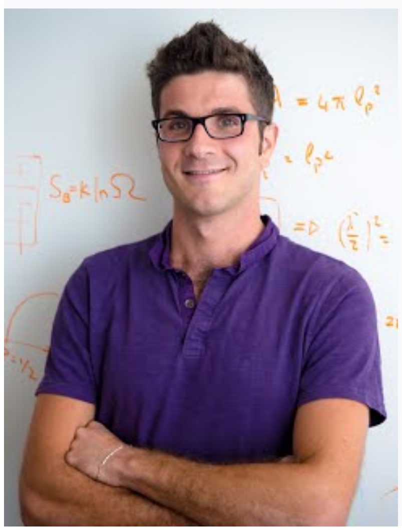
Towards a Theory of Social Dynamics Over Networks

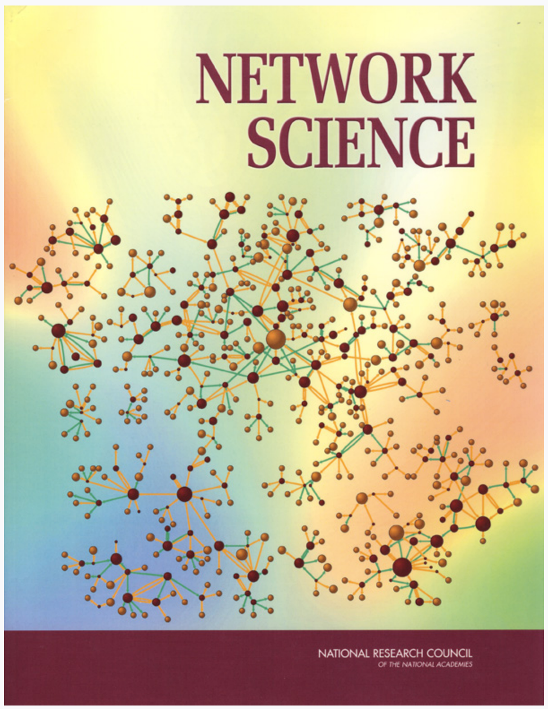
Massimo Franceschetti

University of California San Diego

Advanced Network Sciences



What is network science?



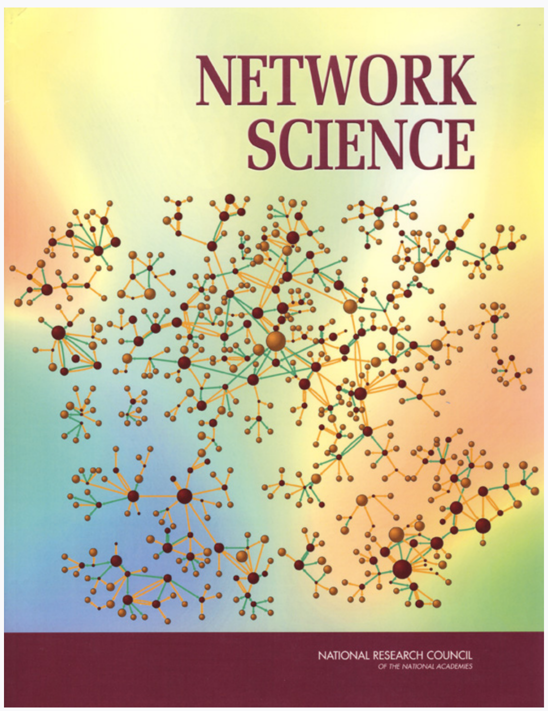
2005

"The study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena."

National Research Council (2005)

What is network science?

- Much research in Network Science on structural properties
- The natural next step: agents interaction



