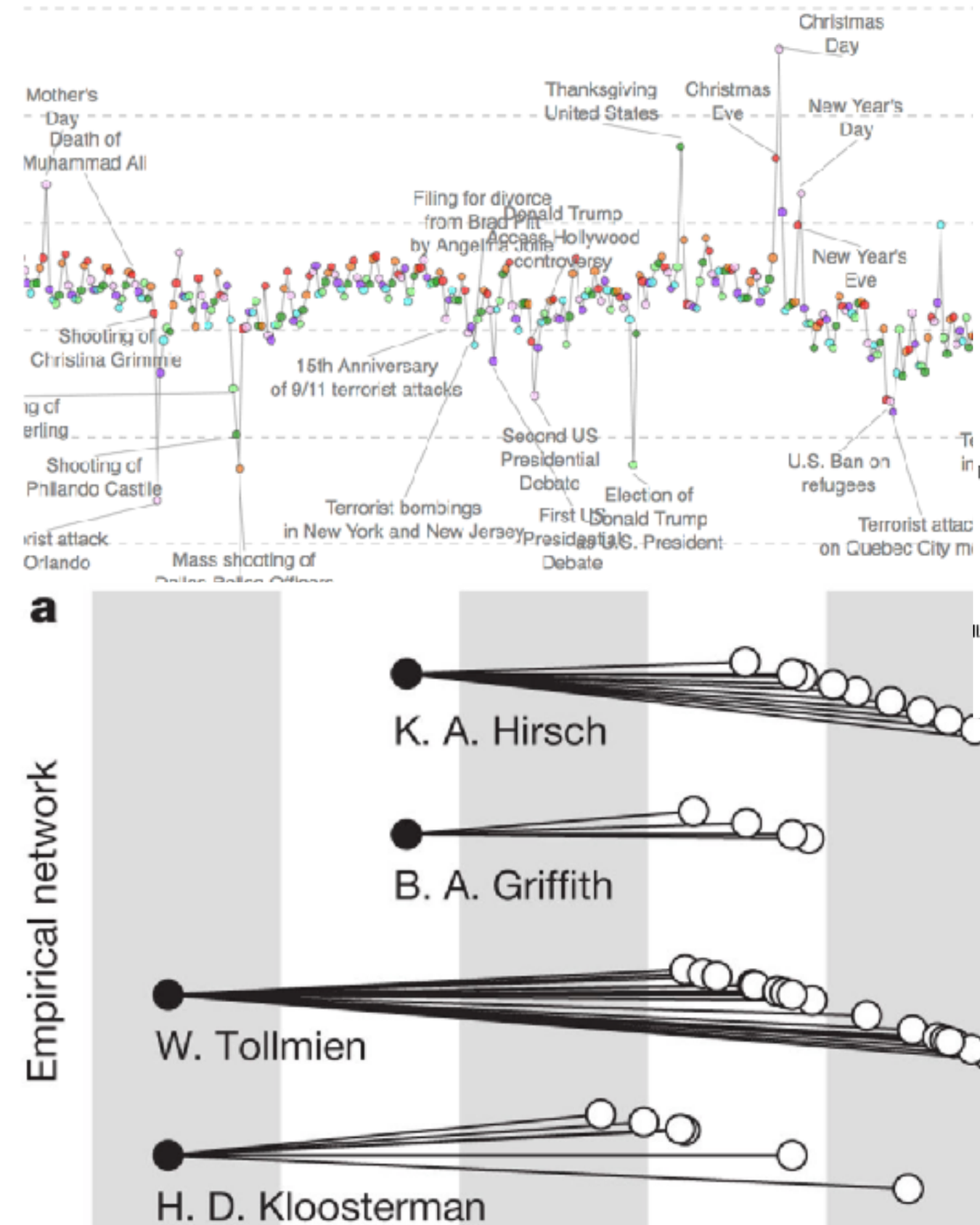
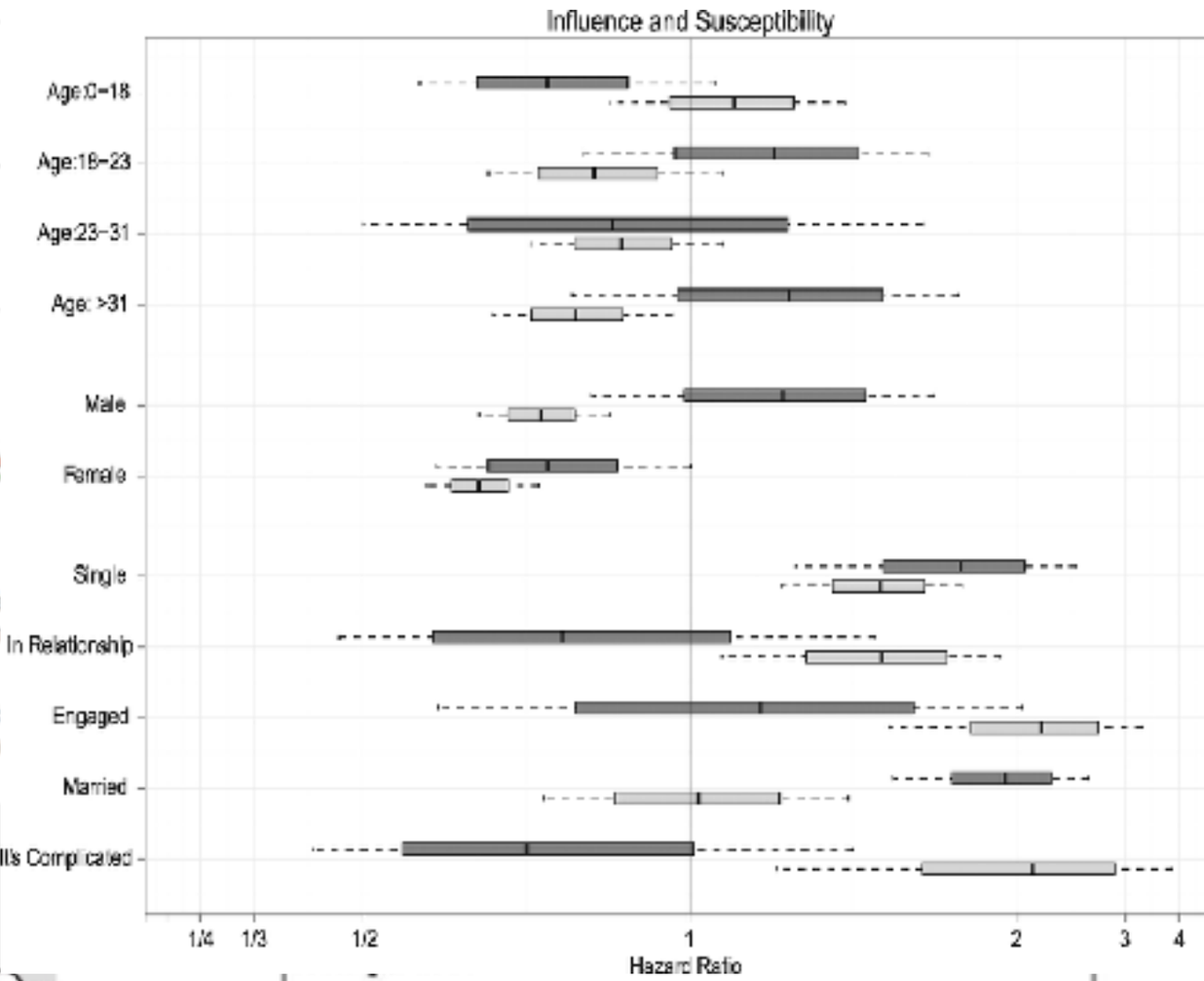
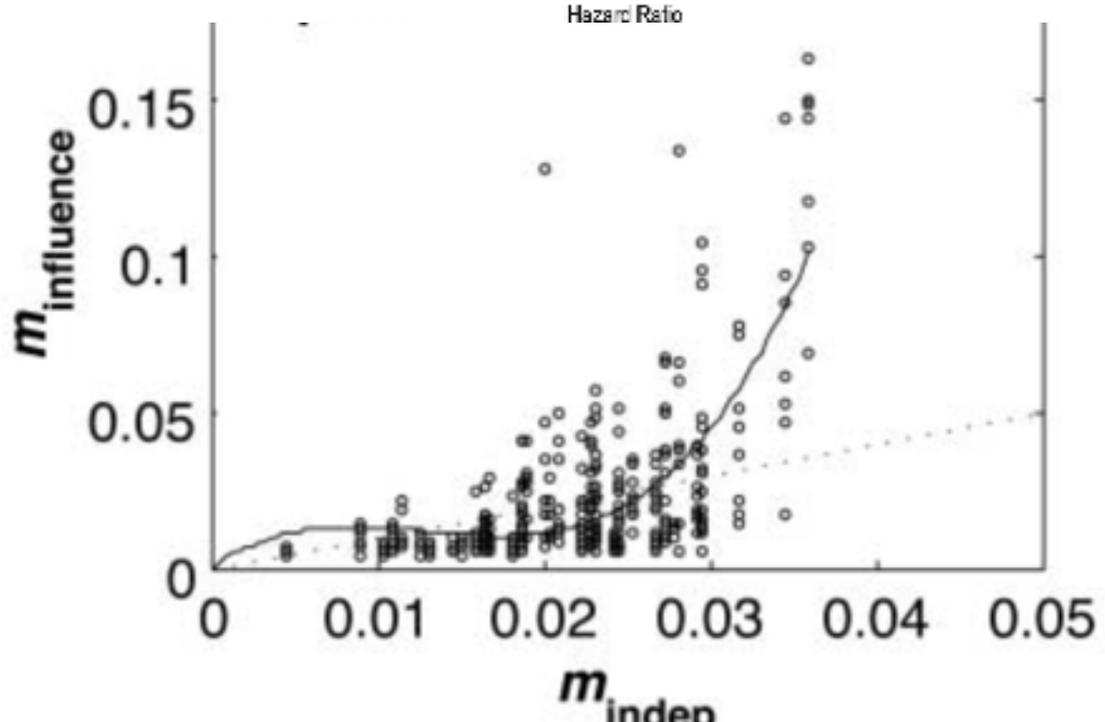
What is Computational Social Science?

Adam Pah

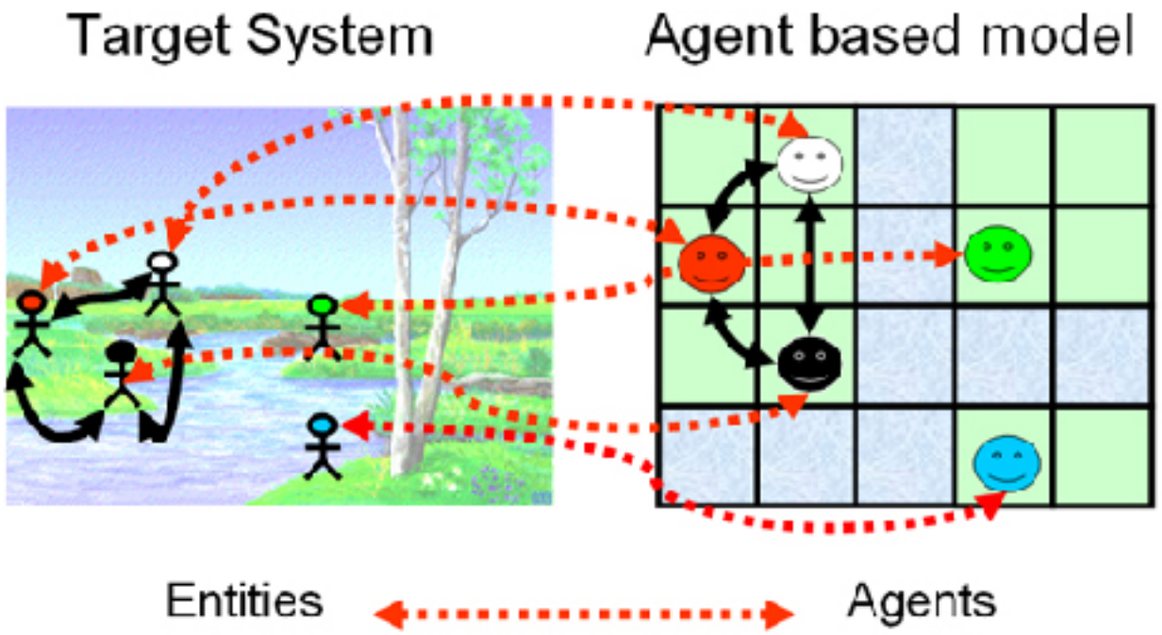
Computational Social Science has enabled significant advances

