

COMPUTATIONAL SOCIAL SCIENCE: EXCITING POTENTIAL AND FUTURE CHALLENGES

Duncan Watts
Microsoft Research

Microsoft
Research

Social Science vs. Social Practice

- Over past 100 years social science has generated tremendous number of theories of individual and collective human behavior but has not produced a cohesive, cumulative, and empirically tested body of theoretical knowledge
 - Exceptions often artificial settings (e.g. mechanism design for auctions).
 - Macro-econ might have been exception pre-financial crisis
 - Nudging, J-PAL possible exceptions, but limited applicability
- Result is that social practice (business, government, policy) is largely uninformed by social science
 - Designing and conducting advertising and/or marketing campaigns
 - Optimizing organizational performance and/or strategy
 - Enhancing collective action and/or resolving conflict
 - Predicting market demand and/or public opinion
 - Managing systemic risk in financial systems
 - Designing communities and cities
 - Allocating development funds

Microsoft
Research

Why the Lack of Progress?

- Individual people are complicated enough, but **social** phenomena involve **many** individuals **interacting** to produce **collective** entities (firms, markets, cultures, political parties, social movements, audiences)
 - “Micro-Macro” problem (aka “Emergence”)
 - Emergent phenomena arise in natural science, but in social science *every* problem of interest involves emergence
- Micro-macro problems are hard to study empirically
 - Difficult to collect observational data about individuals, networks, and populations at same time
 - Even more difficult to do “macro” scale experiments
- Hard to do science when you can’t measure what you’re theorizing about and can’t do experiments

Microsoft
Research

Computational Social Science

- Revolution in digital communication technologies is beginning to life these historical barriers
 - Has dramatically increased the scale, scope *and* granularity of data available to social scientists
 - Email, e-commerce, search, social networking, social media, etc.
 - Web platforms are also lowering the cost, and increasing the speed and scale of experiments
 - Traditional lab-style experiments + Field experiments
- May revolutionize our ability to *study* society
 - Akin to the telescope + collider for social science
 - View as complement to modeling/simulation approach

Microsoft
Research

From Science to Policy?

- Past 15 years have led to remarkable progress in what is possible for social science
 - Also whole new conferences, research centers, journals, etc. on “computational social science,” “network science,” “big data”
- Nevertheless, little progress on matters of social policy
 - Systemic risk in financial systems, dynamics of emerging epidemics, factors affecting cultural change, organizational performance, or political polarization and conflict.
 - Even simpler questions like “when do people change their minds and why”? are still hard to answer except in trivial special cases
- How to close the gap between excitement and results?

Microsoft
Research

Improving Found Data

- Behavioral data (“digital breadcrumbs”) currently collected on many disconnected platforms
 - Facebook for self-reported social networks, Google and Bing for search queries, Amazon and eBay for e-commerce, Nielsen for ratings, various email providers, etc.
- Many questions of interest to social science could be better addressed if these “modes” of behavior could be joined
 - E.g. “Who influences whom” requires (a) individual data, (b) interaction data, (c) behavioral data, (d) attribution data
 - Clear privacy, legal, and technical obstacles to doing this
- Another problem is that digital data is increasingly “algorithmically confounded”
 - Personalization + recommendations bias user behavior
 - Any feature change can impact apparently “social” phenomena
- May need to collect data with research questions in mind
 - Opt-in panels one possibility

Microsoft
Research

Scaling up the Lab

- History of experimental psychology / economics constrained by scale and speed
 - Unit of analysis was individuals or small groups
 - Experiments took ~ 1yr to design and run
- Potentially “Virtual labs” lift both constraints
 - State of the art ~ 5K workers, but in principle could construct subject panel ~ 100K – 1M
 - Could shrink hypothesis-testing cycle to days or hours
- Would open up fundamentally new research designs
 - Could study whole organizations, even “cultures” in the lab
 - Experiments could run for months not minutes
 - Tracking individuals would allow for novel sampling and insights

