# Using the Web to Do Social Science

**Duncan Watts**

***Yahoo! Research***

# What makes Social Science "Social"?

- Social phenomena arise when individuals interact to produce collective entities that have their own attributes, rules etc.
  - Families, firms, markets, associations, societies, cultures
- Yet historically social scientists have tended to study either individual or aggregate behavior, but not both at the same time
  - "microeconomics" vs. "macroeconomics"
  - psychology vs. sociology
- Difficulty is that "micro" → "macro" transition depends on **networks** of information / influence
  - Result is "emergent" behavior:
    - Systemic risk in financial systems, changes in cultural norms, collapse of political regimes

# The Web and Social Science

- Two big reasons for lack of progress
  - Networks hard to measure, especially at scale, over time
  - Can’t do “macro” experiments (i.e. on large groups)
- Very hard to do science when you can’t measure anything or test theories with experiments
- Recent technological advances may lift these historical constraints
  - Email, IM, Social Networking Sites, Online communities, Phone, SMS, Twitter, etc. are generating mountains of **observational network and behavioral** data
  - Also possible to do **human subjects experiments** on a previously unimaginable scale
- These changes are bigger than the WWW proper, but I’ll use the term “Web” loosely to mean the whole gamut.

# Outline For Rest of Talk

- Will describe four projects dating back to 2001, right up until the present
- All four motivated by a general interest in social networks and relevance to other questions of social science
  - Each captures a different aspect of problem
  - Each has also taught us something about using the web to do social science
- Looking forward, can speculate about what might be possible

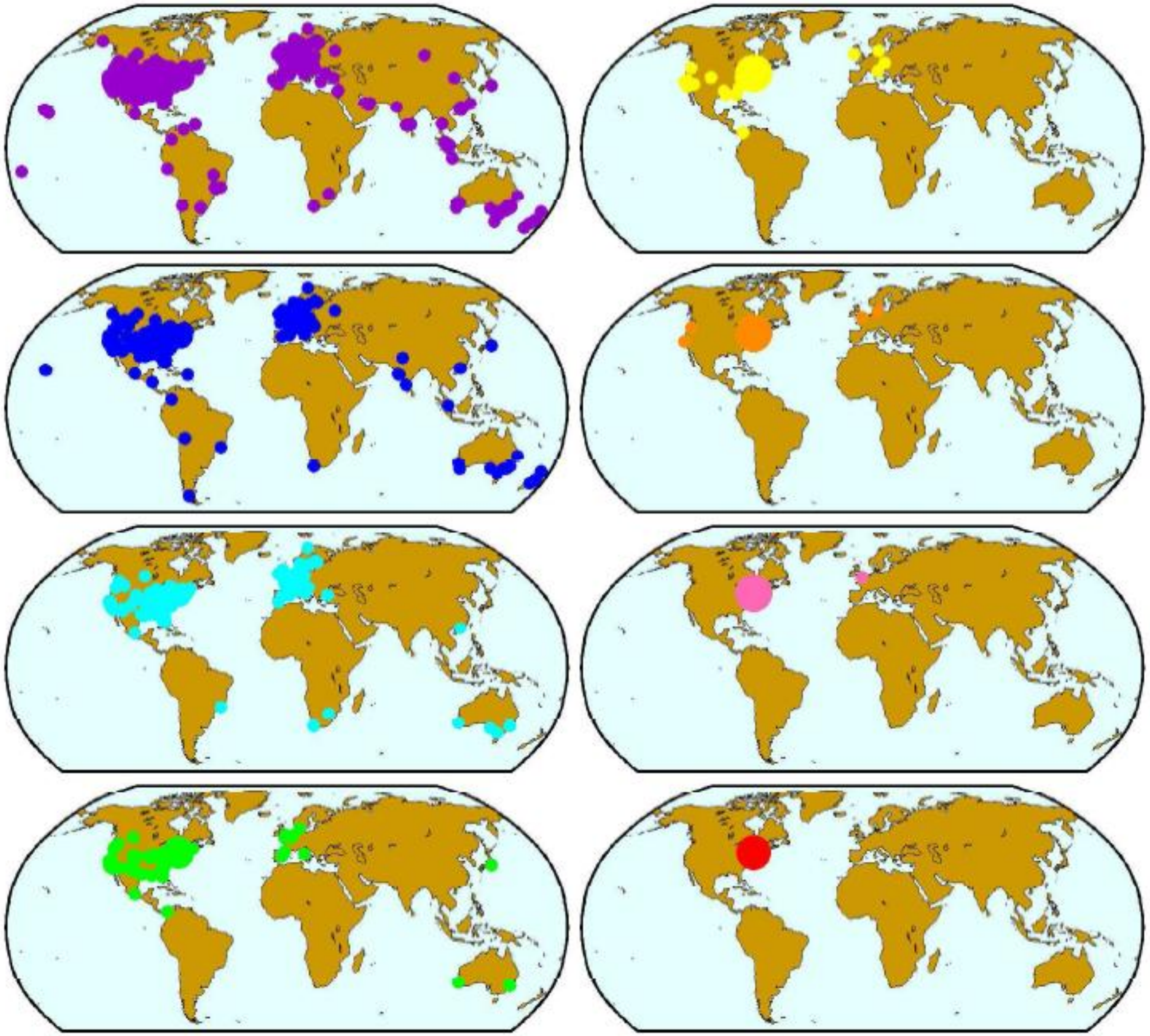
# 1. Small World Experiment

- 1960's: Stanley Milgram and Jeffrey Travers designed first "small world" experiment
  - A single "target" in Boston
  - 300 initial "senders" in Boston and Omaha
  - Each sender asked to forward a packet to a friend who was "closer" to the target
  - The friends got the same instructions
- Protocol generated 300 "letter chains" of which 64 reached the target.
- Found that typical chain length was 6
- Led to the famous "six degrees" phrase

# The Small World Web

- 2001-2002: decided to recreate Travers and Milgram's experiment
  - But using email/web server instead of physical packets
- Whereas Milgram had
  - one target (Boston)
  - 300 chains (Boston and Omaha)
  - 64 reached target
- We used
  - 18 Targets around world
    - A university professor in upstate New York
    - A policeman in Perth, Australia
    - An librarian in Paris
    - A veterinarian in Norway, etc…
  - 24,163 chains passing through 61,168 hands in 166 countries
  - Roughly 400 reached targets

# Chain Progression For One Target



# The “Bored At Work” Network

- Results mostly confirmed Milgram’s findings
  - 6 degrees is surprisingly accurate (median is 7)
  - Dodds et al (Science, 2003); Goel et al (WWW, 2009)
- But we also learned something else:
  - **We managed to run an experiment with over 60,000 participants, on a global scale, at virtually zero cost**
  - BAWN: Millions of people ready to do social science
- What to do next?
  - Small-world experiment not really a “lab” experiment
  - Could we create a “virtual lab” on a web scale?

# 2. Social Influence and Cultural Markets

- In markets for books, music, etc., suspect
  - Individuals influence each other’s choices
  - Social influence leads to inequality and unpredictability
- Wanted to conduct lab experiment in which
  - Can run same process many times under identical conditions, exploring multiple “histories”
  - Can carefully control who is exposed to what
- Problem was that each “history” is a whole distribution of popularity over some set of objects
- Even in a modest experiment, just one history would require at least hundreds of people
  - Entire experiment would require many thousands
  - Far too many to fit in a physical lab
  - What about the web?