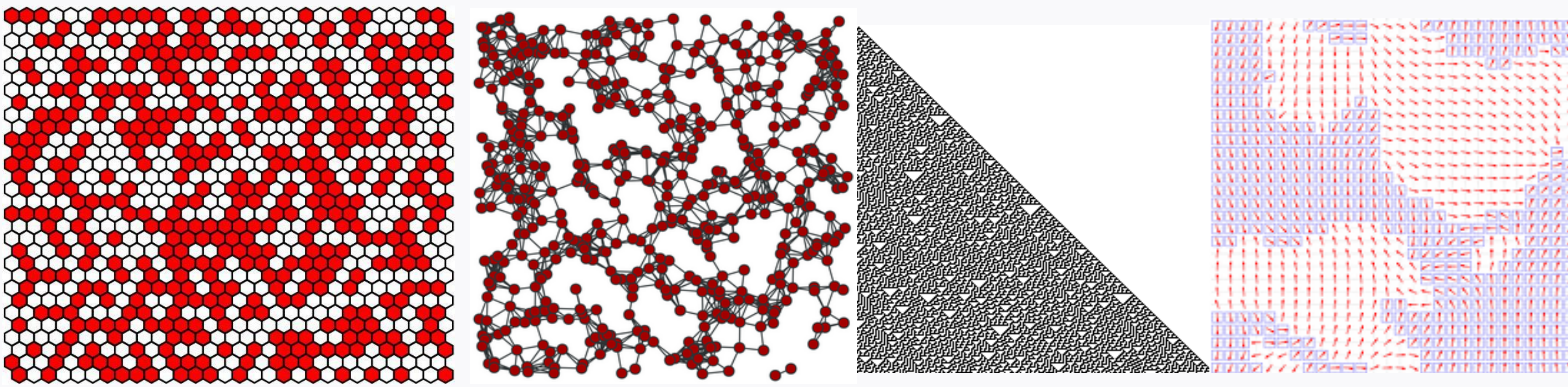
2005

2016

Basic premise

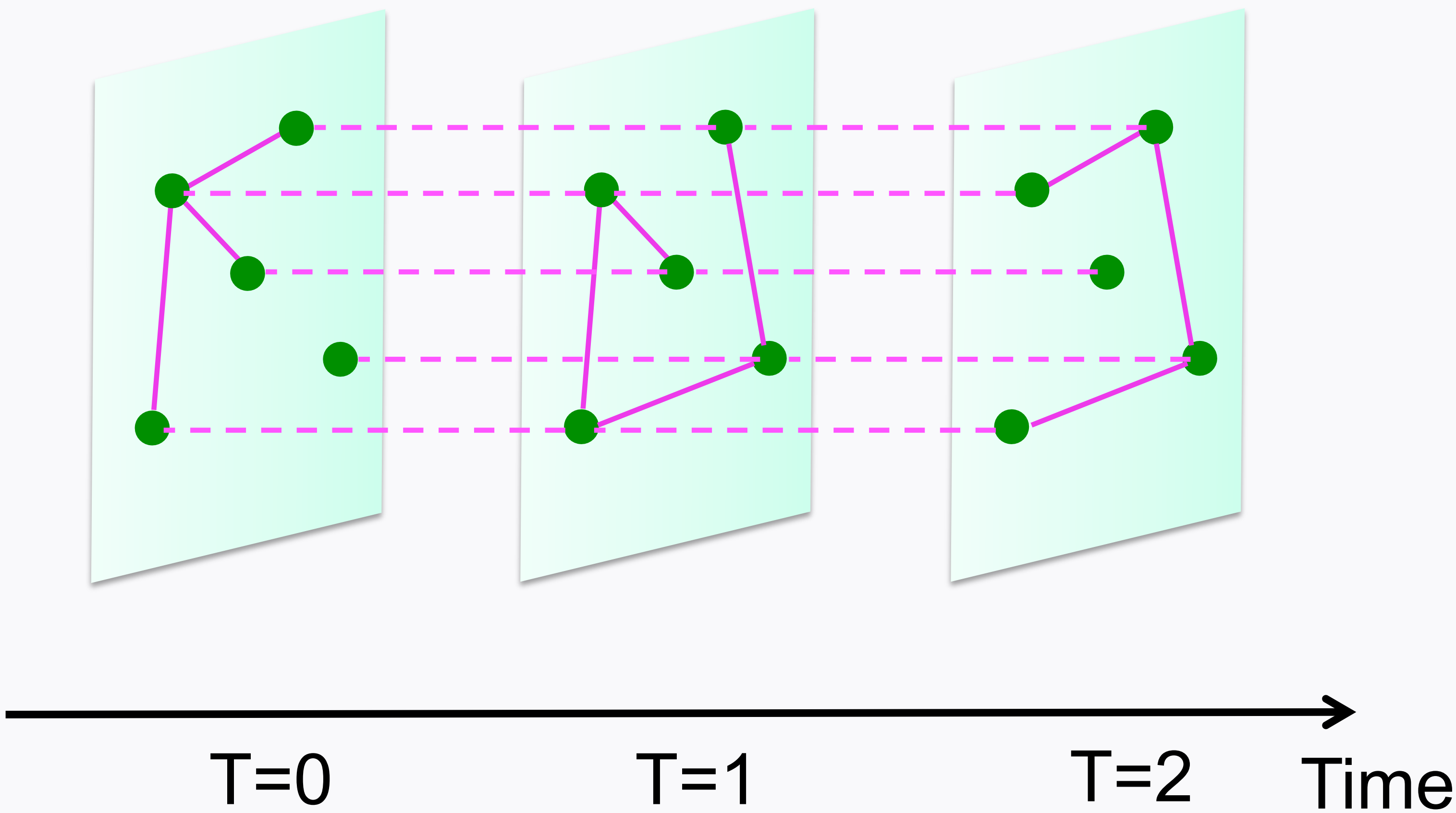
Simple, local rules of social interaction over networks can explain complex, global dynamics

