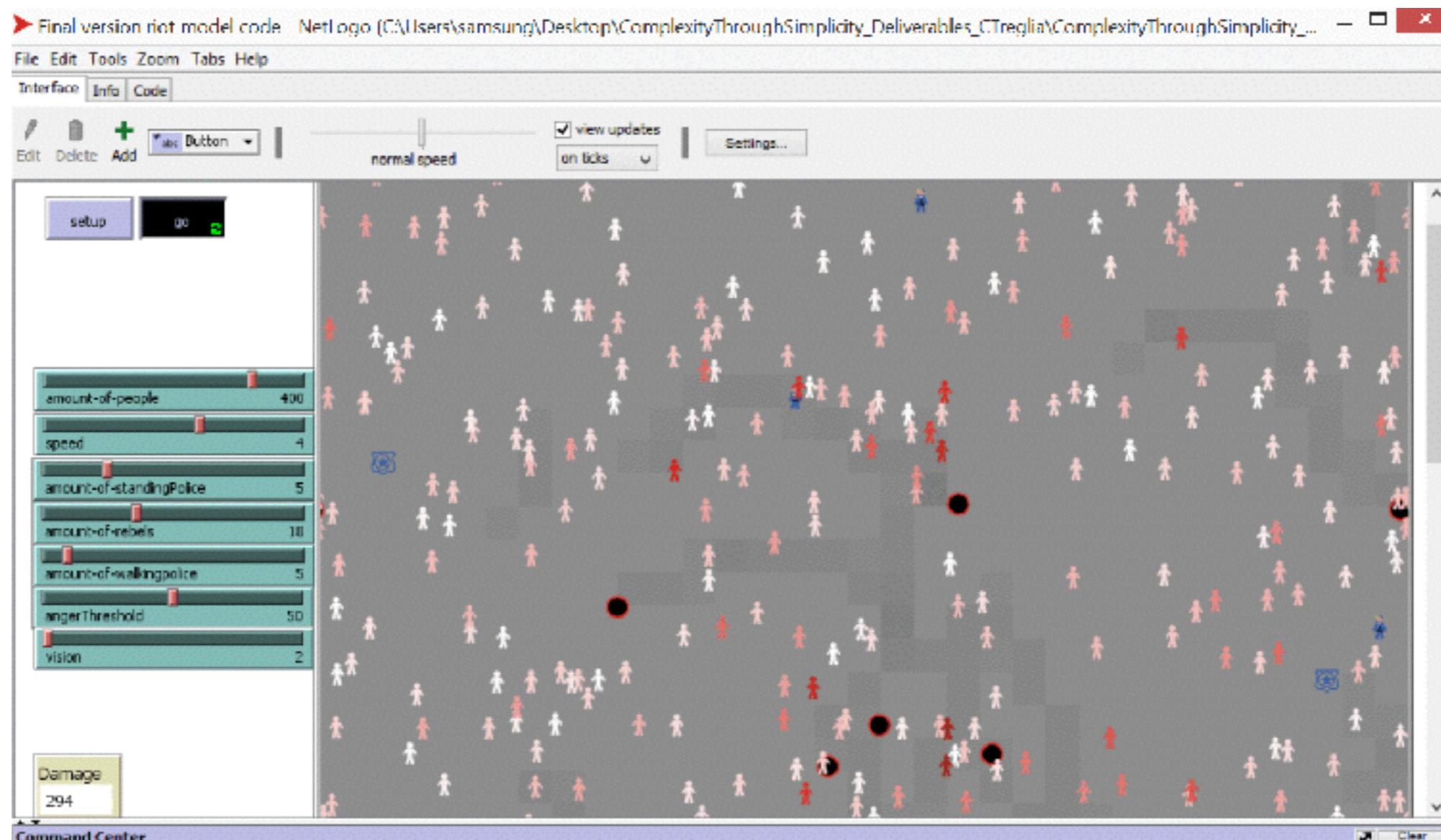
Interaction between entities

Interactions between agents

1990s to mid-2000s



# ABMs were directly analogous to historically ‘recent’ theoretical papers

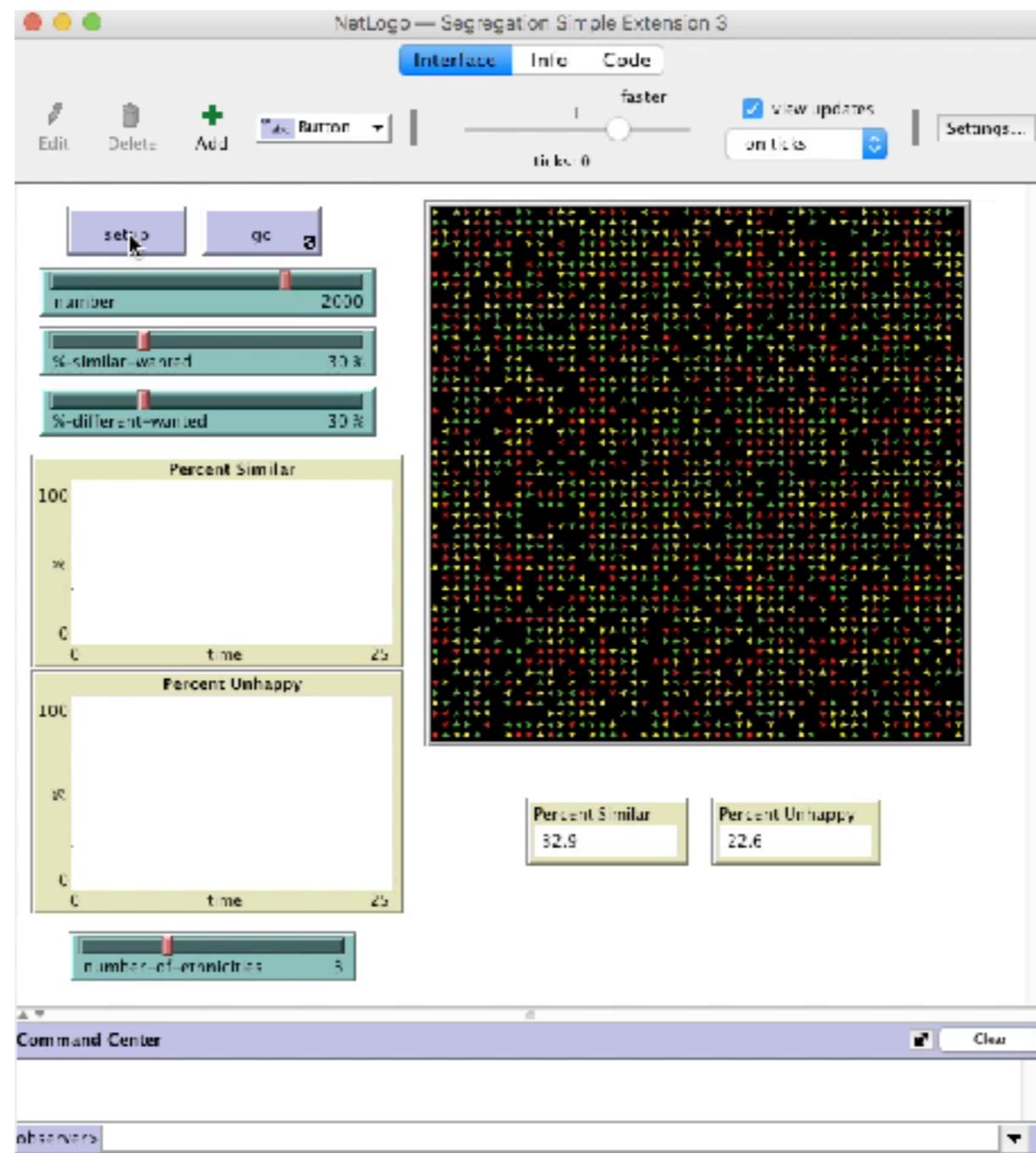


## Threshold Models of Collective Behavior

Mark Granovetter

*The American Journal of Sociology*, Vol. 83, No. 6 (May, 1978), 1420-1443.

ABMs were directly analogous to historically ‘recent’ theoretical papers



## DYNAMIC MODELS OF SEGREGATION†

THOMAS C. SCHELLING

*Journal of Mathematical Sociology* Harvard University  
1971, Vol. 1, pp 143–186

# ABMs are generative models



## Agent

Microscopic/self-similar rules  
Executes faithfully  
(with some possible noise)

**1 agent produces  
nothing of interest**

## System

Interacting agents - not all rules executed at same time  
Or locks/conflicts produced from same rule usage  
Interest is in **macroscopic pattern emergence** (qualitative)



**Difficult to match real-world conditions, so further quantitative usage is limited**

# ABMs are generative models

## Strengths

Agent rules are clear, understandable, and interpretable

Tunable parameters are generally clear  
(i.e. how similar does an agent want its neighbors to be?)

Allows for emergence (i.e. complex system phenomena)

Produces macroscopic patterns that can be contrasted with real world systems

## Drawbacks

Model, by necessity, must simplify reality to be interpretable

Difficult to match real-world initial conditions

Difficult to develop priors on importance of different rules to IRL agents

Only way to quantify importance of individual rules is through matching pattern similarity to real world