Microsoft
Research

Empirically Informed Modeling

- Traditional mathematical or computational modeling
 - Tends to rely on many, often questionable, assumptions
 - Not generally tested in detail against data
- Result is proliferation of models that exist in parallel and are often incompatible with each other
 - Diffusion, cooperation, systemic risk, organizational performance
- New sources of data allow both to test models and also calibrate them
 - Diffusion models tested against observational data
 - Lab experiments used to calibrated agent-based models
- Models complete hypothesis-testing cycle
 - Observations → Models → Lab → Field → Observations

Microsoft
Research

Institutional Innovations

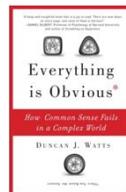
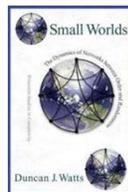
- New platforms and protocols for data management
 - Better coordination of data collection, storage, sharing
 - Recruitment and management of subject pools, field panels
- Collaborative interdisciplinary teams
 - For a given data set, often unclear what the most interesting question is
 - For a given question, often unclear how to collect the right data
- Integrated research designs
 - Coordination across theoretical, experimental and observational studies
- Potentially new research institutions
 - UrbanCCD, CUSP interesting models
 - Public-private partnerships (especially around data sharing)?
 - Janelia Farm for Social Science?
 - J-PAL for the first world?

Microsoft
Research

THANK YOU

<http://research.microsoft.com/en-us/people/duncan/>

<http://everythingisobvious.com>

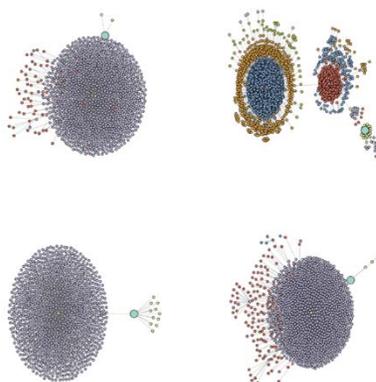
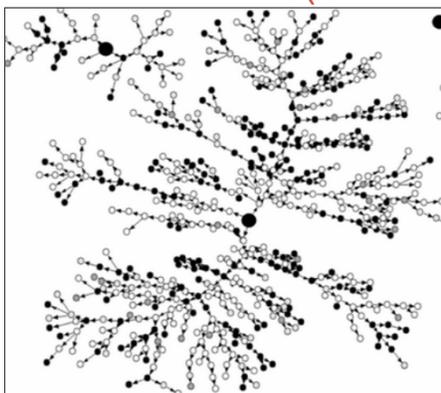


Microsoft
Research

BACKUP SLIDES: ILLUSTRATIVE EXAMPLES OF CSS

Microsoft
Research

How Does Popular Content Get Popular? (Goel et al 2014)



Viral vs. Broadcast?

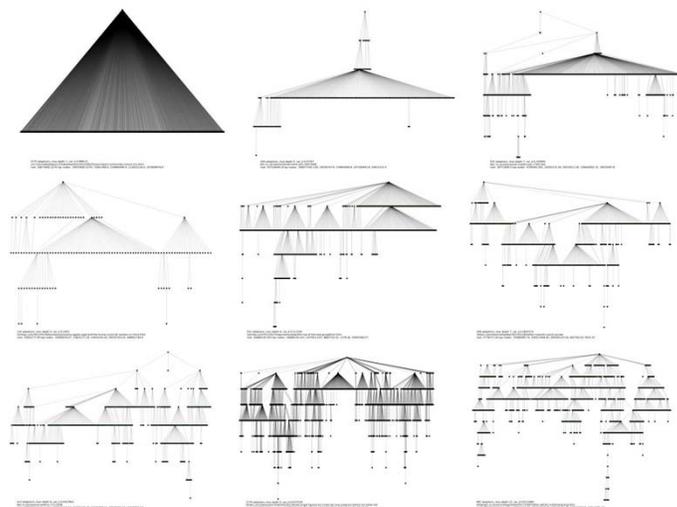
Microsoft
Research

Online Diffusion (Goel et al 2014)

