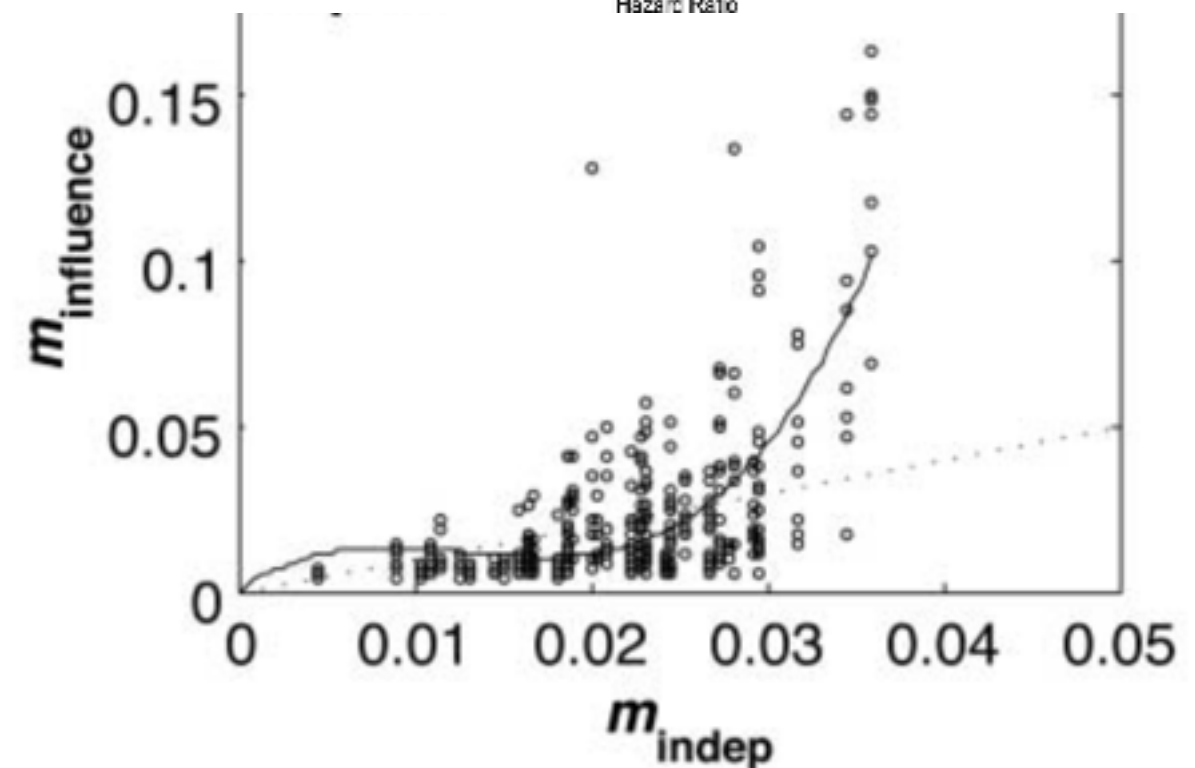
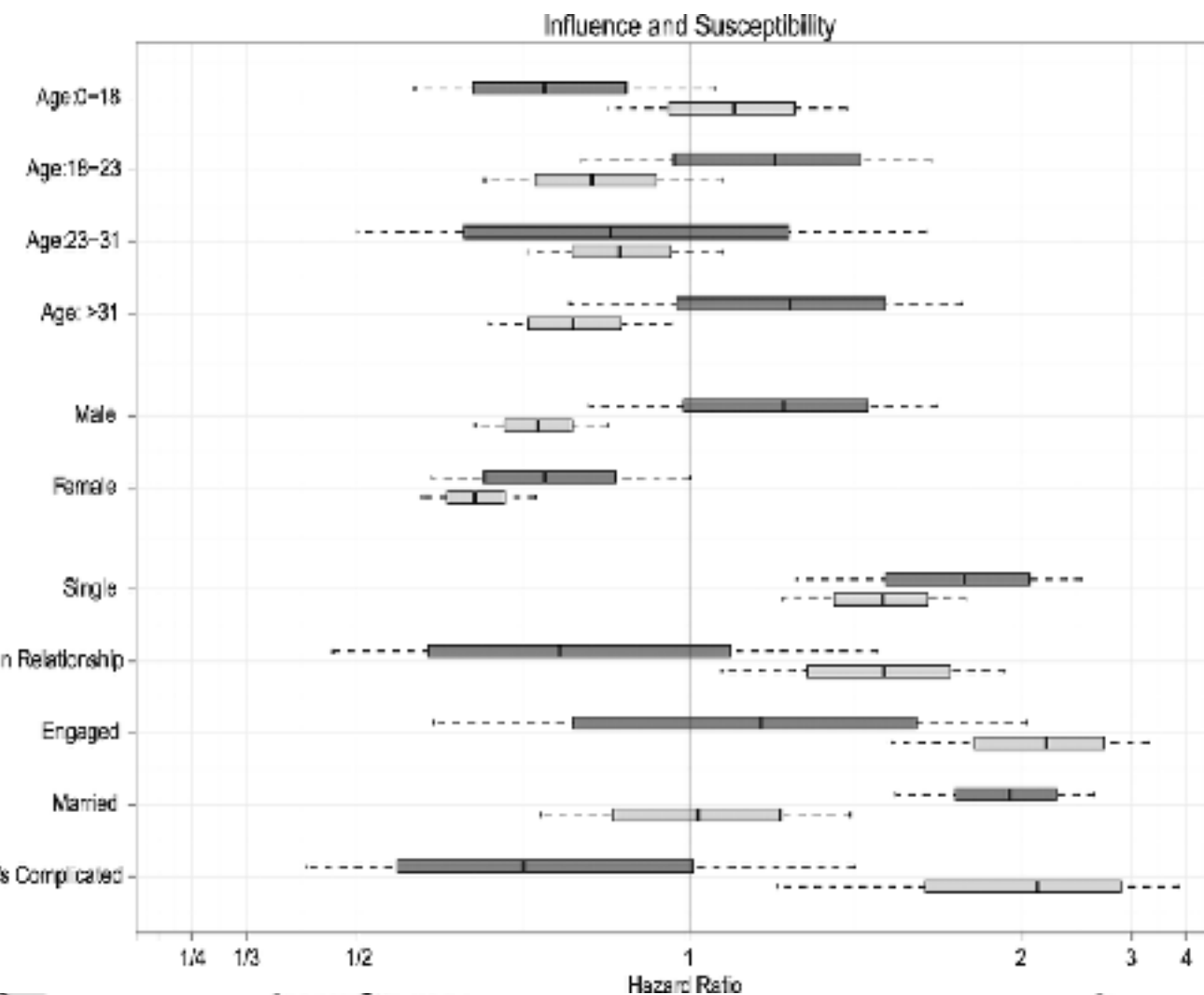
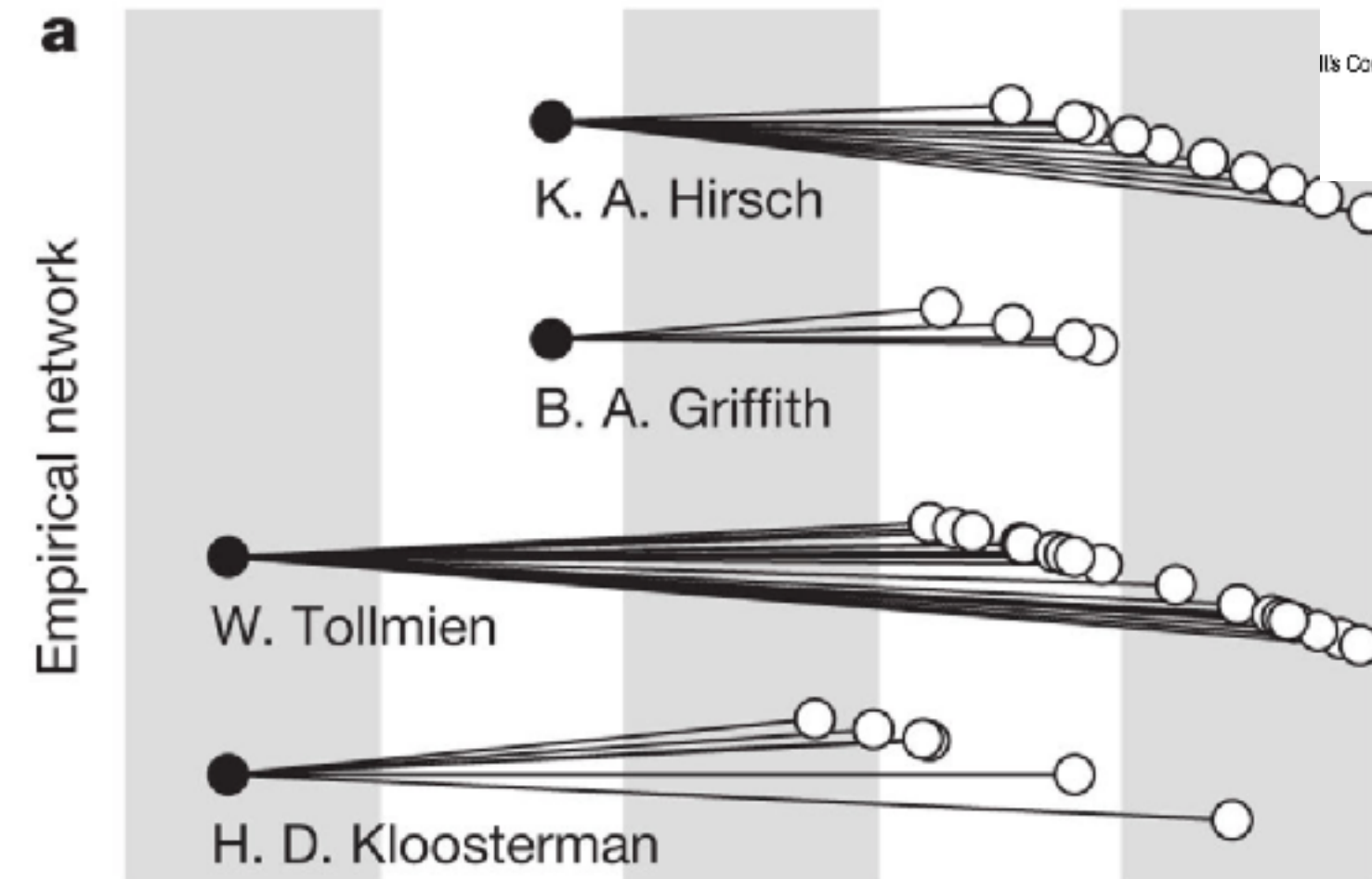
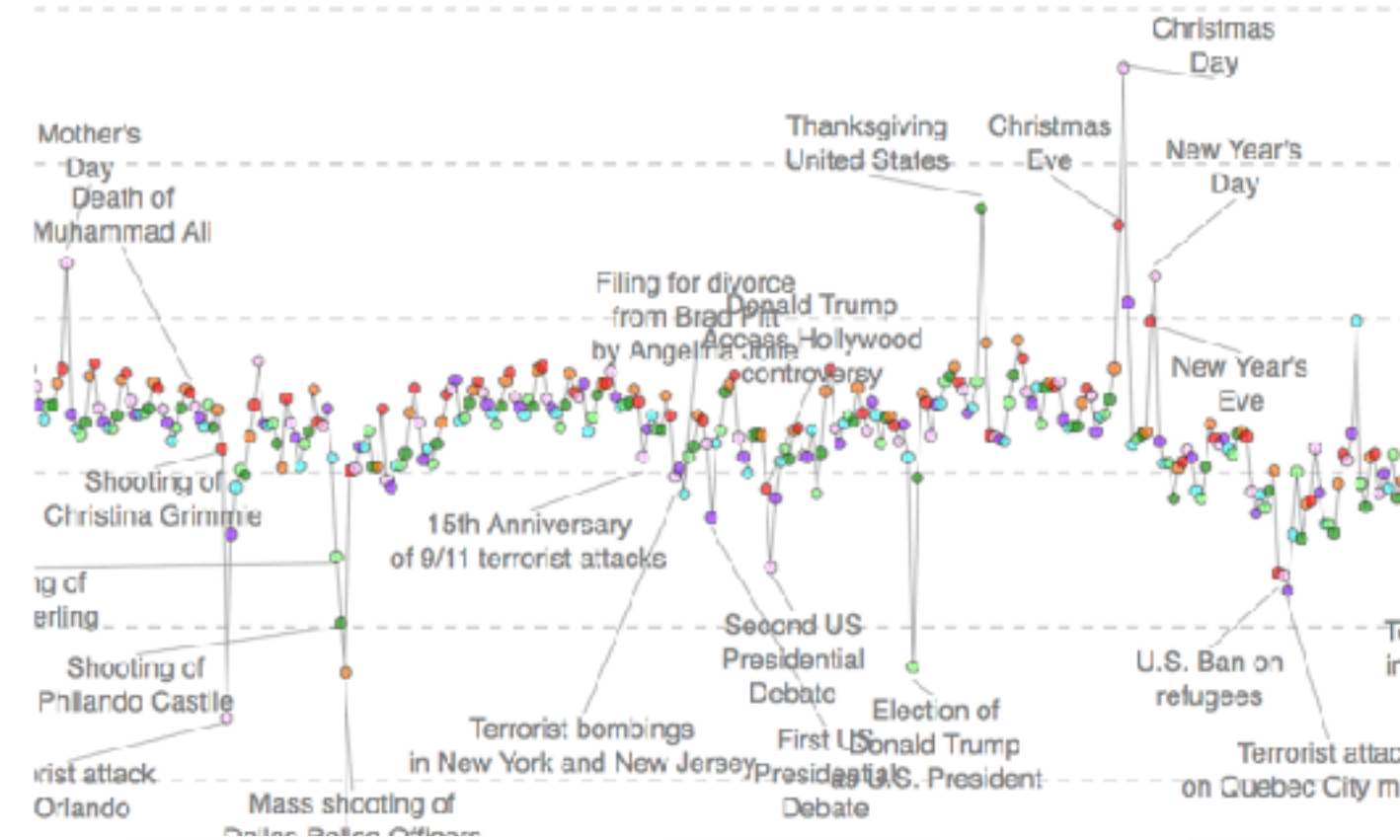
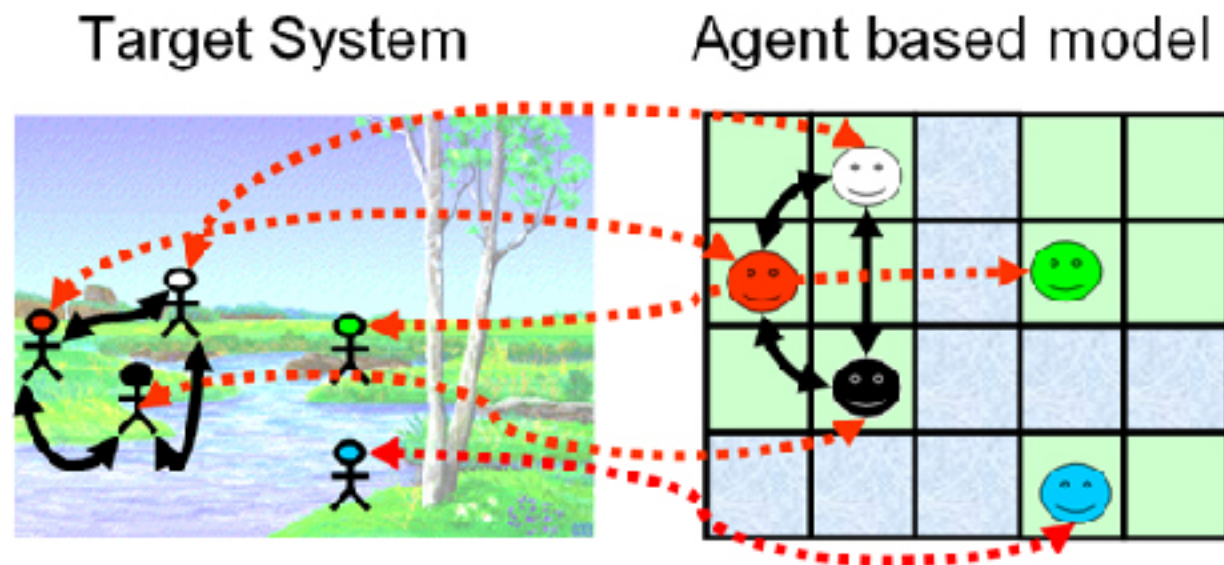