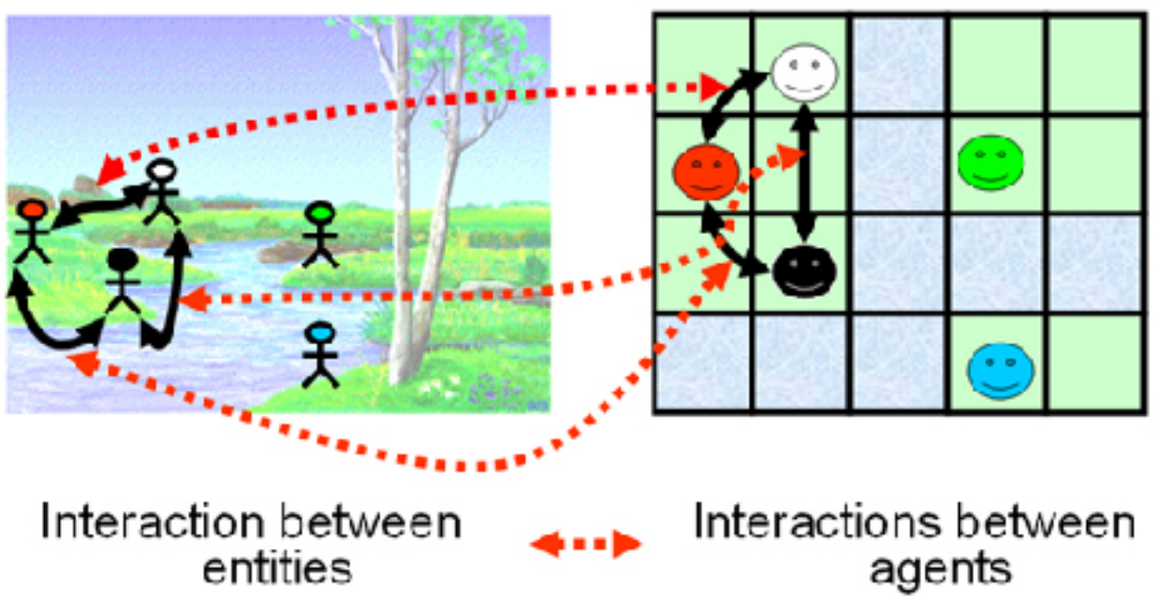




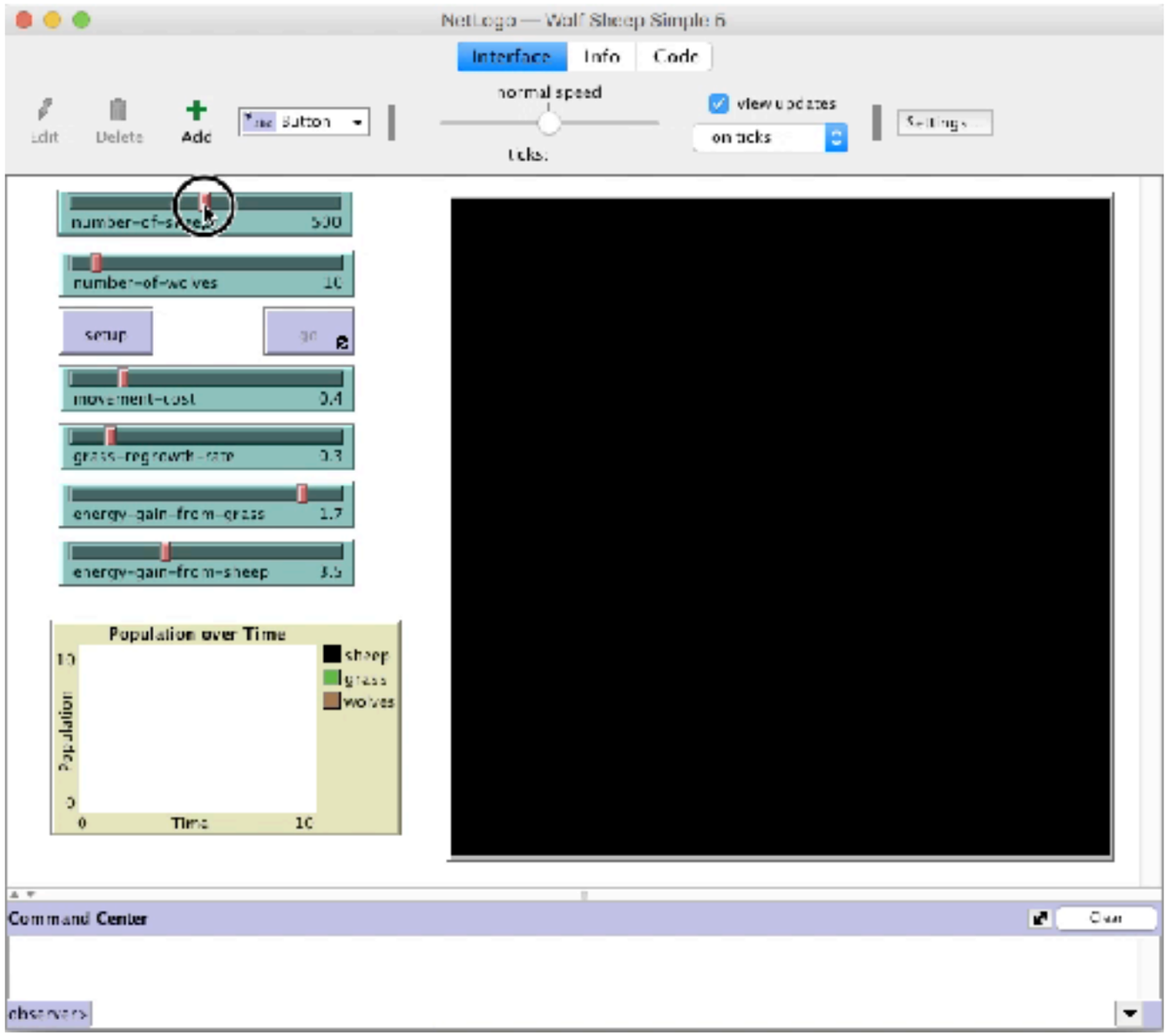


But the definition of computational social science is not static

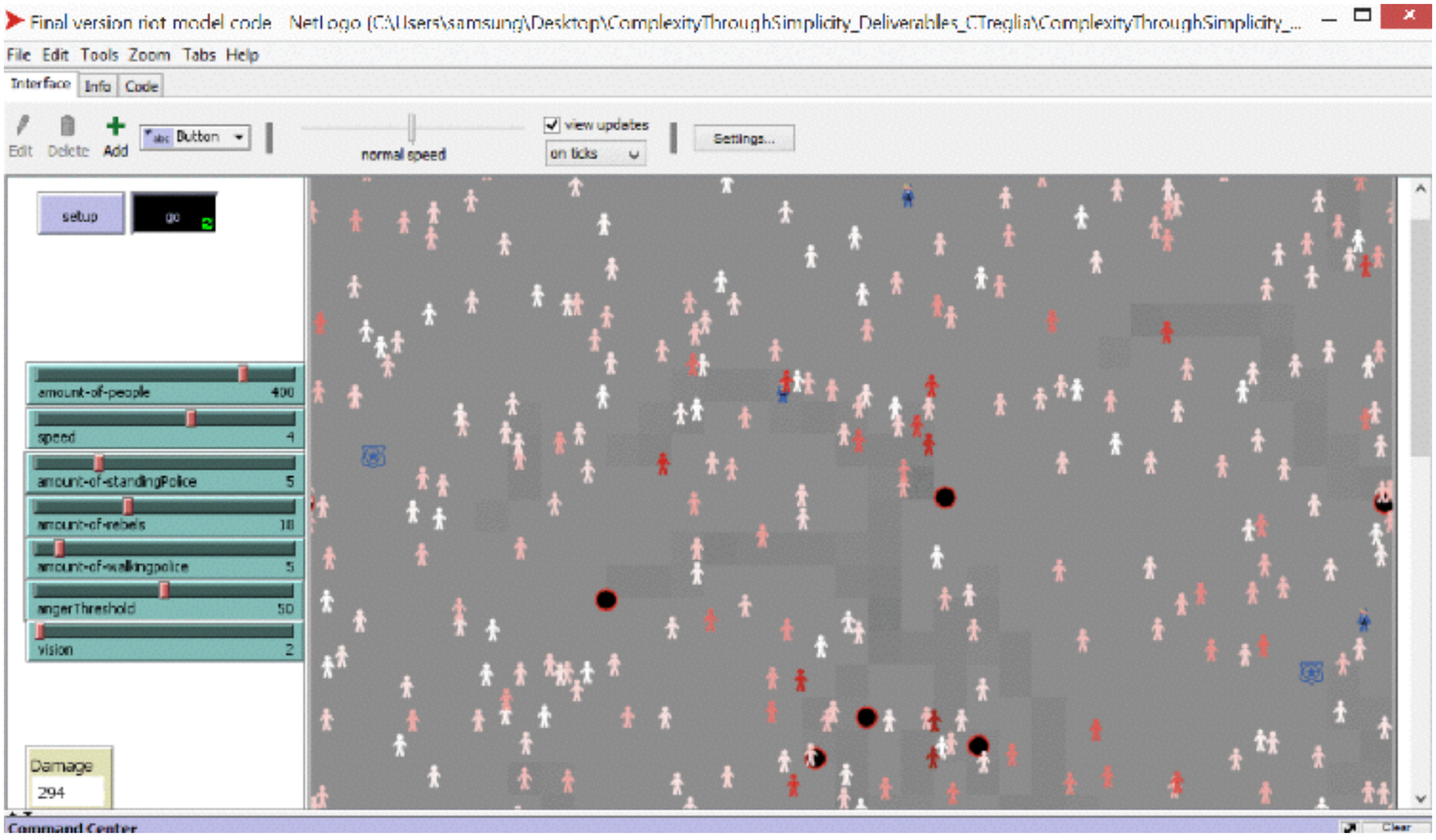




1990s to mid-2000s



ABMs were directly analogous to historically ‘recent’ theoretical papers

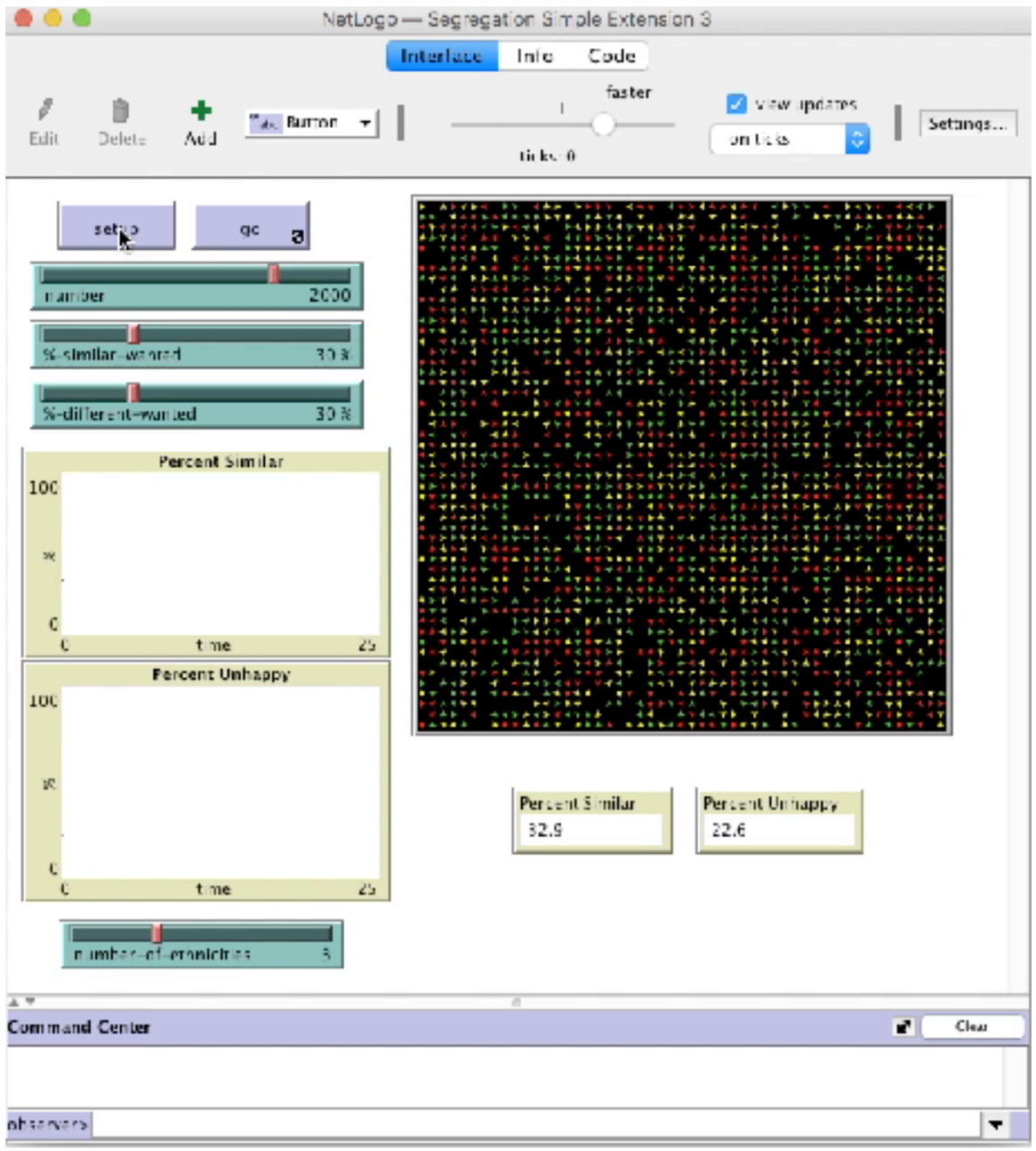


Threshold Models of Collective Behavior

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

Harvard University

Journal of Mathematical Sociology
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