A large, pent-up desire was to increase quantification of rule importance

10 FEBRUARY 2006 VOL 311 SCIENCE [www.sciencemag.org](http://www.sciencemag.org)

## **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>**

**How much do we influence each other?**

listen



rate



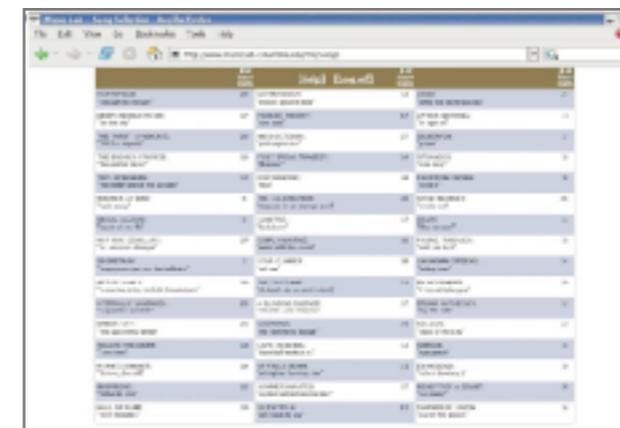
download



# no download count



# download count



listen



rate



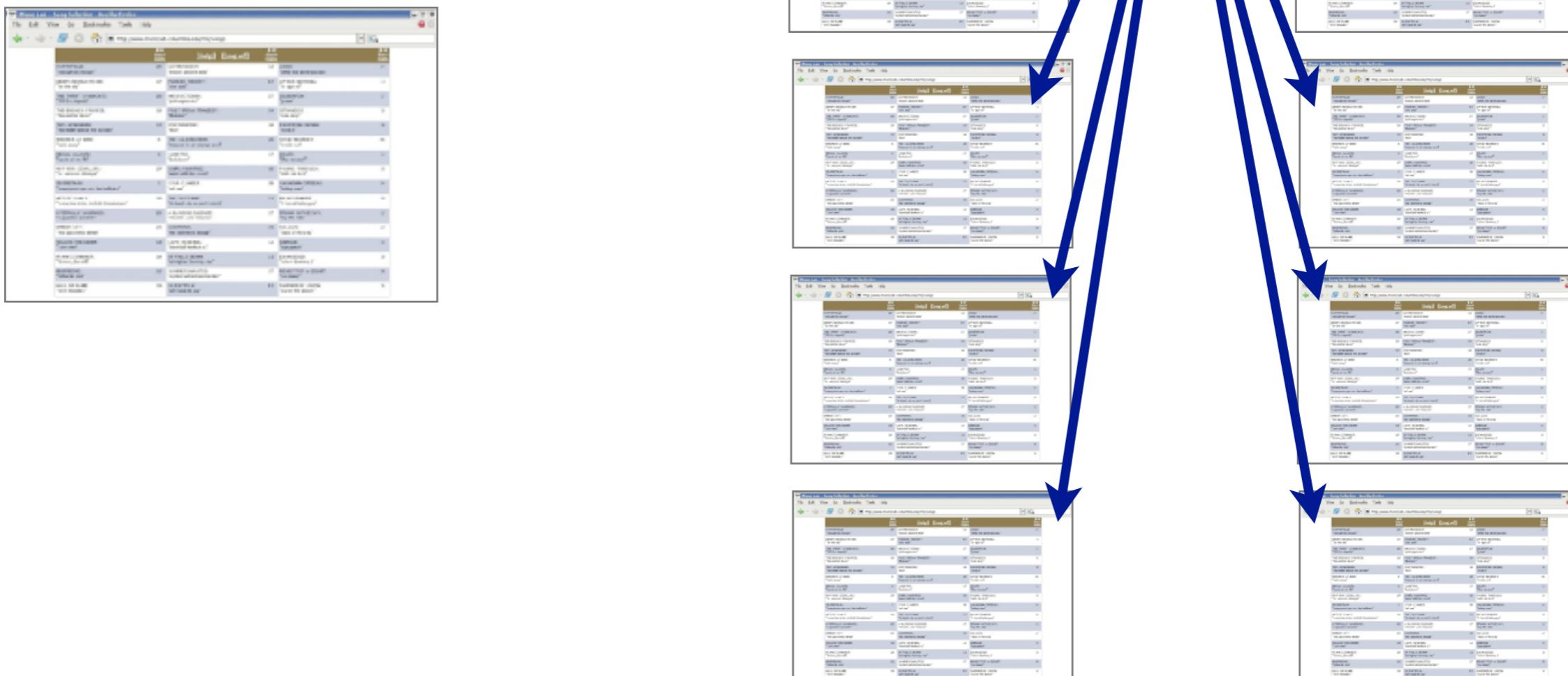
download



MUSIC LAB

# no download count

# download count



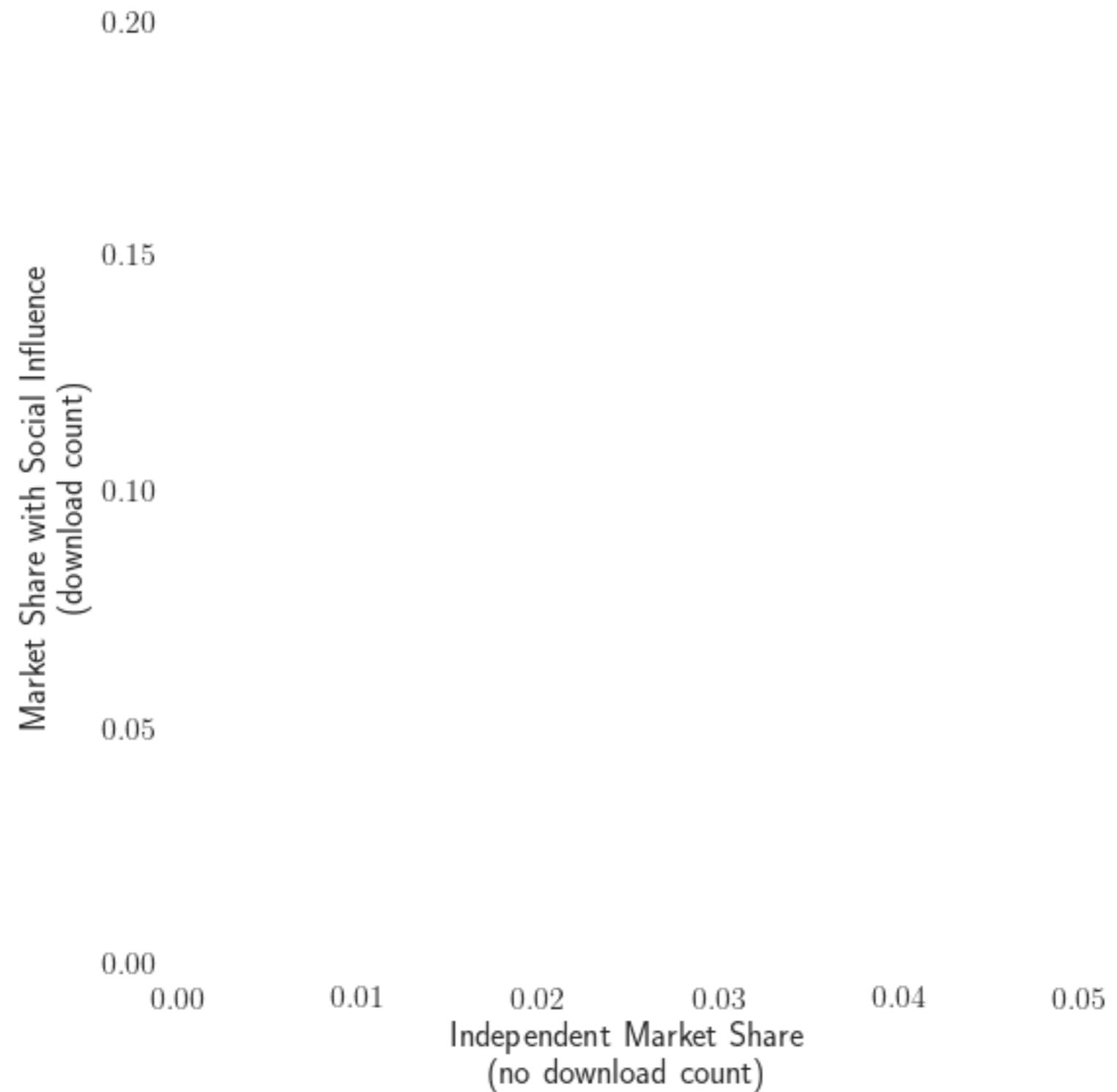
listen ★ rate ★ download ★ MUSIC LAB

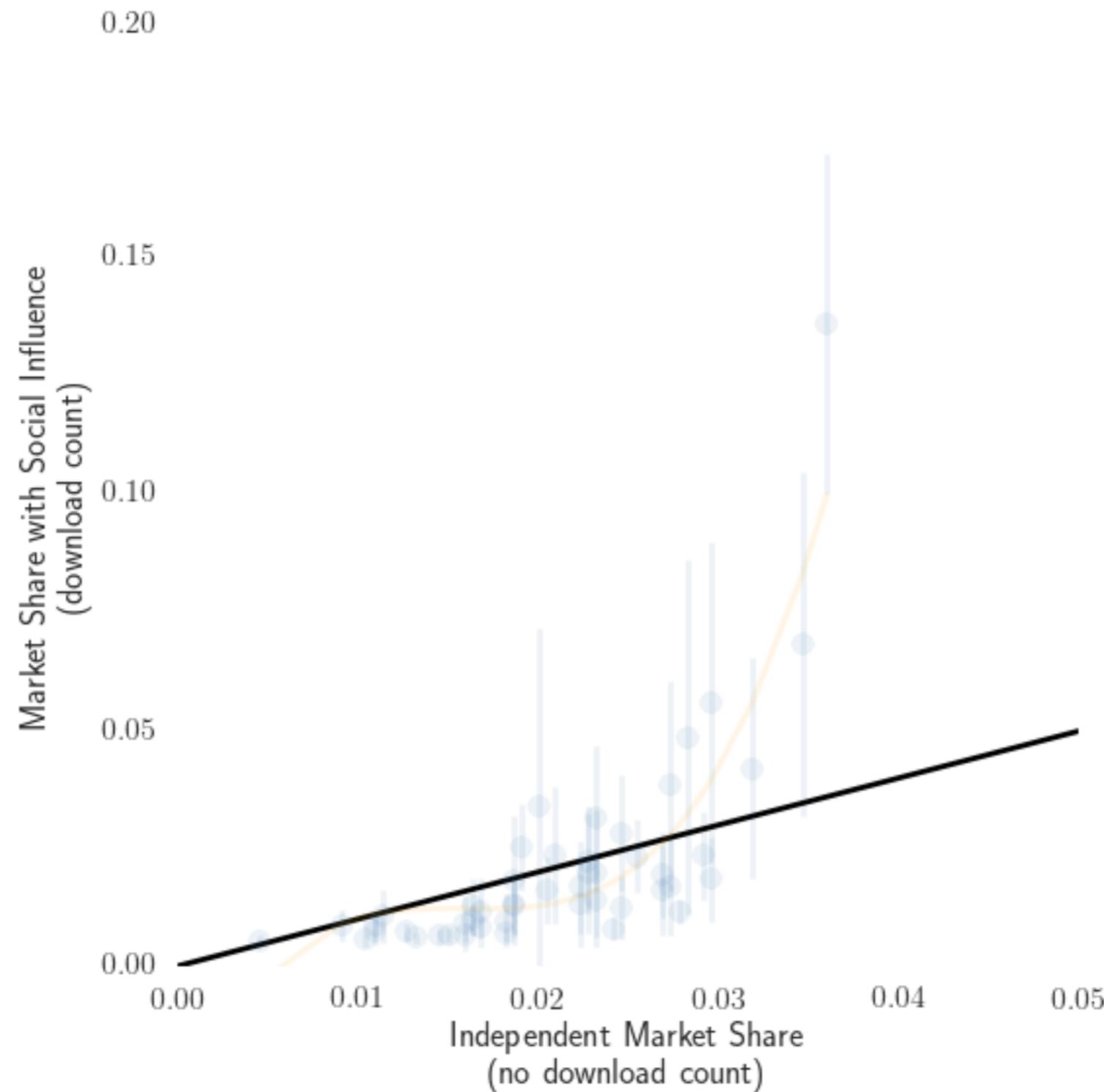
Music Lab – Song Selection - Mozilla Firefox

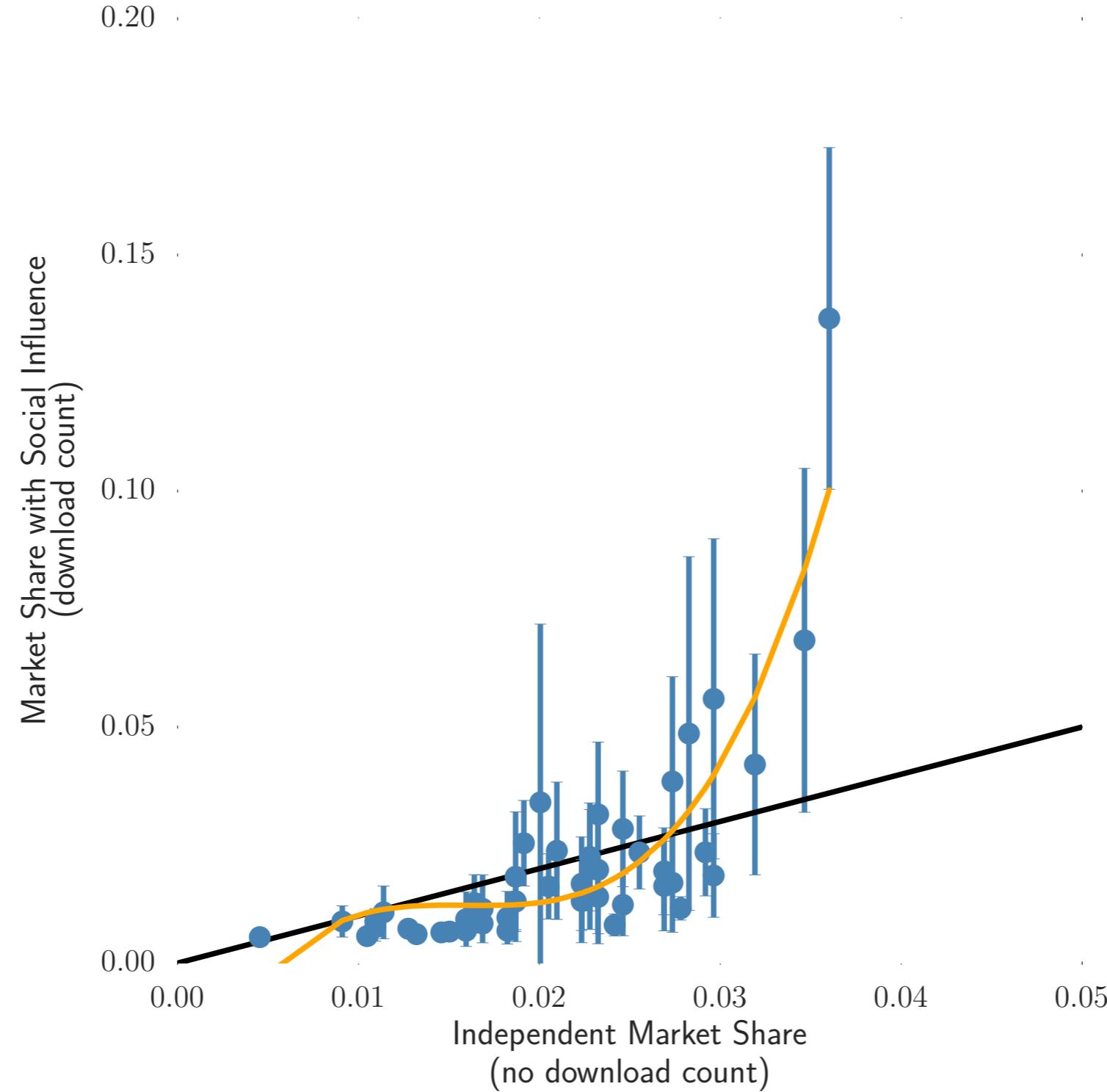
File Edit View Go Bookmarks Tools Help

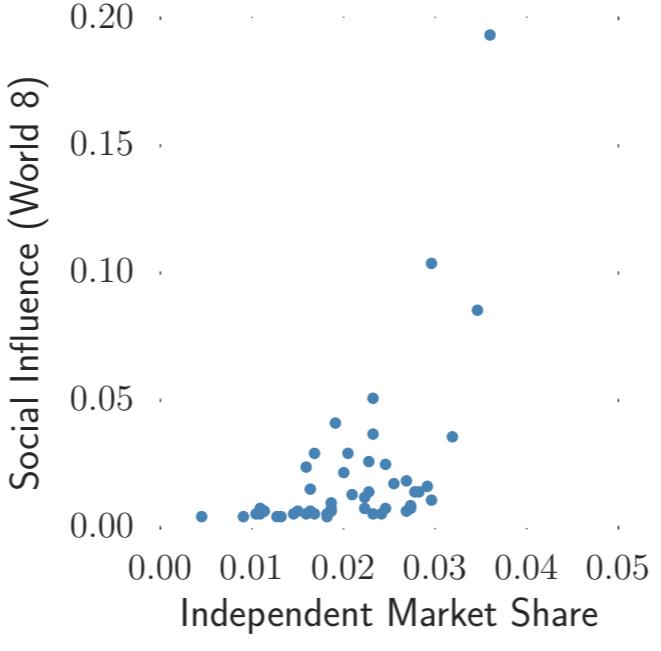
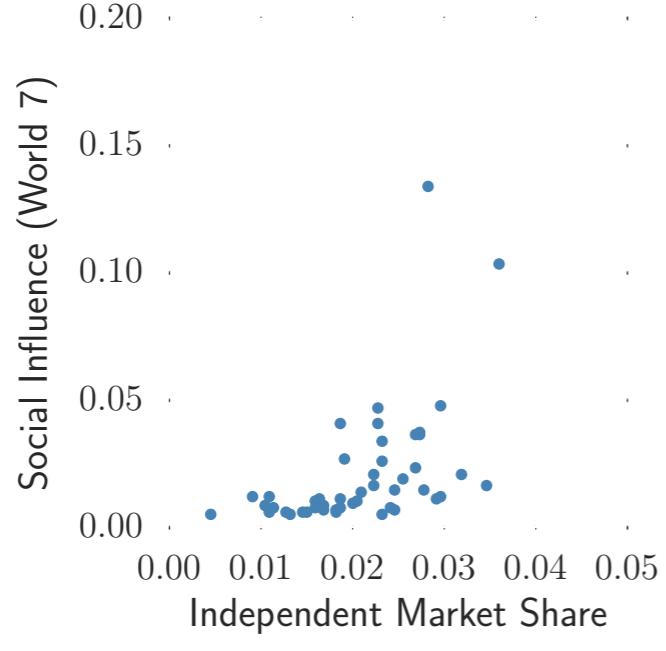
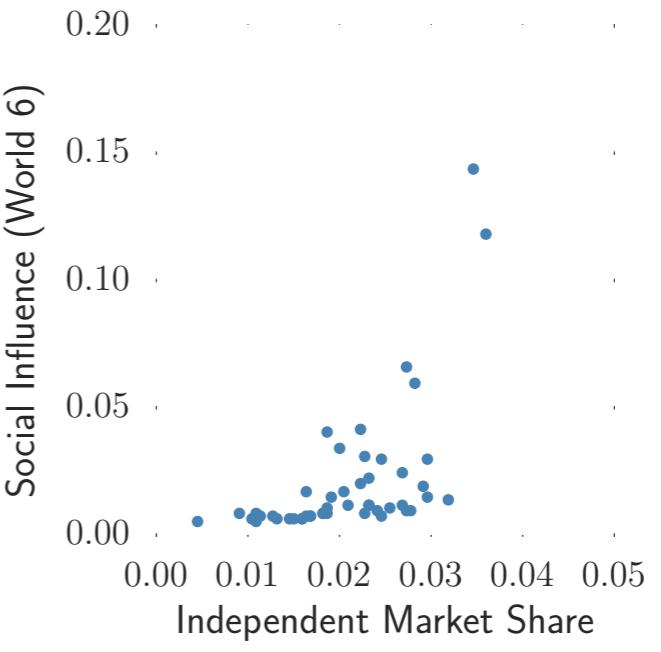
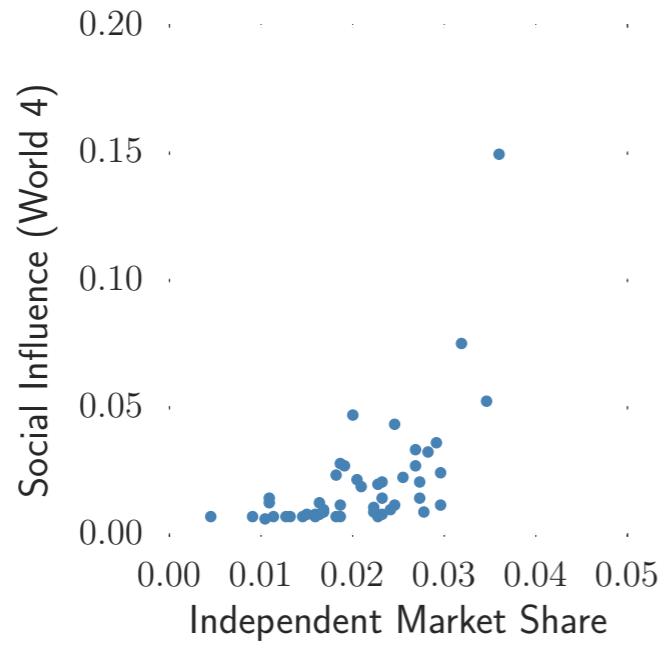
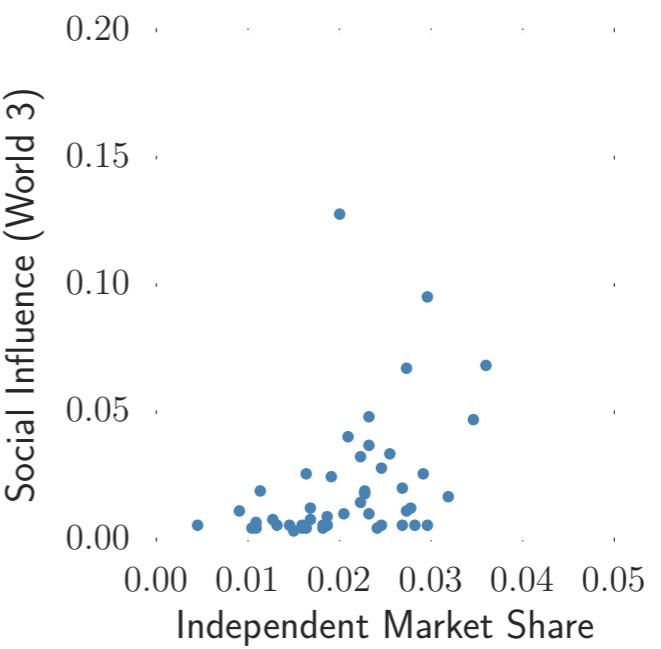
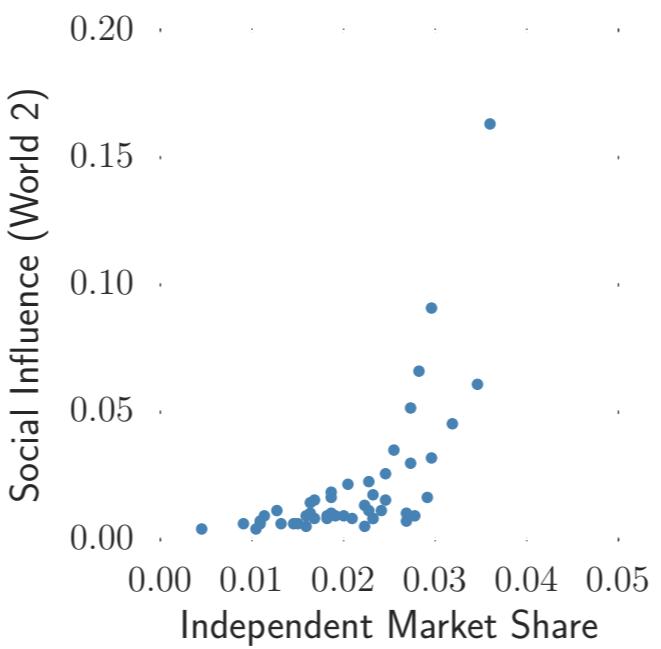
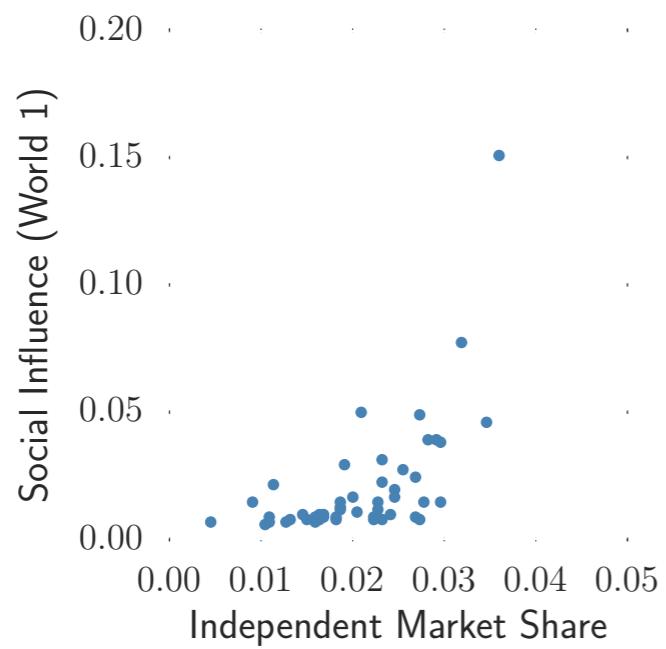
http://www.musiclab.columbia.edu/me/songs