# The Virtual Lab on The Web

# Music Lab: 48 Unknown Bands

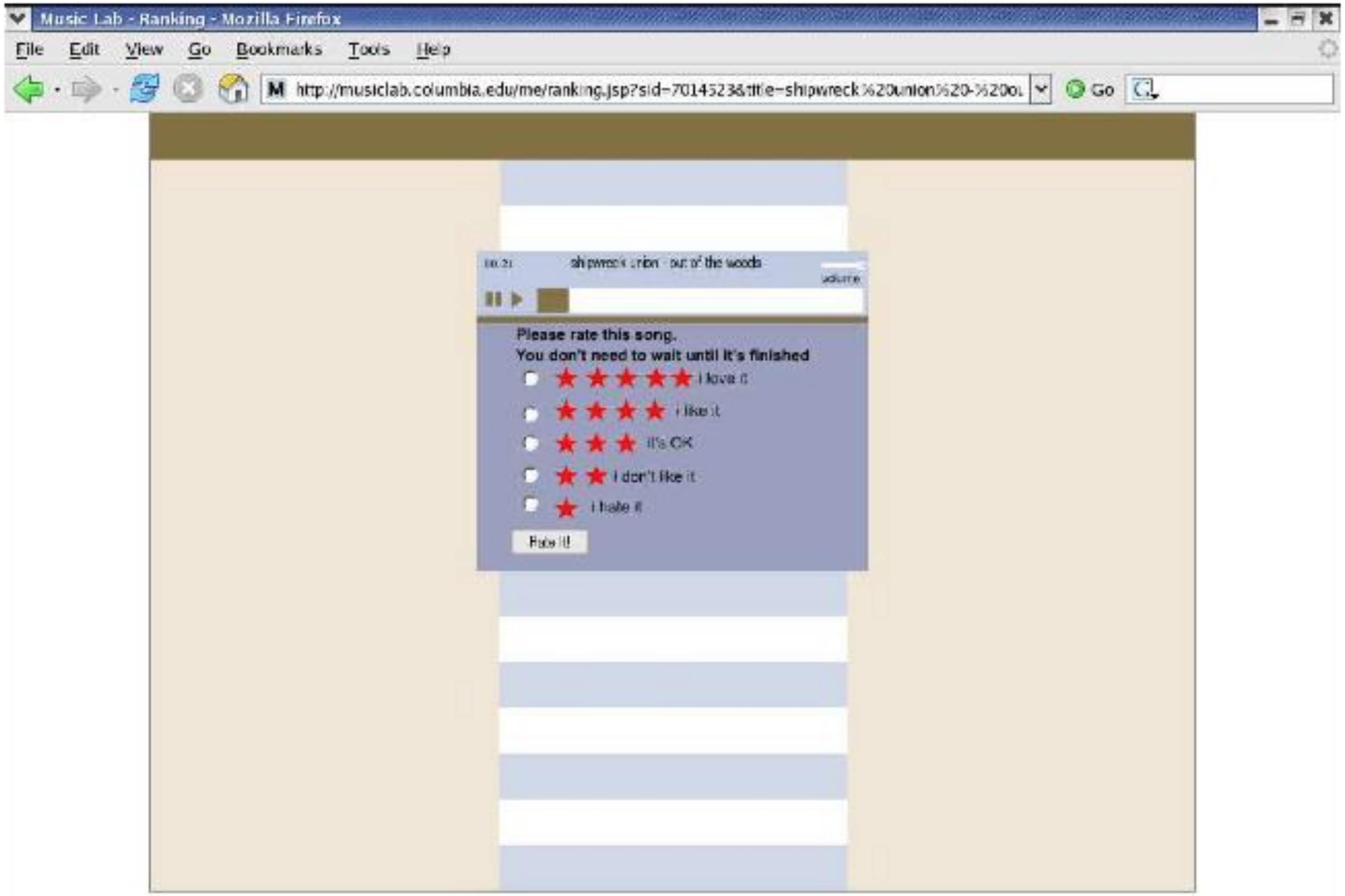
Music Lab -- Song Selection - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

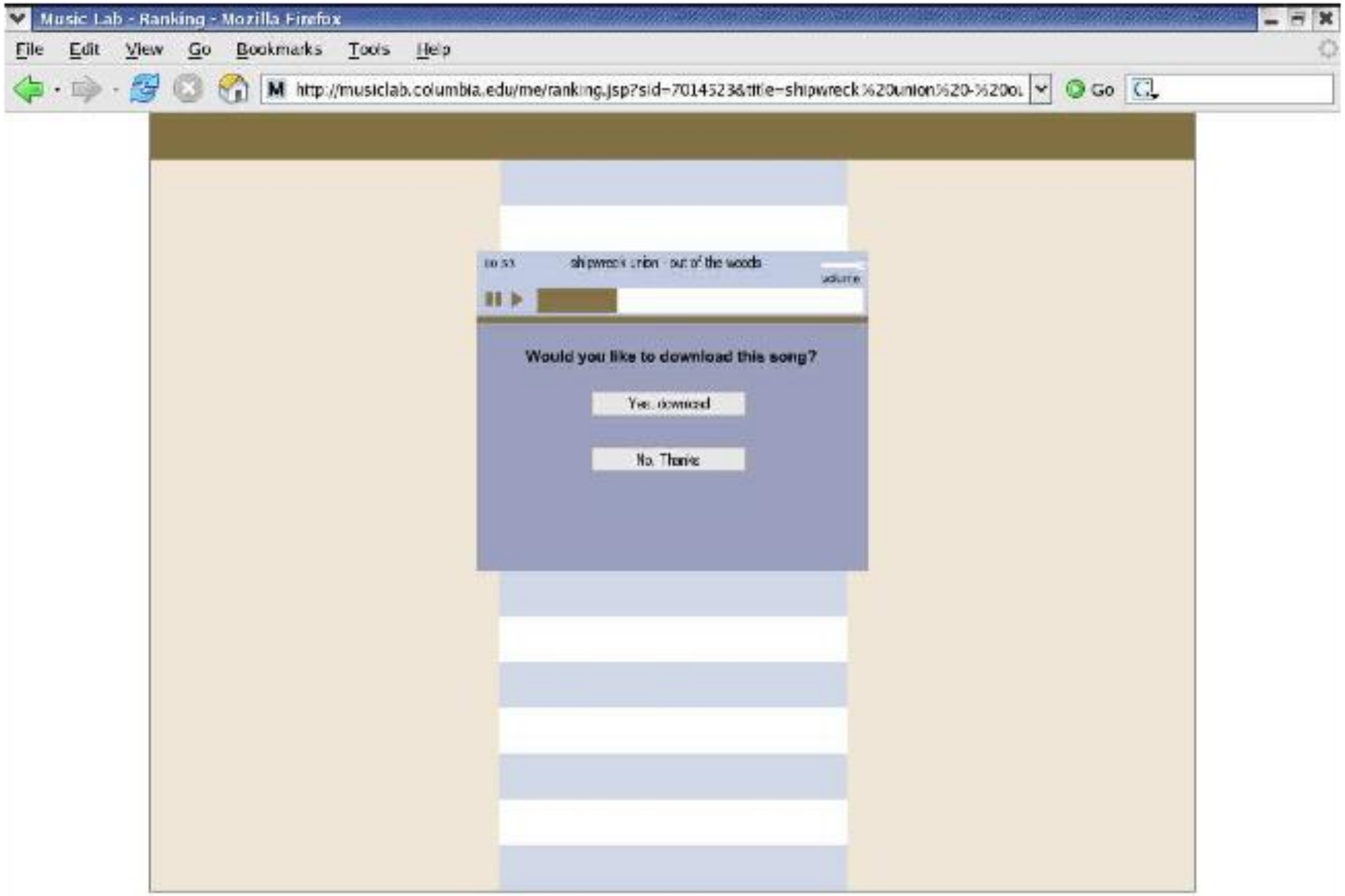
http://www.musiclab.columbia.edu/me/songs