Reminiscent of a theme in physics

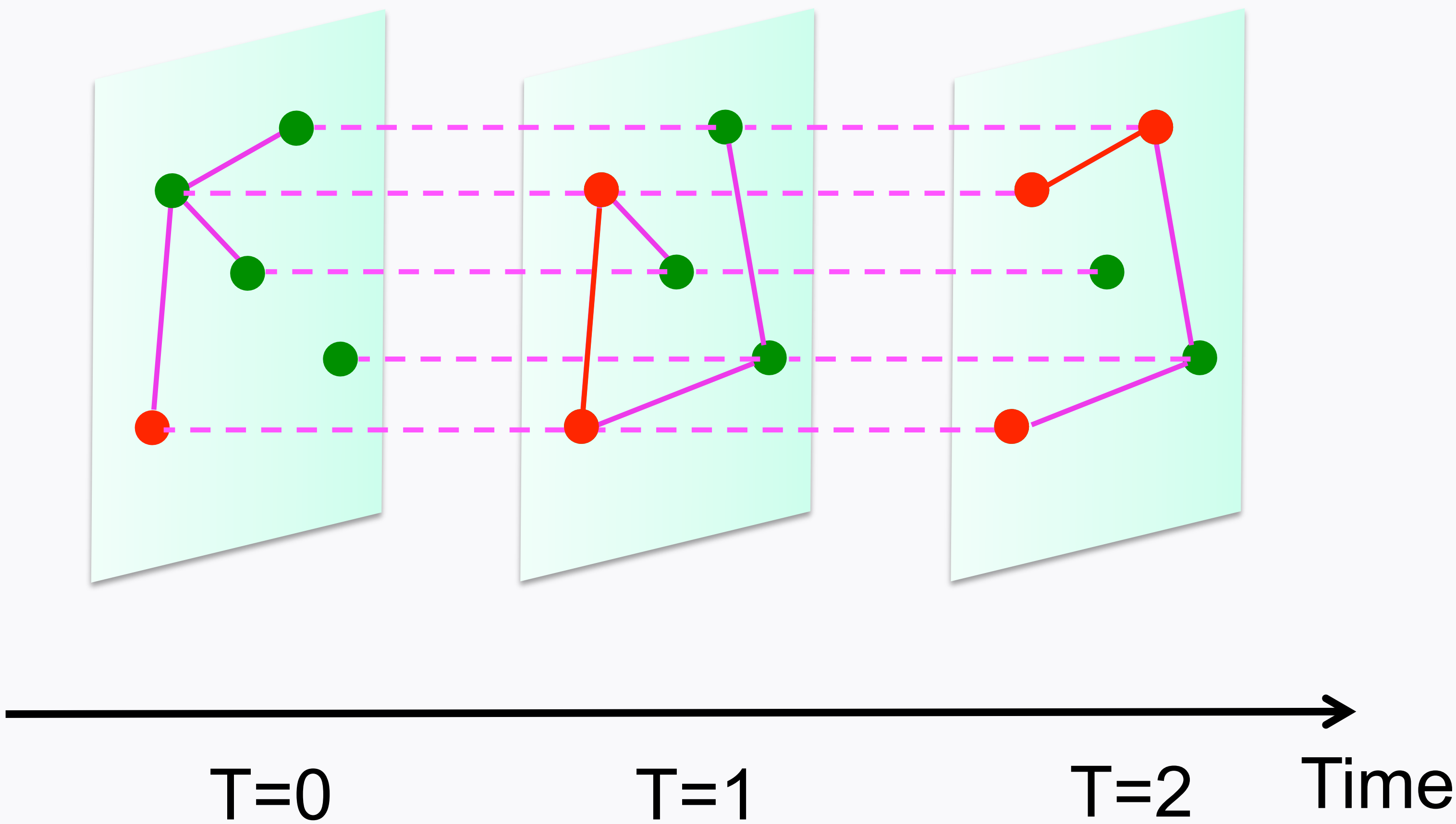


However, algorithmic models enable a complexity analysis generally absent from physical models

Dynamics OF the network



Dynamics ON the network



Human networks

• Behavioral processes for human decision making are driven by algorithmic processes

• Modeling and analysis of these processes can reveal complex network dynamics

Herbert Simon
Nobel laurate, 1978

Topic 1: Social computation

- Real population of heterogeneous, complex agents solving a distributed computation task
- Model as homogeneous, simple agents
- Predictive power

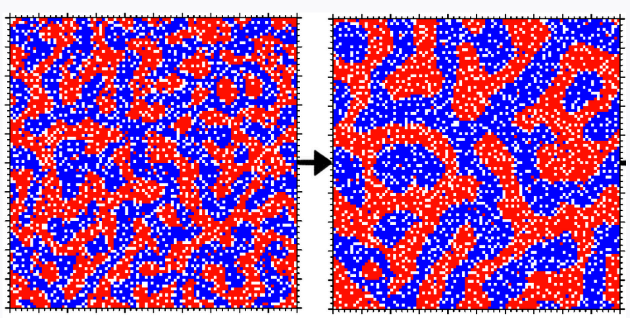
Topic 2: Emotional Contagion

- From information to opinions, and emotions
- Study of expression
- Detect and quantify emotional contagion