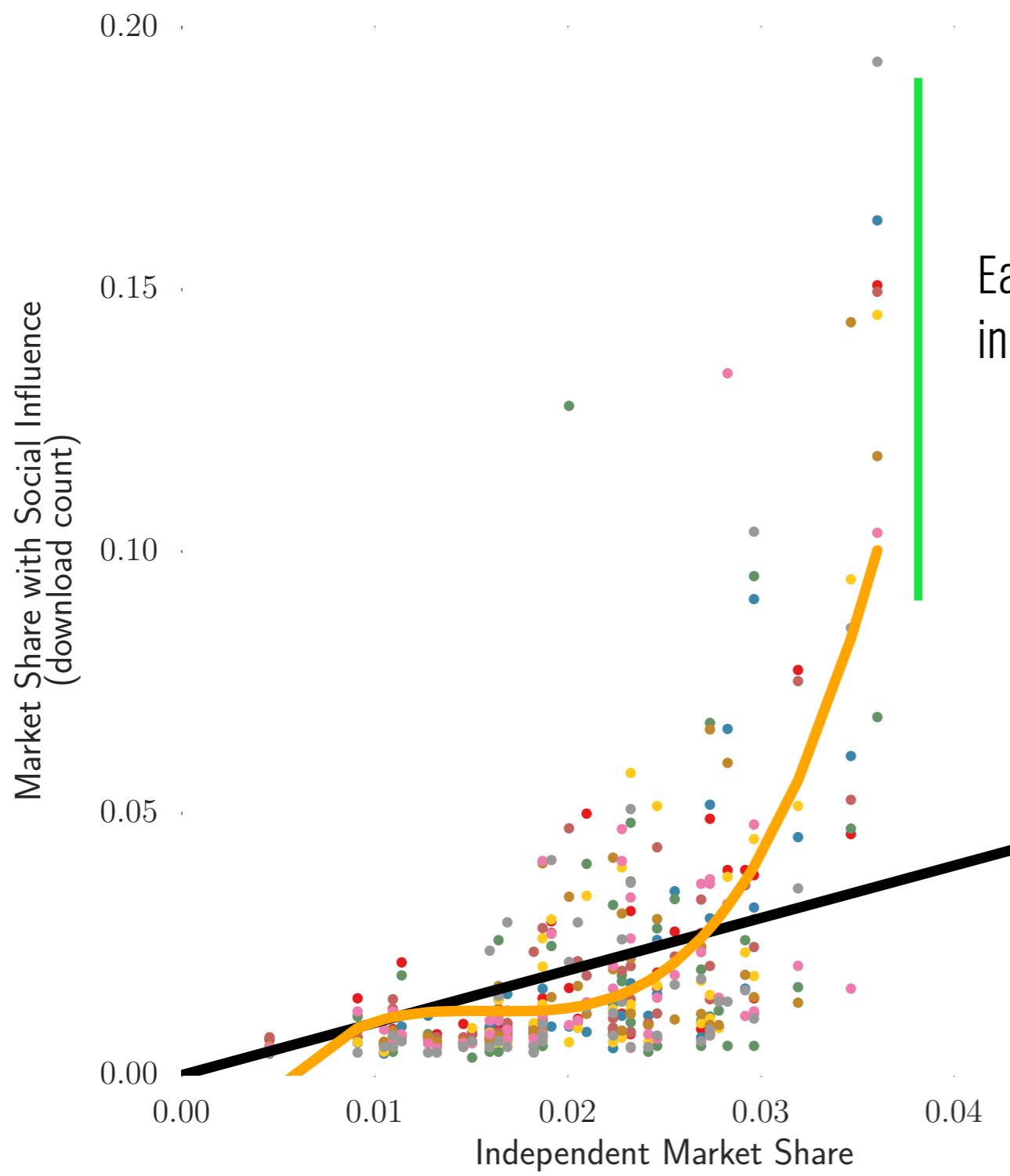
	# of down loads	[Help] [Log off]	# of down loads	# of down loads
HARTSFIELD: "enough is enough"	20	GO MOREDALE: "it does what it has"	12	UNDO: "while the world passes"
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink aggression"	27	SILVERFOX: "grow"
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "Renewal"	14	STRANGER: "one day"
THIS NEW DAWN: "the belief above the answer?"	12	FORTHFACING: "Year"	24	FAR FROM KNOWN: "route 9"
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "wide out"
MORAL HAZARD: "waste of my life"	8	S2METRO: "lockdown"	17	DANTE: "life mystery"
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the crowd"	18	FADING THROUGH: "with me luck"
SECRETARY: "keep your eyes on the balloons"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "till death do us part (I don't)"	31	BY NOVEMBER: "if i could take you"
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miseries"	17	DRAWN IN THE SKY: "top the ride"
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"
SALUTE THE DAWN: "I am here!"	13	CAPE RENEWAL: "baseball warlock v1"	12	SIBRIAN: "eye patch"
RYAN ESSMAKER: "debrief, be still"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "elbert downey jr."
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"



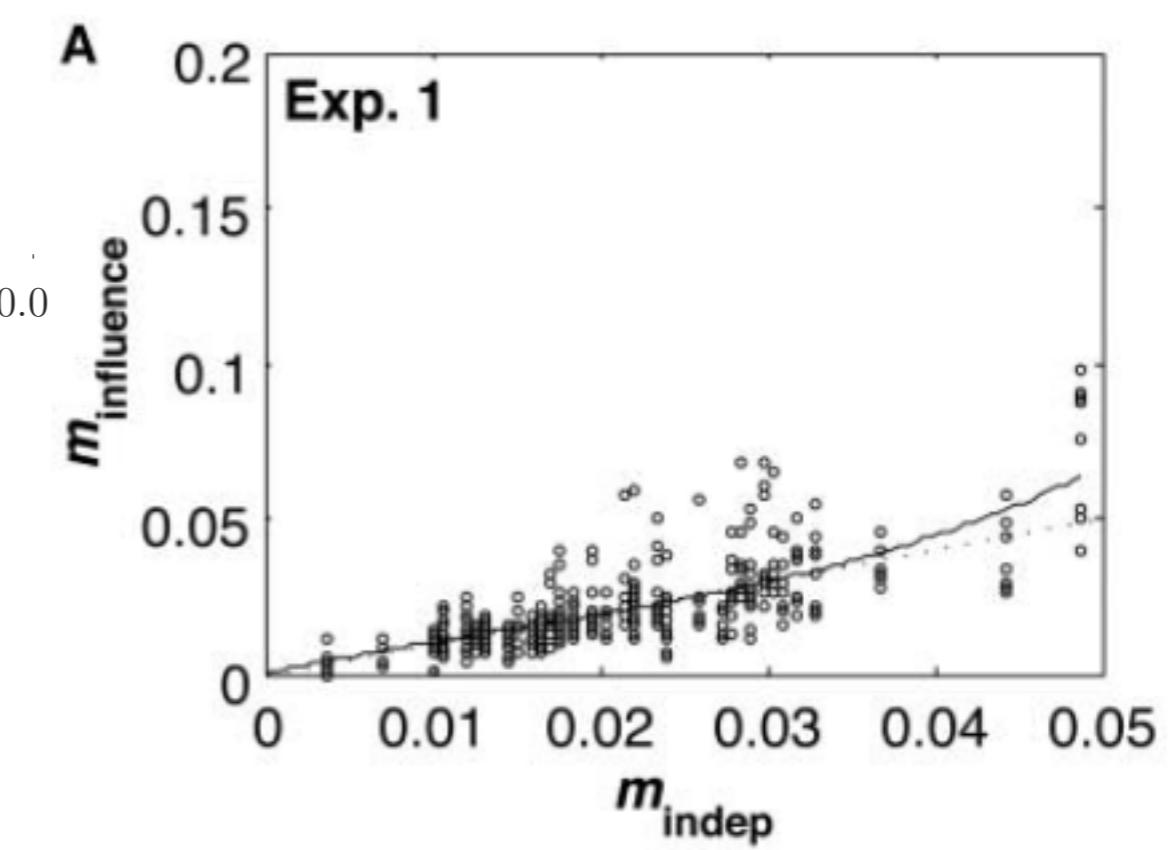








Each column is the same song  
in the 8 different worlds



# **Key advances are shown**

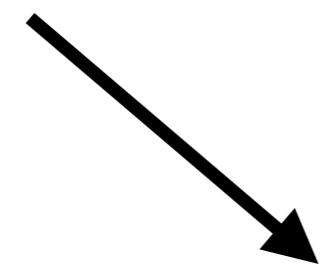
Large-*N* population (32,000 subjects) with rapid collection time-line

Ability to control experimental set-up to isolate effect in a natural setting

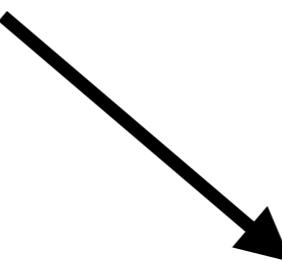
Possibility of less WEIRDs in the subject pool

# Three planks enabled this growth in CSS

Real-time access  
(the internet, sensors)