- Study every tweet containing a link to a video, news story, image, or petition over 12 month period
 - About 1.4B tweets total
 - “Popular” subset ~ 350,000 events
- Also crawled entire “active” follower graph
 - ~65M users, > 10B edges
- To focus now on “popular” content, consider only URLs that receive > 100 tweets, roughly 1 in 3,000 events
 - > 1,000 tweets roughly 1 in 1,000,000 events
 - If want thousands of large events, need ~ 1B observations!
- Measure “structural virality” of cascades:
 - Construct tree of all retweets, retweets of retweets etc.
 - Compute average all-pairs path length on these trees

Microsoft
Research

Diversity of Structural Virality (Goel et al 2014)



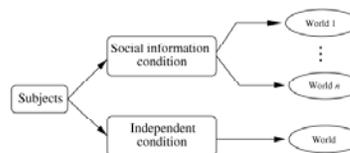
Microsoft
research

Music Lab

Salganik, Dodds, Watts (Science, 2006)

- Why are ‘hits’ in cultural markets
 - Much more successful than average
 - Yet so hard to predict?
- Conducted experiment on social influence and market dynamics
 - 14,000 participants chose between 48 songs by unknown bands
 - Randomly assigned to ‘social influence’ and ‘independent’ conditions
- Social influence simultaneously increased inequality and unpredictability
 - Markets “construct” preferences as well as reveal them

Subject	World	Condition	Rank	Score
1	World 1	Social information	1	0.95
2	World 1	Social information	2	0.85
3	World 1	Social information	3	0.75
4	World 1	Social information	4	0.65
5	World 1	Social information	5	0.55
6	World 1	Social information	6	0.45
7	World 1	Social information	7	0.35
8	World 1	Social information	8	0.25
9	World 1	Social information	9	0.15
10	World 1	Social information	10	0.05
11	World 1	Social information	11	0.05
12	World 1	Social information	12	0.05
13	World 1	Social information	13	0.05
14	World 1	Social information	14	0.05
15	World 1	Social information	15	0.05
16	World 1	Social information	16	0.05
17	World 1	Social information	17	0.05
18	World 1	Social information	18	0.05
19	World 1	Social information	19	0.05
20	World 1	Social information	20	0.05
21	World 1	Social information	21	0.05
22	World 1	Social information	22	0.05
23	World 1	Social information	23	0.05
24	World 1	Social information	24	0.05
25	World 1	Social information	25	0.05
26	World 1	Social information	26	0.05
27	World 1	Social information	27	0.05
28	World 1	Social information	28	0.05
29	World 1	Social information	29	0.05
30	World 1	Social information	30	0.05
31	World 1	Social information	31	0.05
32	World 1	Social information	32	0.05
33	World 1	Social information	33	0.05
34	World 1	Social information	34	0.05
35	World 1	Social information	35	0.05
36	World 1	Social information	36	0.05
37	World 1	Social information	37	0.05
38	World 1	Social information	38	0.05
39	World 1	Social information	39	0.05
40	World 1	Social information	40	0.05
41	World 1	Social information	41	0.05
42	World 1	Social information	42	0.05
43	World 1	Social information	43	0.05
44	World 1	Social information	44	0.05
45	World 1	Social information	45	0.05
46	World 1	Social information	46	0.05
47	World 1	Social information	47	0.05
48	World 1	Social information	48	0.05
49	World 1	Social information	49	0.05
50	World 1	Social information	50	0.05
51	World 1	Social information	51	0.05
52	World 1	Social information	52	0.05
53	World 1	Social information	53	0.05
54	World 1	Social information	54	0.05
55	World 1	Social information	55	0.05
56	World 1	Social information	56	0.05
57	World 1	Social information	57	0.05
58	World 1	Social information	58	0.05
59	World 1	Social information	59	0.05
60	World 1	Social information	60	0.05
61	World 1	Social information	61	0.05
62	World 1	Social information	62	0.05
63	World 1	Social information	63	0.05
64	World 1	Social information	64	0.05
65	World 1	Social information	65	0.05
66	World 1	Social information	66	0.05
67	World 1	Social information	67	0.05
68	World 1	Social information	68	0.05
69	World 1	Social information	69	0.05
70	World 1	Social information	70	0.05
71	World 1	Social information	71	0.05
72	World 1	Social information	72	0.05
73	World 1	Social information	73	0.05
74	World 1	Social information	74	0.05
75	World 1	Social information	75	0.05
76	World 1	Social information	76	0.05
77	World 1	Social information	77	0.05
78	World 1	Social information	78	0.05
79	World 1	Social information	79	0.05
80	World 1	Social information	80	0.05
81	World 1	Social information	81	0.05
82	World 1	Social information	82	0.05
83	World 1	Social information	83	0.05
84	World 1	Social information	84	0.05
85	World 1	Social information	85	0.05
86	World 1	Social information	86	0.05
87	World 1	Social information	87	0.05
88	World 1	Social information	88	0.05
89	World 1	Social information	89	0.05
90	World 1	Social information	90	0.05
91	World 1	Social information	91	0.05
92	World 1	Social information	92	0.05
93	World 1	Social information	93	0.05
94	World 1	Social information	94	0.05
95	World 1	Social information	95	0.05
96	World 1	Social information	96	0.05
97	World 1	Social information	97	0.05
98	World 1	Social information	98	0.05
99	World 1	Social information	99	0.05
100	World 1	Social information	100	0.05