# **What is Computational Social Science?**

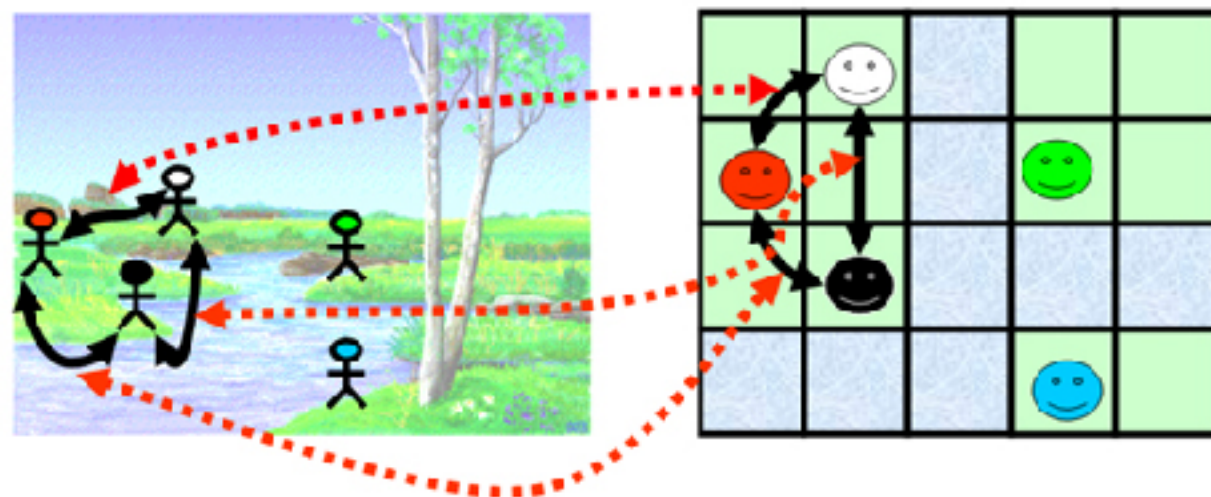
# Computational Social Science has enabled significant advances