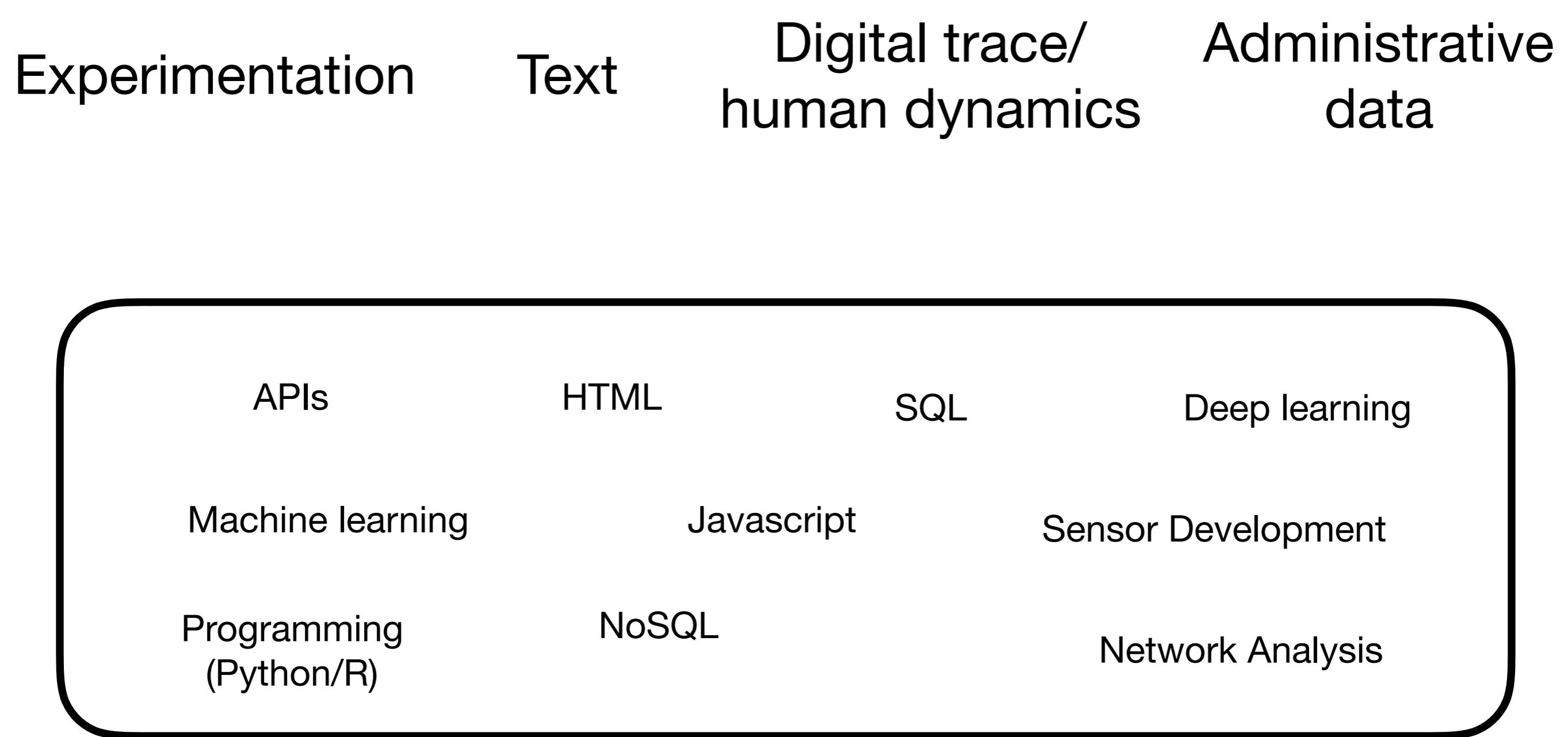


Big Data

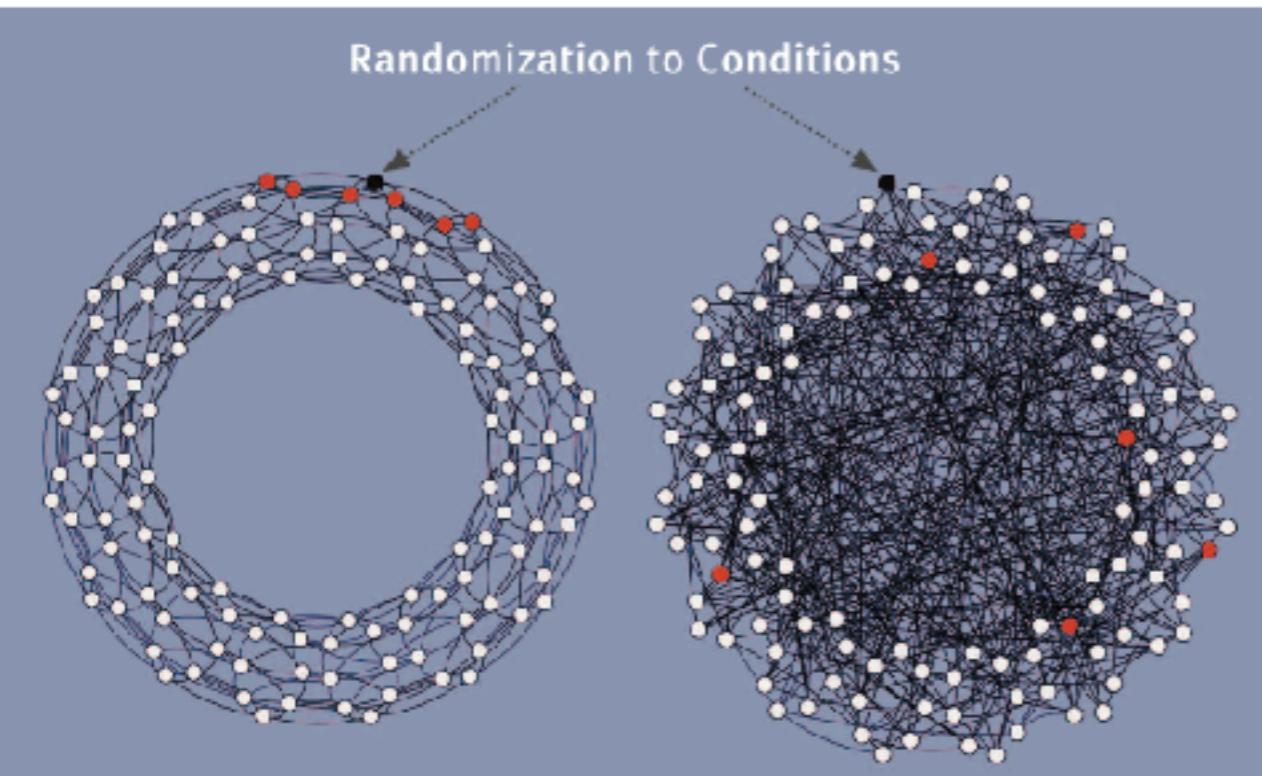


Algorithmic Maturity

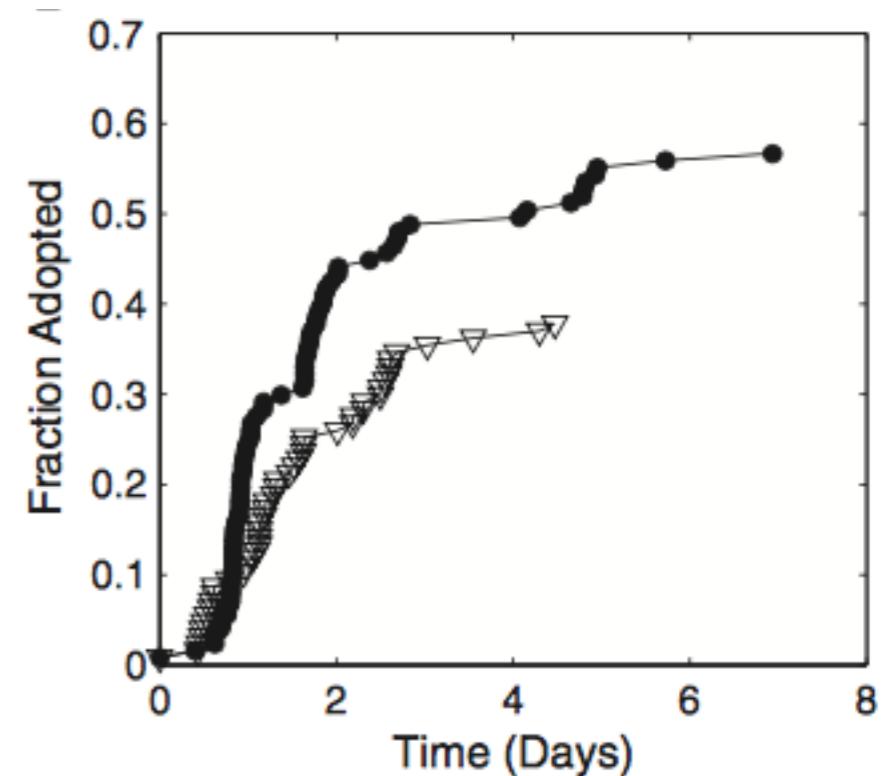
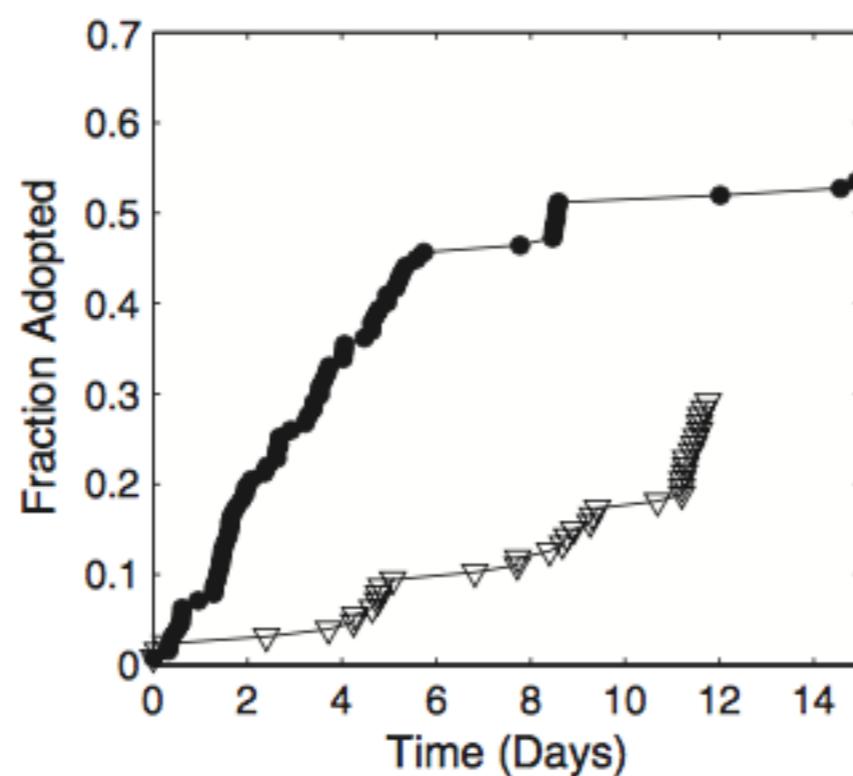
# Resulting in different research programmes



# Experimentation



**The Spread of Behavior in an Online Social Network Experiment**  
Damon Centola, *et al.*  
*Science* **329**, 1194 (2010);  
DOI: 10.1126/science.1185231