Network epidemics

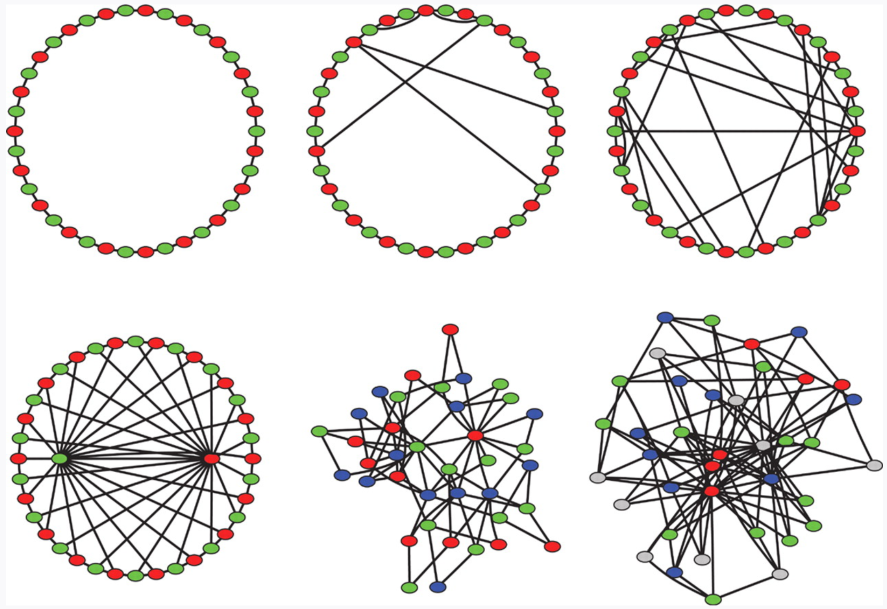
Predicting and containing epidemic risk using social networks data

Models of segregation



Characterize how local decisions can have global outcomes

Social computation via coordination games

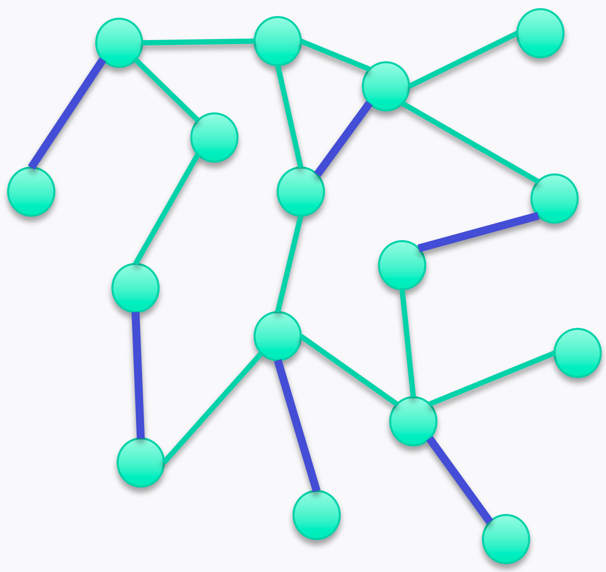


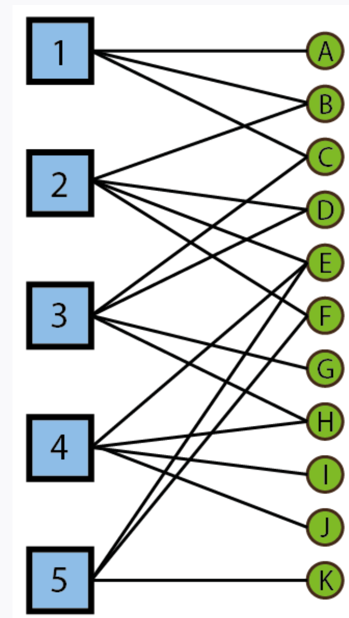
Kearns et al. (Science 2006, Comm. ACM 2012)

- Coloring and consensus games
- No attempt to model human behavior
- Focus on what network structures facilitate a solution

Coordination games over networks

Coviello, et al. (PLOS ONE 2013, IEEE Trans. CNS 2016)