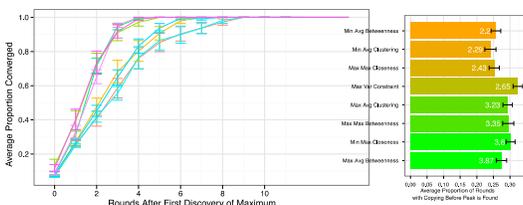
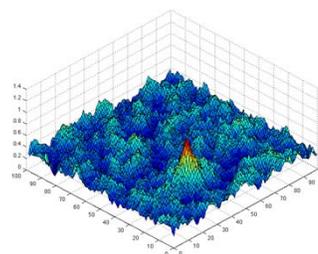


Microsoft Research

Collaborative Learning in Networks

Mason and Watts (2012)

- Networks of N=16 individuals collaboratively searching a fitness landscape
 - Efficient (short path length) networks distributed information faster
 - Also resulted in less copying, more exploration



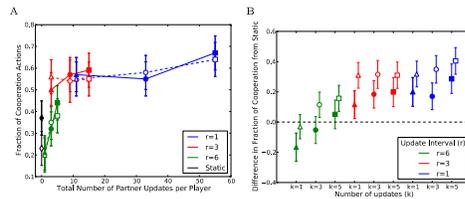
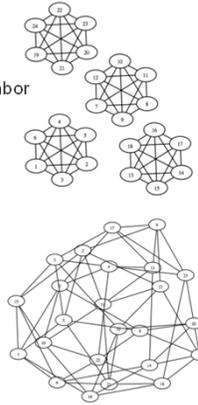
Microsoft Research

Cooperation on Dynamic Networks

Wang, Suri, Watts (2012)

- Studied repeated prisoner's dilemma on a network (n=24)
- Every r rounds, players allowed to make up to k updates
 - Delete an edge from an existing neighbor or propose an edge to a new neighbor
 - Proposed ties had to be accepted, but deletions were unilateral
- Example: for r = 3, k = 5, have:

1	2	3	rewire	4	5	6	rewire	7	8	9	rewire	10	11	12
			5 links				5 links				5 links			
- Studied r = 1, 3, 6; k = 1, 3, 5
 - Studied two initial substrates (random and cliques)
 - Also studied static networks as controls
 - ~4 trials per treatment, so 80 experiments in total
- Found that rewiring matters (a lot!) but substrate doesn't
 - Consistent with earlier exps on static networks (Suri and Watts 2011)



Microsoft Research