| | # of downloads | [Help] [Log off] | # of downloads | | # of downloads |
|---|---|---|---|---|---|
| HARTSFIELD: "enough is enough" | 20 | GO MOREDCAI: "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 agression" | 27 | SILVERFOX: "gnaw" | 17 |
| THE BROKEN PROMISE: "the end in friend" | 19 | POST BREAK TRAGEDY: "florence" | 14 | STRANGER: "one drop" | 10 |
| THIS NEW DAWN: "the belief above the answer" | 12 | FORTHFADING: "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 | 52METRO: "lockdown" | 17 | DANTE: "lifes mystery" | 14 |
| NOT FOR SCHOLARS: "as seasons change" | 27 | SIMPLY WAITING: "went with the count" | 16 | FADING THROUGH: "wish 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: "til 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 miracles" | 17 | DRAWN IN THE SKY: "tap the ride" | 12 |
| EMBER SKY: "this upcoming winter" | 25 | SUM RANA: "the bolshevik boogie" | 15 | SELSIUS: "stars of the city" | 22 |
| SALUTE THE DAWN: "i am error" | 13 | CAPE RENEWAL: "baseball warlock v1" | 12 | SIBRIAN: "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 |

# Subjects can listen to songs (via streaming) and rate them…

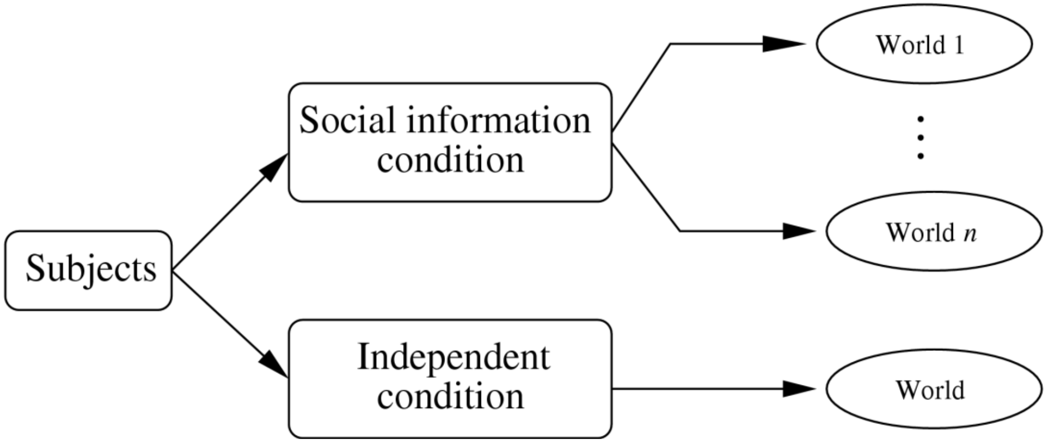


# At which point they can also elect to download song

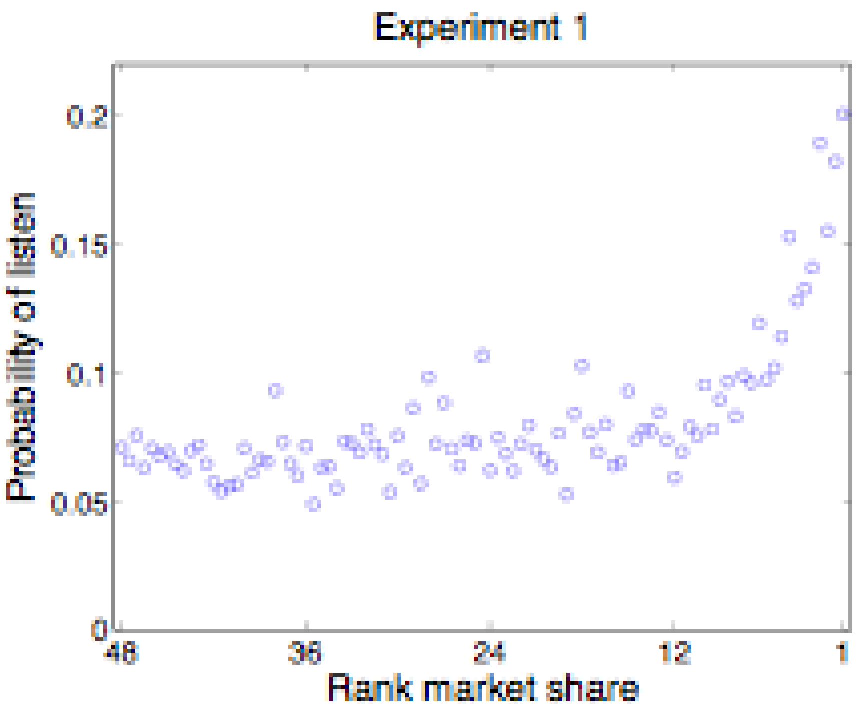


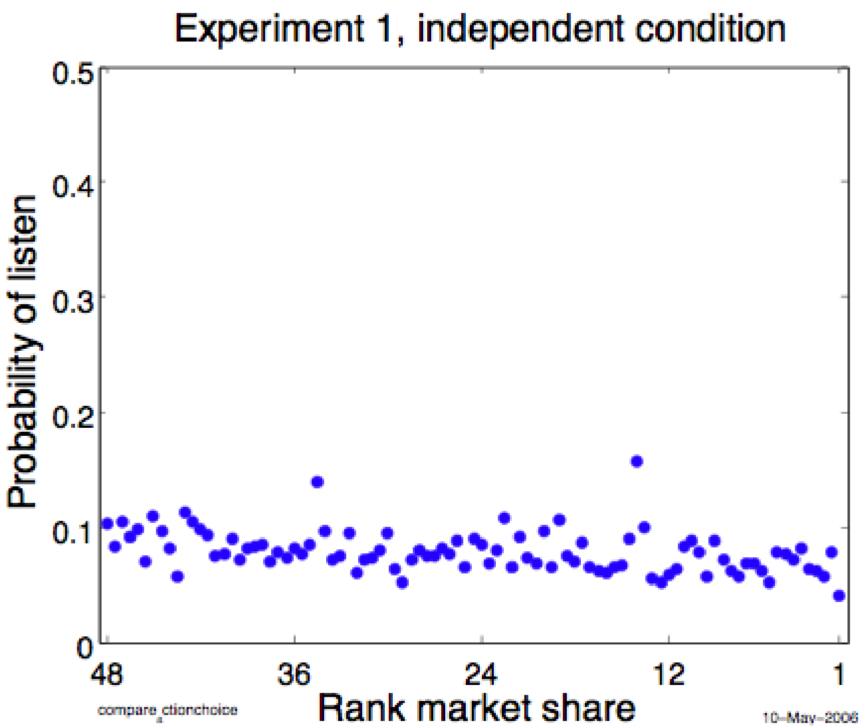
# Experimental Design

- As subjects arrive, they are allocated randomly into one of two conditions
  - Independent
    - Just see the names of bands, and songs
  - Social Influence
    - In addition see number of previous downloads
    - Weak signal, can choose to ignore it
  - In addition, social influence condition is broken into eight (8) "worlds"
    - Arriving subjects in each world can see downloads of previous participants in that world only



# Social Influence at Micro Level:




$$m_i = d_i \Big/ \sum_{k=1}^{S} d_k$$

$m_i$ = market share for song i

$d_i$ = downloads of song i

# Properties at Macro Level

1. Inequality of Success
   - Measure with Gini Coefficient (G)
     - Expected difference in market share $m_i$ between two randomly chosen songs, normalized to the range [0,1]
       - G = 0 (all objects have equal market share)
       - G = 1 (one object has entire market)
2. Unpredictability of Success
   - Quantity (U) defined as average difference in market share of songs across *R* realizations of the world
   - Also normalized between [0,1]
   - Result is measure of “inherent unpredictability”
     - i.e. if you know market share in one world, how much could you predict about another?