# But the definition of computational social science is not static

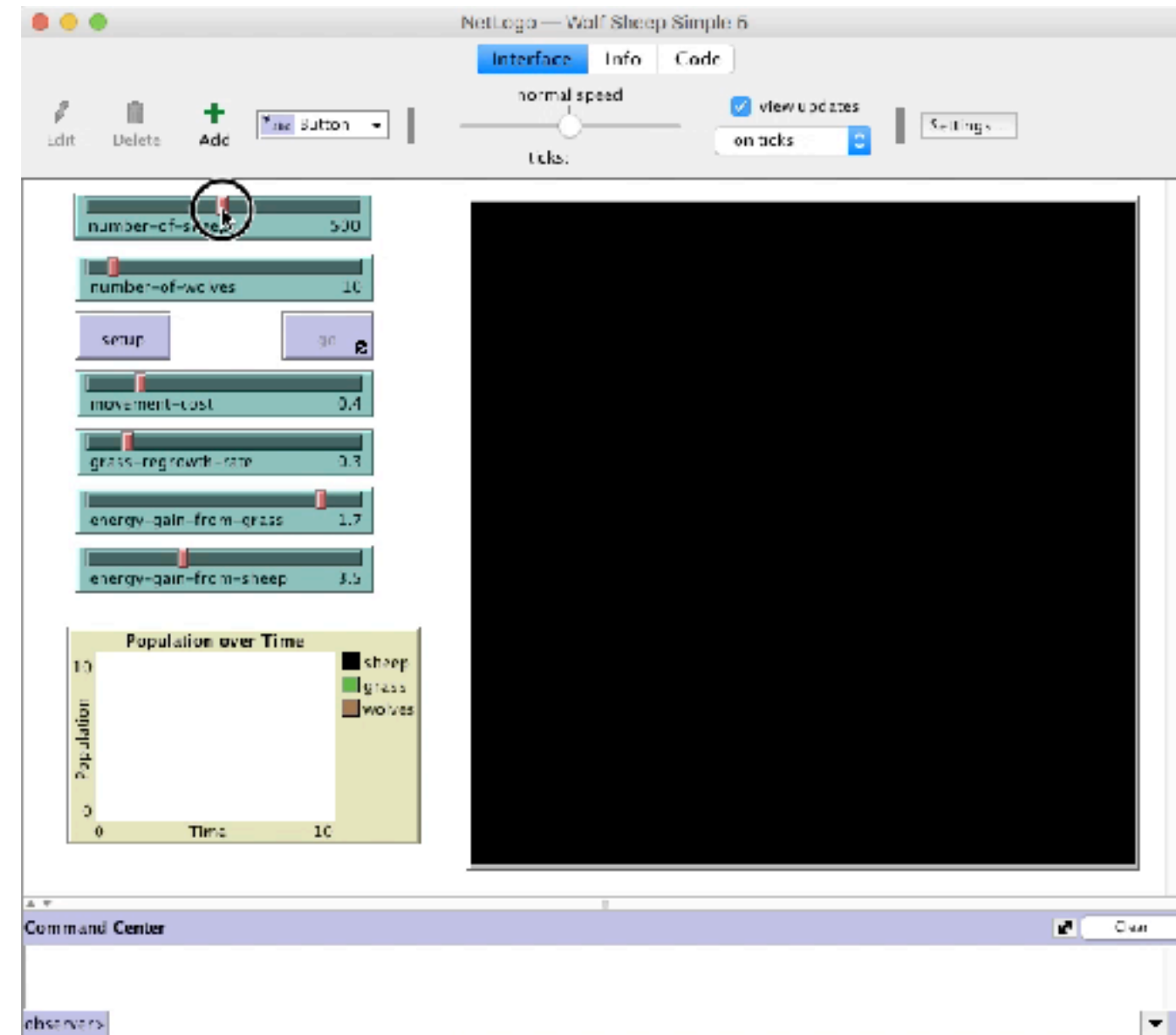


Entities      Agents

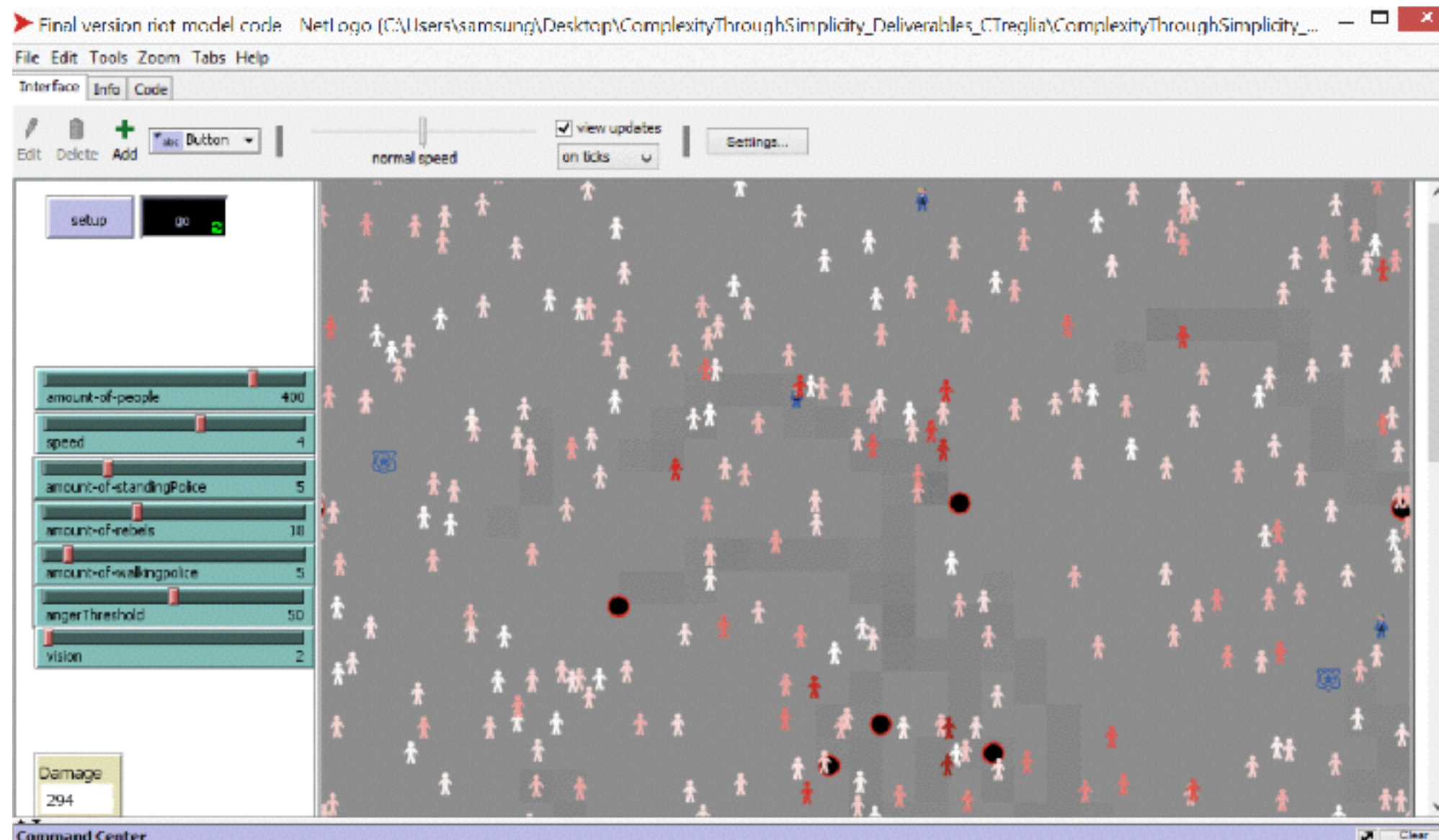


Interaction between entities      Interactions between agents

1990s to mid-2000s



ABMs were directly analogous to historically 'recent' theoretical papers



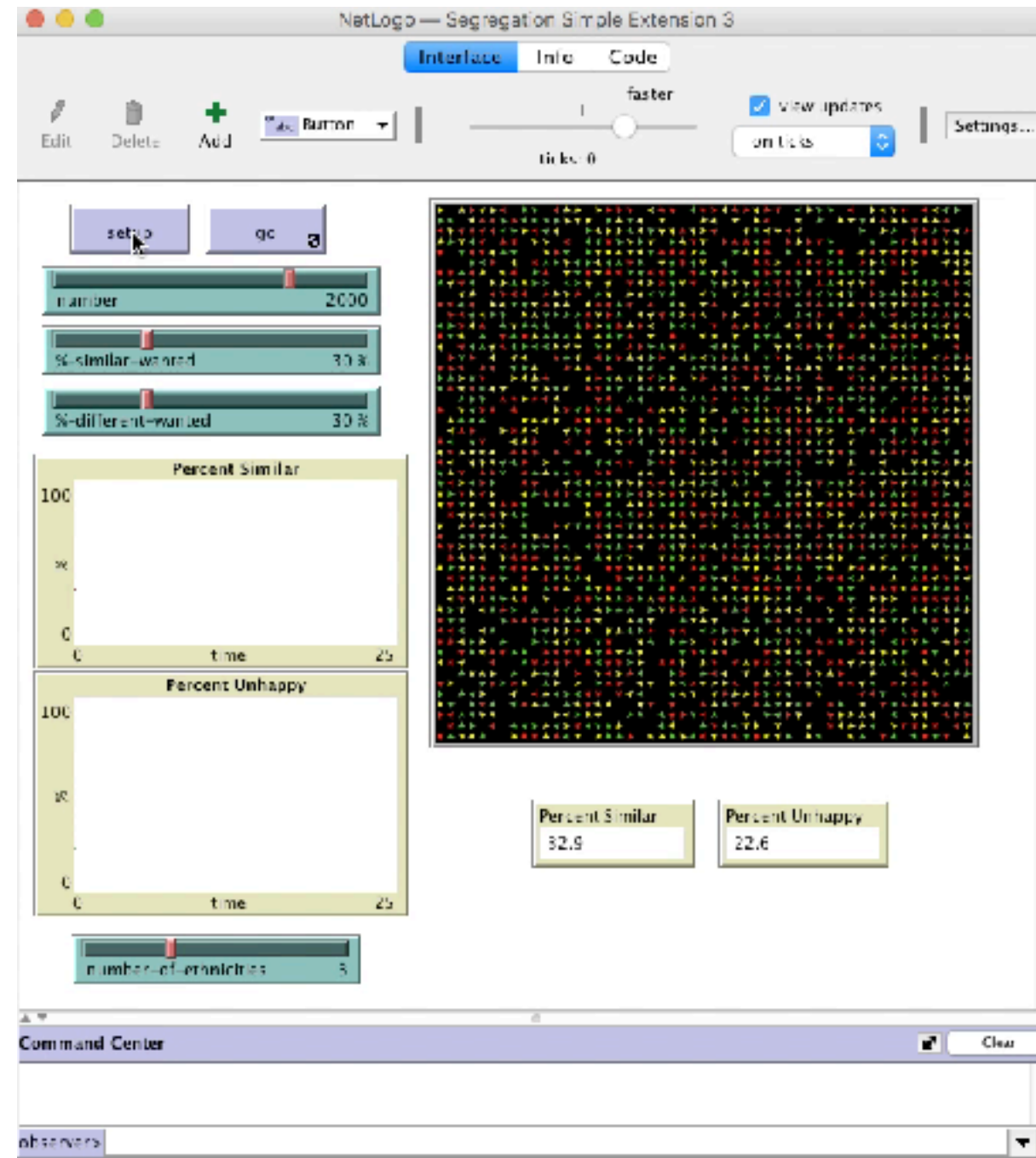
## Threshold Models of Collective Behavior

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 ★ MUSIC LAB

download count



A screenshot of a music library table. The table has columns for song title, artist, album, genre, and a 'Downloads' column. The 'Downloads' column contains numerical values representing the number of times each song has been downloaded.

Song Title	Artist	Album	Genre	Downloads
1. The Sound of Silence	Simon & Garfunkel	Bridge Over Troubled Water	Folk	10
2. Hotel California	Eagles	Hotel California	Rock	15
3. Billie Jean	Michael Jackson	Thriller	Pop	20
4. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance	12
5. Smells Like Teen Spirit	Nirvana	Bleach	Rock	8
6. Billie Jean	Michael Jackson	Thriller	Pop	18
7. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance	14
8. Smells Like Teen Spirit	Nirvana	Bleach	Rock	9
9. Billie Jean	Michael Jackson	Thriller	Pop	19
10. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance	13
11. Smells Like Teen Spirit	Nirvana	Bleach	Rock	7
12. Billie Jean	Michael Jackson	Thriller	Pop	17
13. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance	11
14. Smells Like Teen Spirit	Nirvana	Bleach	Rock	6
15. Billie Jean	Michael Jackson	Thriller	Pop	16
16. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance	10
17. Smells Like Teen Spirit	Nirvana	Bleach	Rock	5
18. Billie Jean	Michael Jackson	Thriller	Pop	15
19. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance	9
20. Smells Like Teen Spirit	Nirvana	Bleach	Rock	4

no download count



A screenshot of a music library table, identical to the one above, but with the 'Downloads' column removed. The table contains the same song titles, artists, albums, and genres, but lacks the numerical download counts.

Song Title	Artist	Album	Genre
1. The Sound of Silence	Simon & Garfunkel	Bridge Over Troubled Water	Folk
2. Hotel California	Eagles	Hotel California	Rock
3. Billie Jean	Michael Jackson	Thriller	Pop
4. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance
5. Smells Like Teen Spirit	Nirvana	Bleach	Rock
6. Billie Jean	Michael Jackson	Thriller	Pop
7. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance
8. Smells Like Teen Spirit	Nirvana	Bleach	Rock
9. Billie Jean	Michael Jackson	Thriller	Pop
10. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance
11. Smells Like Teen Spirit	Nirvana	Bleach	Rock
12. Billie Jean	Michael Jackson	Thriller	Pop
13. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance
14. Smells Like Teen Spirit	Nirvana	Bleach	Rock
15. Billie Jean	Michael Jackson	Thriller	Pop
16. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance
17. Smells Like Teen Spirit	Nirvana	Bleach	Rock
18. Billie Jean	Michael Jackson	Thriller	Pop
19. Stayin' Alive	Bee Gees	Saturday Night Fever	Dance
20. Smells Like Teen Spirit	Nirvana	Bleach	Rock

listen



rate



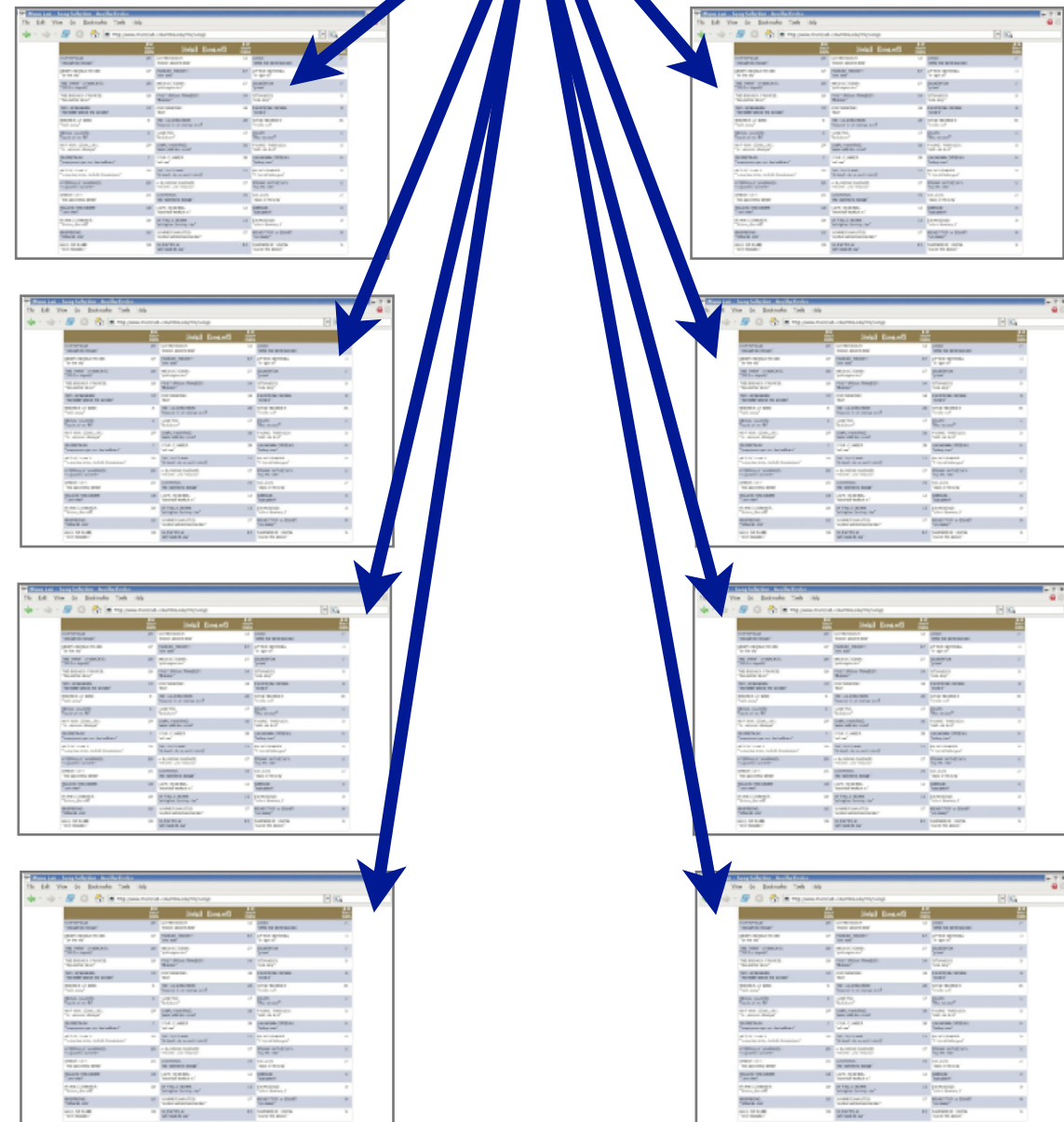
download




download count

no download count

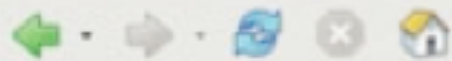
name	album	artist	genre	download count
1. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
2. Mr. Tambourine Man	The Sound of Silence	Simon & Garfunkel	Folk	1
3. The Boxer	The Boxer	Simon & Garfunkel	Folk	1
4. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
5. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
6. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
7. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
8. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
9. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1
10. The Sound of Silence	The Sound of Silence	Simon & Garfunkel	Folk	1



listen ★ rate ★ download ★ 

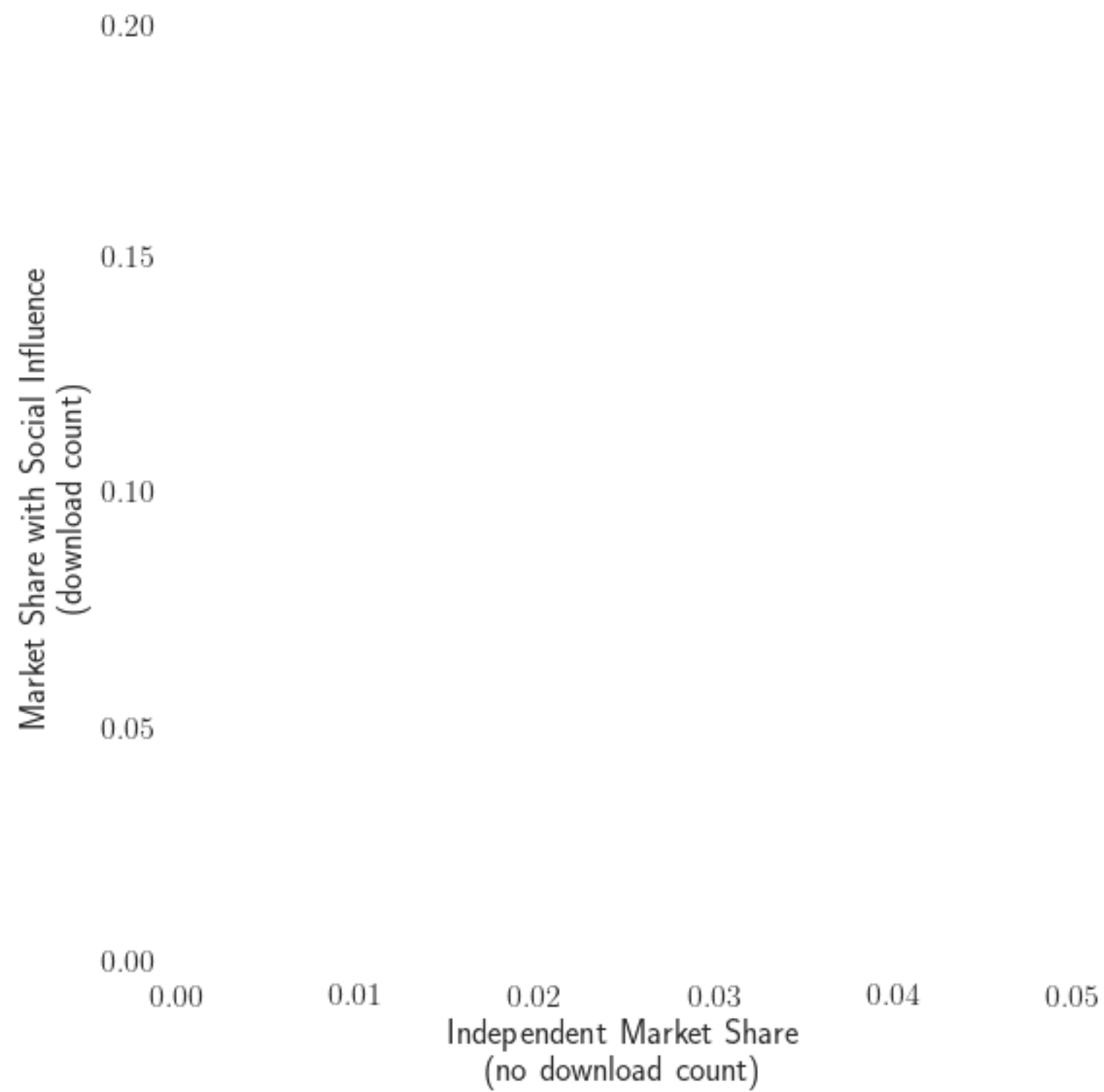
Music Lab – Song Selection - Mozilla Firefox