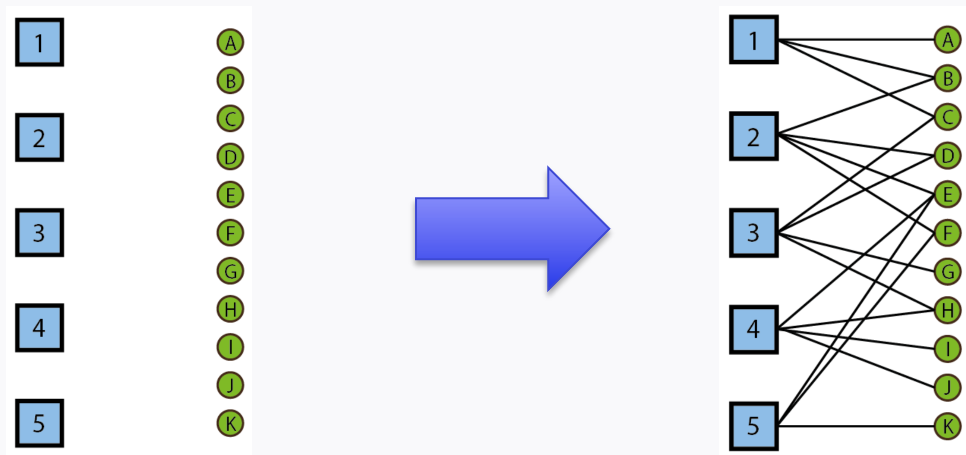




- Matching game
- Group membership game
- Focus on algorithmic game dynamics

Group membership task

Leaders and followers form a bipartite communication network



Each agent has a view of its neighborhood only

- ℓ has to build a team of c_ℓ followers
- Can join a single team at any time

Lab experiments

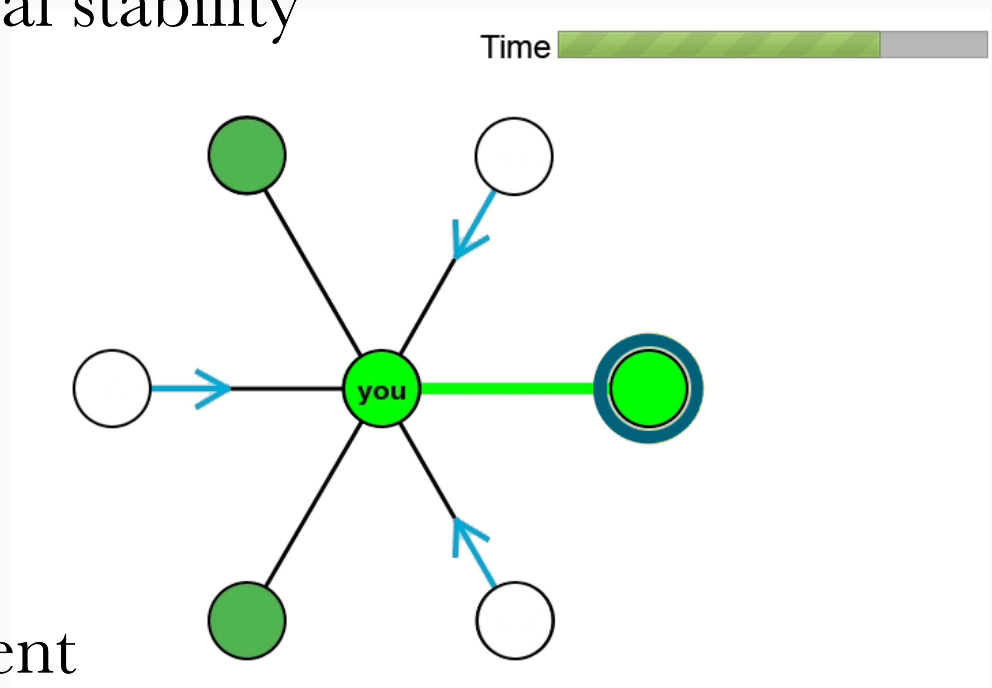
Lab experiments

Each user controls one node through a computer interface

Common goal: reach global stability

5min

$1



36 games over 10 different
networks of 16 nodes each

Algorithmic model

Leader

IF (team size $< c_\ell$) **THEN**

with probability p

select follower f at random (prefer unmatched)