# Four experiments (N = 27,267)

| Influence | Population 1 | Population 2 |
|---|---|---|
| Weak | Experiment 1 (*N* = 7,149) | |
| Strong | Experiment 2 (*N* = 7,192) | Experiment 3 (*N* = 2,930) |
| Deception | | Experiment 4 (*N* = 9,996) |

# Music Lab Results

- Individuals are influenced by their observations of the choices of others
  - The stronger the social signal, the more they are influenced
- Collective decisions are also influenced
  - The popular songs are more popular (and unpopular songs are less popular)
  - However, which particular songs become the popular ones becomes harder to predict
- The paradox of social influence is that
  - Individuals have more information on which to base choices
  - But the collective choice (i.e. what becomes popular) reveals less and less about individual preferences
  - Salganik Dodds, and Watts (*Science* 2006)
- Manipulating social influence not so easy
  - Can create self-fulfilling prophecies at level of individual songs, but not for entire market
  - Salganik and Watts (*SPQ*, 2008)

# 3. Networks and Diffusion

- Small World Experiment explored structure of global social networks and Music Lab explored social influence
  - But influence in real life <u>diffuses</u> through networks
- In recent years, attention has focused on so-called "**influencers**," who exert disproportionate influence over others
  - Oprah Winfrey and Books
  - Sarah Jessica Parker and Shoes
  - Jeff Jarvis and Dell Customer Service
- Marketers love this idea:
  - Find the influencers and they will do all the work!
- Problem is: influencers generally identified after the fact, when to be useful marketers need to identify them ex-ante
  - No evidence that they can do this consistently

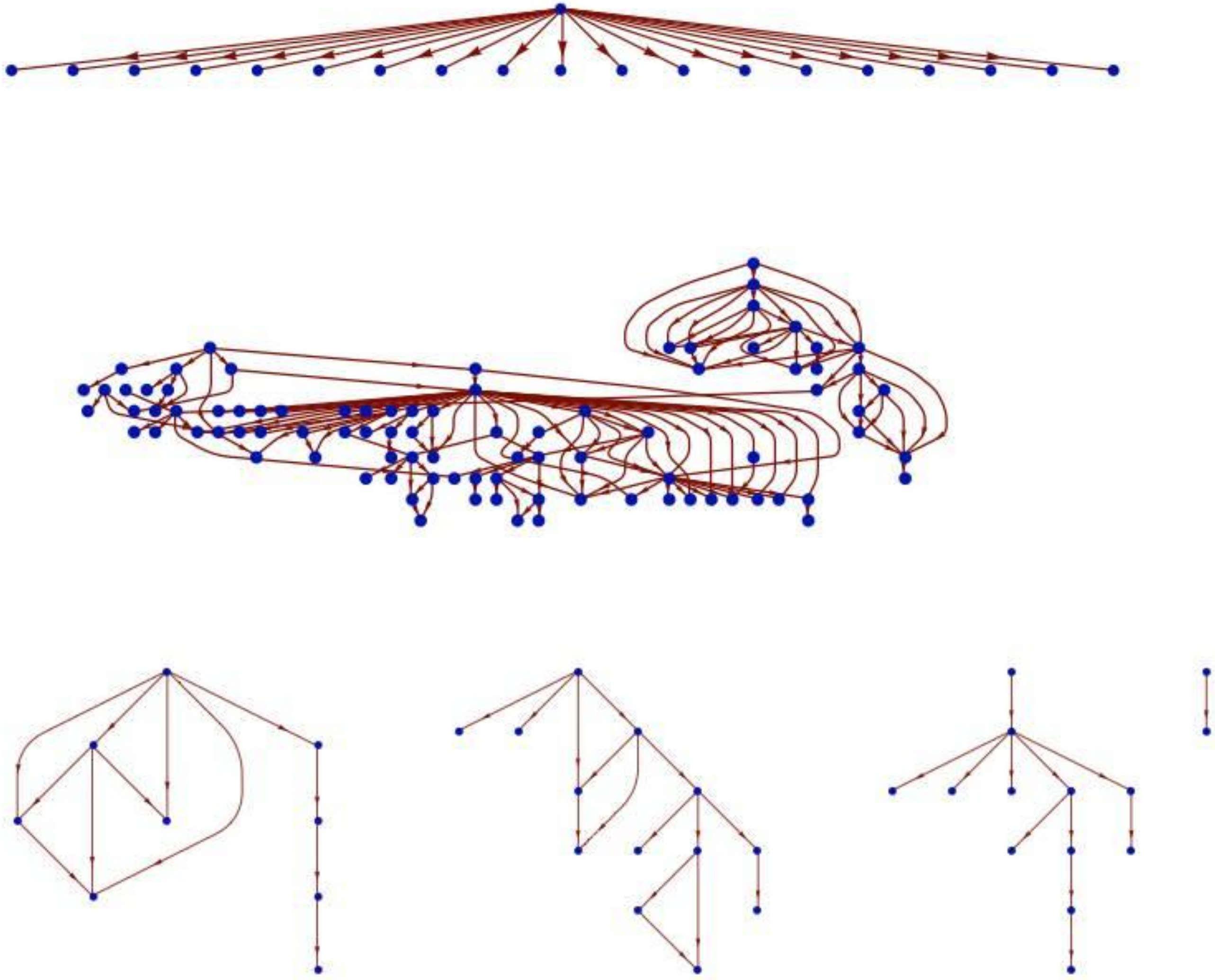
# Influencers on Twitter

- Twitter is ideally suited to study influencers
  - Fully-observable network of “who listens to whom”
  - URL shorteners enable us to track diffusion of unique pieces of information
  - Millions of diffusion “events”
- Bakshy et al (*WSDM*, 2011)
- 1B public twitter posts between 9/15/09-11/15/09
- 90M posts containing bit.ly URL’s from 1.6M users
  - We’ll focus on this subset
- Crawled 56M Twitter users, 1.7B edges