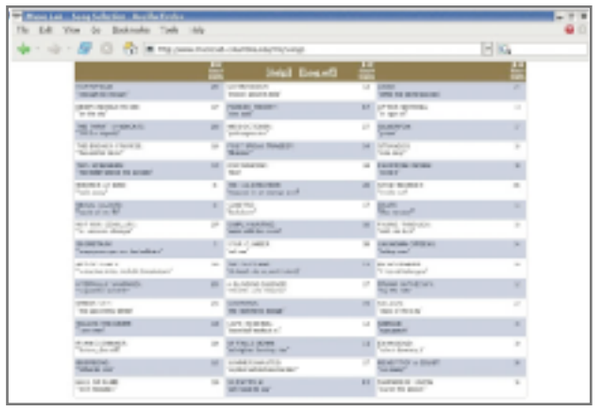
Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market

Matthew J. Salganik,[1,2*] Peter Sheridan Dodds,[2*] Duncan J. Watts[1,2,3*]

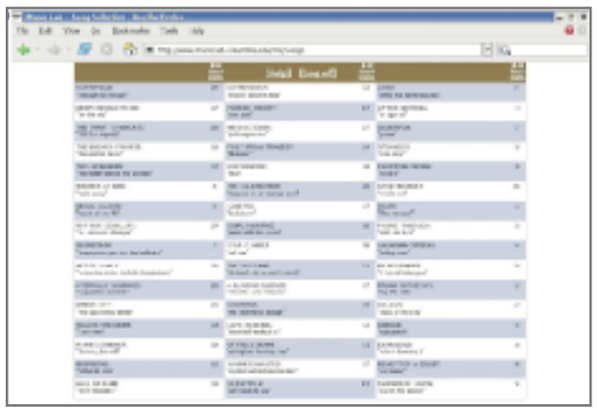
How much do we influence each other?