# Experimentation

NATURE | LETTER



日本語要約

## A 61-million-person experiment in social influence and political mobilization

Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler

**a**

**Informational message**

Today is Election Day

Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

I Voted

What's this? • close

0 1 1 5 5 3 7 6

People on Facebook Voted

**Social message**

Today is Election Day

Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

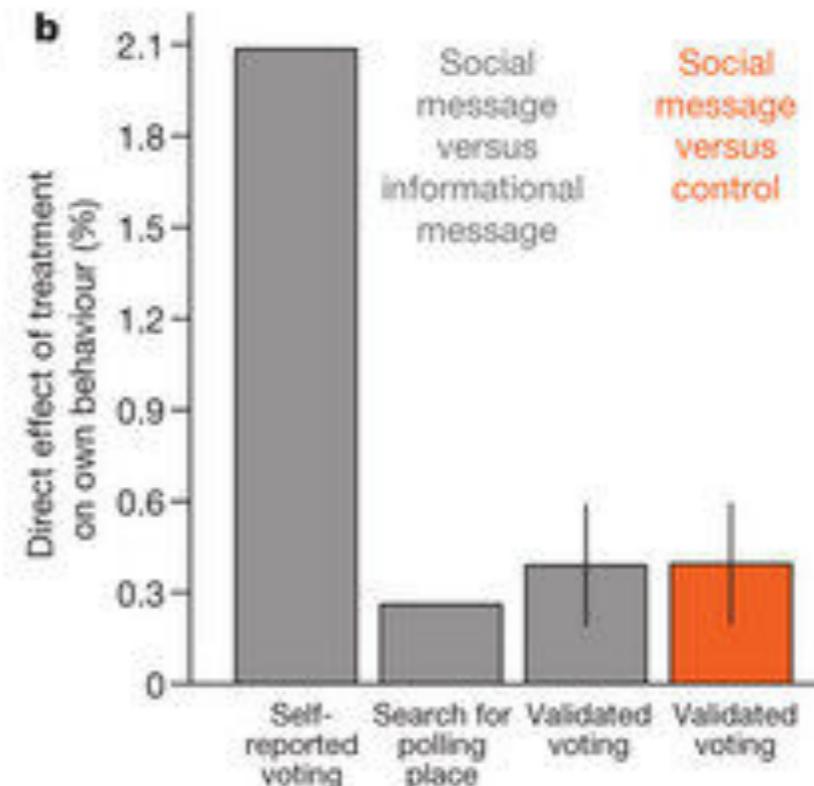
I Voted

What's this? • close

0 1 1 5 5 3 7 6

People on Facebook Voted

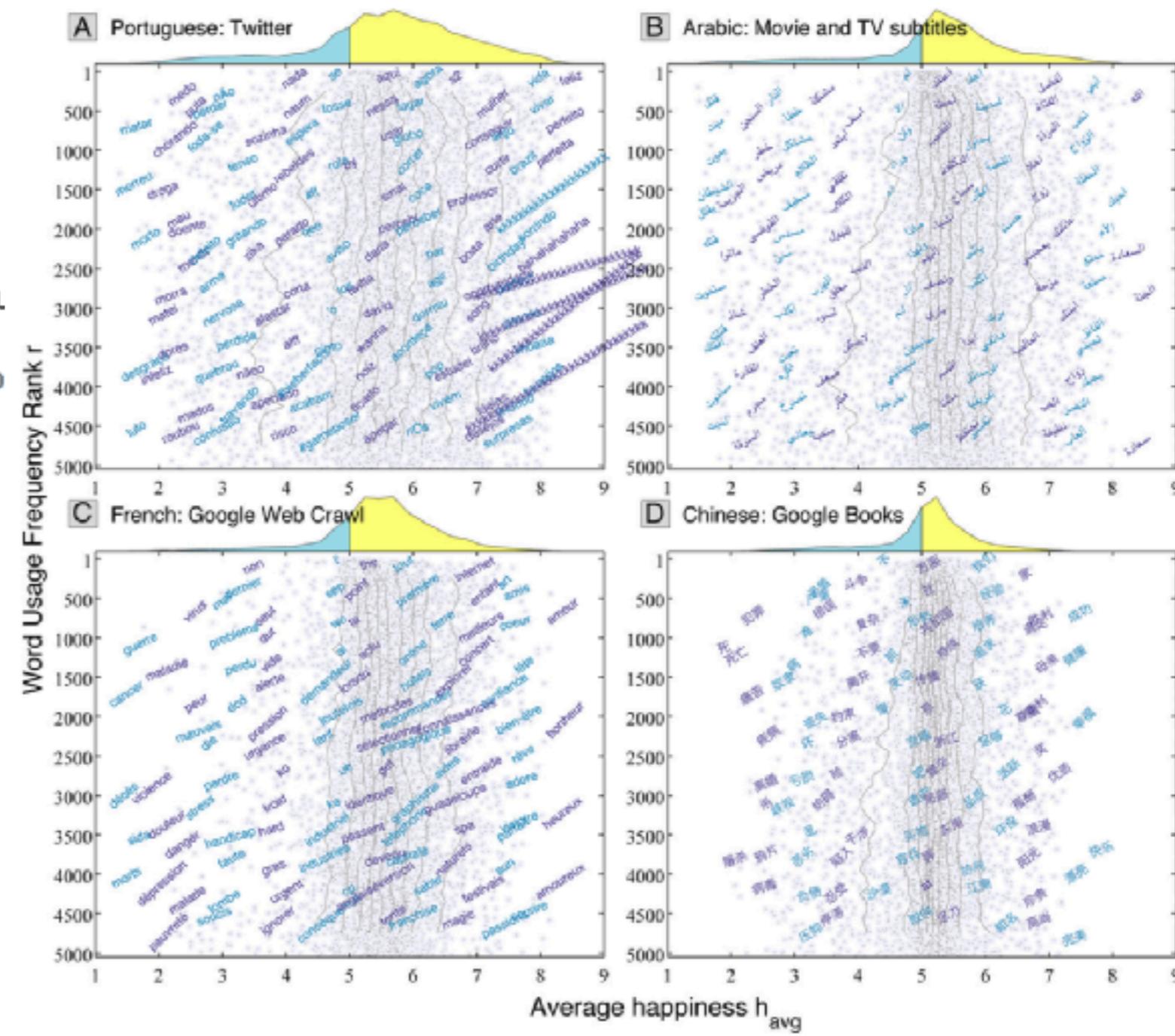
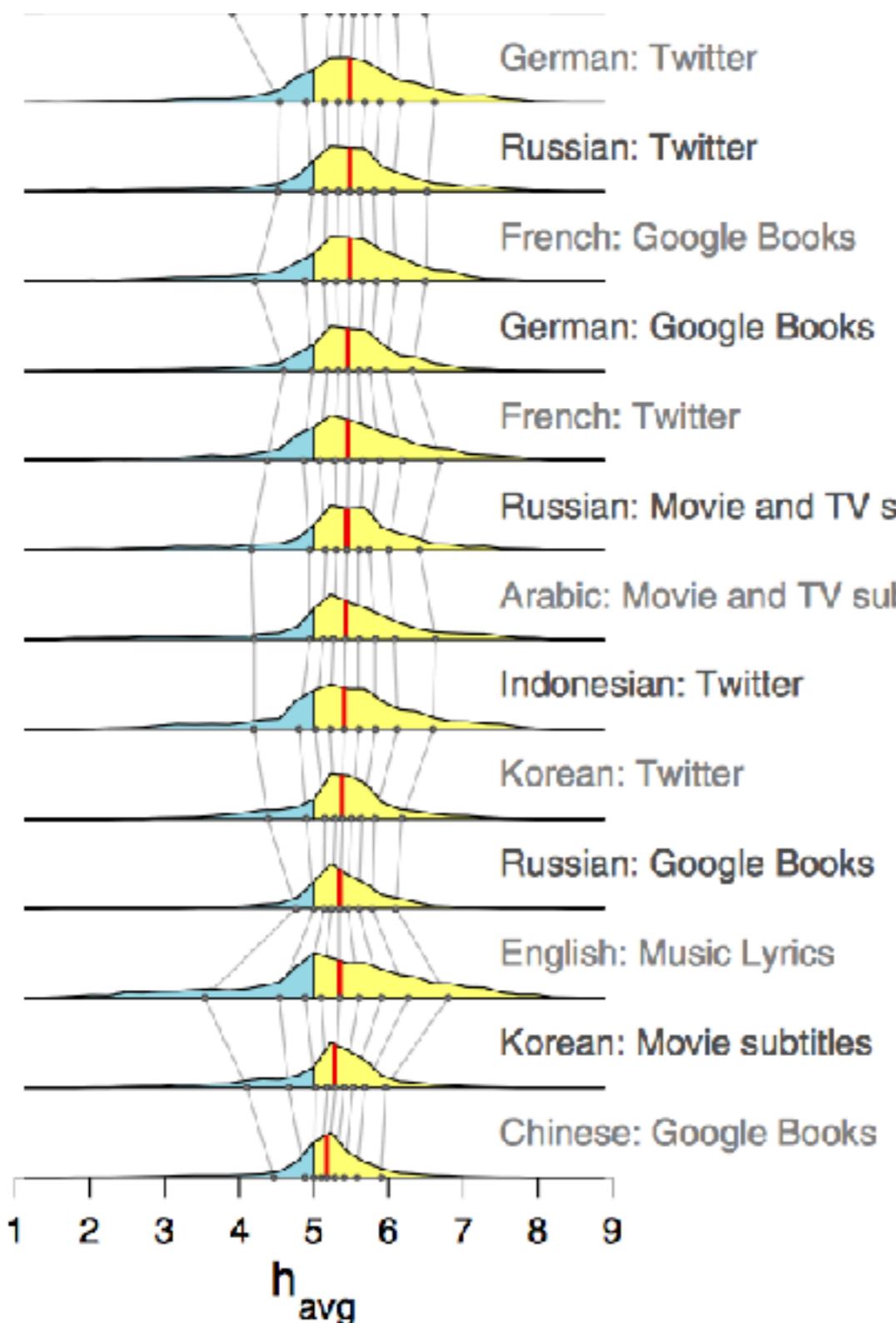
Jaime Settle, Jason Jones, and 18 other friends have voted.



# Text

# Human language reveals a universal positivity bias

Peter Sheridan Dodds<sup>a,b,1</sup>, Eric M. Clark<sup>a,b</sup>, Suma Desu<sup>c</sup>, Morgan R. Frank<sup>c</sup>, Andrew J. Reagan<sup>a,b</sup>, Jake Ryland Williams<sup>a,b</sup>, Lewis Mitchell<sup>d</sup>, Kameron Decker Harris<sup>e</sup>, Isabel M. Kloumann<sup>f</sup>, James P. Bagrow<sup>a,b</sup>, Karine Megerdoomian<sup>g</sup>, Matthew T. McMahon<sup>g</sup>, Brian F. Tivnan<sup>b,g,1</sup>, and Christopher M. Danforth<sup>a,b,1</sup>



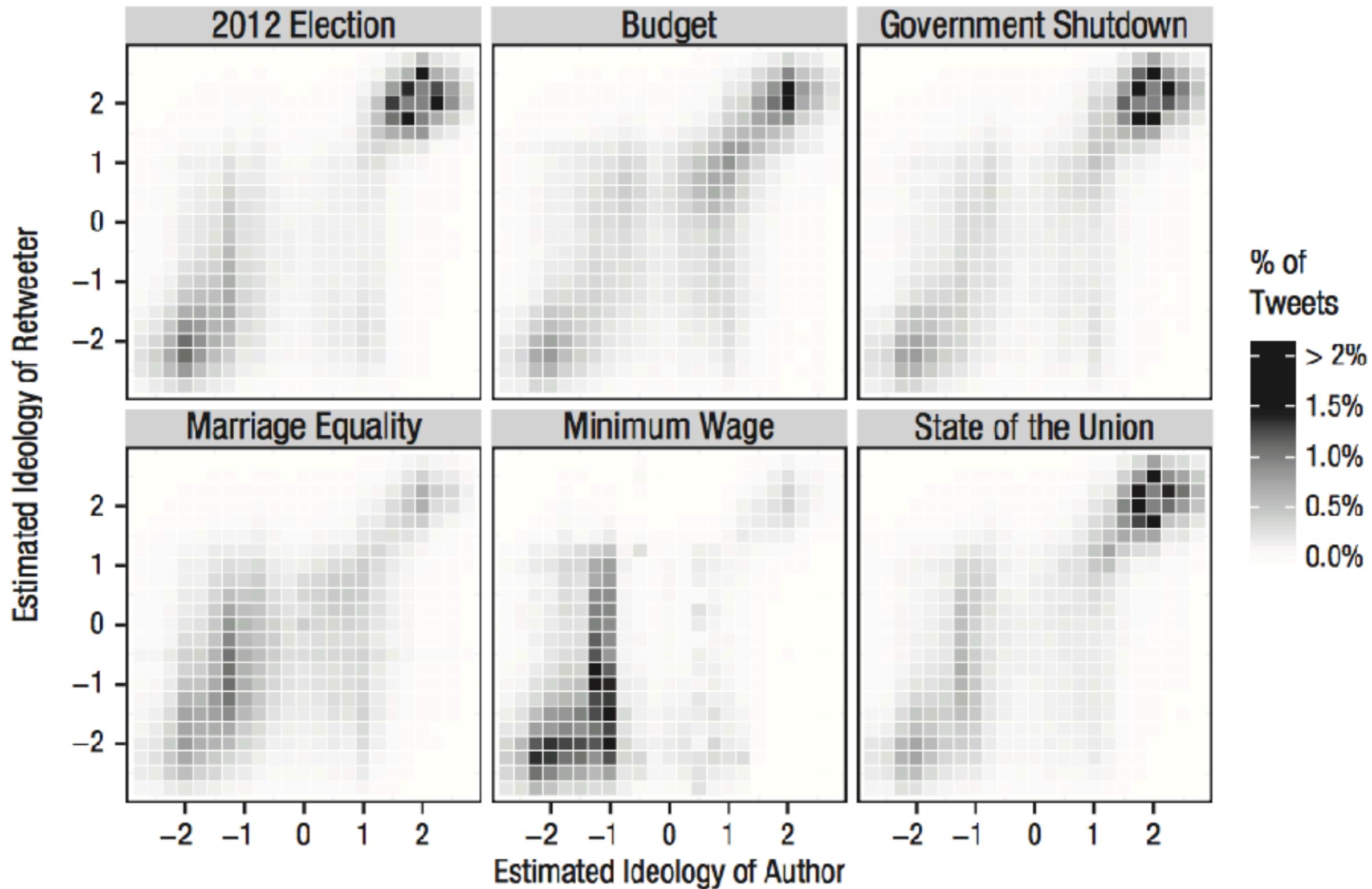
# Text

## Tweeting From Left to Right

*Is Online Political Communication More Than an Echo Chamber?*

Pablo Barberá , John T. Jost , Jonathan Nagler , more...

First Published August 21, 2015 | Research Article

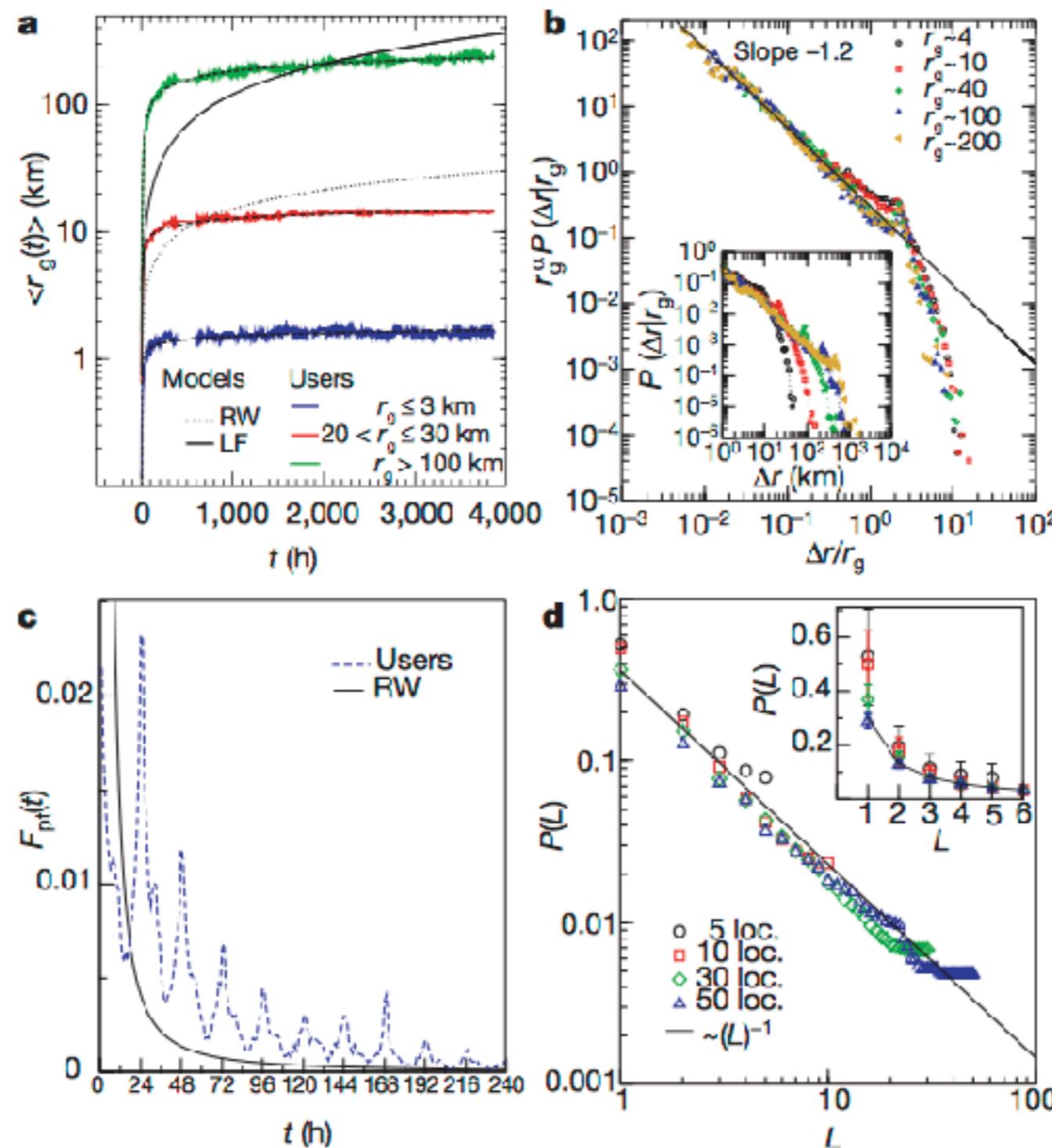


# Digital Trace

## Understanding individual human mobility patterns

Marta C. González<sup>1</sup>, César A. Hidalgo<sup>1,2</sup> & Albert-László Barabási<sup>1,2,3</sup>