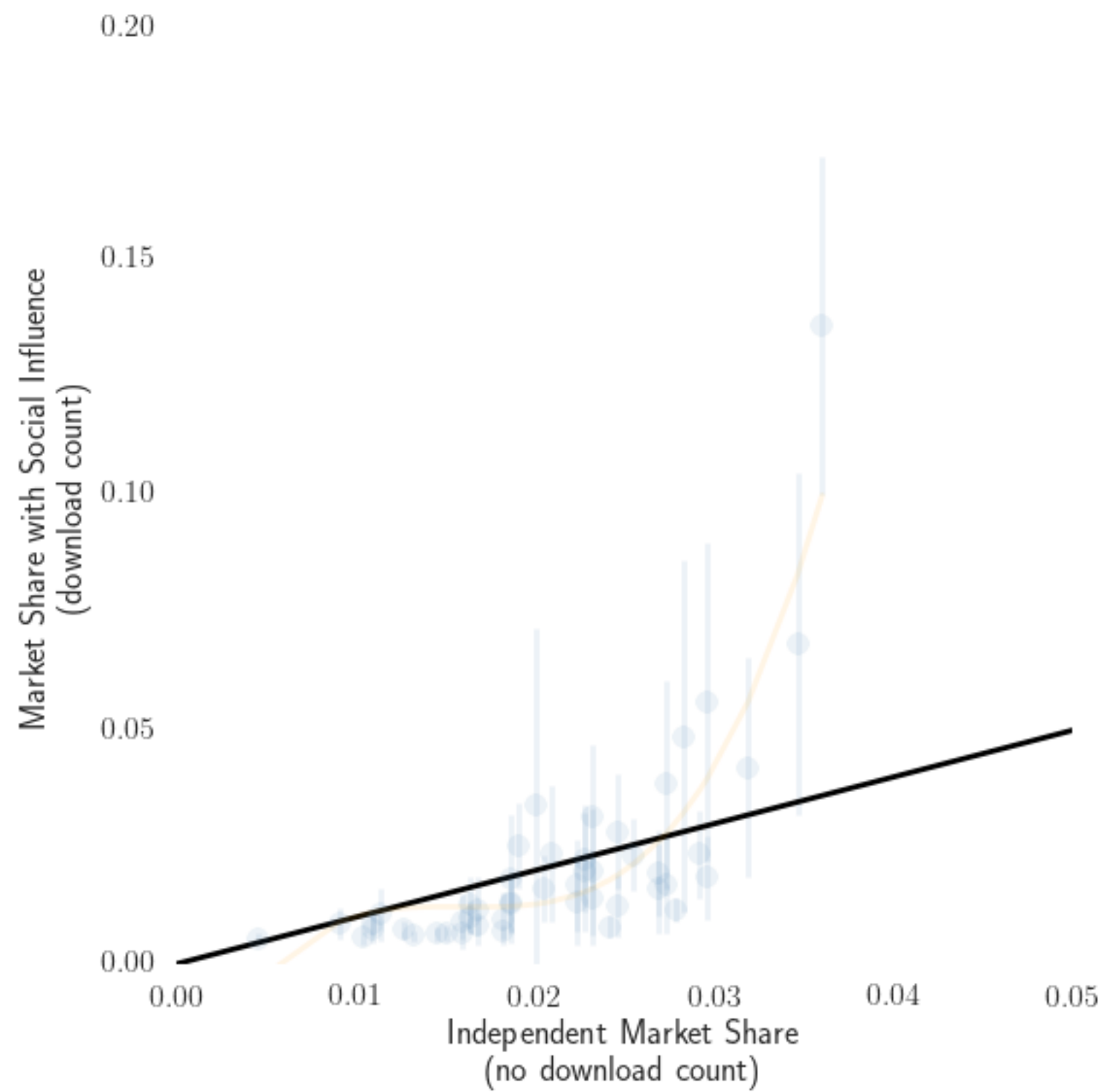
File Edit View Go Bookmarks Tools Help



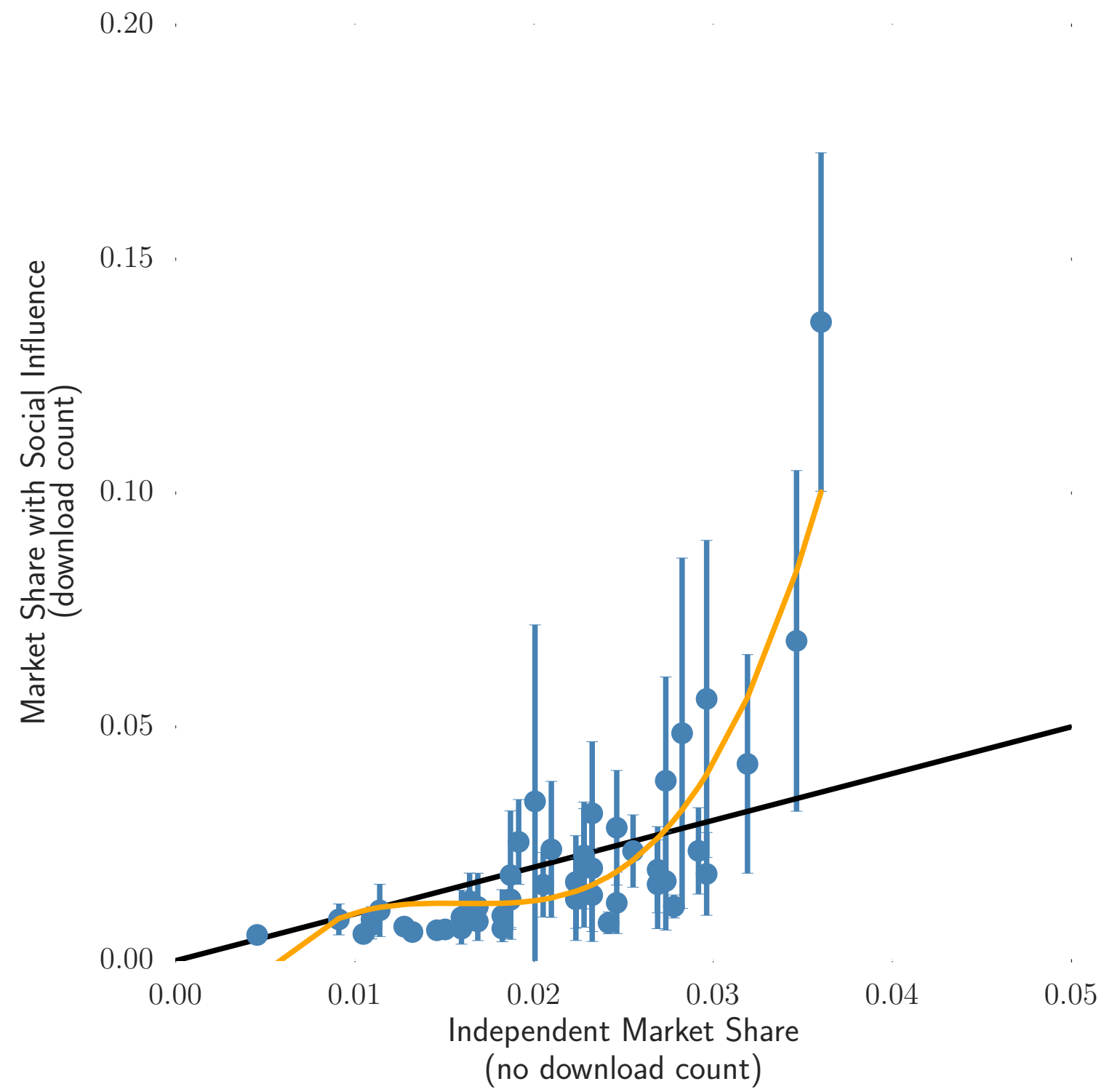
http://www.musiclab.columbia.edu/me/songs