listen ★ rate ★ download ★ MUSIC LAB

download count



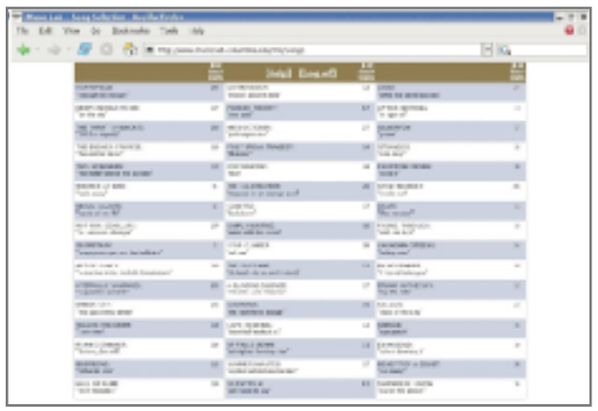
no download count



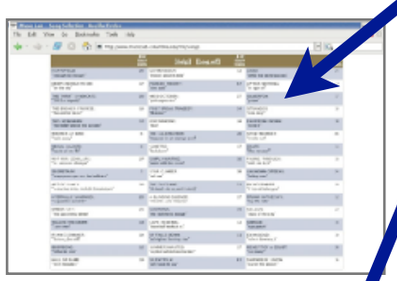
Salganik, Dodds, and Watts, 2006

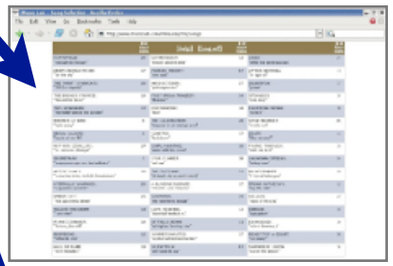
listen ★ rate ★ download ★ MUSIC LAB

no download count



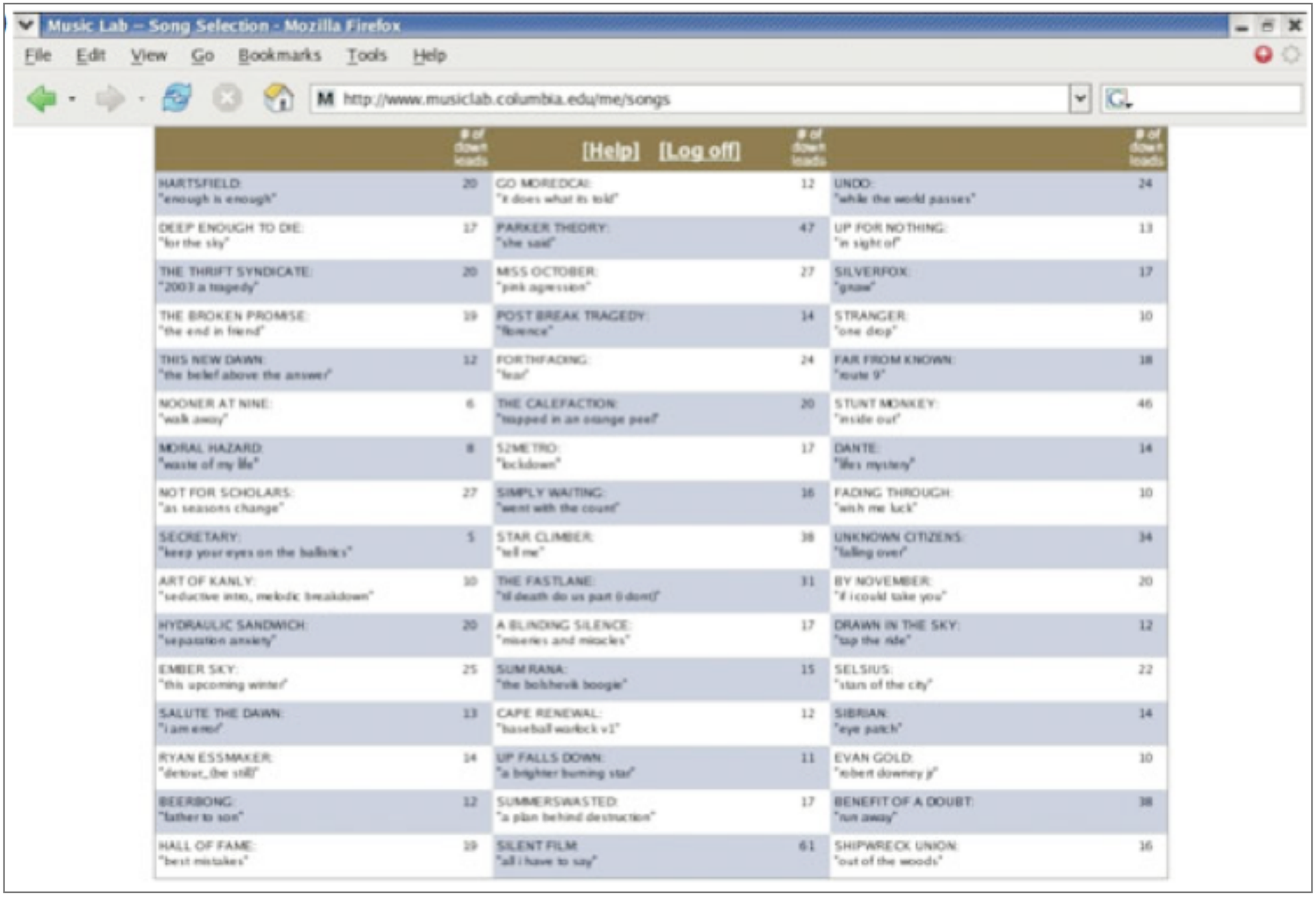
download count



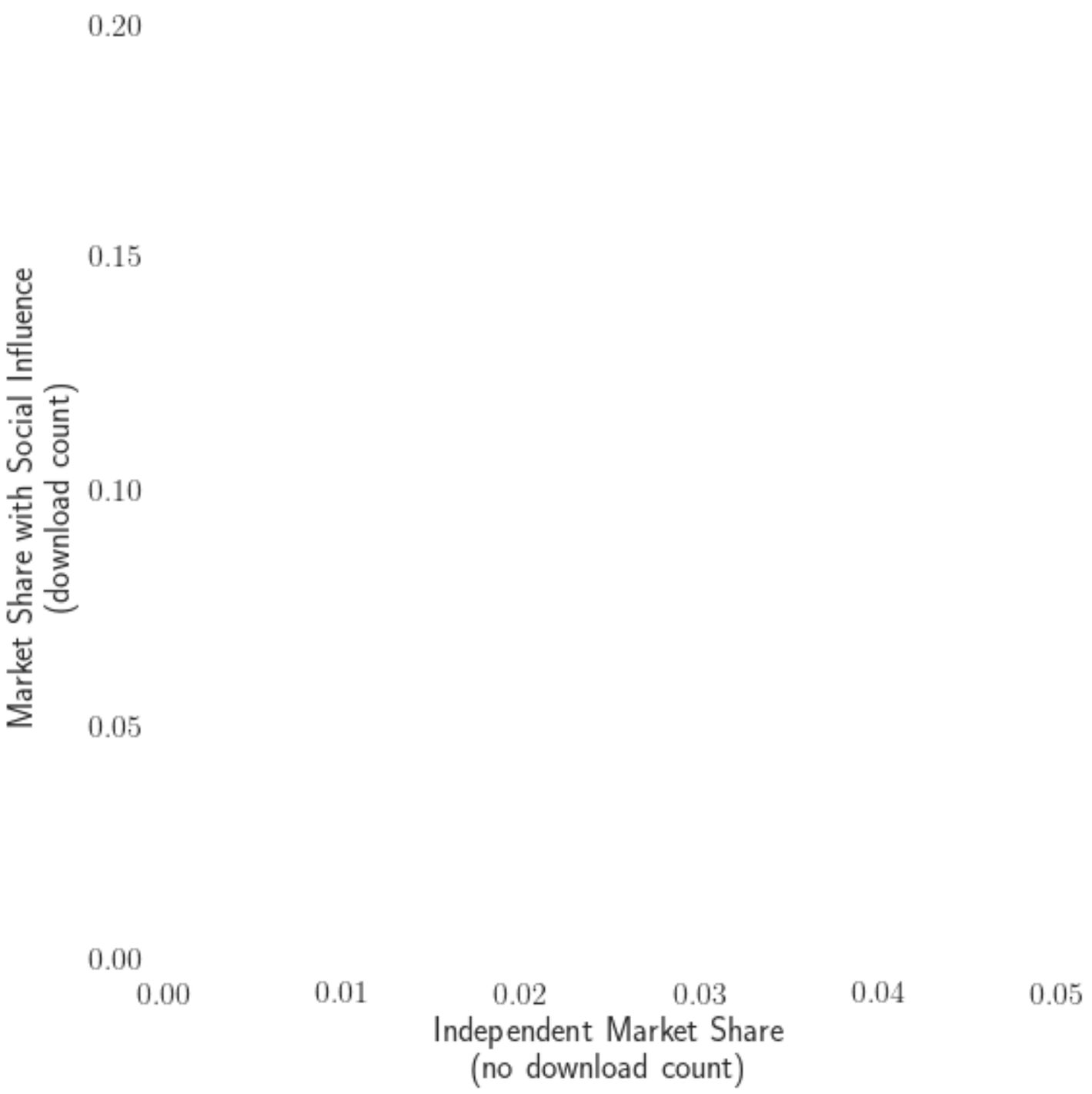


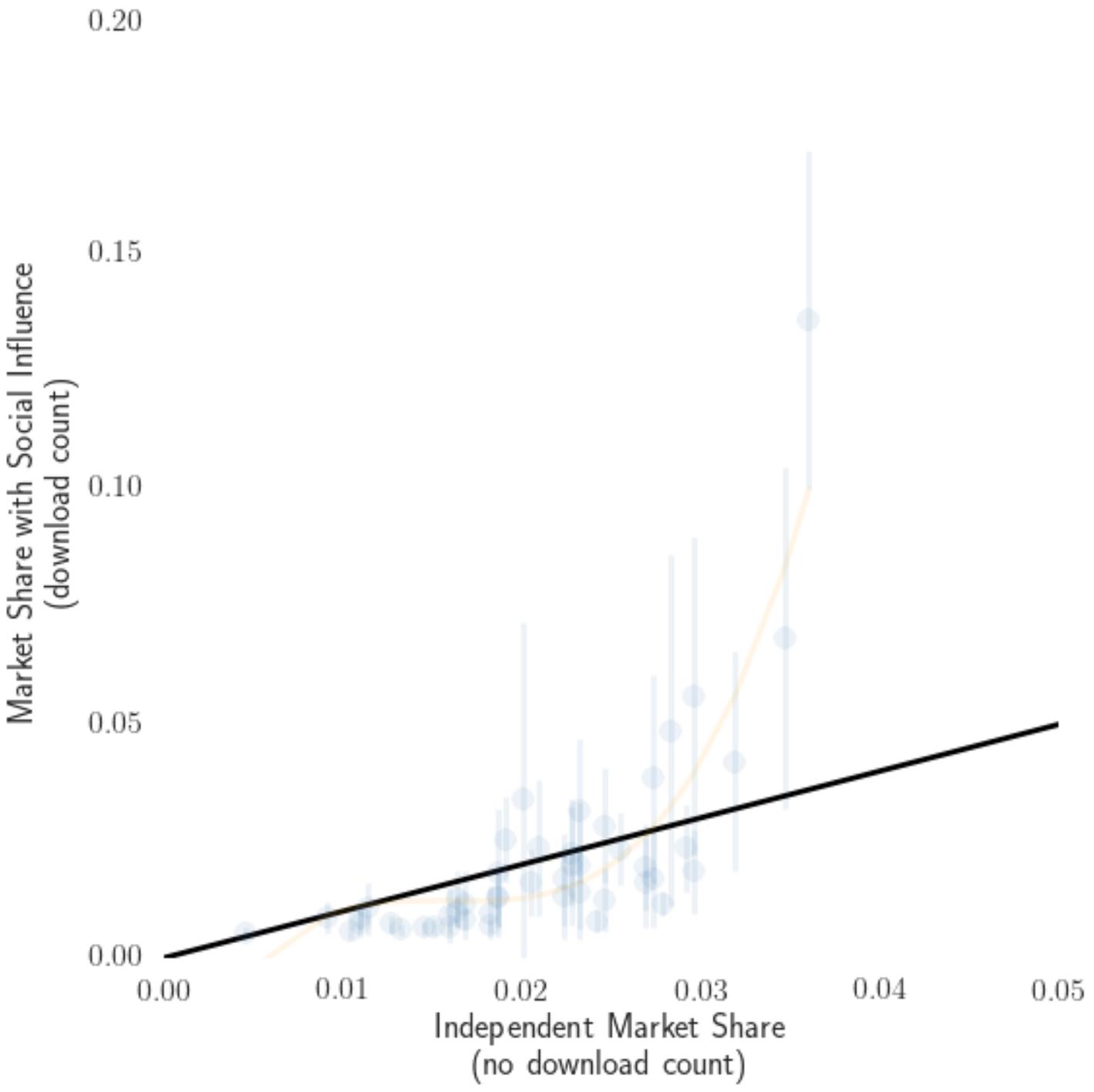
Salganik, Dodds, and Watts, 2006

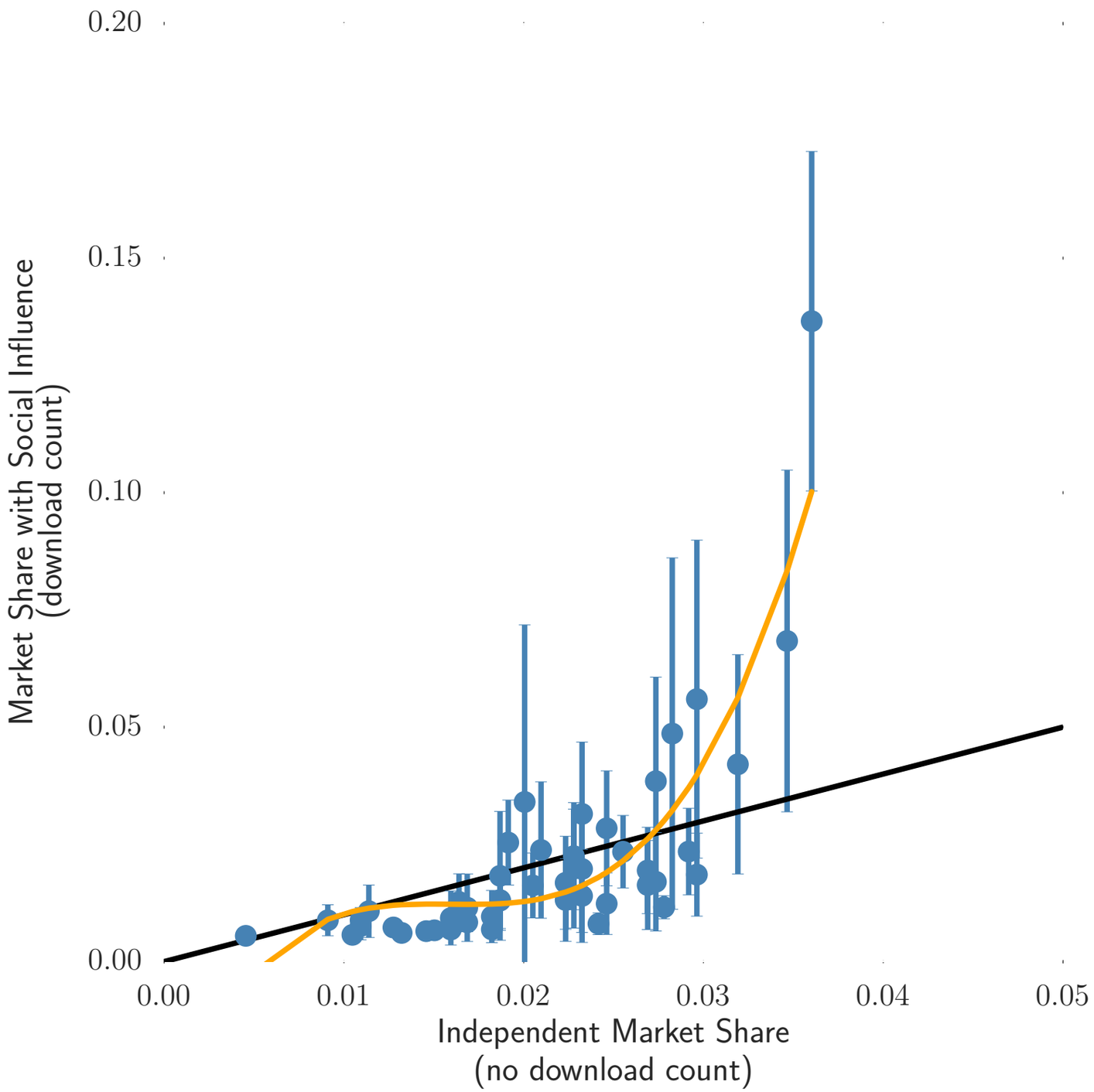
listen ★ rate ★ download ★ MUSIC LAB

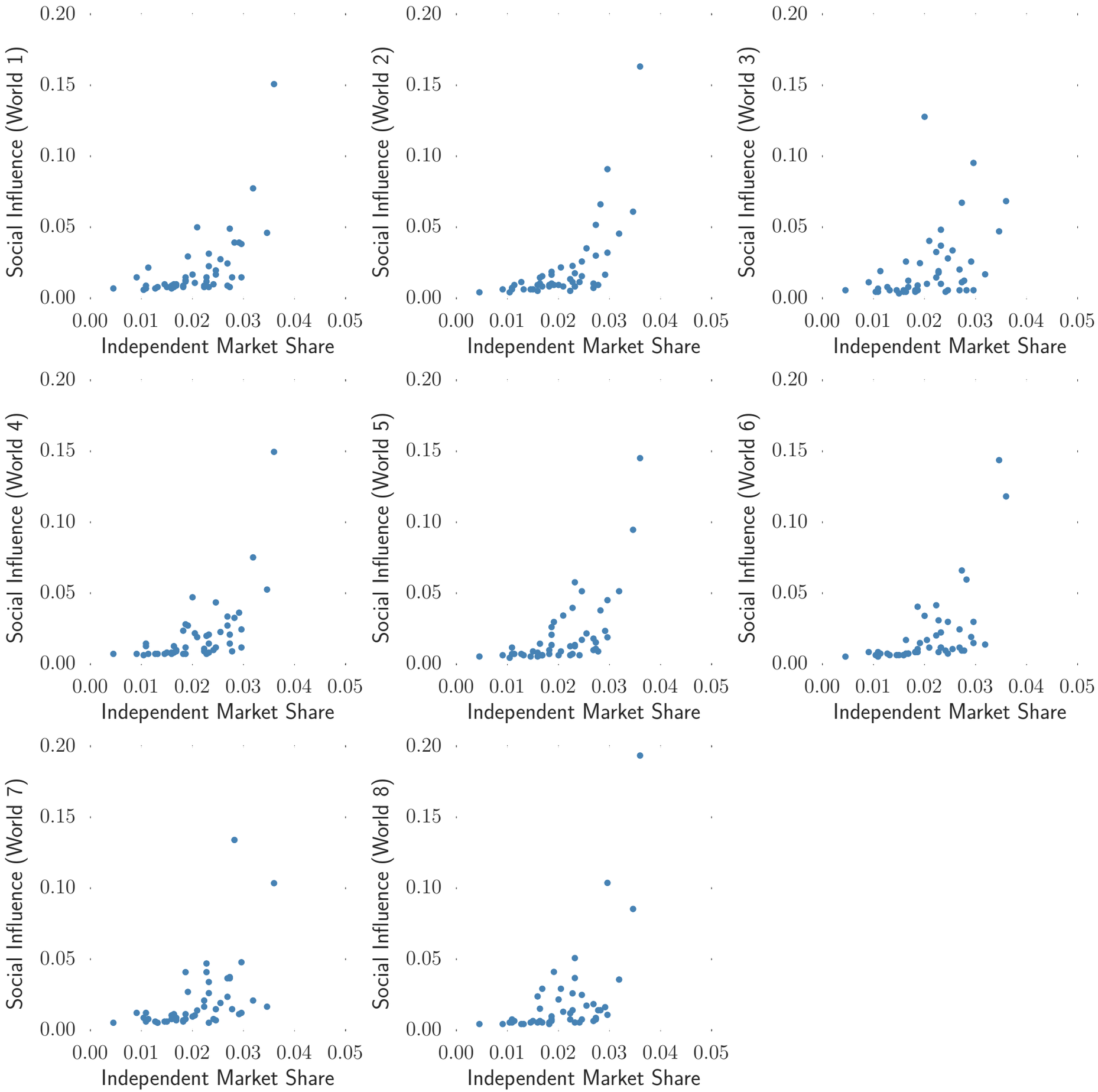


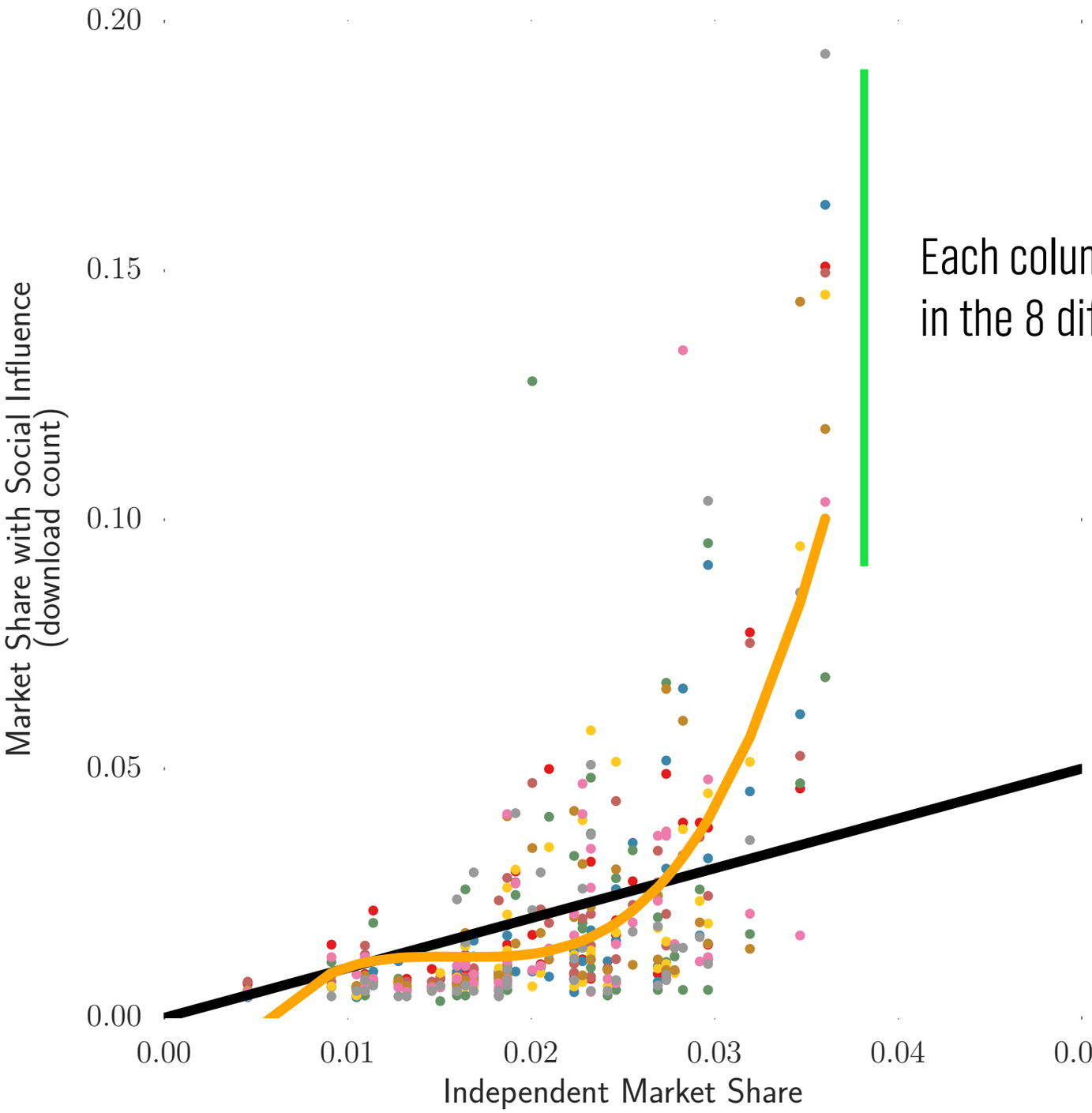
Salganik, Dodds, and Watts, 2006

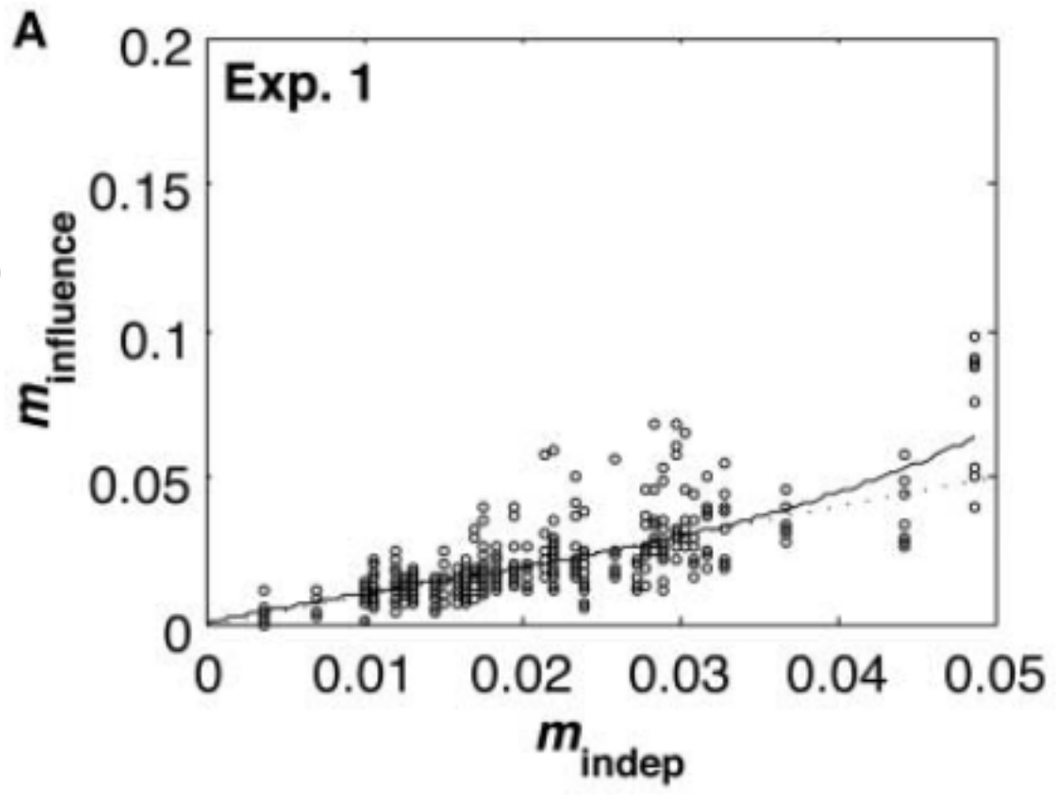

Market Share with Social Influence
(download count)
0.20
0.15
0.10
0.05
0.00
0.00
0.01
0.02
0.03
0.04
0.05
Independent Market Share
(no download count)


0.20
0.15
0.10
0.05
0.00
Market Share with Social Influence
(download count)
0.00
0.01
0.02
0.03
0.04
0.05
Independent Market Share
(no download count)


0.20
0.15
0.10
0.05
0.00
0.00
0.01
0.02
0.03
0.04
0.05
Market Share with Social Influence