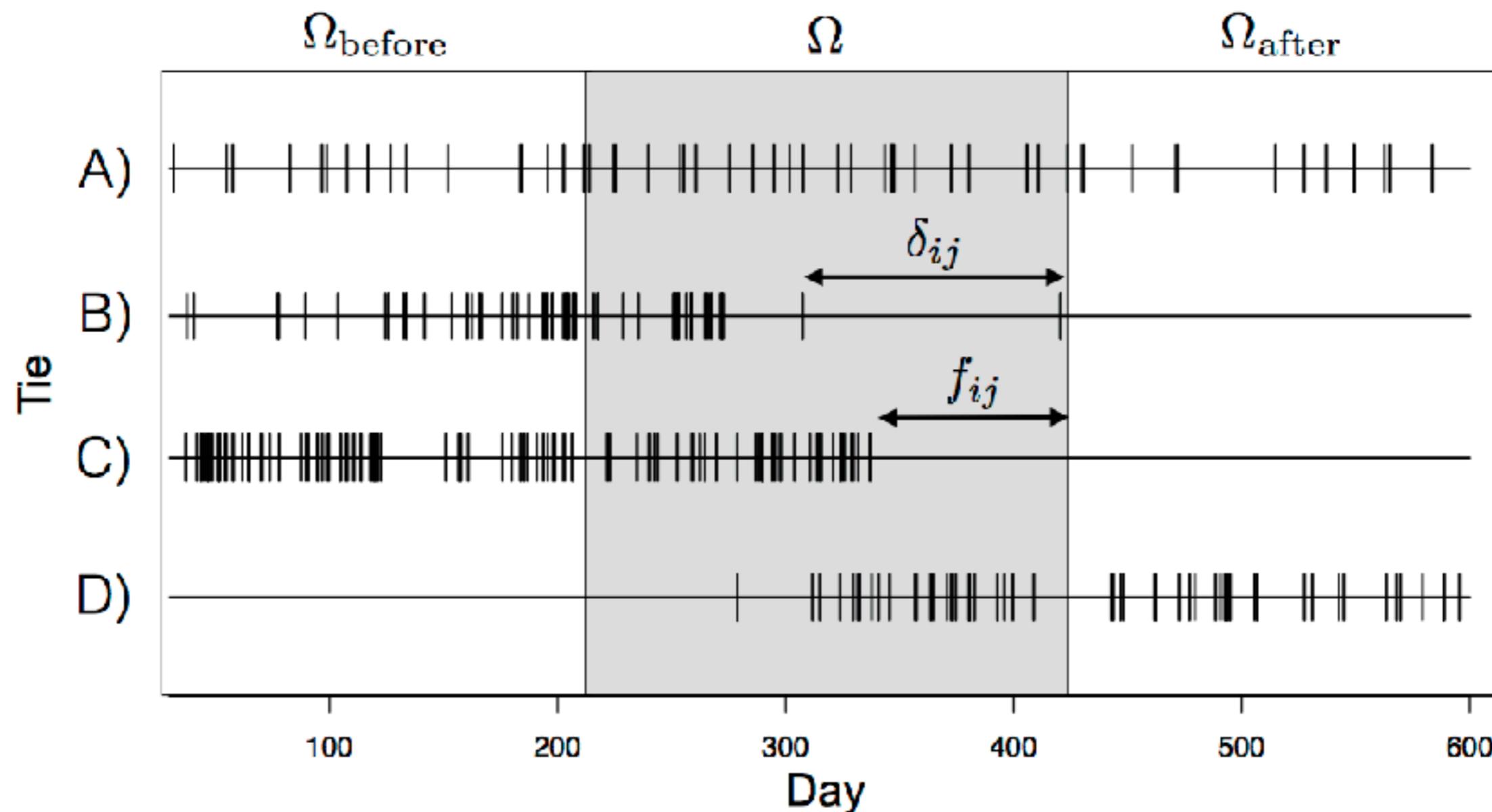
Vol 453 | 5 June 2008 | doi:10.1038/nature06958



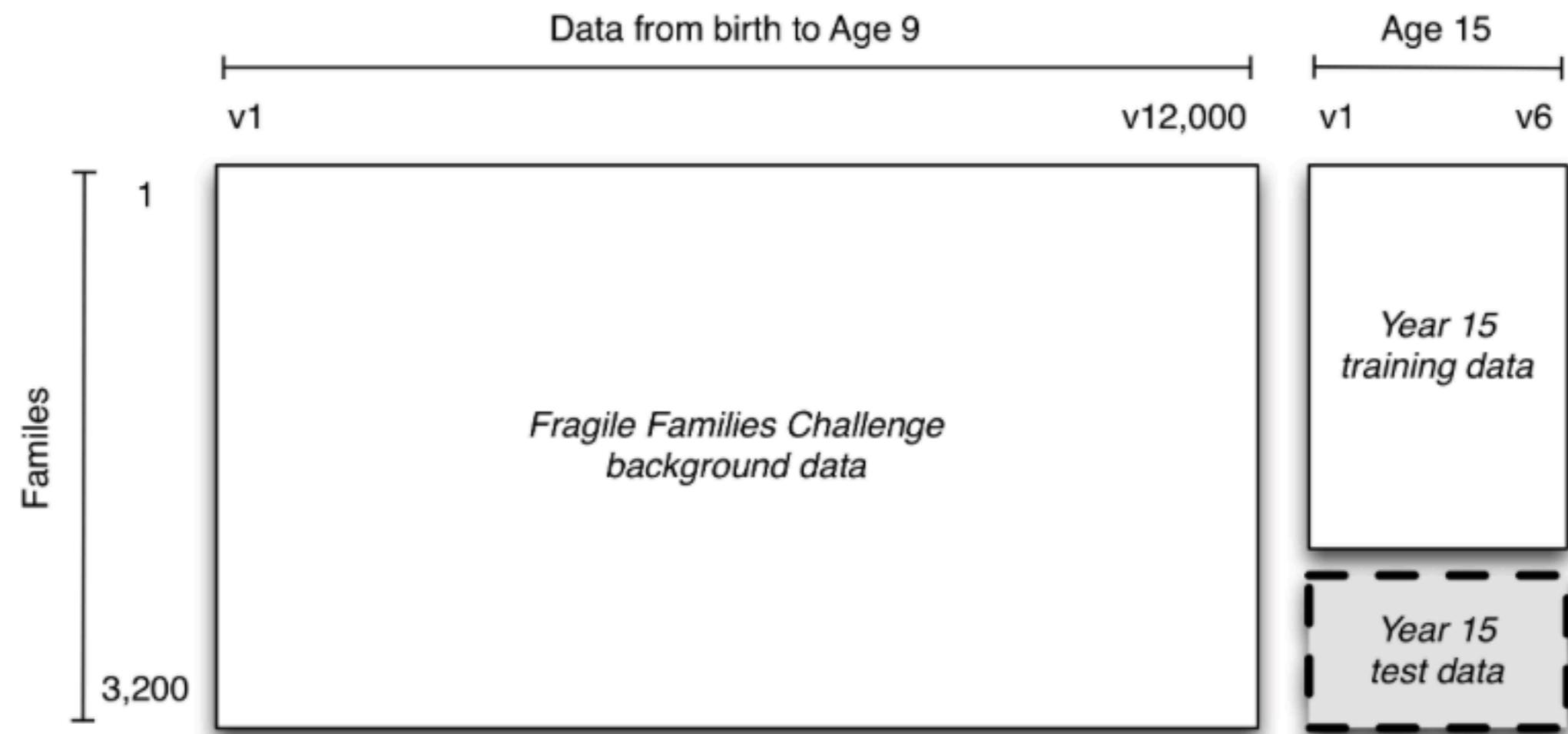
# Digital Trace

## Temporal patterns behind the strength of persistent ties

Henry Navarro<sup>1</sup>, Giovanna Miritello<sup>1,2</sup>, Arturo Canales<sup>3</sup>, Esteban Moro <sup>1\*</sup>

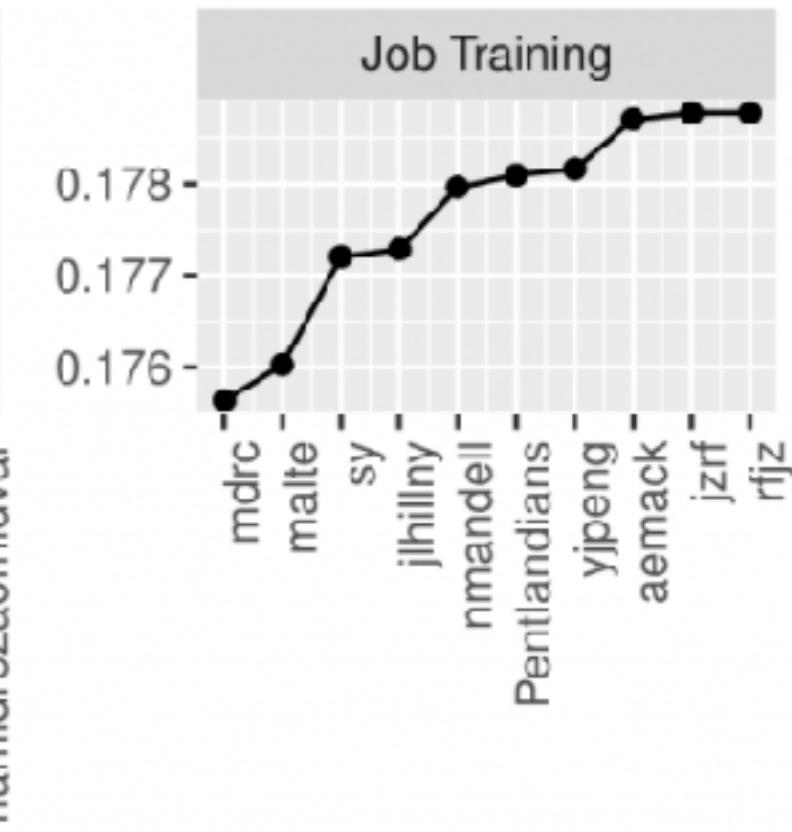
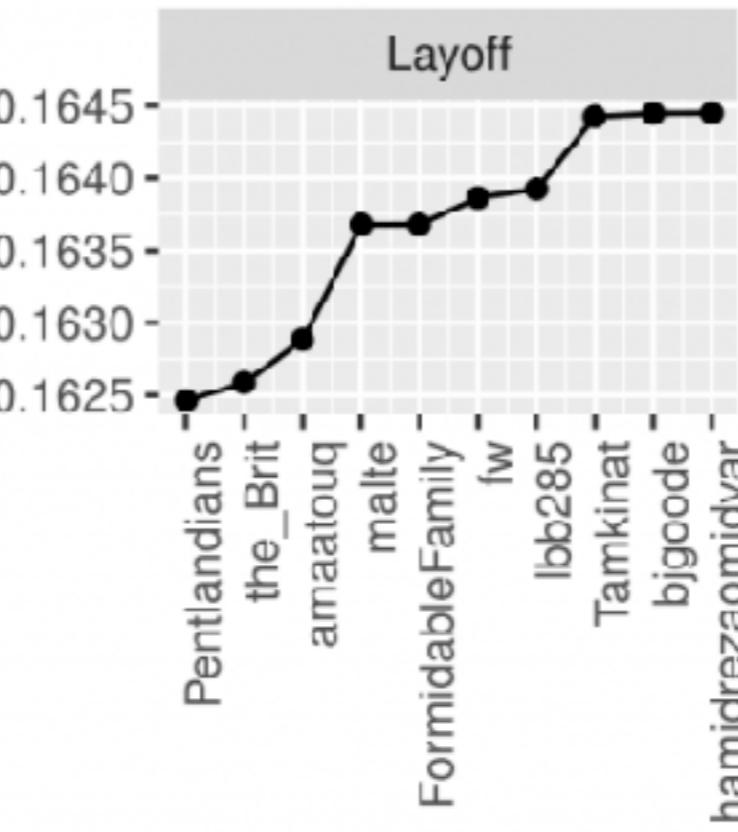
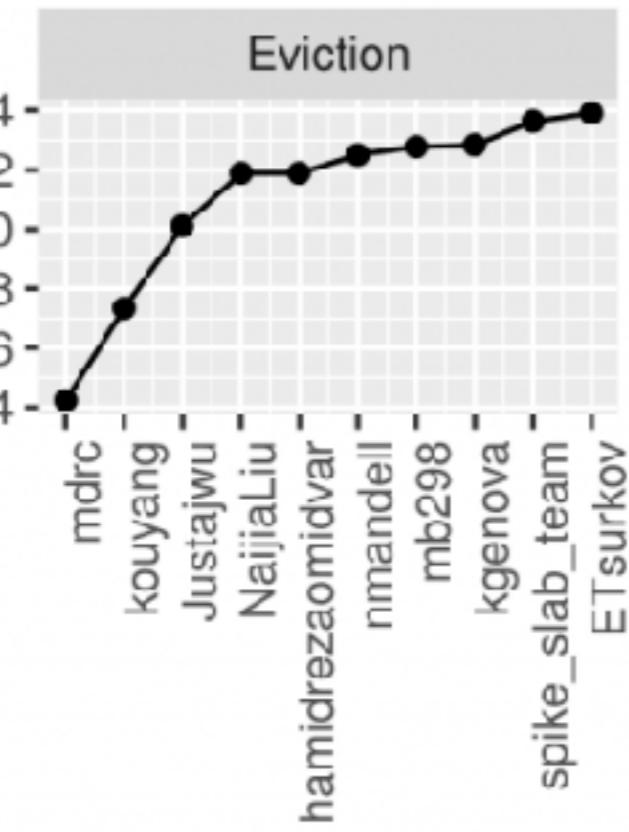
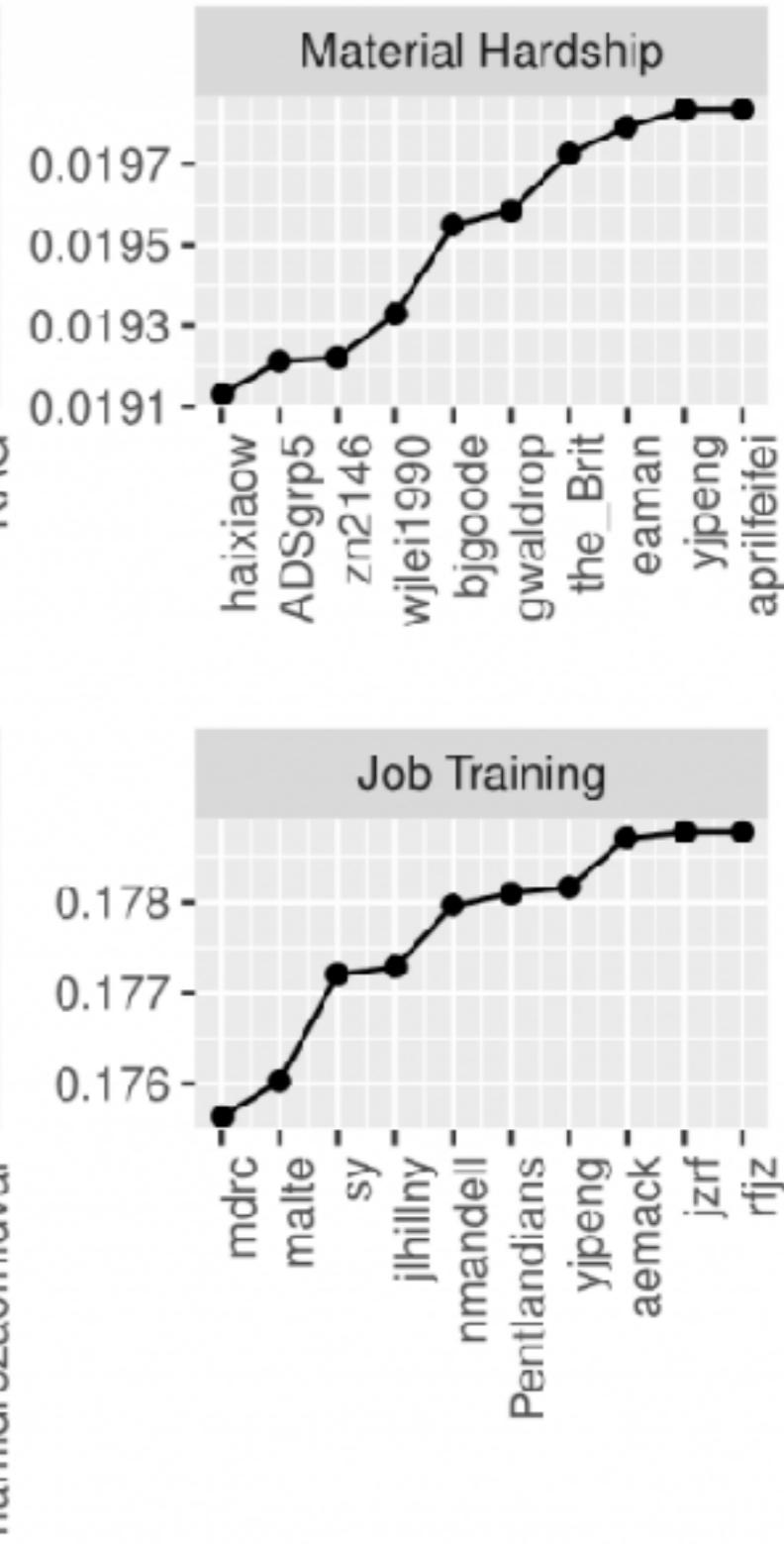
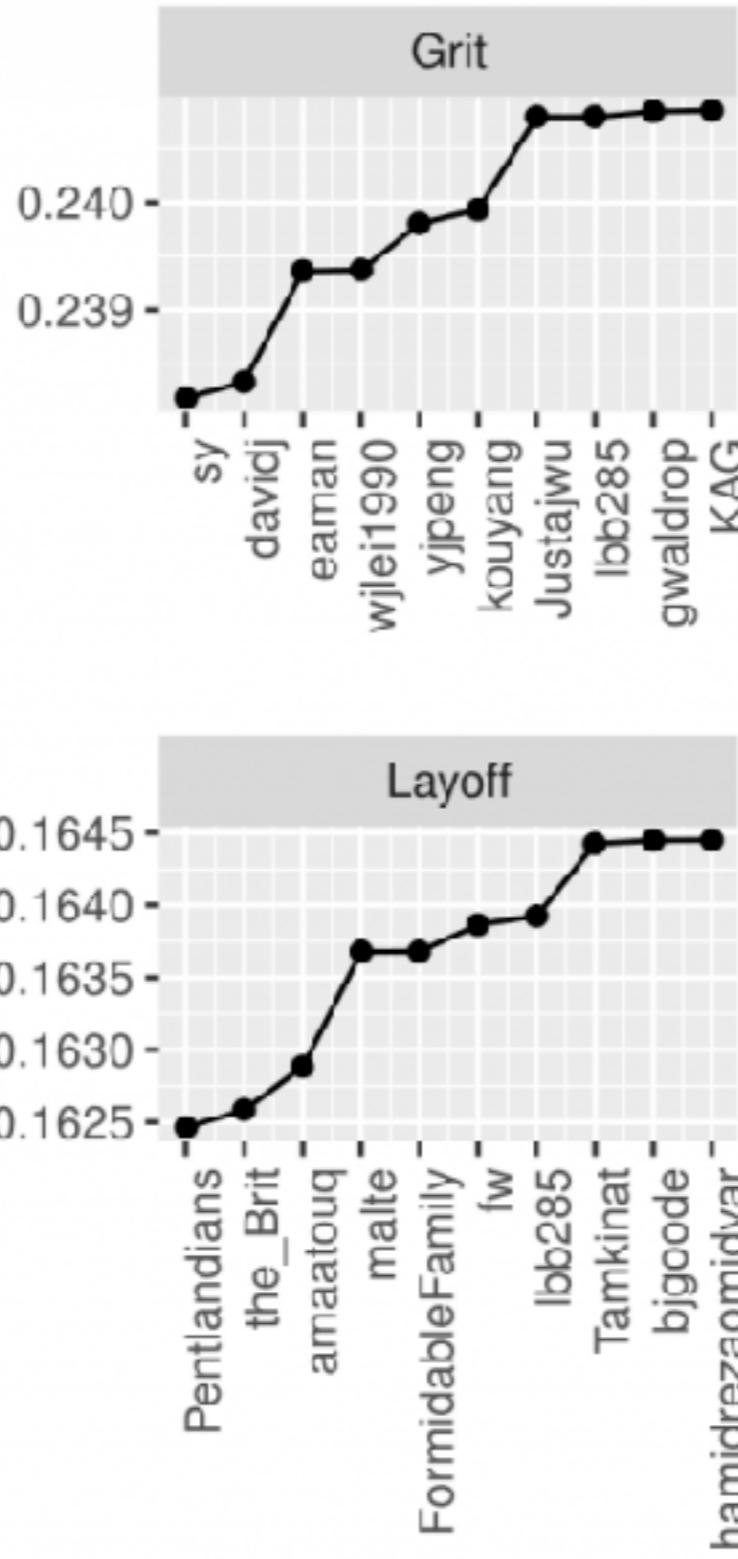
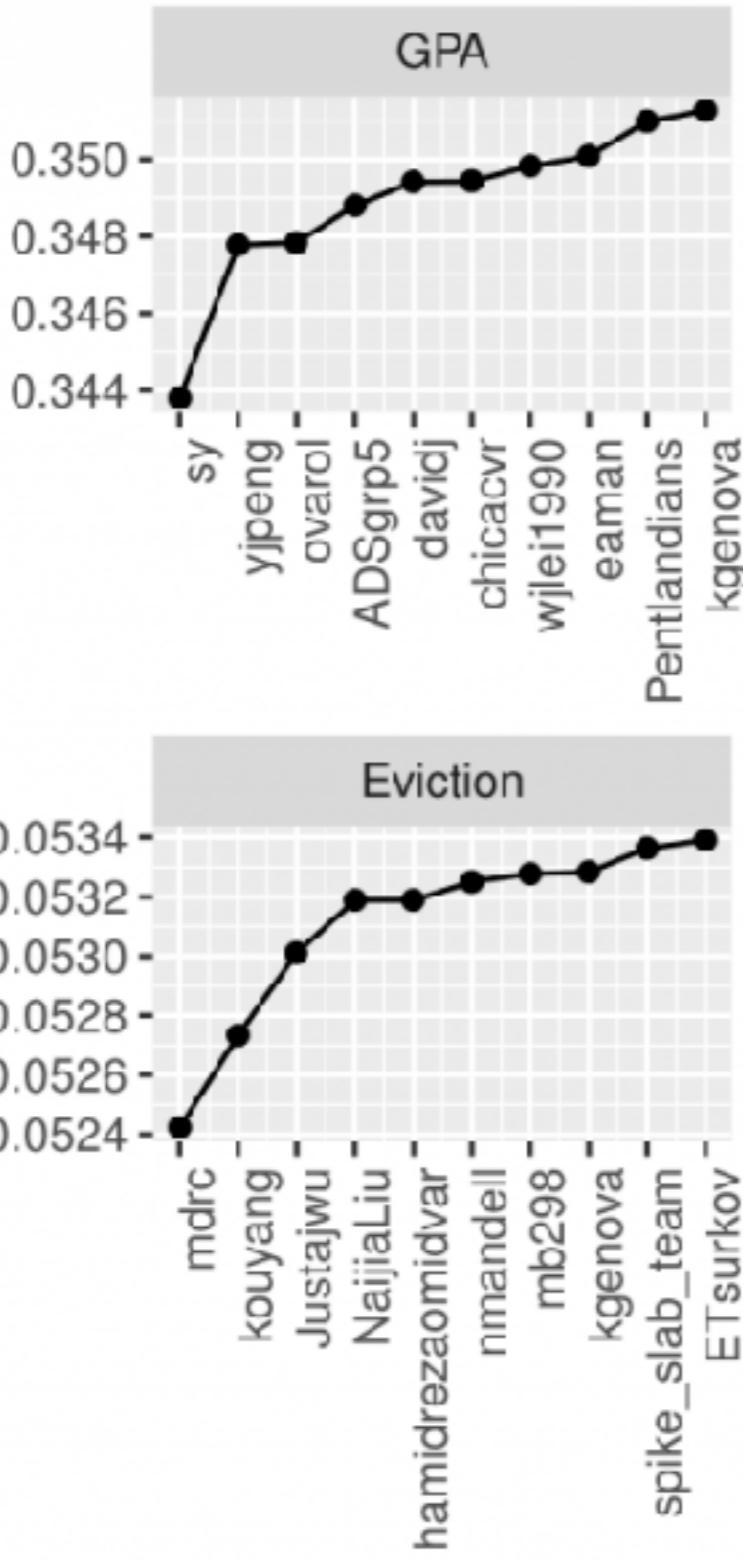