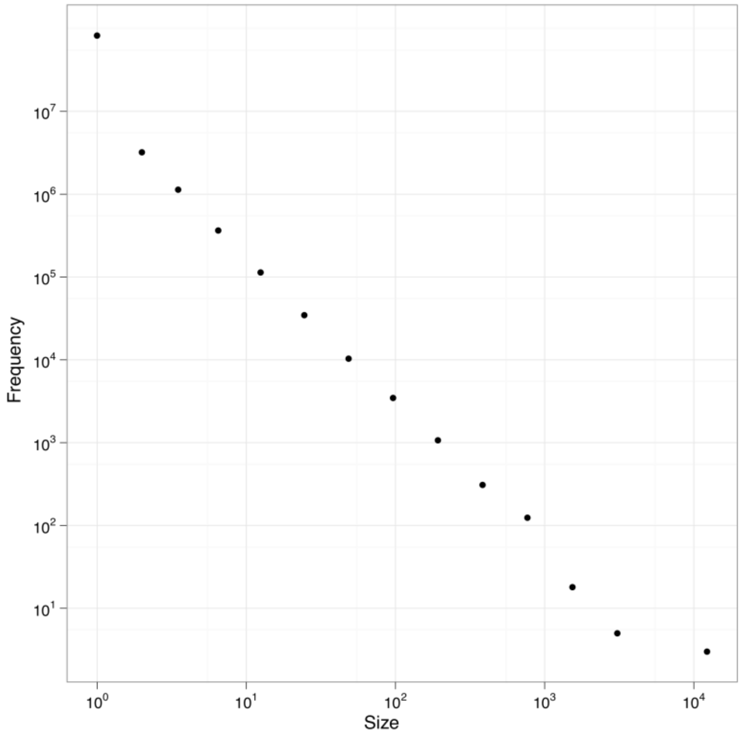
# Computing Influence on Twitter

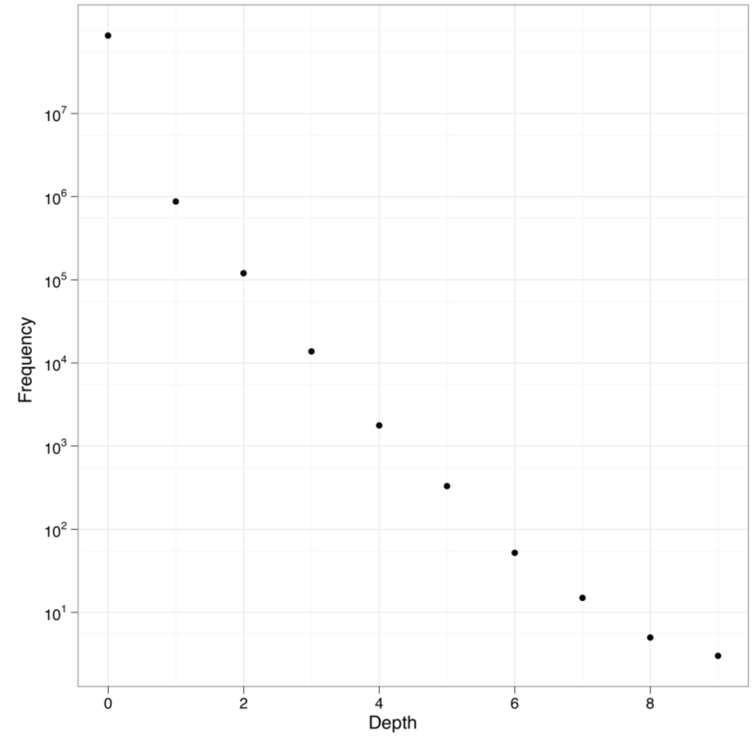
- An individual "seed" user tweets a URL (here we consider only bit.ly)
- For every follower who subsequently posts same URL (whether explicit "retweet" or not), seed accrues 1 pt
- Repeat for followers-of-followers, etc. to obtain total influence score for that "cascade"
- Average individual influence score over all cascades
  - Highly conservative measure of influence, as it requires not only seeing but acting on a tweet
  - Click-through would be good, but not available to us

# Cascades on Twitter



# Cascade Distribution Highly Skewed



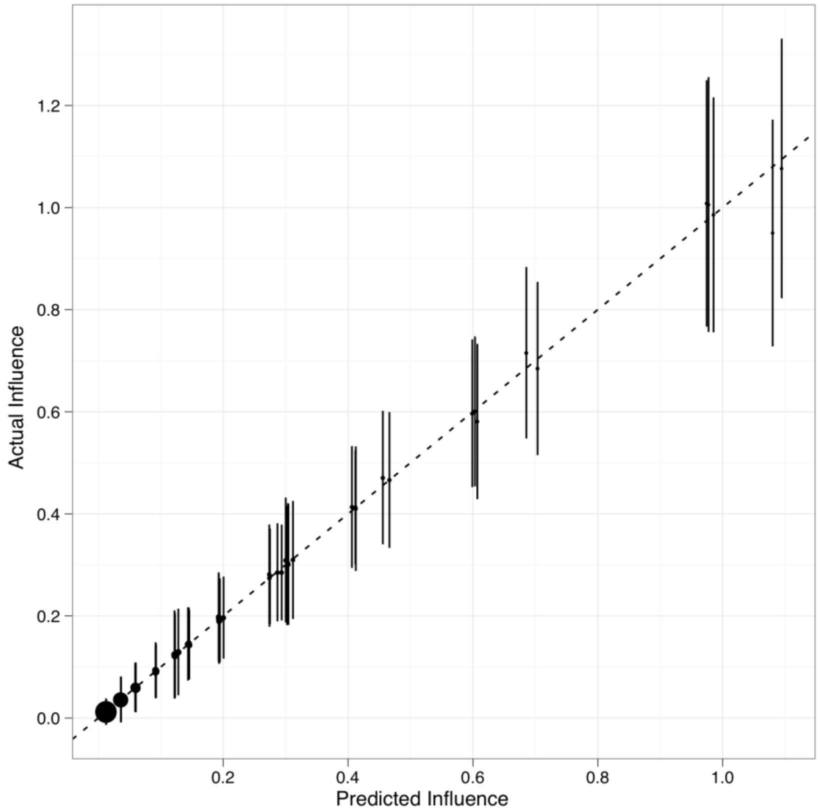

- Almost all cascades are small and shallow
    - Average size = 1.14; Median size = 1
- A tiny fraction reach thousands and propagate up to 8 hops

# Predicting Influence

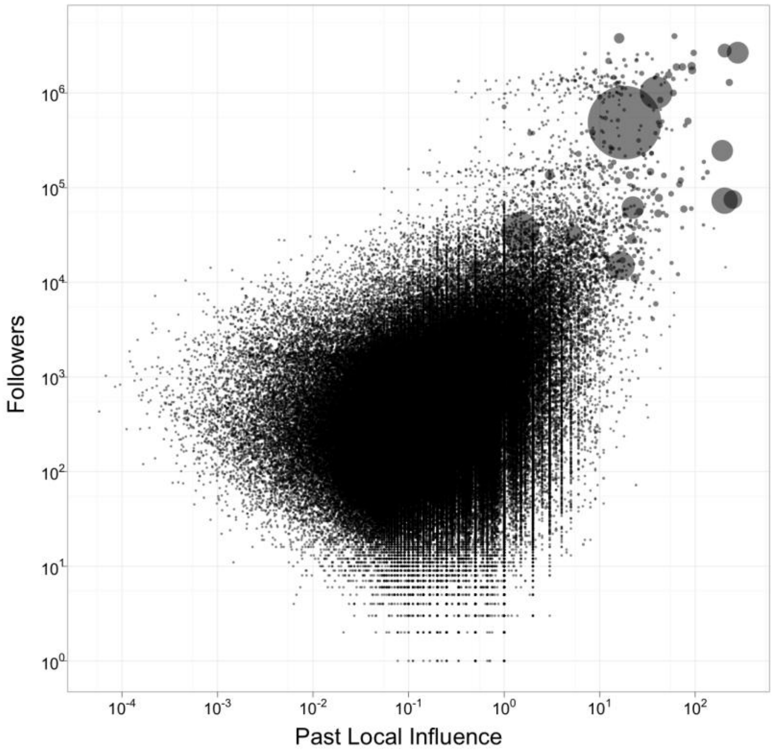
- Objective is to predict log (influence score) for future cascades as function of
  - Log # Followers, log # Friends, log # Reciprocated Ties
  - log # Tweets, Time of joining
  - Log (past influence score)
- Fit data using regression tree
  - Recursively partitions feature space
  - Piecewise constant function fit to mean of training data in each partition
  - Nonlinear, non-parametric
    - Better calibrated than ordinary linear regression
  - Use five-fold cross-validation
    - For each fold, estimate model on training data, then evaluate on test data
    - Every user gets included in one test set