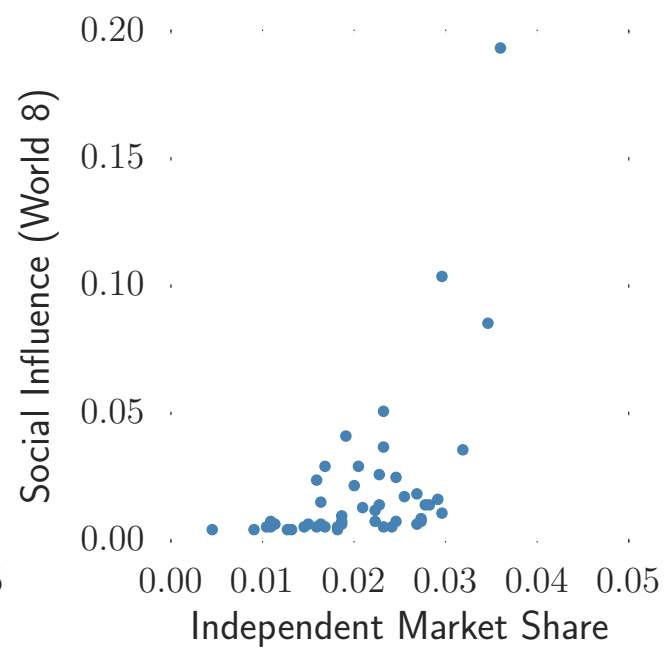
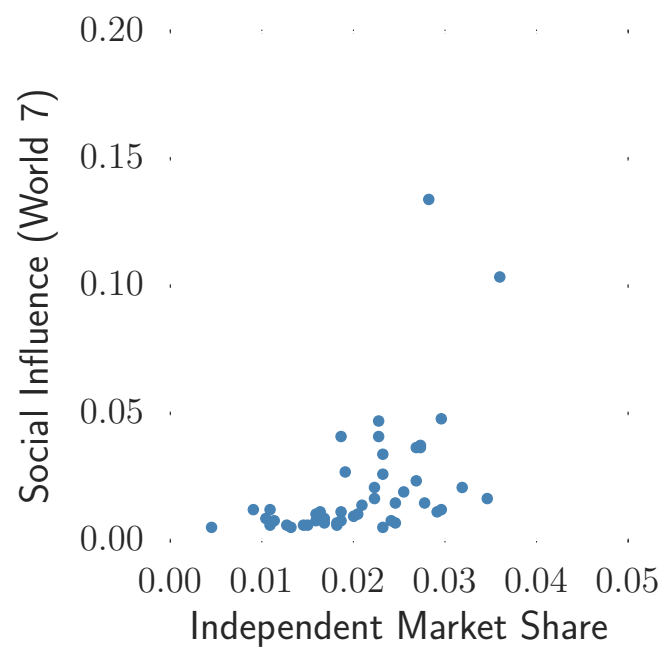
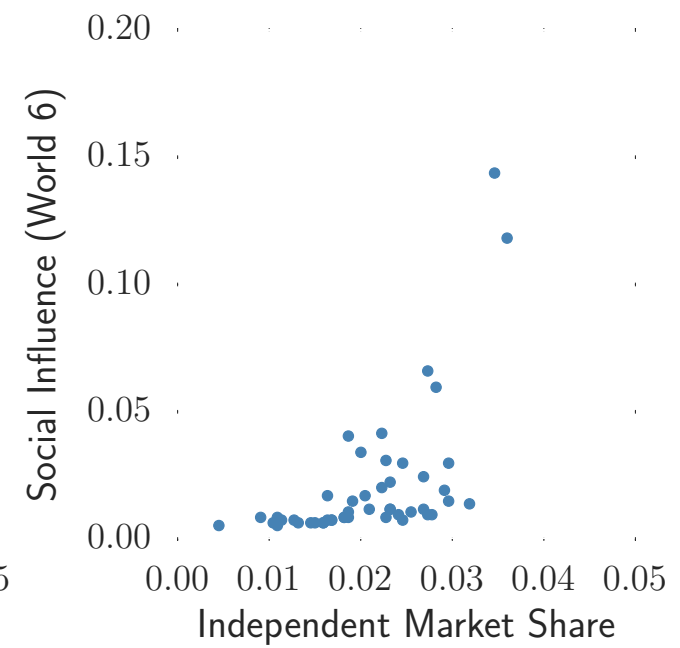
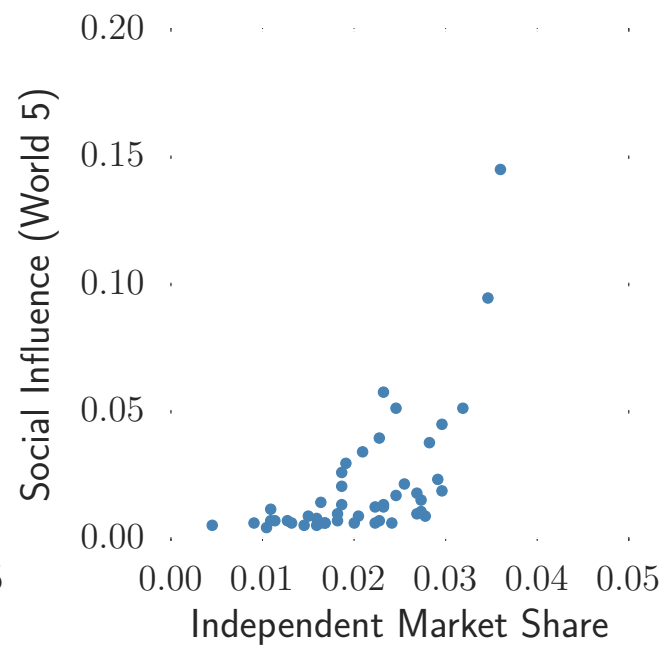
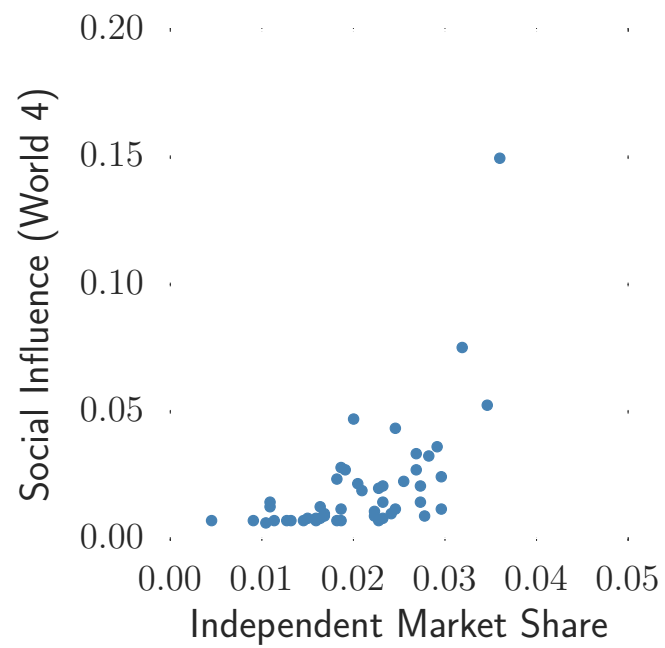
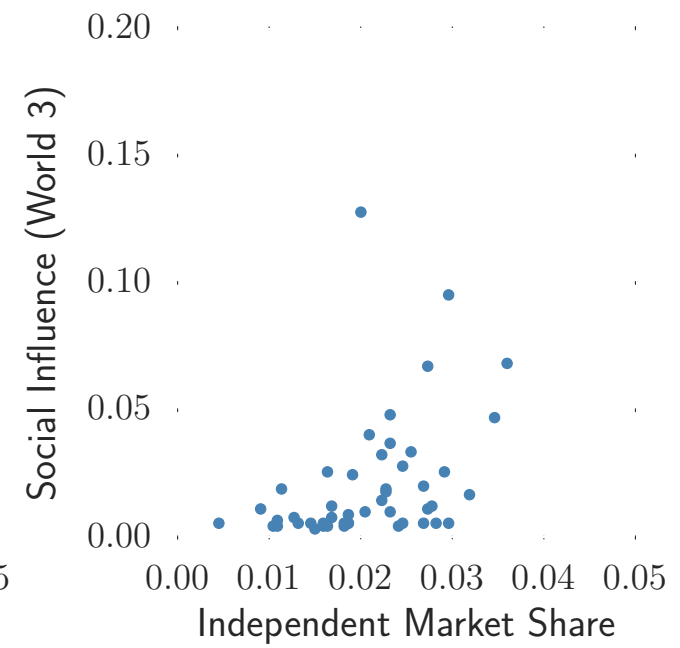
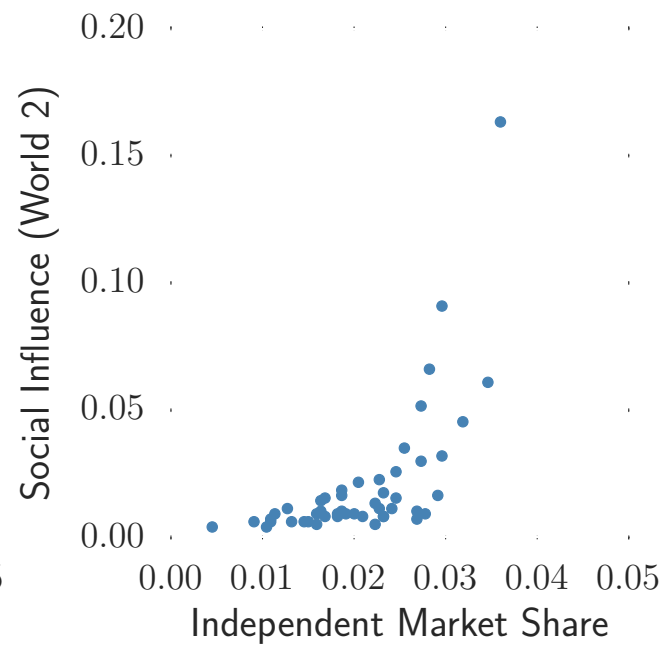
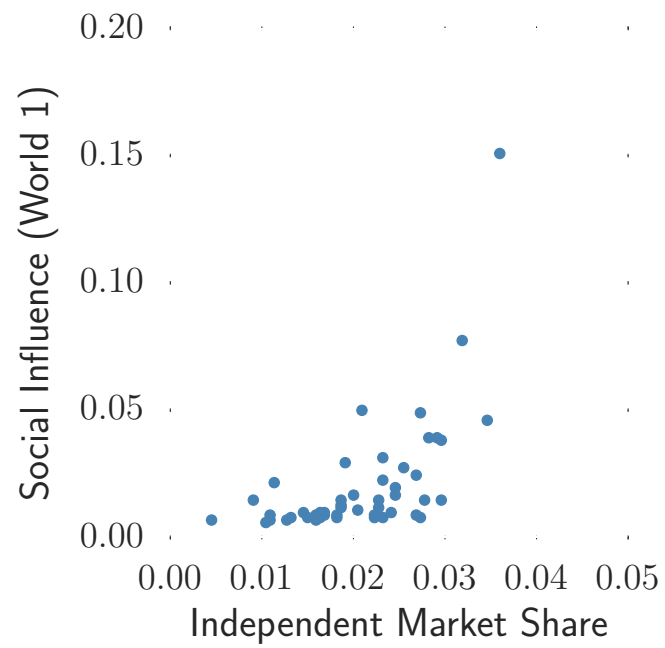
	# of down loads	[Help] [Log off]	# of down loads	# of down loads	
HARTSFIELD: "enough is enough"	20	GO MORECAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink aggression"	27	SILVERFOX: "grow"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "fluency"	14	STRANGER: "one drop"	10
THIS NEW DAWN: "the belief above the answer"	12	FORTHFACING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
MORAL HAZARD: "waste of my life"	8	S2METRO: "lockdown"	17	DANTE: "lives mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "with me luck"	10
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "if death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and misdeeds"	17	DRAWN IN THE SKY: "top the side"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshhevik boogie"	15	SELSIUS: "stars of the city"	22
SALUTE THE DAWN: "i am emel"	13	CAPE RENEWAL: "baseball worklock v1"	12	SIBIRIAN: "eye patch"	14
RYAN ESSMAKER: "detour, be still"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16

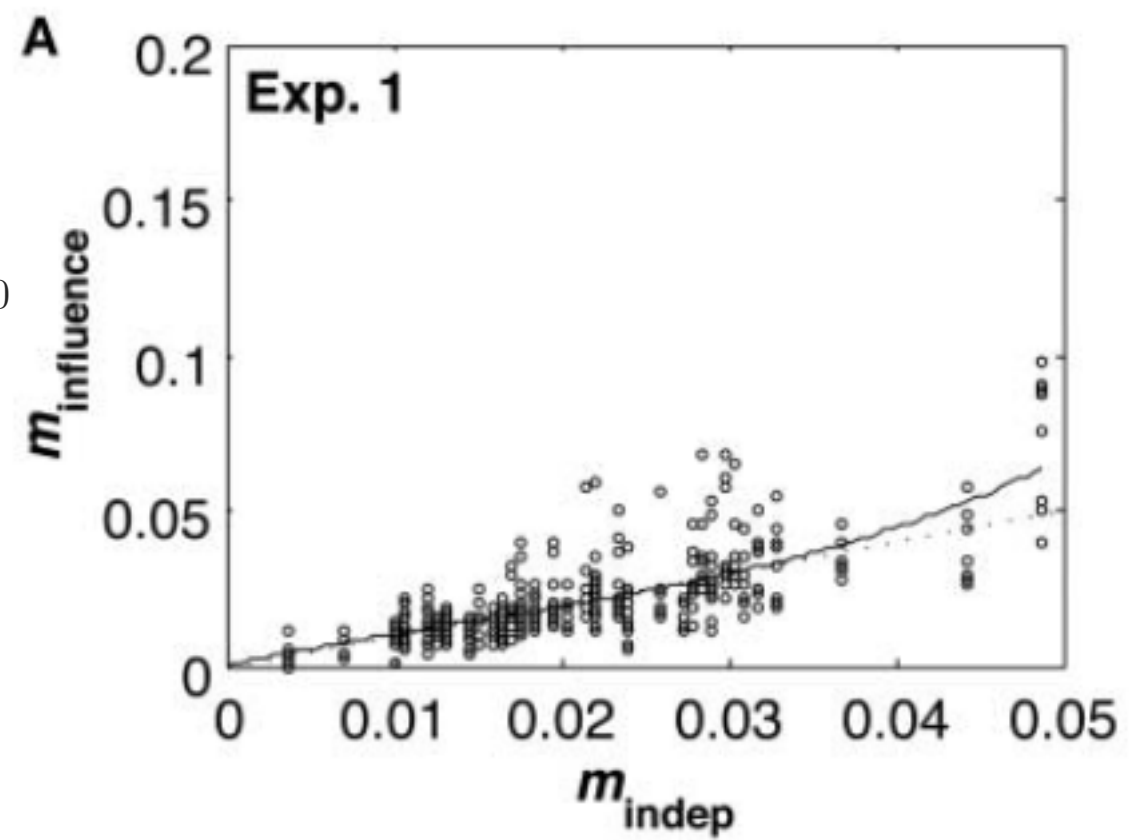
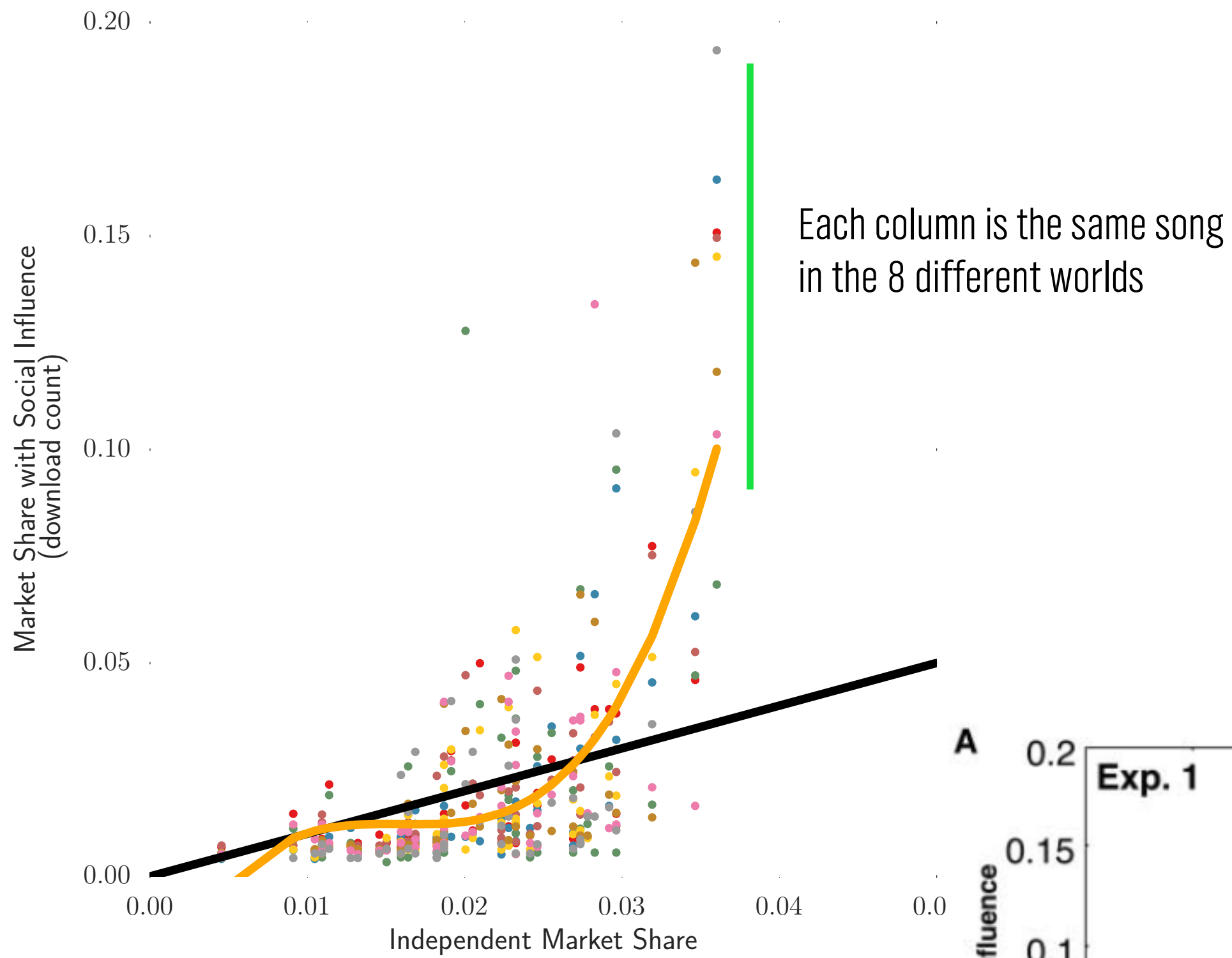












# Key advances are shown

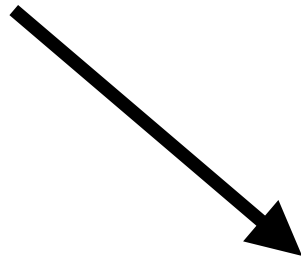
Large- $N$  population (32,000 subjects) with rapid collection time-line