(download count)
Independent Market Share
(no download count)


Social Influence (World 1)
Social Influence (World 2)
Social Influence (World 3)
Social Influence (World 4)
Social Influence (World 5)
Social Influence (World 6)
Social Influence (World 7)
Social Influence (World 8)
0.20
0.15
0.10
0.05
0.00
0.00 0.01 0.02 0.03 0.04 0.05
Independent Market Share


0.20
0.15
0.10
0.05
0.00
Market Share with Social Influence
(download count)
0.00
0.01
0.02
0.03
0.04
0.0
Independent Market Share
Each column is the same song
in the 8 different worlds


A
Exp. 1
0.2
0.15
0.1
0.05
0
$m_{influence}$
0
0.01
0.02
0.03
0.04
0.05
m_{indep}

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)

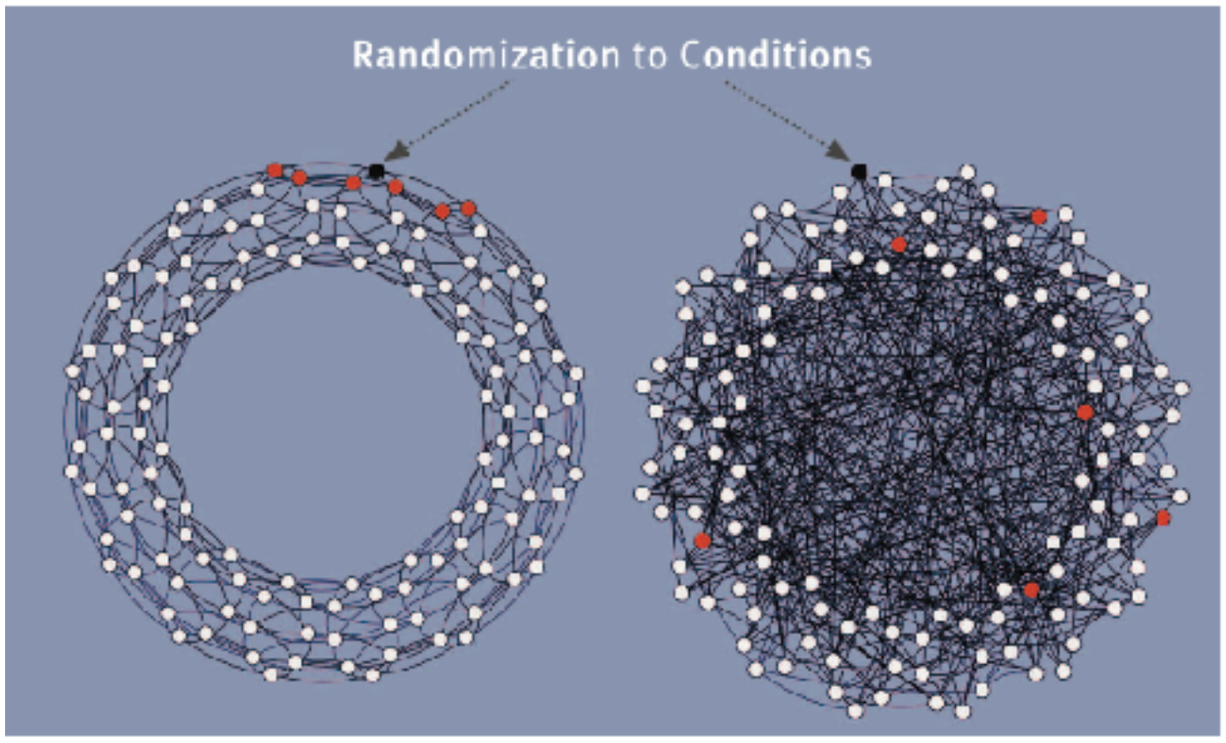
Big Data

Algorithmic Maturity

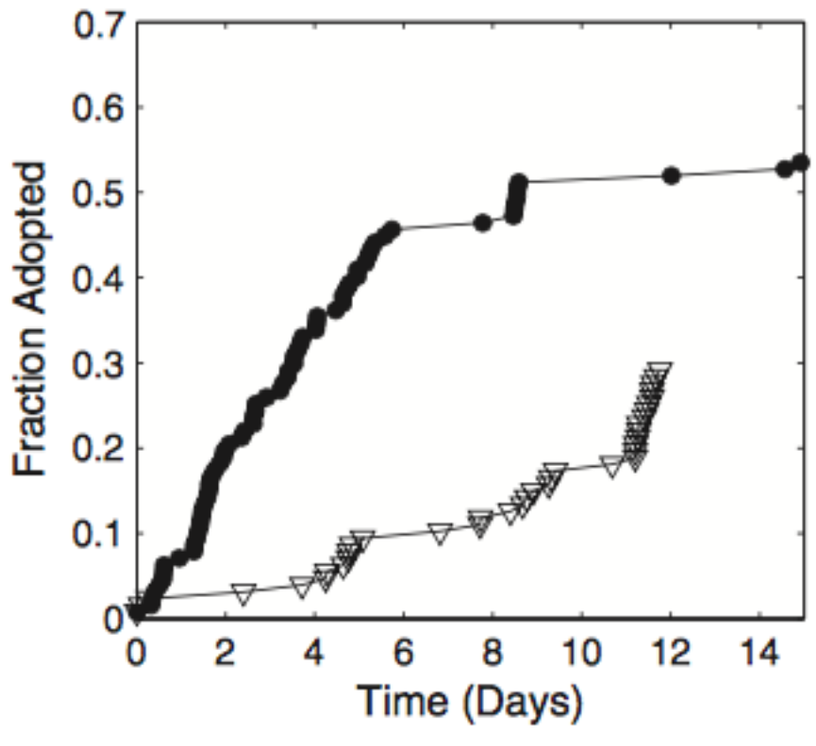
Resulting in different research programmes

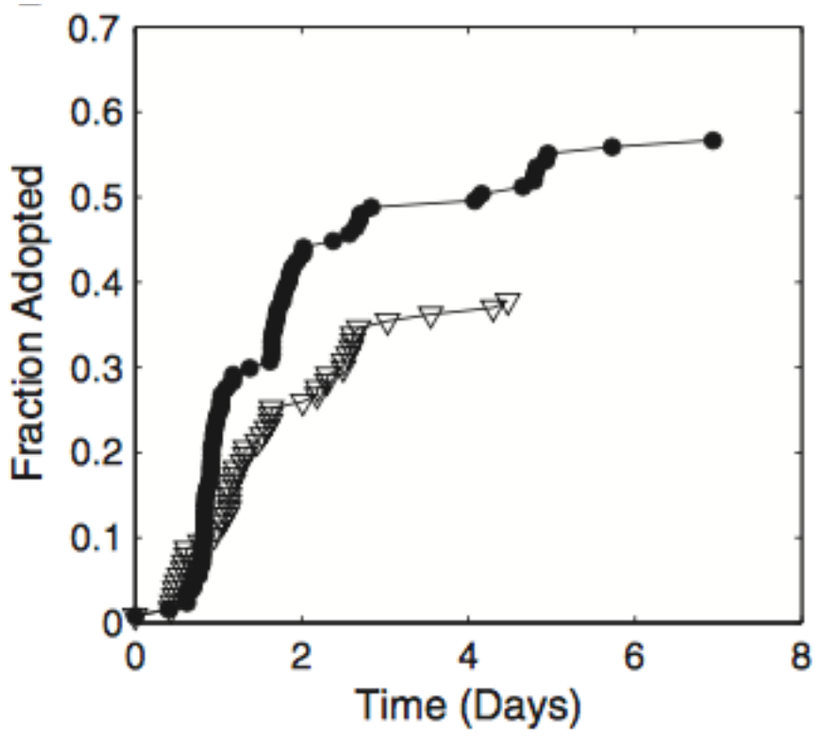
Experimentation	Text	Digital trace/ human dynamics	Administrative data
APIs	HTML	SQL	Deep learning
Machine learning	Javascript		Sensor Development
Programming (Python/R)	NoSQL		Network Analysis