send "team-join" request to f

Follower

IF ($\exists$ incoming "team-join" request) **THEN**

choose one at random

join corresponding team with probability q

Algorithmic model

Memoryless

Local information

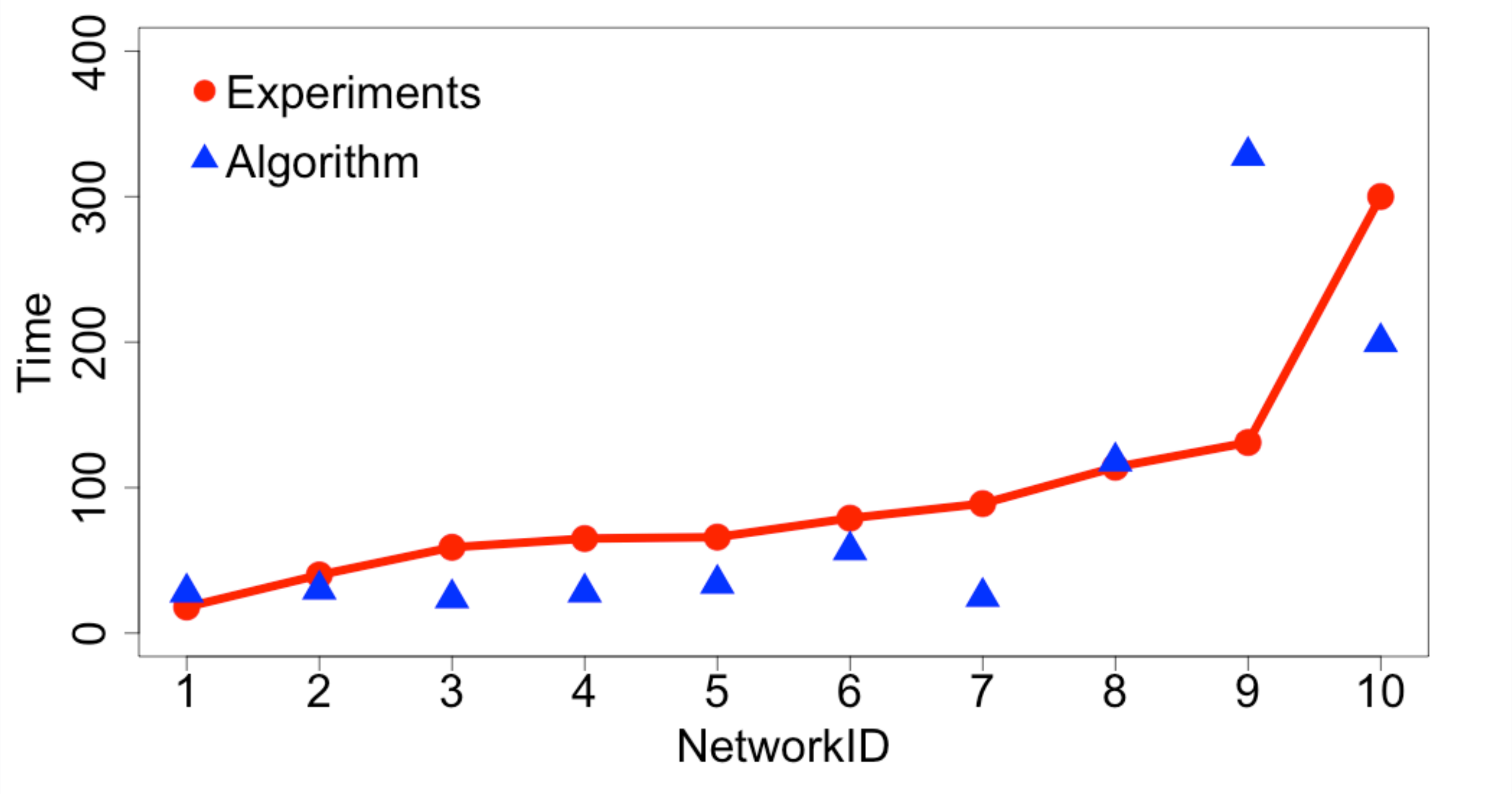
Self-stabilizing

1-bit messages

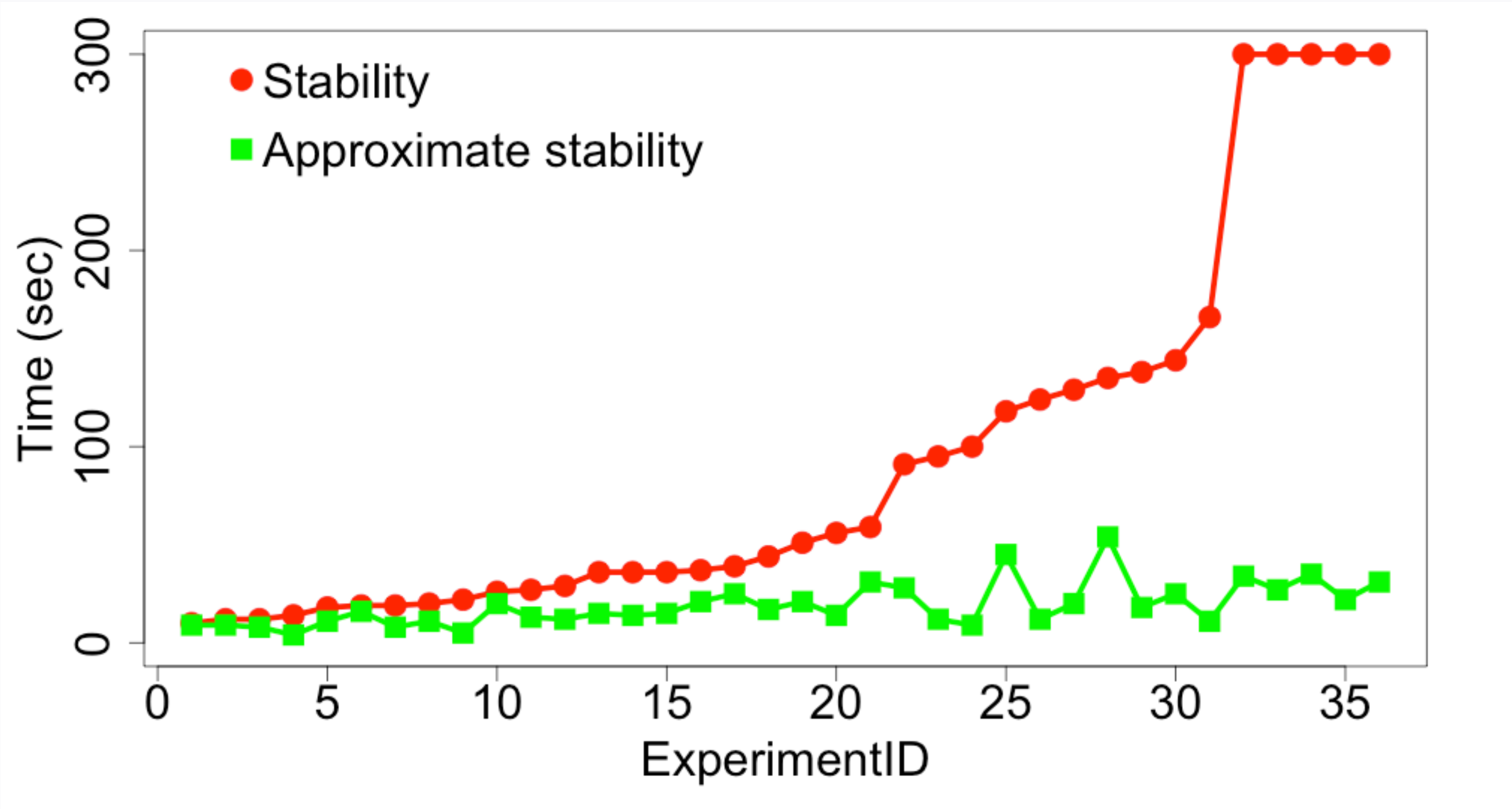
Leaders pursue local stability