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

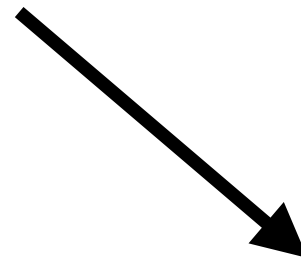
Possibility of less WEIRDs in the subject pool

# 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

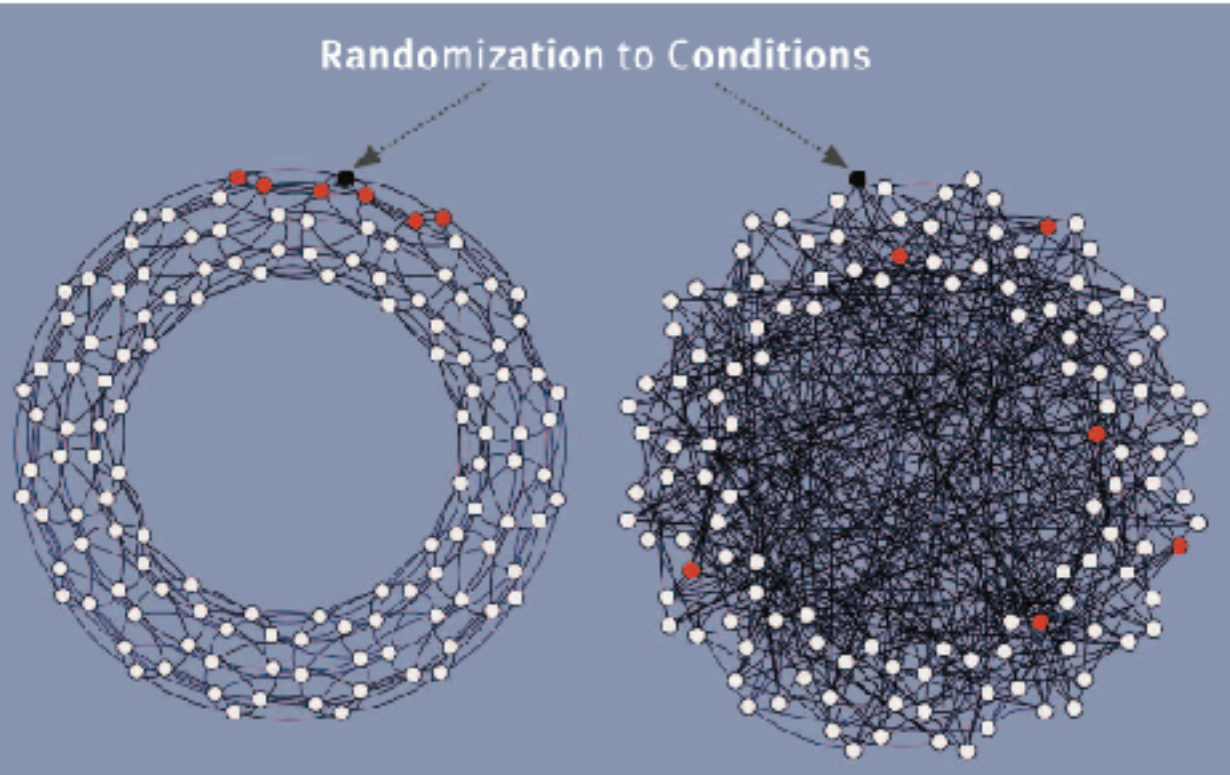
Sensor Development

Programming  
(Python/R)

NoSQL

Network Analysis

# Experimentation

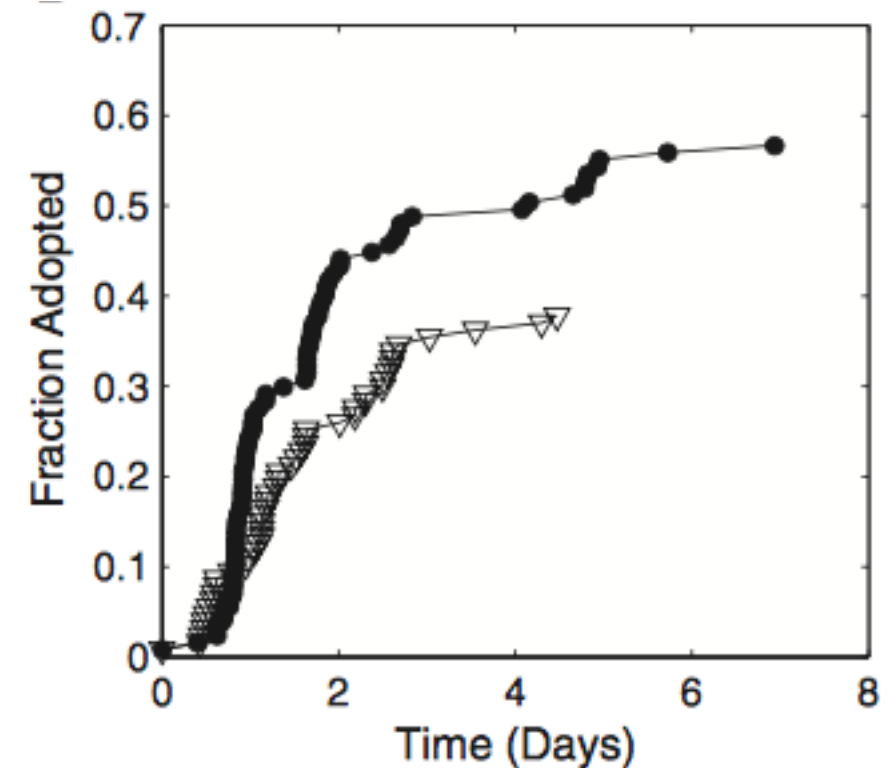
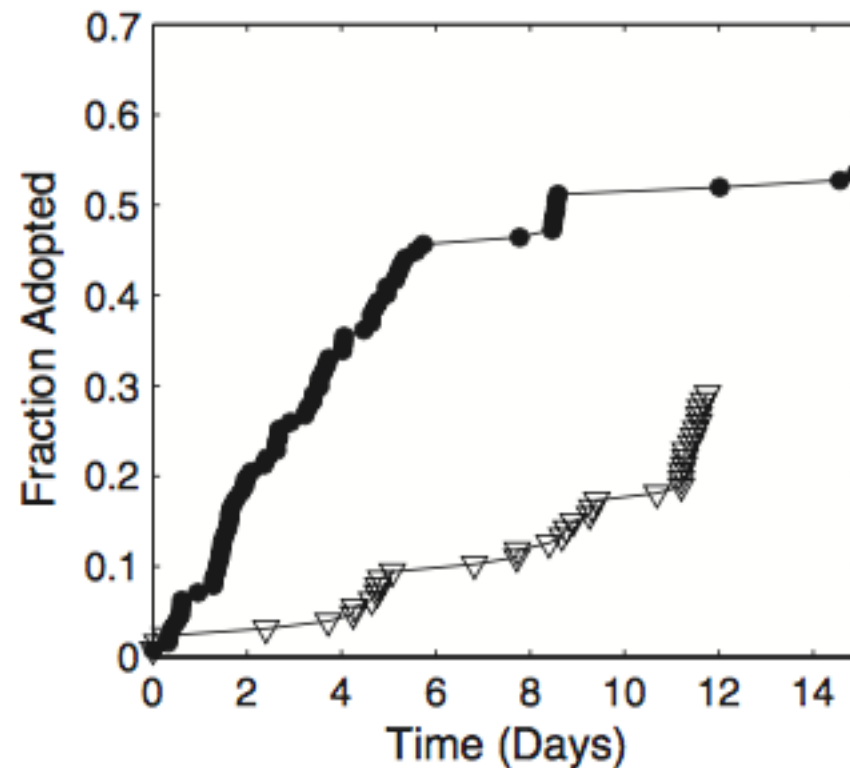


## The Spread of Behavior in an Online Social Network Experiment

Damon Centola, *et al.*

*Science* **329**, 1194 (2010);