Experimentation



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





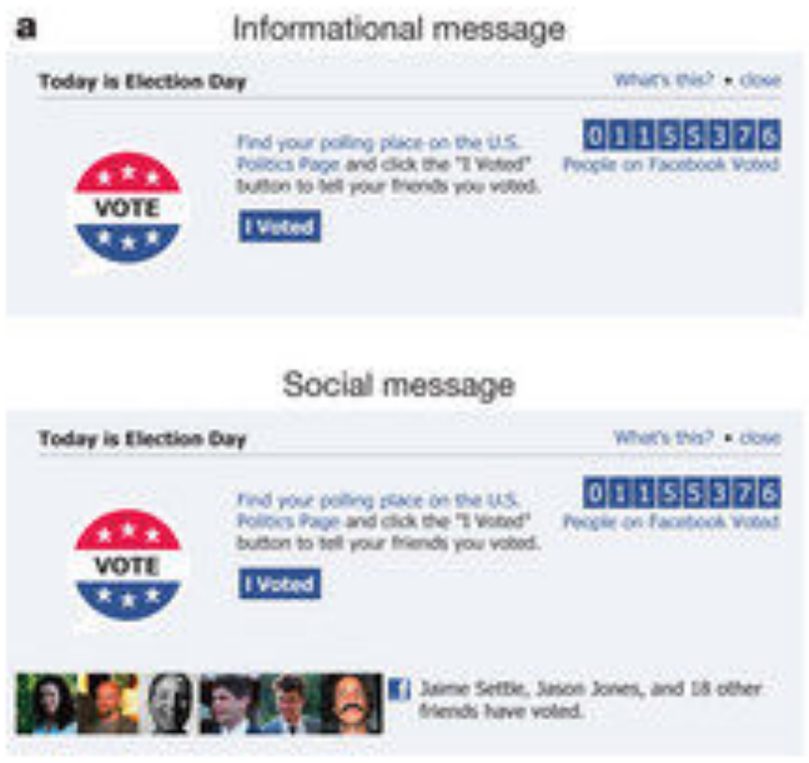
Experimentation

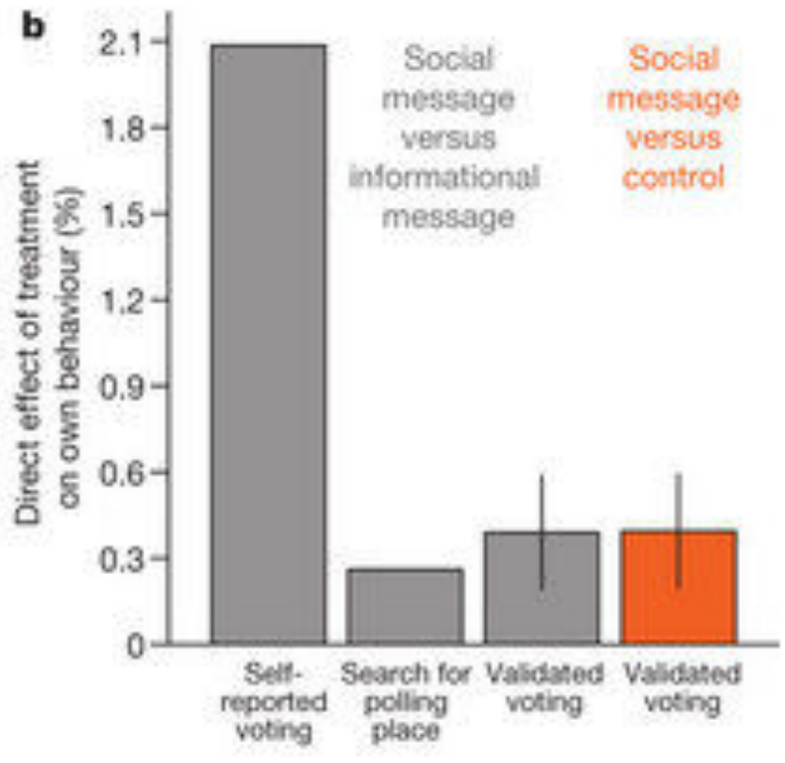
NATURE | LETTER

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

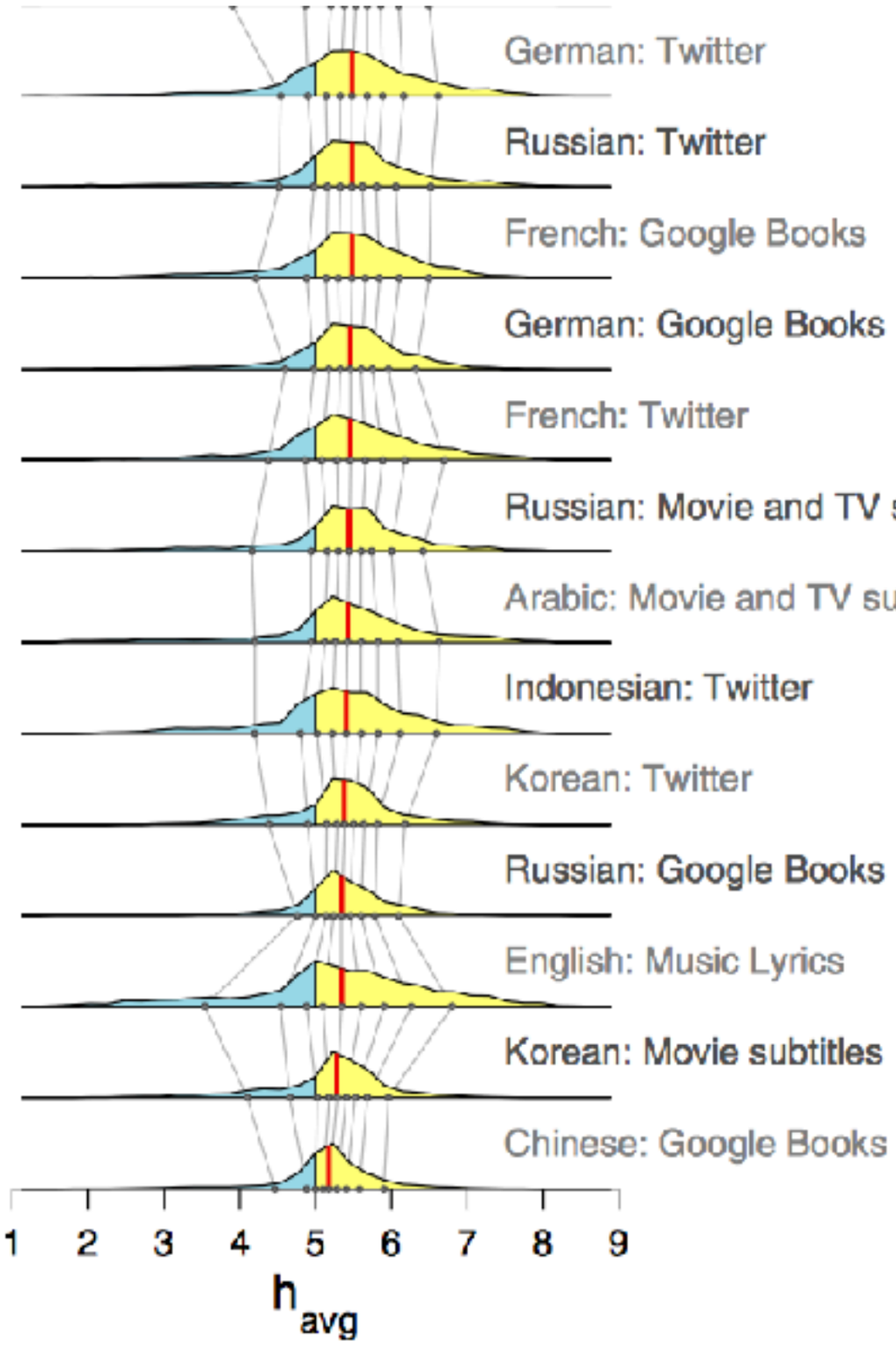


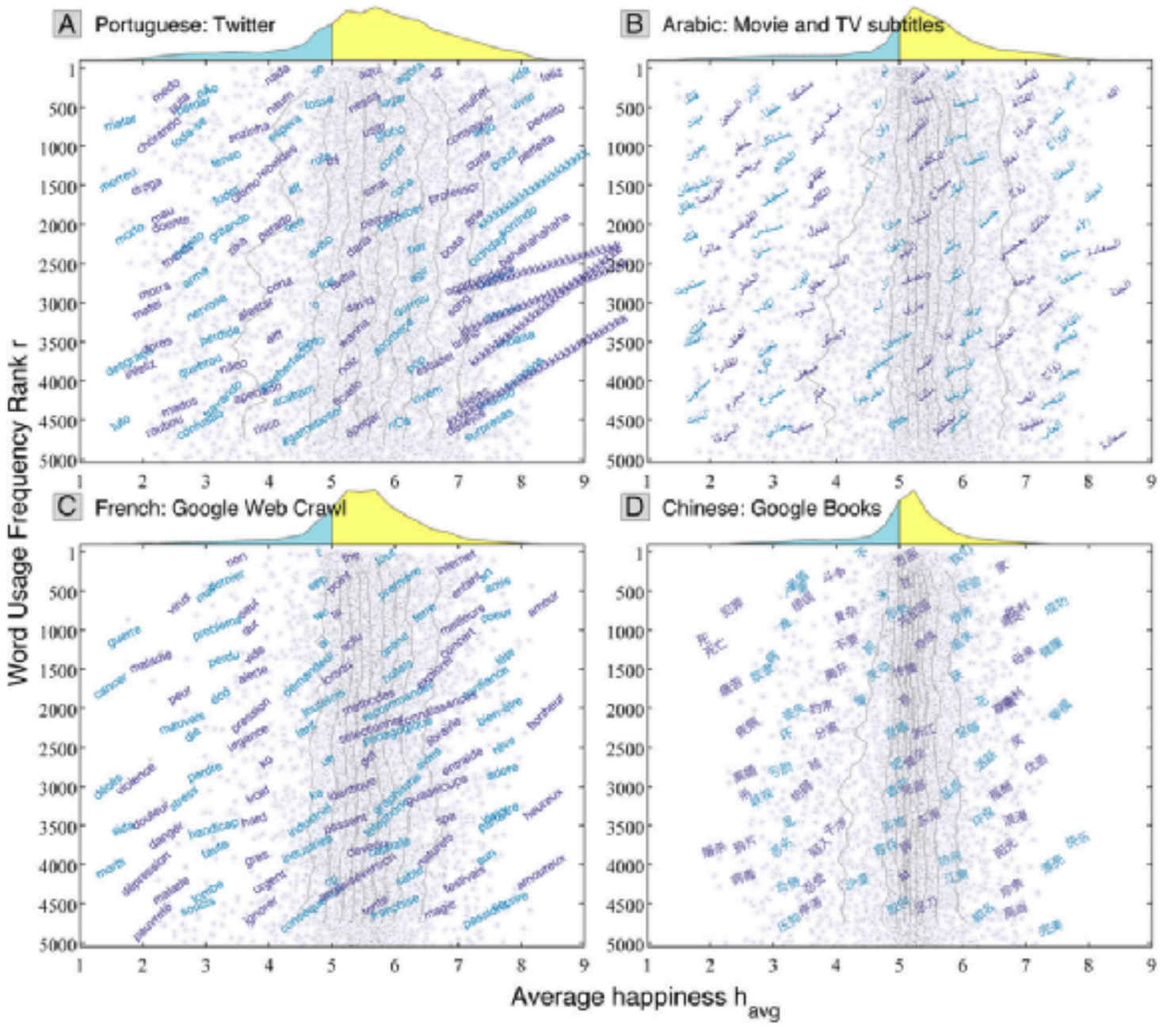


Text

Human language reveals a universal positivity bias