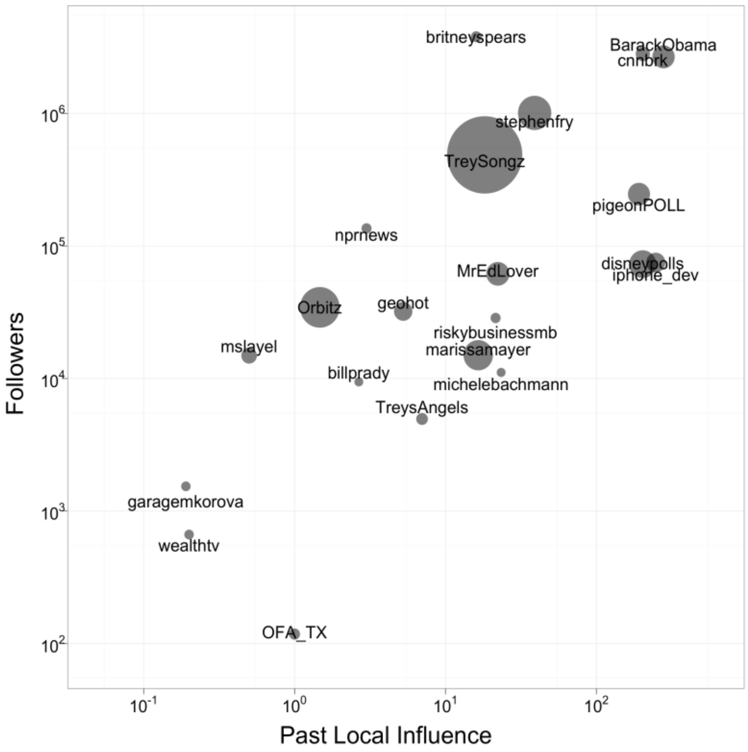
# Results

- Only two features matter
  - Past local influence
  - # Followers
- Surprisingly, neither # tweets nor # following matter
- Model is well calibrated
  - average predicted close to average actual within partitions
- But fit is poor ($R^2 = 0.34$)
  - Reflects individual scatter
- Also surprisingly, content doesn't help



# Who are the Influencers?




Circles represent individual seeds (sized by influence)

# Necessary but not sufficient

- Seeds of large cascades share certain features (e.g., high degree, past influence)
- However, many small cascades share those features, making “success” hard to predict at individual level
- Common problem for rare events
  - School shootings, Plane crashes, etc.
  - Tempting to infer causality from “events,” but causality disappears once non-events accounted for
- Lesson for marketers:
  - Individual level predictions are unreliable, even given “perfect” information
- Fortunately, can target *many* seeds, thereby harnessing average effects

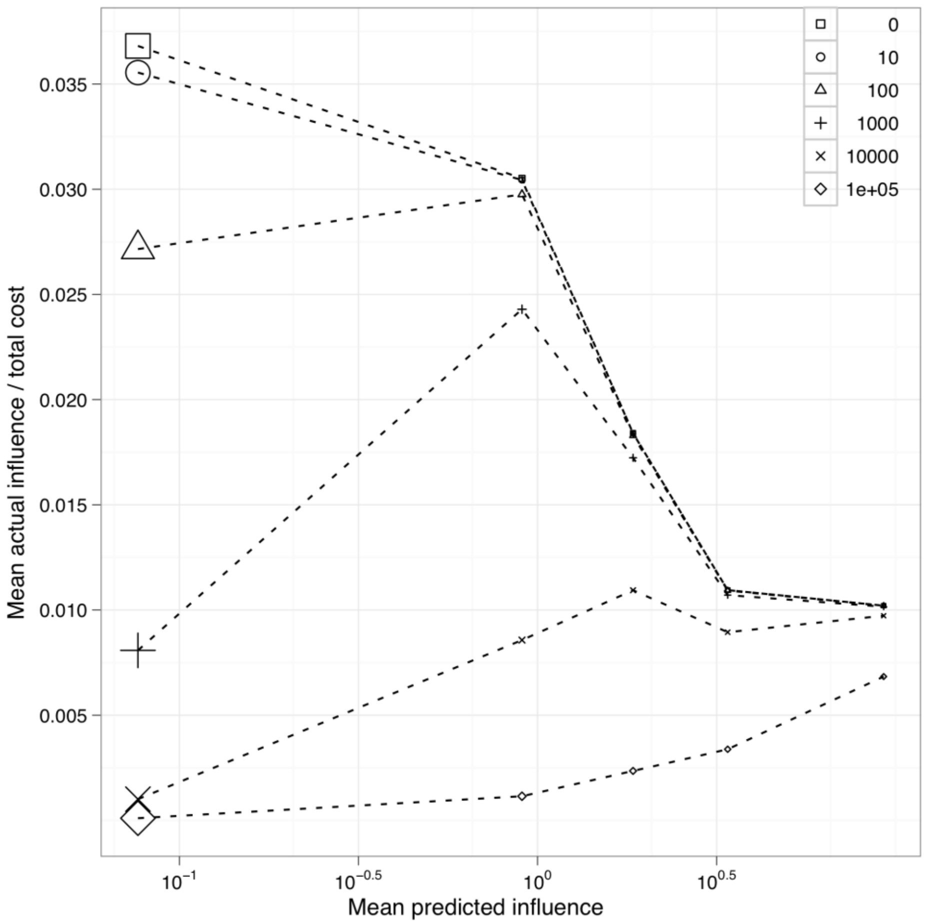
# Should Kim Kardashian Be Paid $10,000 per Tweet?

- On average, some types of influencers are more influential than others
  - Many of them are highly visible celebrities, etc. with millions of followers
  - But these individuals may also be very expensive (i.e. Kim Kardashian)
- Assume the following cost function
  - $c_i = c_a + f_i * c_f$, where $c_a$ = acquisition cost; $c_f$ = per-follower cost
  - Also $c_a = a * c_f$, where a expresses cost of acquiring individual users relative to sponsoring individual tweets
- Should you target:
  - A small # of highly influential seeds?
  - A large # of ordinary seeds with few followers?
  - Somewhere in between?

# “Ordinary Influencers” Dominate

- Assume $c_f$ = \$0.01
  - Equivalent to paying \$10K per tweet for user with 1M followers
- When $c_a$ = \$1,000, ($a$ = 100,000) highly influential users are most cost effective
- But for lower ratios, most efficient choice can be individuals who influence at most one other

Influence per Follower



# 4. Networked Experiments

- Twitter study included both networks and influence
  - But not an actual experiment; so no causal statements possible
- Would like to study impact of network structure in an experimental setting, akin to Music Lab
  - Kearns et al have studied “networked games” in specially equipped labs
    - Graph coloring, consensus, exchange, etc.
    - In all cases, network structure important
  - Can these experiments also be run on the Web?
- Unlike Small World Experiment and Music Lab, subjects must play synchronously
    - Easy to solve in a lab, but not on the web

# Cooperation on Networks

- Question of why presumptively selfish people cooperate one of the most studied in all of social science
  - Dates back to Hobbes.
  - Elinor Ostrom: 2009 Nobel in Economics
- A standard model of problem is “linear public goods” game (also common pool resource, voluntary contribution mechanism)
  - On each of N “rounds”, each member of a group given an endowment
  - Members choose how much of their endowment to contribute
  - Contributions multiplied by some constant (hence “linear”)
  - Then redistributed equally to all members
- Situation poses a dilemma
  - Everyone better off if everyone contributes than if nobody does
  - Individuals better off if everyone else contributes and they don’t