DOI: 10.1126/science.1185231



# Experimentation

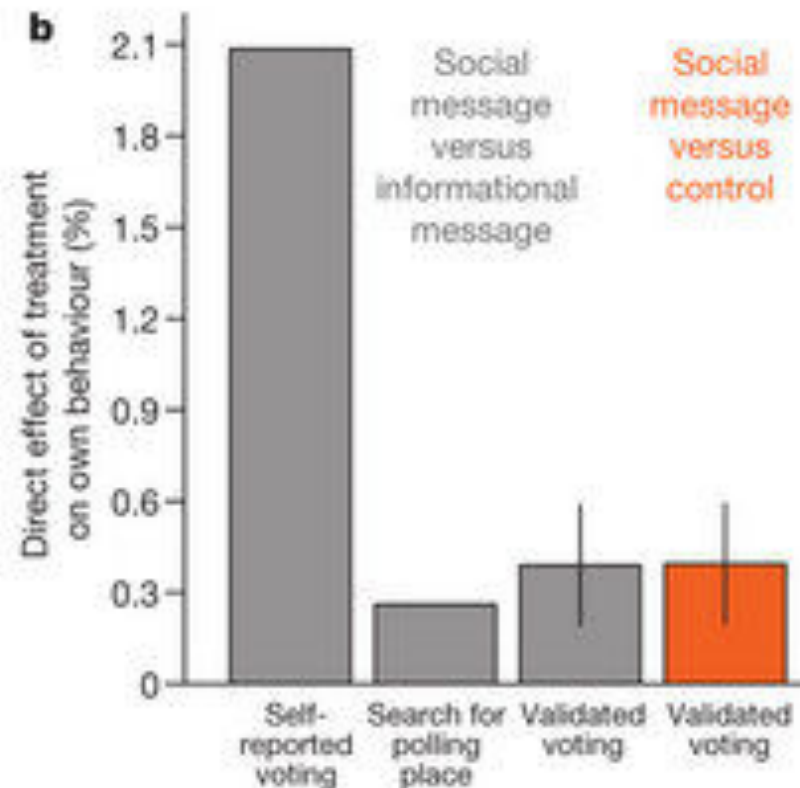
NATURE | LETTER



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

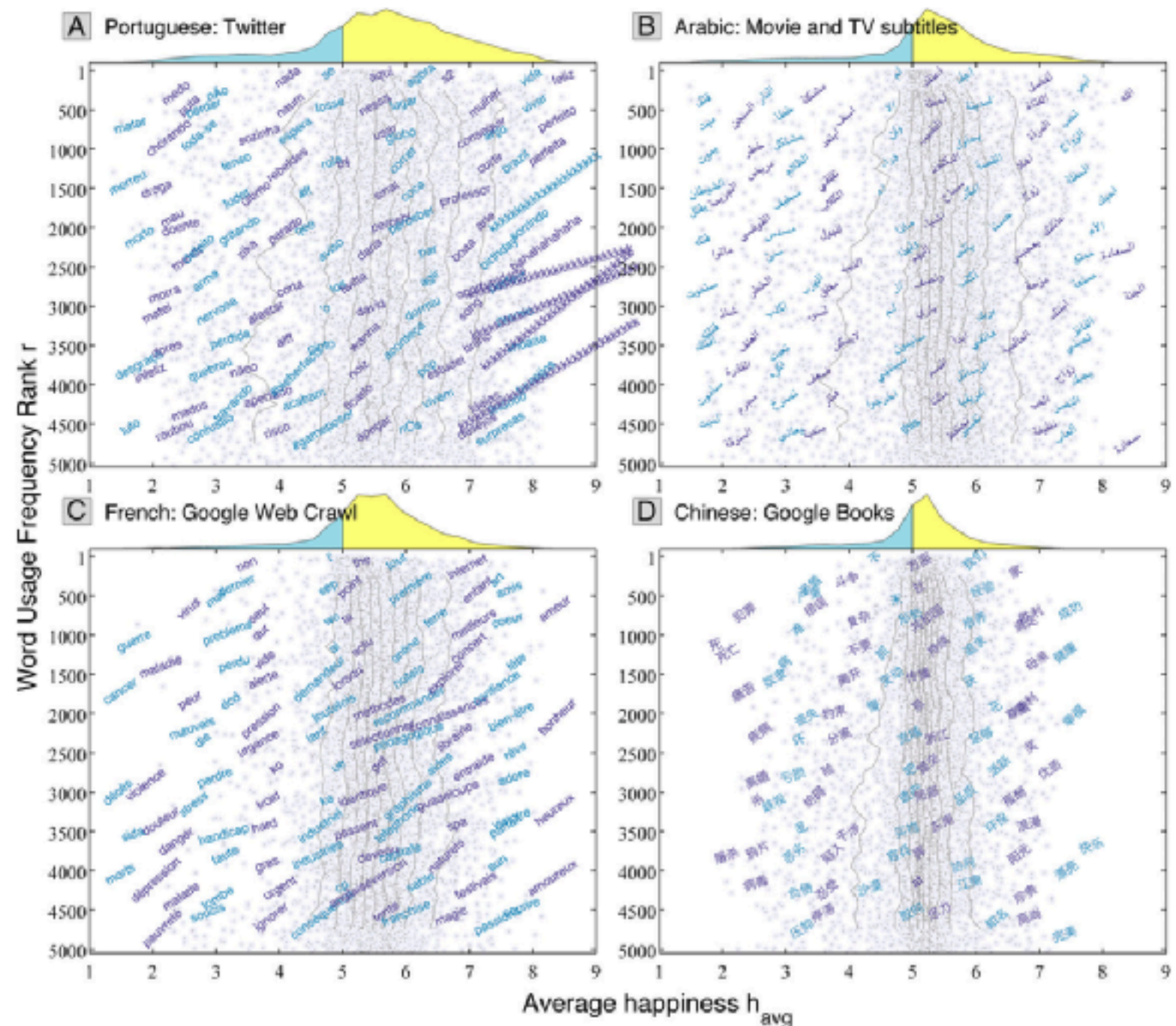
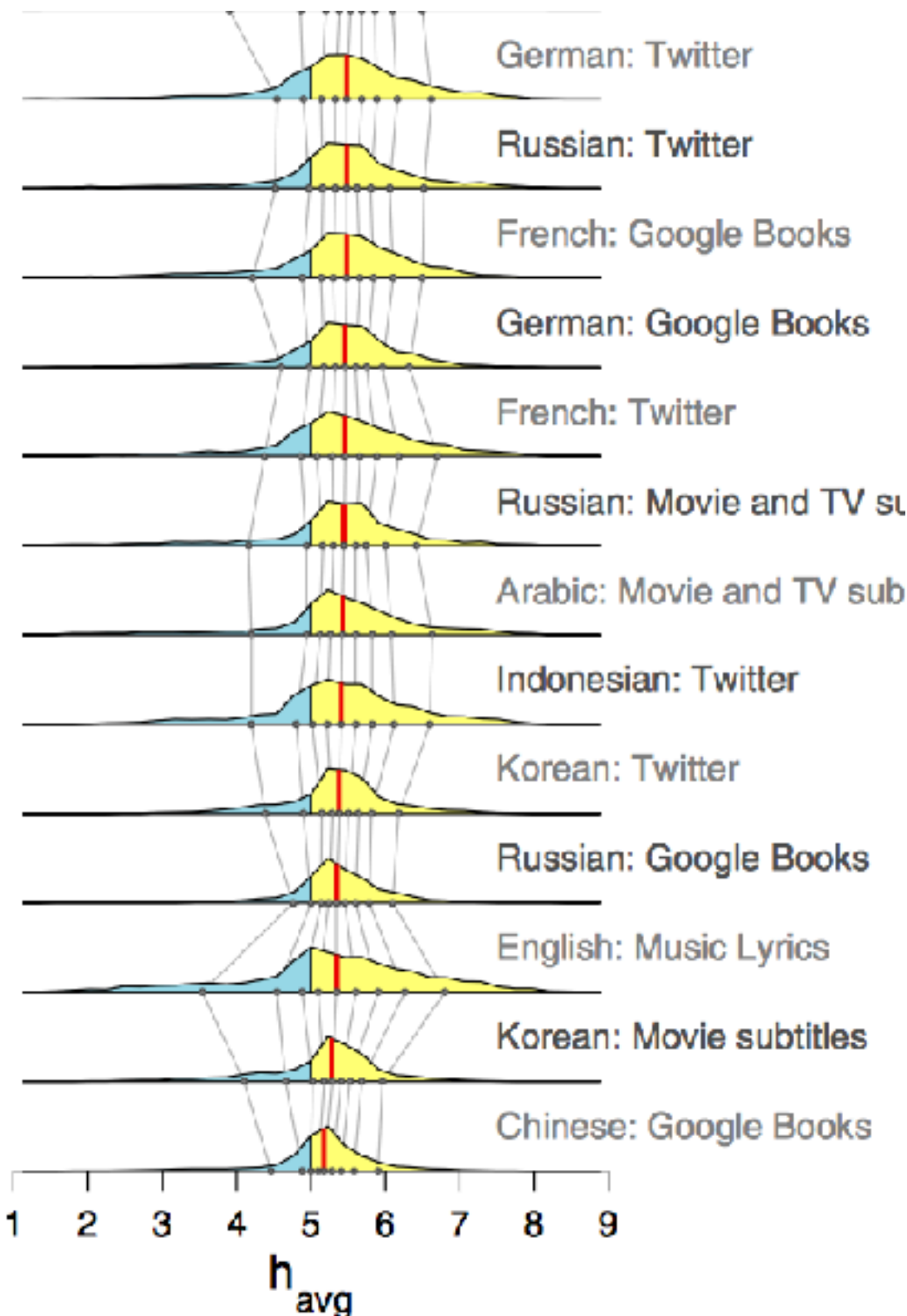




# Text

## Human language reveals a universal positivity bias

Peter Sheridan Dodds<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 Megerdooomian<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>



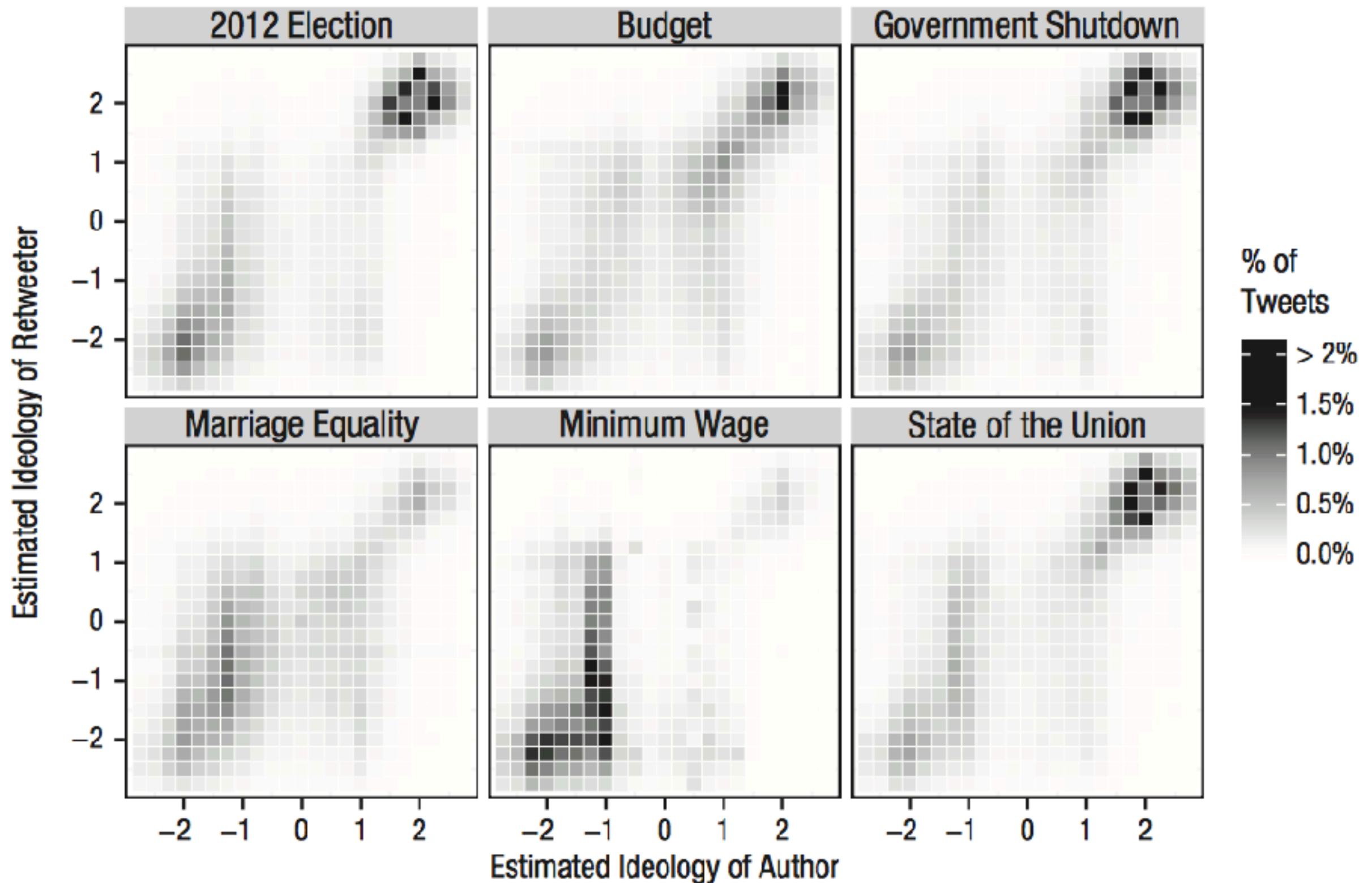
# Text

## 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



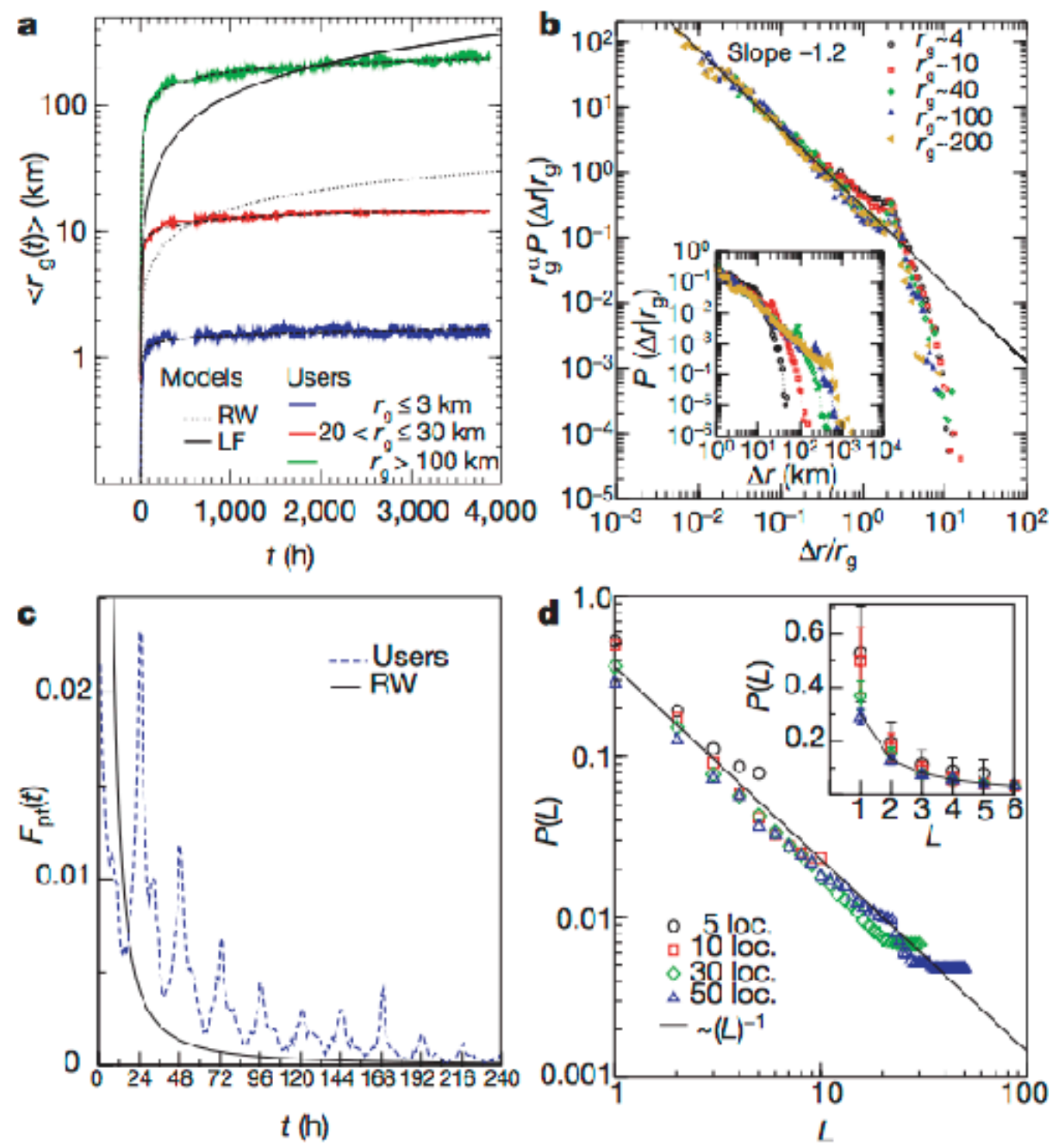


# Digital Trace

## 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>

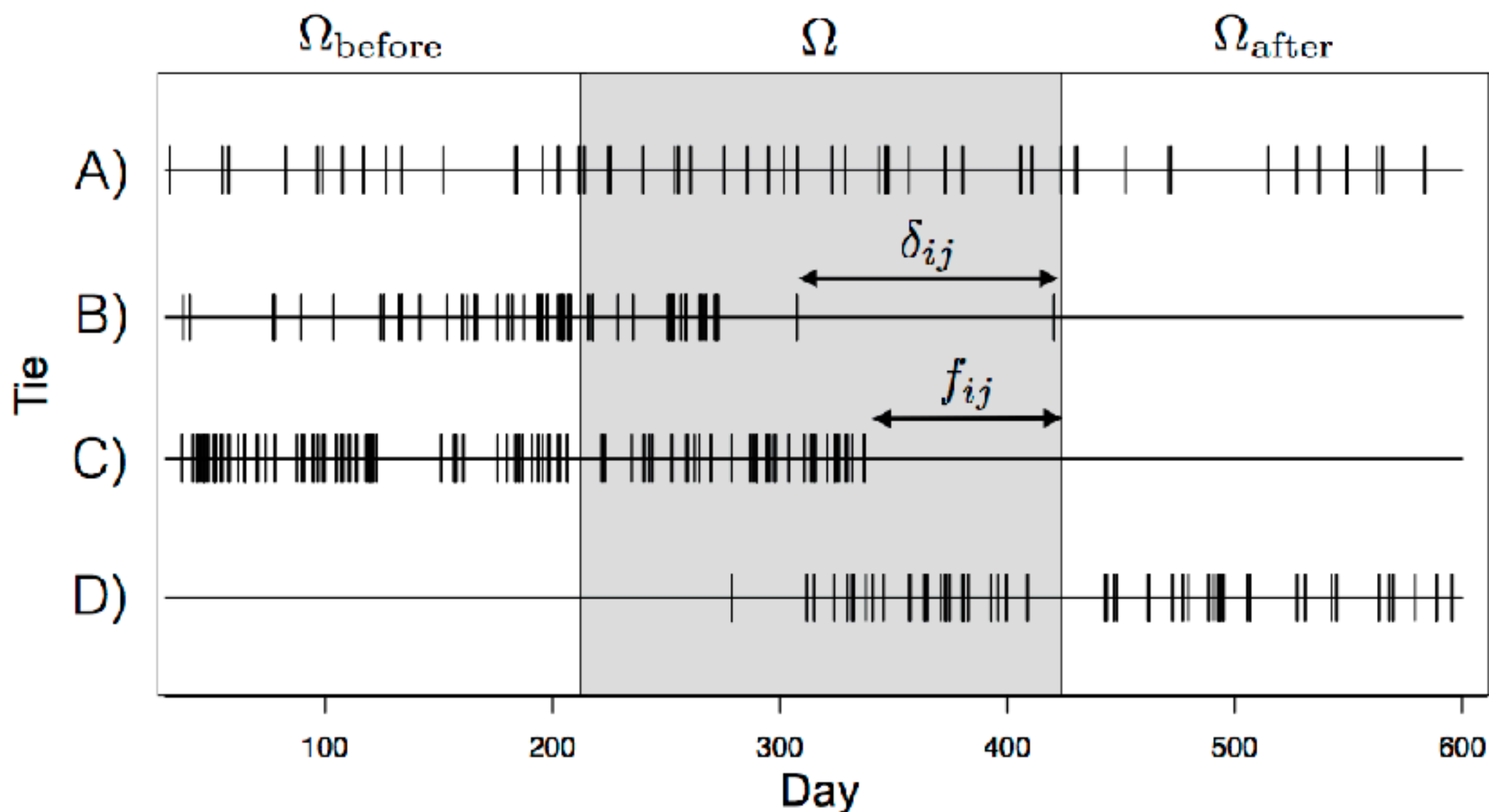
Vol 453 | 5 June 2008 | doi:10.1038/nature06958