# Administrative data



# Administrative data

Mean squared error in holdout



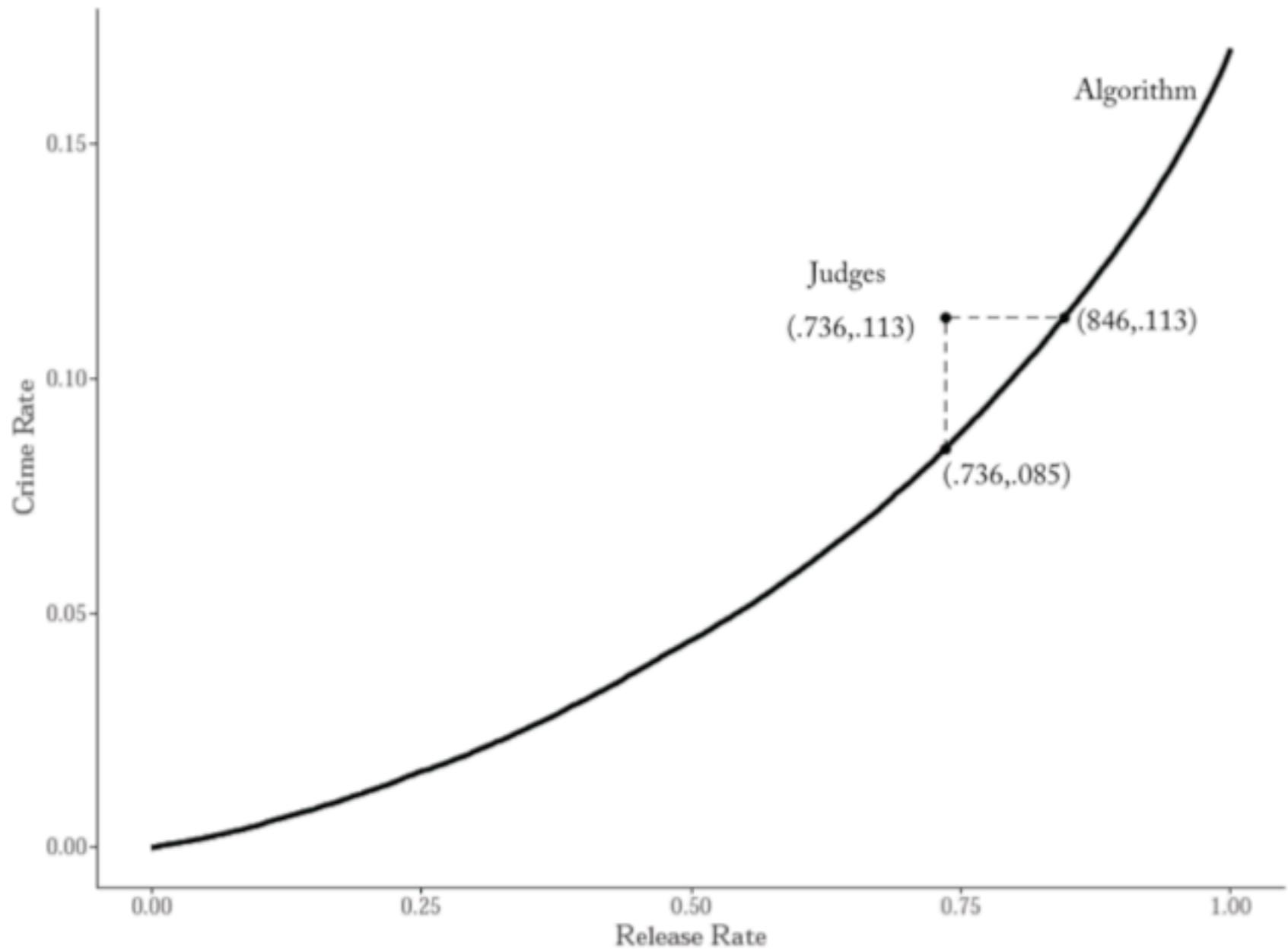
Top 10 submissions

# Administrative data

## Human Decisions and Machine Predictions\*

Jon Kleinberg<sup>†</sup>  
Himabindu Lakkaraju<sup>‡</sup>  
Jure Leskovec<sup>§</sup>  
Jens Ludwig<sup>¶</sup>  
Sendhil Mullainathan<sup>||</sup>

January 23, 2017



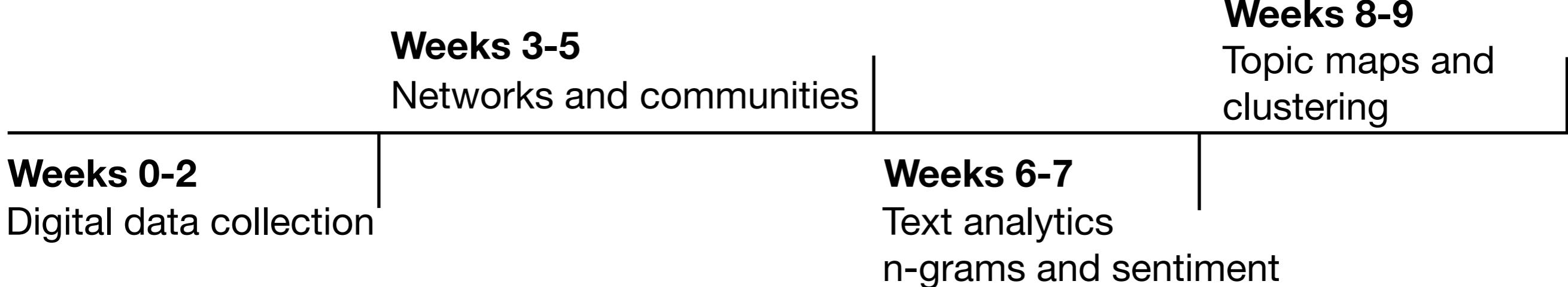
# Administrative data

COMBINING SATELLITE IMAGERY AND MACHINE  
LEARNING TO PREDICT POVERTY

NEAL JEAN, MARSHALL BURKE, MICHAEL XIE, W. MATTHEW DAVIS, DAVID B. LOBELL,  
STEFANO ERMON

# Where does KPHD 540 fit?

## Winter Quarter



## Spring Quarter

**Weeks 0-9:**  
Independent research project

**Focus:** gaining topic expertise and **developing self-sufficiency**  
for future independent research in topic areas

# KPHD 540 Grade Breakdown

<b>Assignments</b> During Winter Quarter Submitted through Vocareum Will include code review	60%
<b>Participation</b> During Winter Quarter In-class and Yellowdig	7%
<b>Final Project</b> During Sprint Quarter	33%

Everyone will receive a grade of 'I' at end of Winter Quarter

All grades will be revised to letter grades at end of Spring Quarter

Drop date is February 16th for class - **there is NO dropping during Spring Quarter**

Programming help/hints (either through auto-grading on Vocareum or assignment starter material) will not continue past Week 1.

# KPHD 540 Contract

Drop date is February 16th for class

There is **NO DROPPING** during Spring Quarter

This class is not designed to teach you how to program

Programming help/hints (via Vocareum auto-grading or assignment templates) will exist until the start of week 2 to help you get into the ‘groove’

After that there will be no “auto-help” to faithfully simulate the rest of the class so you can make an informed decision before the drop date.