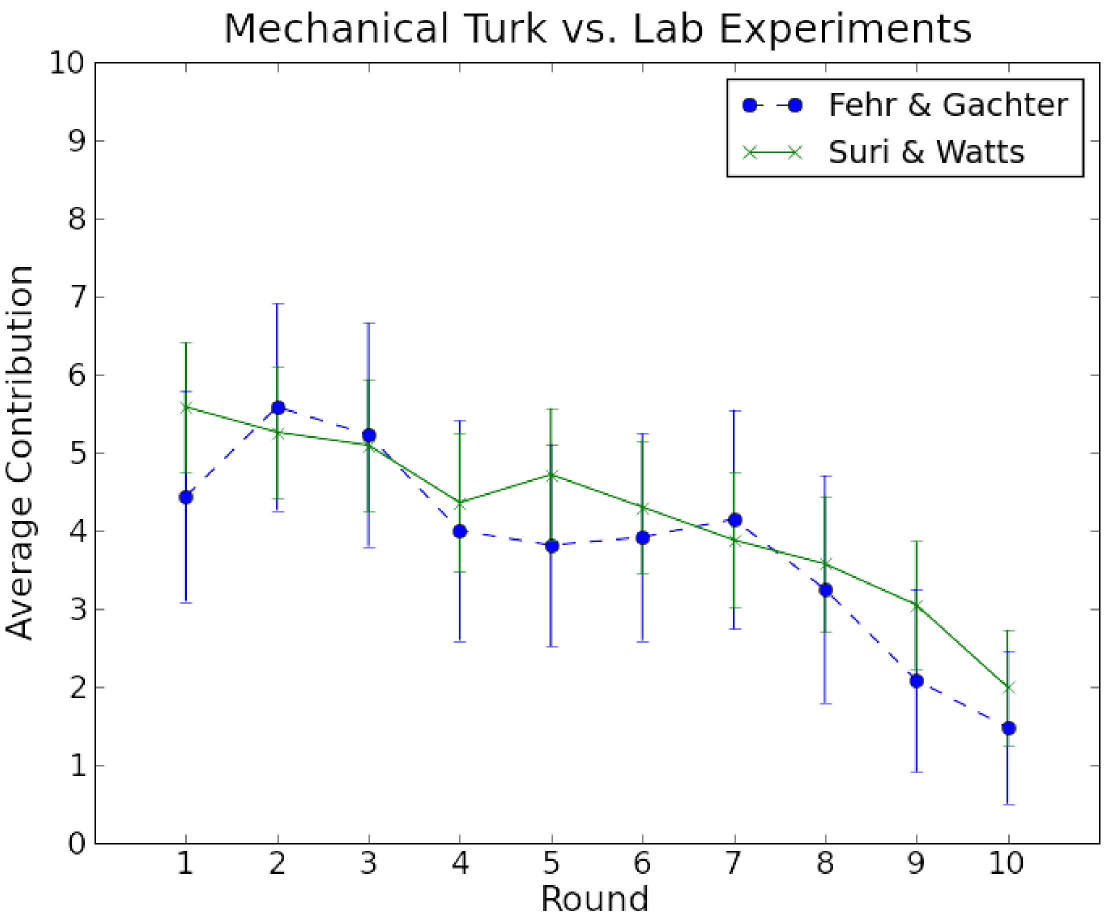
# Public Goods Games

- Public Goods games have been studied extensively in experiments (Fehr/Gachter)
  - Contributions start out high and end low
  - Players will pay to punish non-contributors
  - Punishment increases contributions
- But all these experiments are for “groups” in which everyone plays everyone (N=4)
  - How do these results depend on network structure?

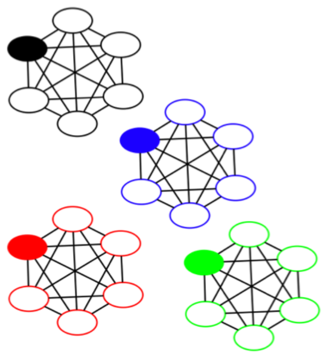
# Amazon's Mechanical Turk

- AMT originally designed for "crowd sourcing"
  - "Requestors" post "HITS" (human intelligence tasks)
  - "Turkers" accept HITS for piece rates
- Increasingly used by behavioral scientists as a virtual lab for conducting human subjects experiments
  - Mason and Suri (2010) recently written a handbook for running experiments on AMT
- Solving the synchronous play problem:
  - In series of preliminary games, recruited a panel of ~ 100 players
  - Notified them in advance of games
    - Scheduled up to four sessions per day
    - Games of 24 players fill up in ~ 2 mins
- Also used preliminary experiments to calibrate AMT

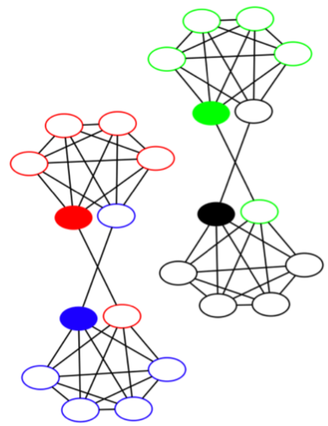
# Comparison with Physical Lab Results (F&G, AER, 2000)



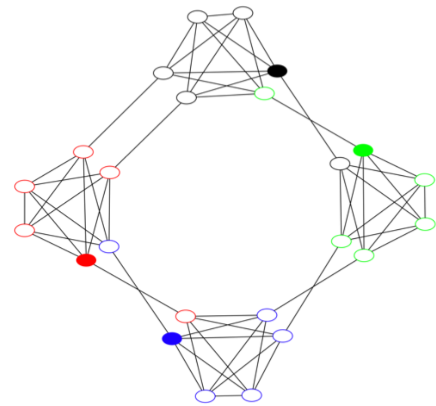
# Networks (N = 24, k = 6)