Followers provide randomization

Average solving times



Human networks experiments



Hypothesis

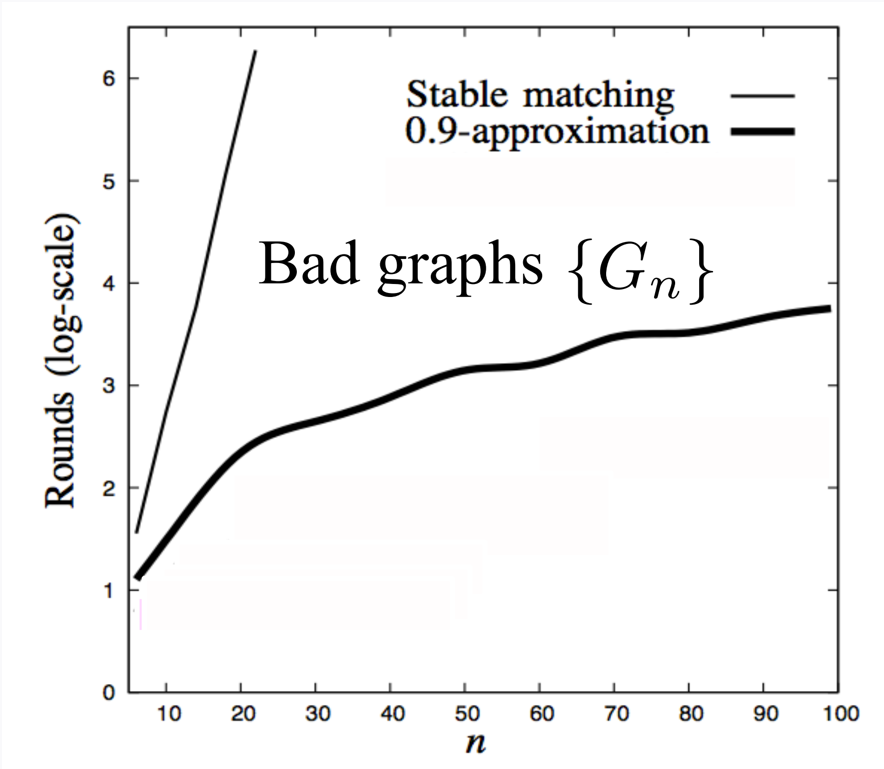
A good solution is always found quickly,

But it can take a long time to improve it to the optimum

Theorem

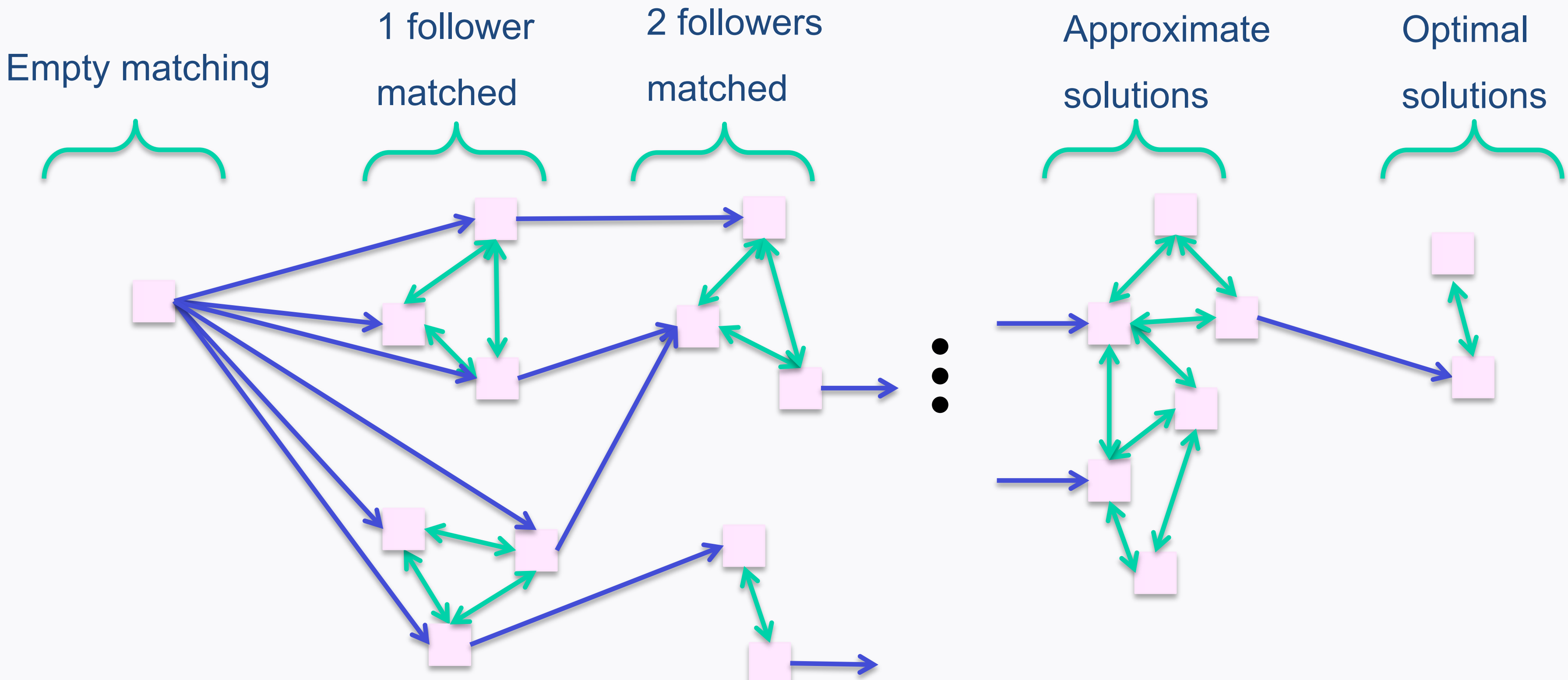
$$\forall \text{ graphs } T(n) = O(\Delta^{1/\epsilon} n) \text{ w.h.p.}$$