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





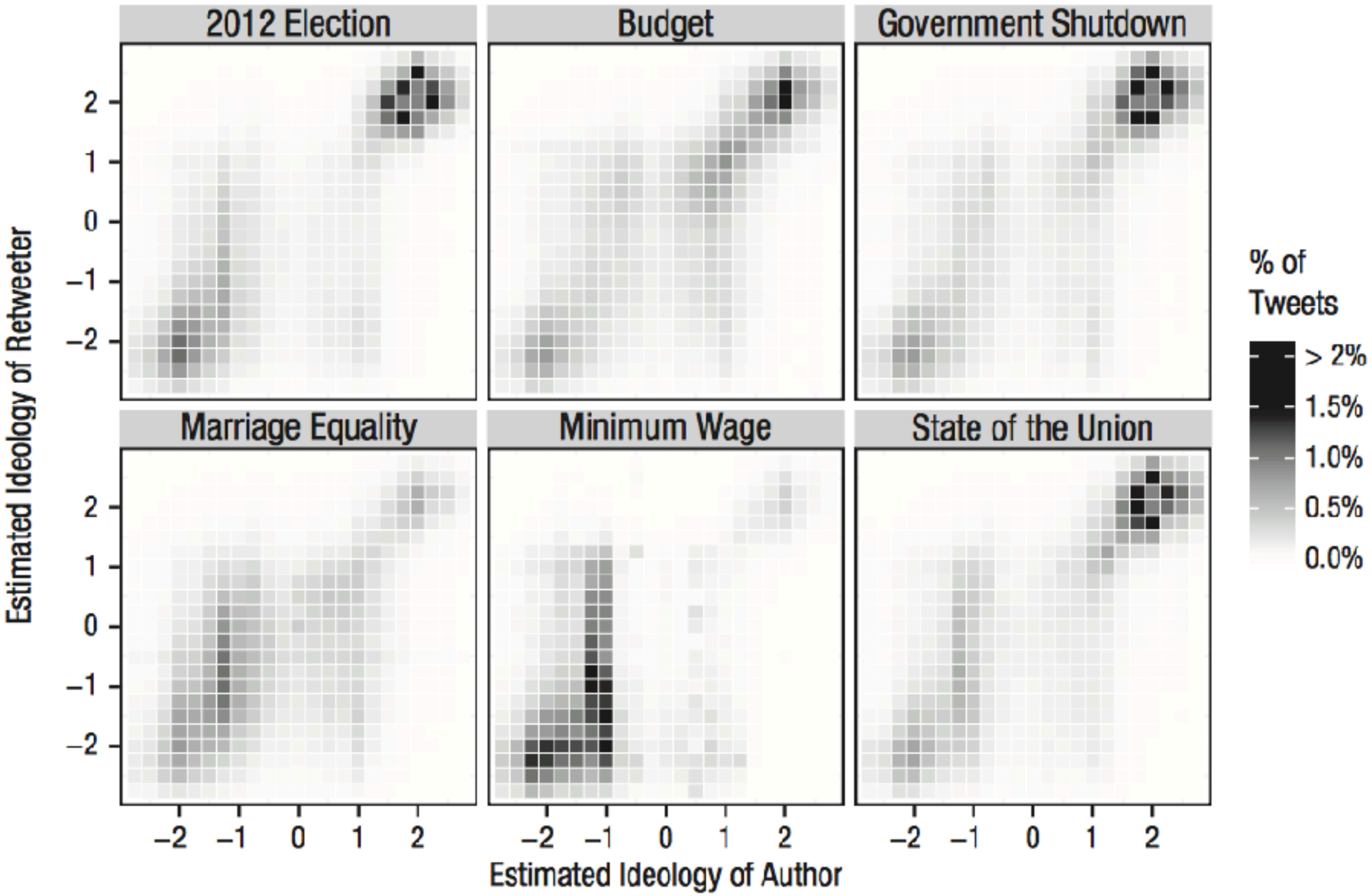
Text

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

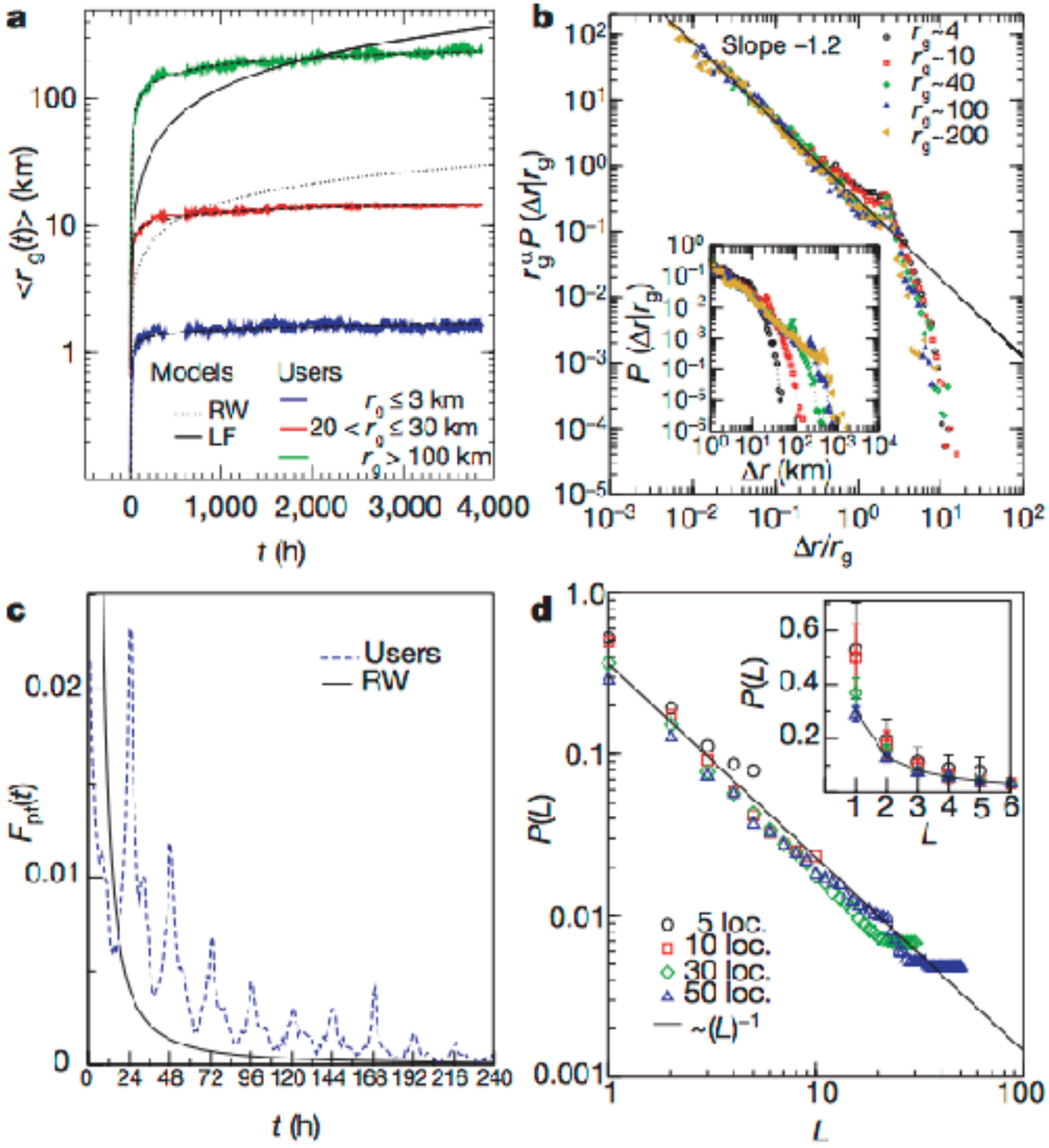


Digital Trace

Understanding individual human mobility patterns

Marta C. González[1], César A. Hidalgo[1,2] & Albert-László Barabási[1,2,3]

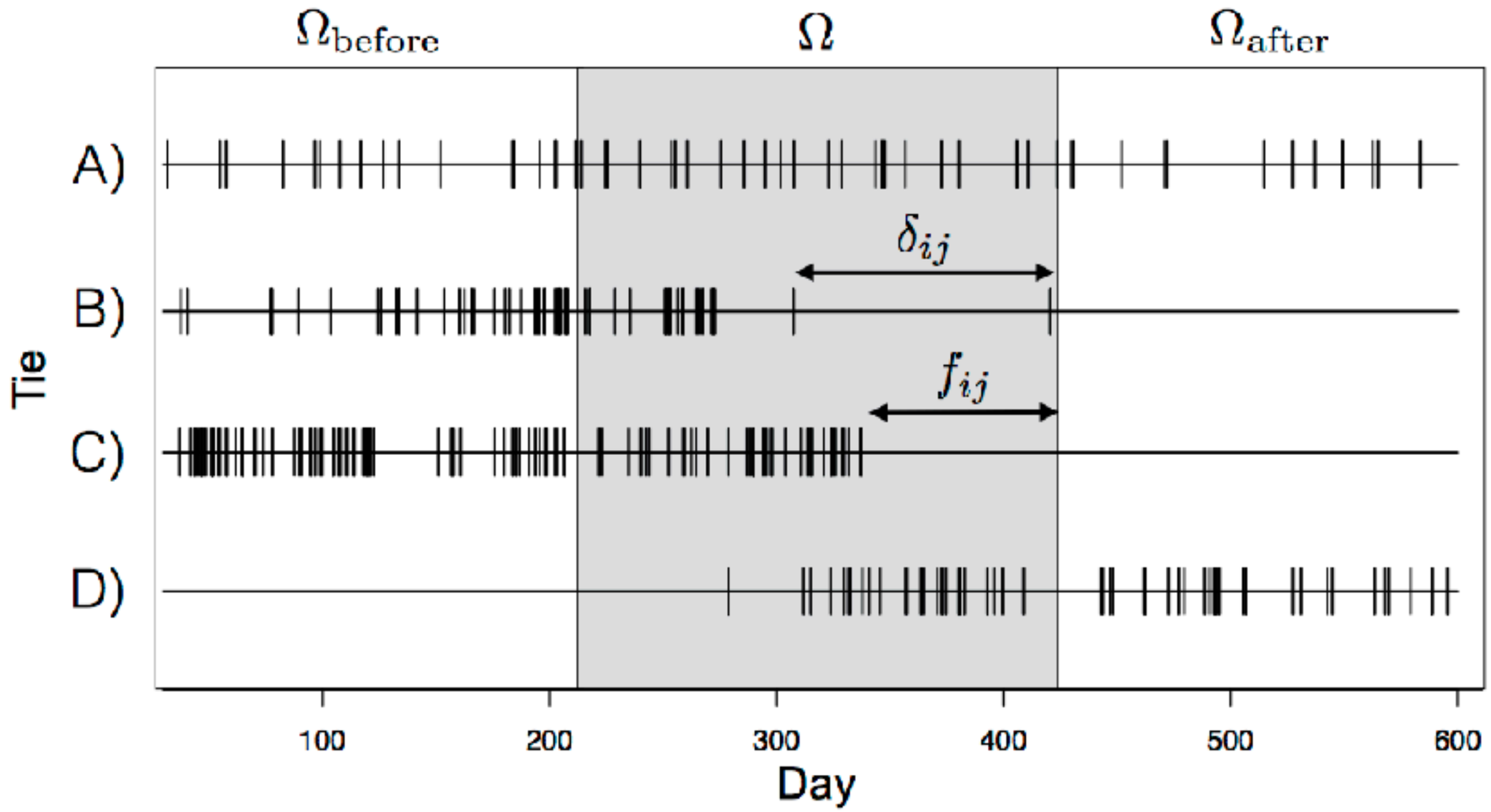
Vol 453 | 5 June 2008 | doi:10.1038/nature06958