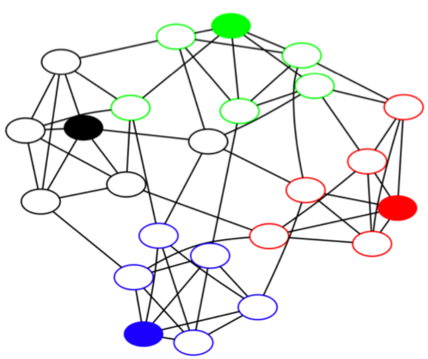
(A) Cliques



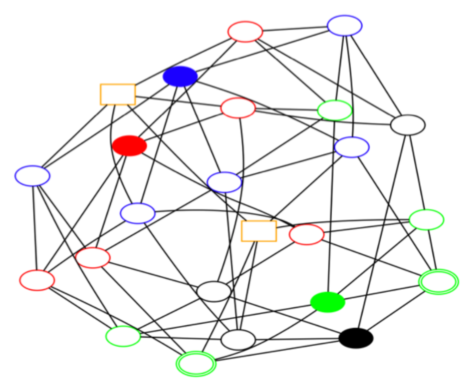
(B) Paired Cliques



(C) Cycle



(D) Small World

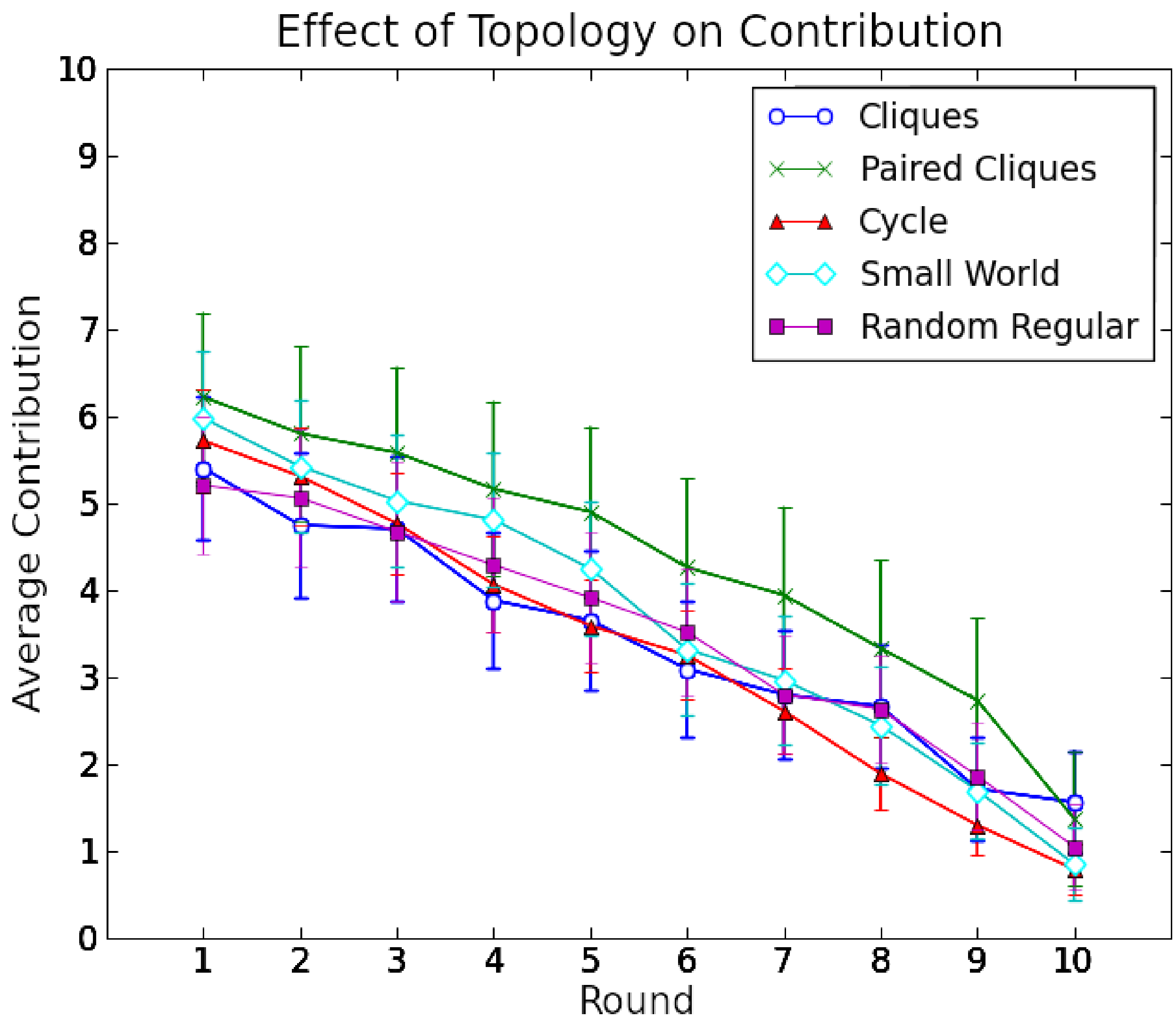


(E) Random Regular

# Network Stats

| | Cliques | Paired Cliques | Cycle | Small World | Random Regular |
|---|---|---|---|---|---|
| Clustering Coefficient (C) | 1.00 | 0.80 | 0.60 | 0.41 | 0.09 |
| Average Path Length (L) | 1.00 | 1.81 | 2.54 | 2.23 | 2.01 |
| Diameter (D) | ∞ | ∞ | 5 | 4 | 3 |
| ROI | 1.04 | 1.09 | 1.38 | 0.80 | 1.00 |

# Surprisingly (to us) Networks Don't Seem to Affect Aggregate Contributions



# When Do Networks Matter?

- Hypothesized importance of networks comes from intuition that cooperation is conditional
  - "I'll cooperate if at least X of my neighbors do"
  - Works better when cooperators interact preferentially
  - Cooperation should be contagious
- By introducing fully-contributing agents in various configurations, we found that
  - Players do cooperate conditionally, but overall bias towards swamps everything
  - No evidence of positive contagion
- Raises interesting theoretical question of when networks matter and when they don't
  - Contrast with coordination / anti-coordination games

# Advantages of the AMT Lab

- Overall, we have now run 113 experiments with N=24 players
  - Up to 20 experiments per week
- Cost of roughly $1 per player per game
  - An order of magnitude cheaper than physical lab
- Greater speed and lower cost allows us to speed up hypothesis-testing cycle
- Also allows us try more variations
  - Different information conditions (friends of friends etc.)
  - Nonlinear production function, rewiring, etc.
  - Larger N, more topologies

# Where have we come in 8 years?

- Small-World experiment large (global) scale network experiment, but no control
- Music Lab medium scale, and no network, but better control
- Twitter, large-scale network with diffusion, but not experimental
- Public Goods games are genuine controlled, web-based, networked games, but scale is still small
  - Main constraint is the size of our panel (now about 100).

# Where are we going now?

- Build “virtual labs” on the web for running all manner of macro economic and sociological experiments
  - Hope to construct large standing panel
- Run field experiments, akin to bucket testing of ads / search results
  - David Reiley at Yahoo! has already done some experiments with brand advertising
- Ultimately, combine study of real world networks (e.g. Mail/IM) with experimental science

# Computational Social Science?

- The web promises to dramatically improve
  - Our ability to measure individual level behavior and interactions on a massive scale in real time
  - Our ability to run “macro sociological” lab experiments and field experiments
- There is a long way to go to from existing studies to the “big” questions of social science
  - Systemic risk, economic development, terrorism
- Nevertheless, technological innovations have revolutionized science in the past (Telescope, Microscope)
- Could Web/Internet revolutionize social science?
  - Will at least provide lots of interesting problems for computer scientists!

# References


*General*

D. J. Watts. A 21st Century Science. *Nature*, 445, p. 489 (2007)

*Small World Experiment*

P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *Science*, 301, 827-829 (2003).

Sharad Goel, Roby Muhamad, and D. J. Watts. "Social Search in 'Small-World' Experiments" *Proceedings of the 18th international conference on World Wide Web,* 701-710 (2009)

*Music Lab*

M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311, 854-856 (2006).

Matthew J. Salganik and Duncan J. Watts. "Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market." *Social Psychology Quarterly* 71:338-355 (2008)

Matthew J. Salganik and Duncan J. Watts. "Web-based experiments for the study of collective social dynamics in cultural markets." *Topics in Cognitive Science*, 1, 439-468 (2009).

*Twitter*

E. Bakshy, J. Hofman, W. Mason, and D. J. Watts. "Identifying 'Influencers' on Twitter" *WSDM* (2011)

D. J. Watts and P. S. Dodds. "Networks, influence, and public opinion formation." *Journal of Consumer Research*, 34(4), 441-458 (2007).

D. J. Watts and J. Peretti. Viral marketing in the real world. *Harvard Business Review* (May, 2007)

*Networked Public Goods*

S.Suri and D. J. Watts "A Study of Cooperation and Contagion in Networked Public Goods Experiments" (submitted)

W. Mason and S. Suri. "A Guide to Running Experiments on AMT" (submitted)