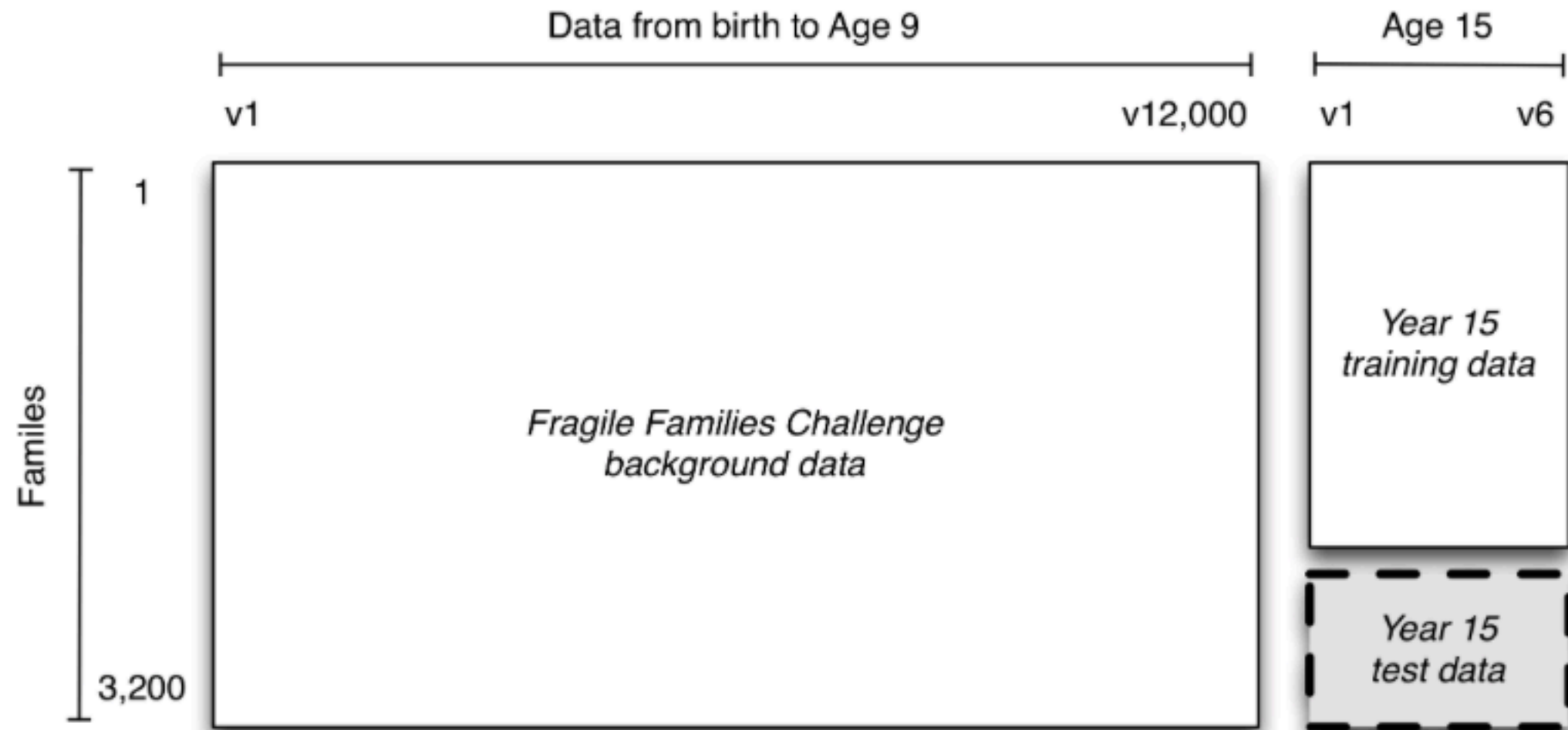
# Digital Trace

## Temporal patterns behind the strength of persistent ties

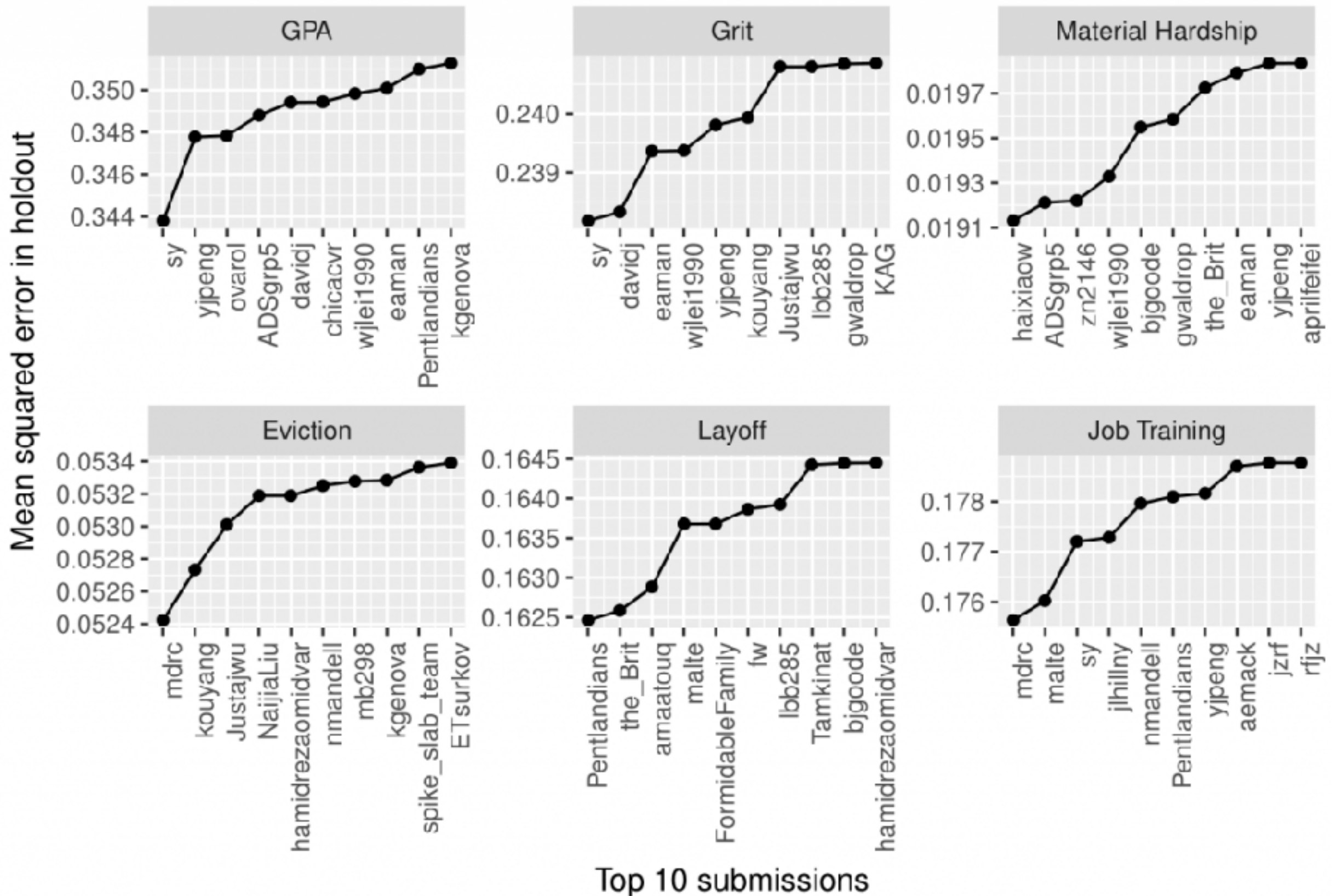
Henry Navarro<sup>1</sup>, Giovanna Miritello<sup>1,2</sup>, Arturo Canales<sup>3</sup>, Esteban Moro <sup>1\*</sup>



# Administrative data



# Administrative data

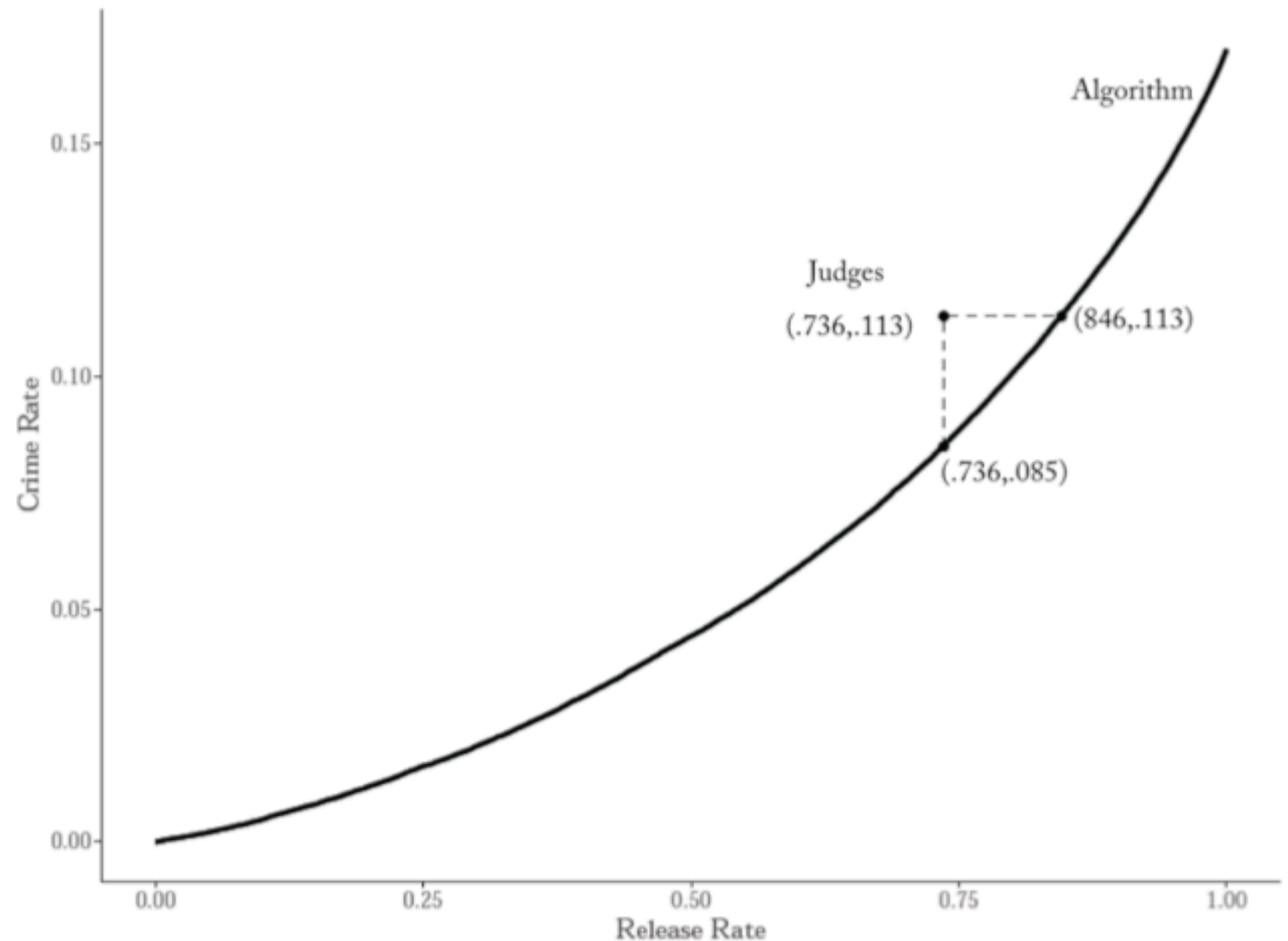


# Administrative data

## 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



# **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 7-11**

Text

## **Weeks 4-6**

Digital Data Collection  
& Processing

## **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.