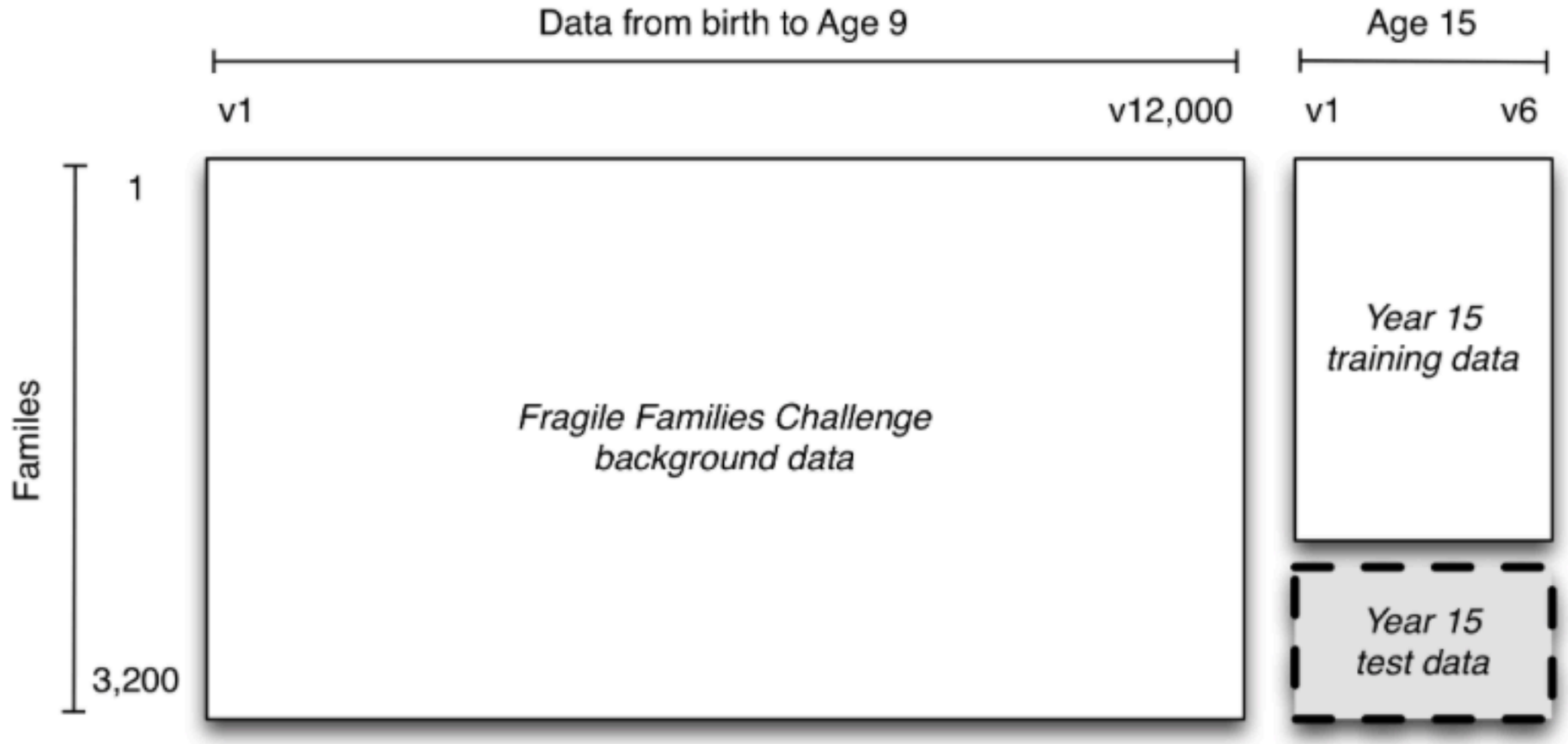
Digital Trace

Temporal patterns behind the strength of persistent ties

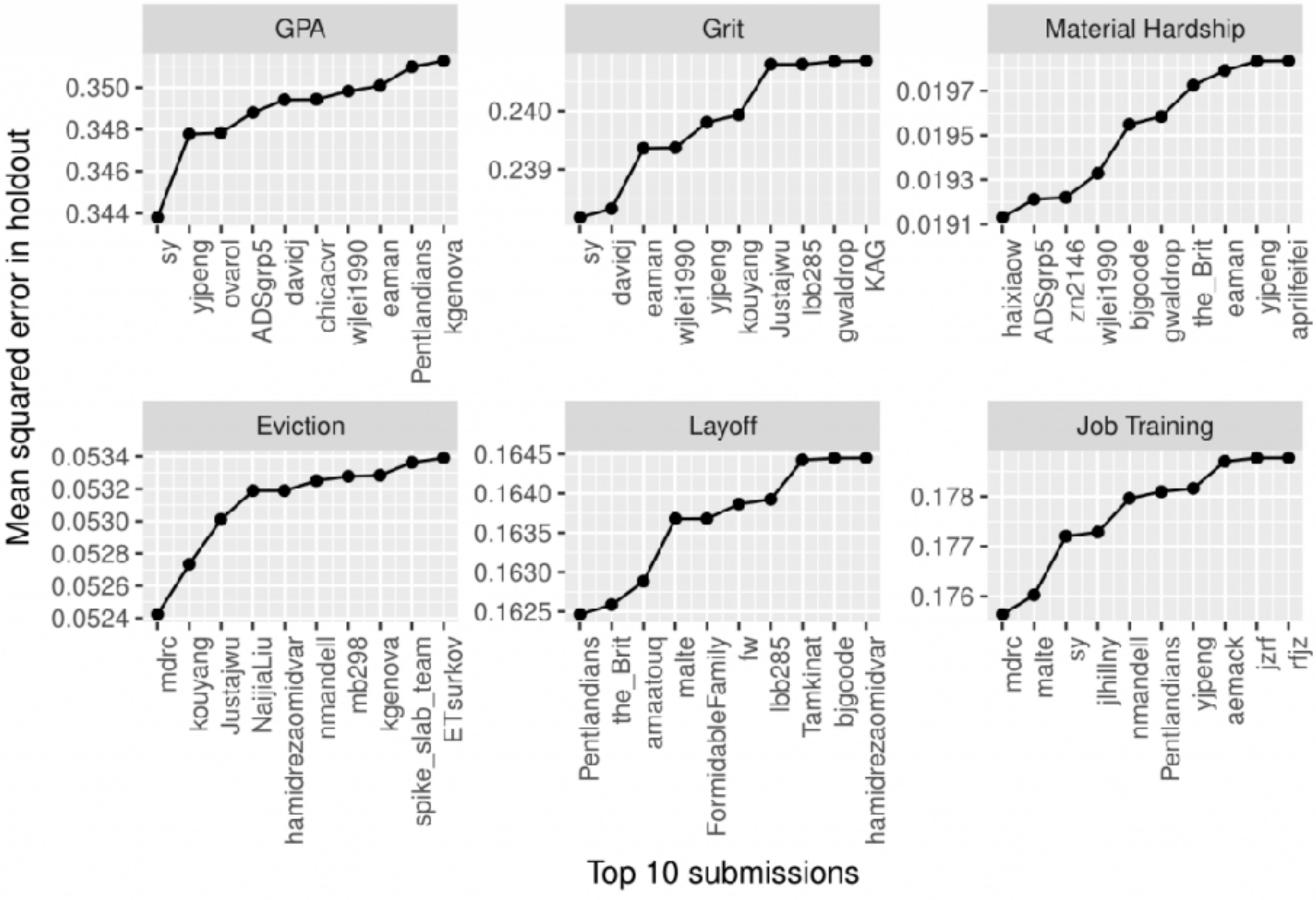
Henry Navarro[1], Giovanna Miritello[1,2], Arturo Canales[3], Esteban Moro [1*]



Administrative data



Administrative data

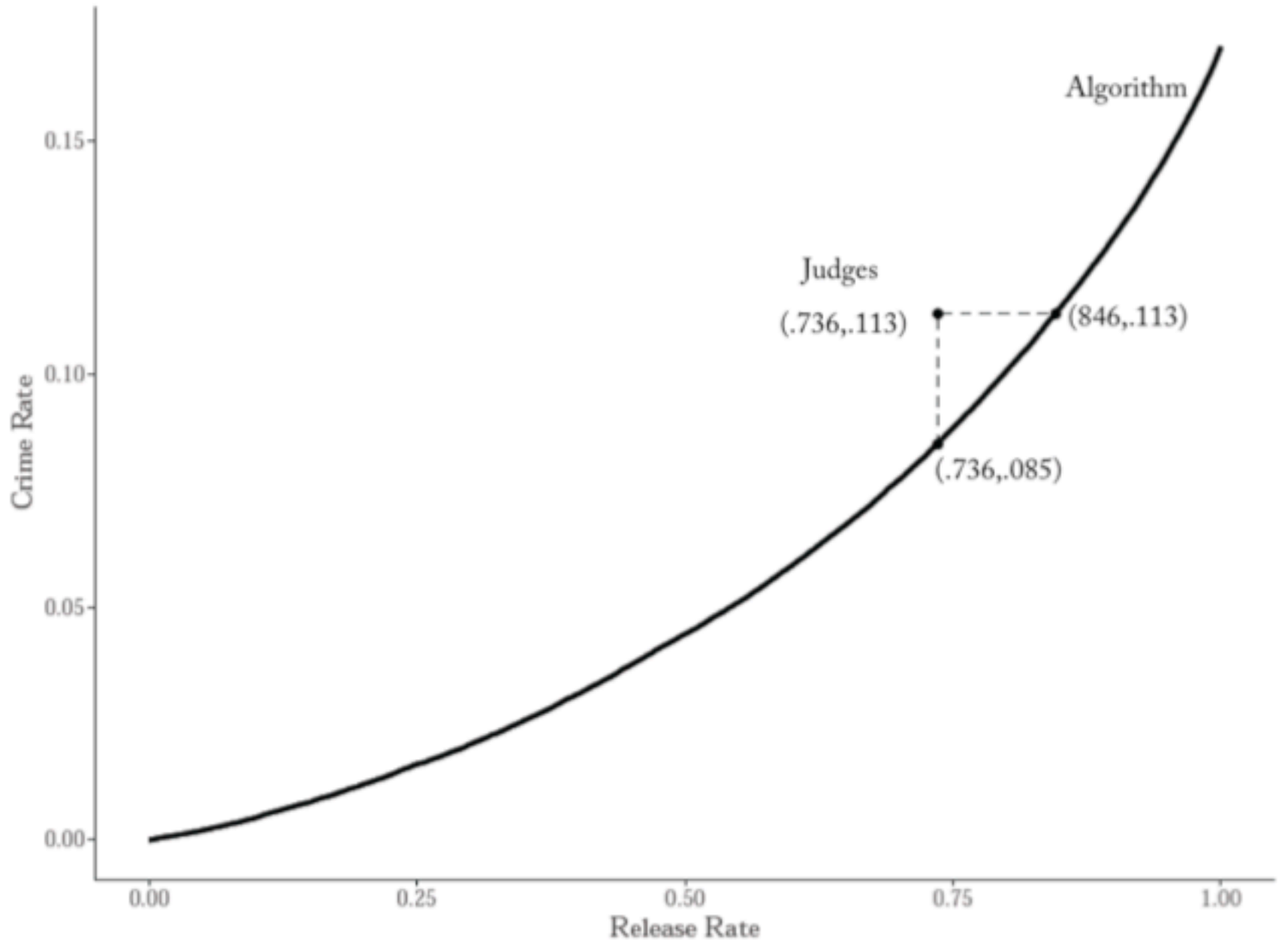


Administrative data

Human Decisions and Machine Predictions*

Jon Kleinberg †
Himabindu Lakkaraju‡
Jure Leskovec§
Jens Ludwig¶
Sendhil Mullainathan‖

January 23, 2017



Administrative data

The modern vision —
CSS is for both qualitative and quantitative research

Qualitative Research

- Accelerates data collection if digitally available (text, image, video)
- Data mining to produce broad-based qualitative features
  (i.e. key words/concepts/associations and evolution over time)

Expands breadth of data for qualitative insights

The modern vision — CSS is for both qualitative and quantitative research

Quantitative Research

- Make use of novel datasets
- Increase predictive power
- Generate features at scale in big data for use in standard modeling frameworks

Increasing depth of insights in larger data settings
Novel datasets/insights from increased scale

What's our plan?

Weeks 1-3
Python Fundamentals

Weeks 4-6
Digital Data Collection
& Processing

Weeks 7-11
Text

Weeks 12-14
Networks

Deliverables

Weeks 1-5
Coding assignments
Flex that coding muscle!

Weeks 4-14
Final Project Work

Grade Breakdown

Assignments 5 Programming Assignments	**40%**
Participation Paper presentation In-class discussion	**10%**
Final Project Project Proposal Final Paper	**10%** **40%**

Programming assignments focus on completeness and readability

Final Project is a write up on research of your own choosing that leverages class concepts and skills. Grade based on quality of work.

About me

BS in Molecular Biotechnology, Phd in “Interdisciplinary Biology”

Dissertation on using network analytics to map organismal metabolisms, develop predictive algorithms, and evaluate database reliability

Postdoctoral work in public health/health informatics/implementation science with a focus on electronic health records data and physician adoption of innovations

Research tracks on violent behavior in the US at the population level and international terrorism at the organizational level

Current Primary Focus:

Developing natural language models and ontologies to decode litigation events in court records

Systematically analyze litigation event outcomes in the US federal courts and assess variation due to local procedure, judge assignment, party variation, etc.