# **Who am I?**



....besides a weirdo  
with a large collection of tiny hats

I'm one of these

# MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

## MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

## PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g., R
- ★ Databases: SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative



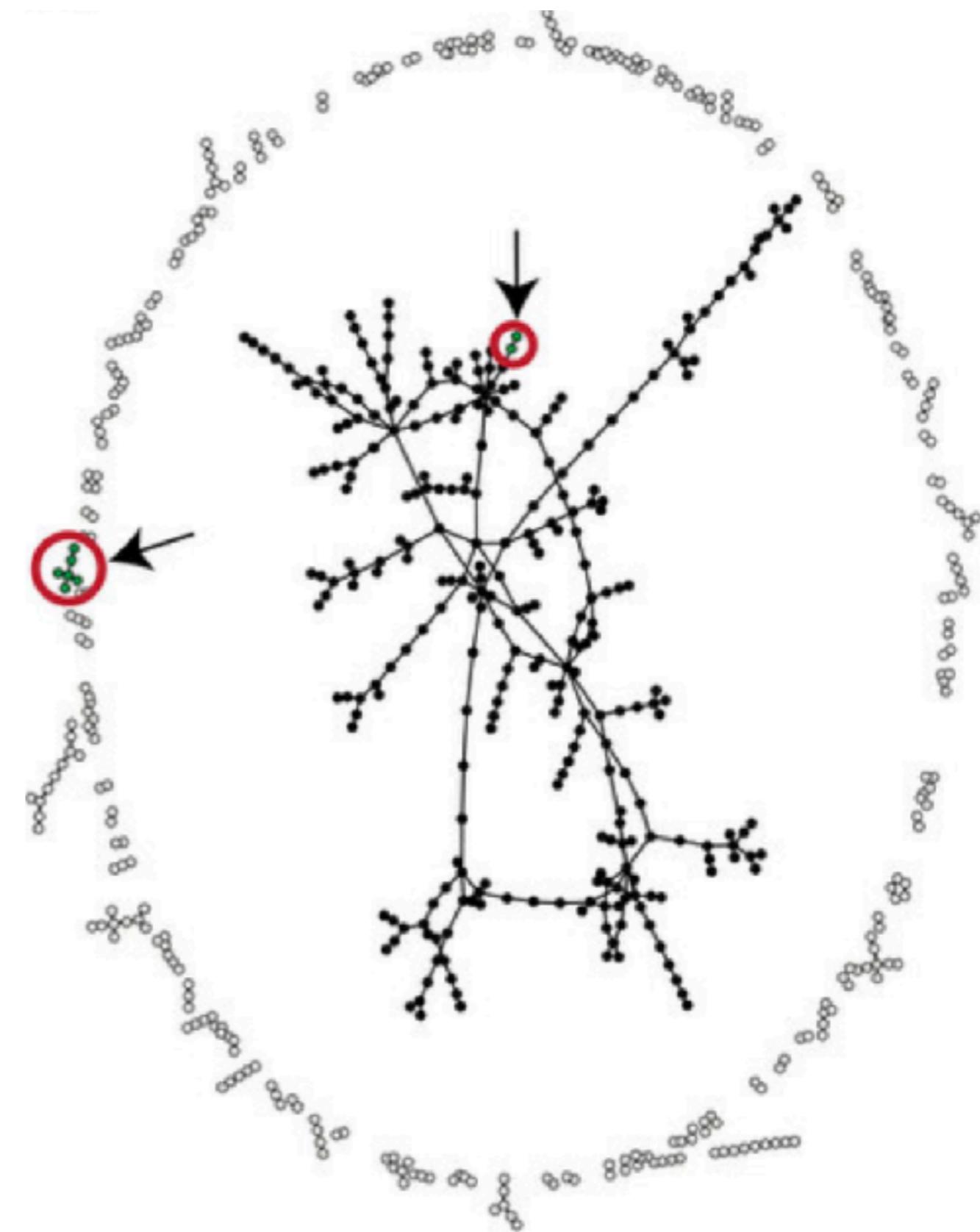
## COMMUNICATION & VISUALIZATION

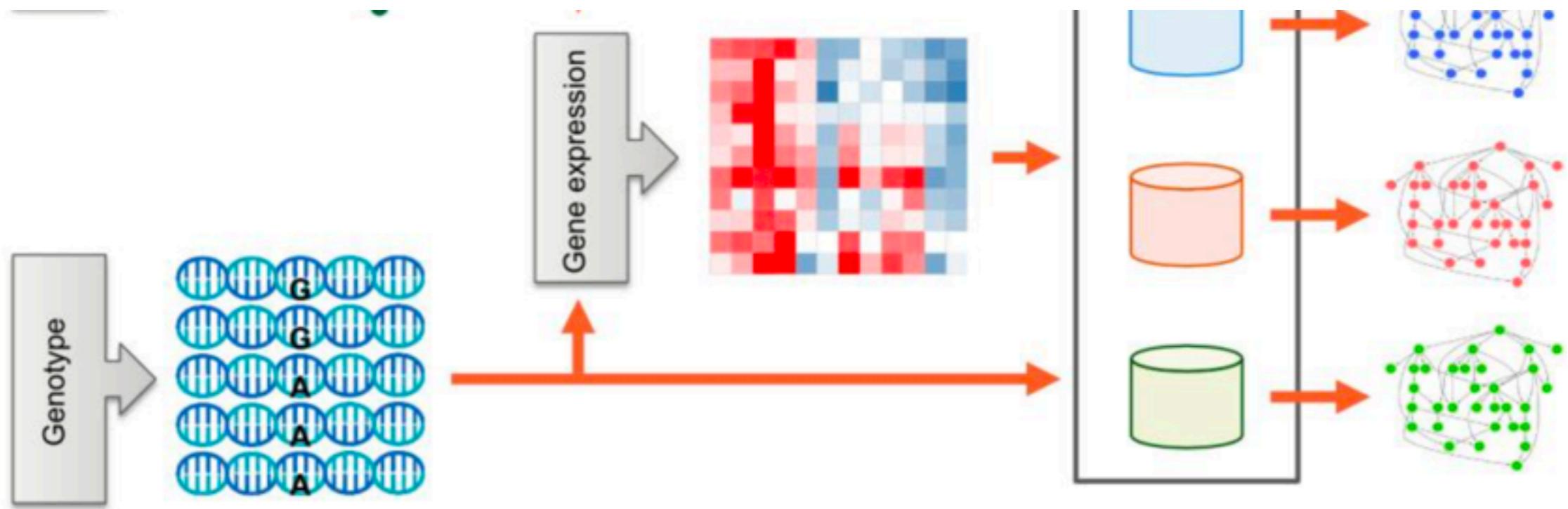
- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any visualization tools e.g. Flare, D3.js, Tableau

But I started first  
genetically modifying  
these



And then I learned  
how to be a data  
scientist



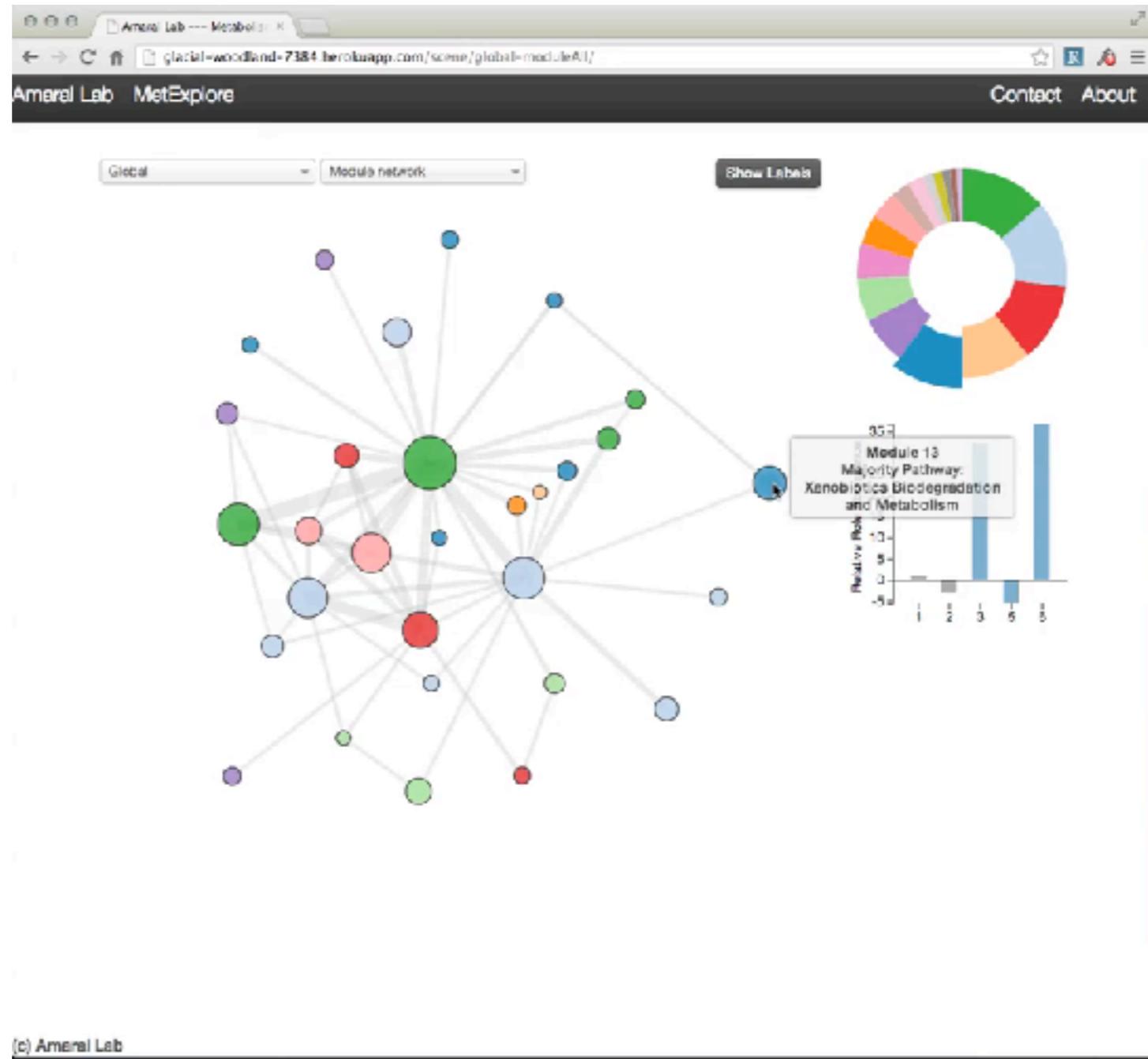


Took a break to  
learn machine learning

# datascope



and took another break  
to work

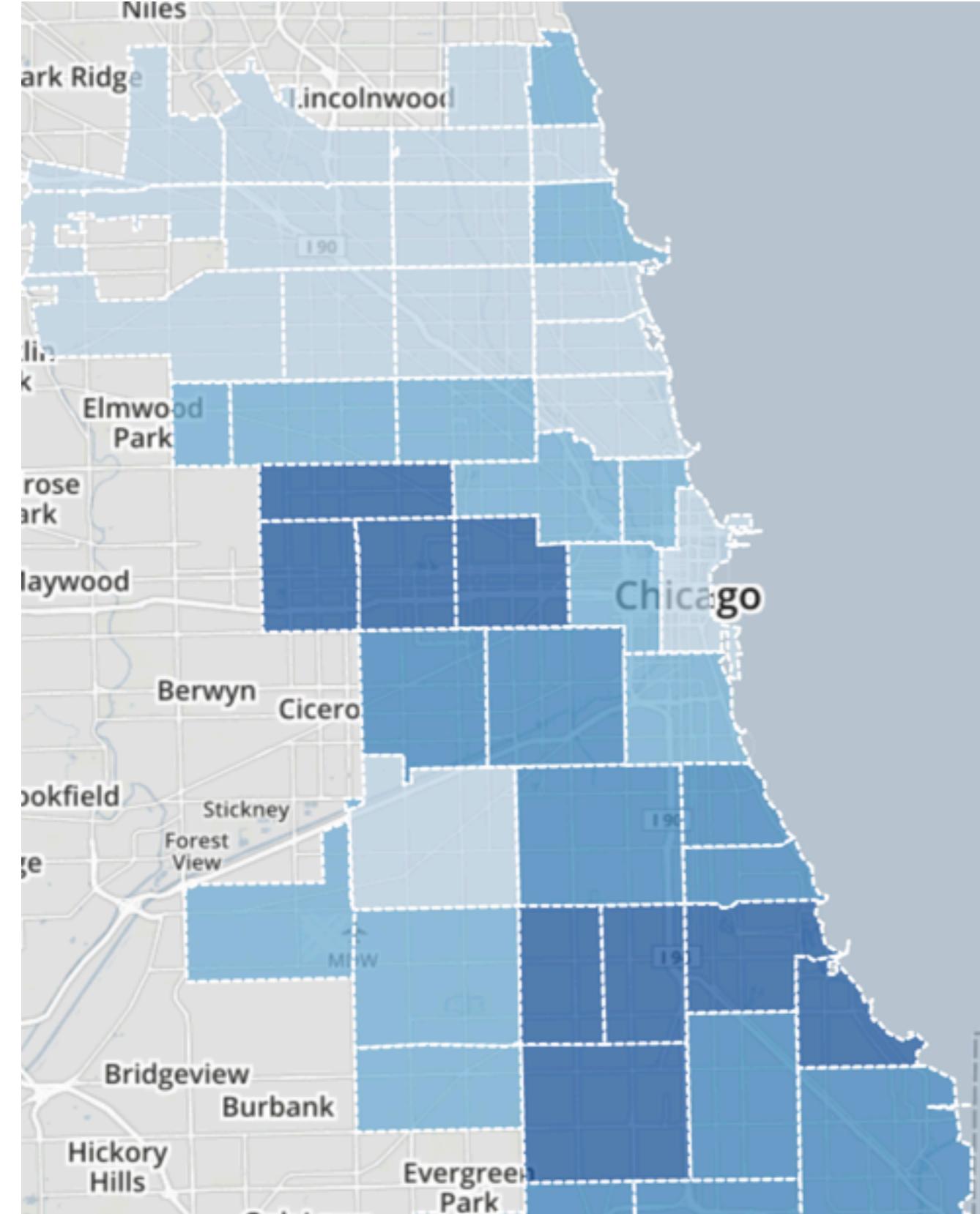


Finished by figuring out  
how to communicate  
complex data

and then went to work  
trying to find railcars



got bored and moved  
to public health



Estimated asthma prevalence in Chicago for adults aged 18-89 based on aggregated Electronic Health Record (EHR) data from a selection of healthcare institutions from 2006 through 2012.

Source: CHITREC | [Data](#) | [i](#)

broke into sociology from epidemiology

## Economic insecurity and the rise in gun violence at US schools

A. R. Pah<sup>1,2</sup>, J. Hagan<sup>3,4</sup>, A. L. Jennings<sup>5</sup>, A. Jain<sup>5</sup>, K. Albrecht<sup>3</sup>, A. J. Hockenberry<sup>5</sup> and L. A. N. Amaral<sup>1,5-7\*</sup>



# My main interests

- Generalized contagion in behavioral time series
- Robust adjustments for individual performance over longitudinal data
- Missing information in social networks
- Impact of social cohesion on human conflict