$$\exists \text{ graph: } T(n) = \Omega(\exp(n)) \text{ w.h.p.}$$



Analysis

State evolution is a Markov chain over one-to-many matchings



Summary

Simple models of distributed computation can predict the performance of real populations solving computational problems over networks

Global dynamics of complex agents with possibly diverse strategies can be well described by simple synthetic agents with uniform strategies

Advocate usage of simple algorithmic models to investigate a wider variety of social computation tasks

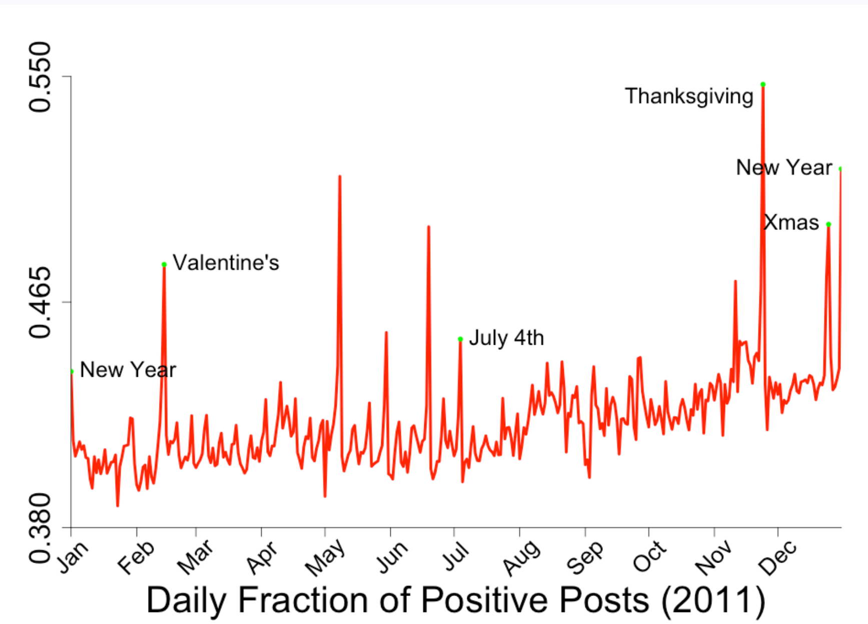
Detecting emotional contagion

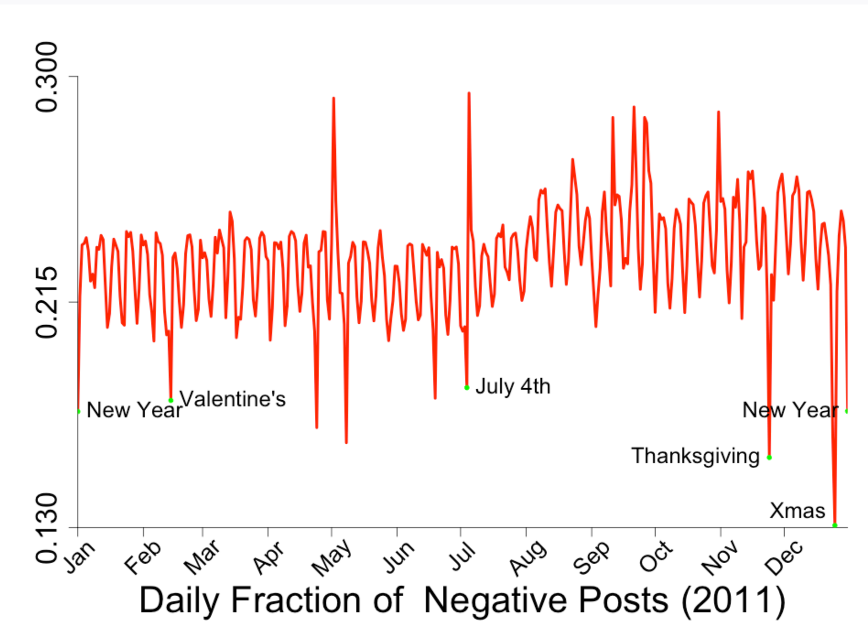
Linguistic word count

Status updates (posts): undirected expression

Classify semantic content of posts using LIWC

Count the fraction of posts with a word from a given semantic category





Experimental approach

Kramer, et al. (PNAS 2014)

Experimental treatment → User's expression → Friends' expression

Research at Facebook

By Mike Schroepfer, Chief Technology Officer

…We should have done differently. For example, we should have considered other, non-experimental ways to do this research…

The Washington Post

Angry mood manipulation subjects interview with Facebook…

The New York Times

Facebook promises deeper review of user research…

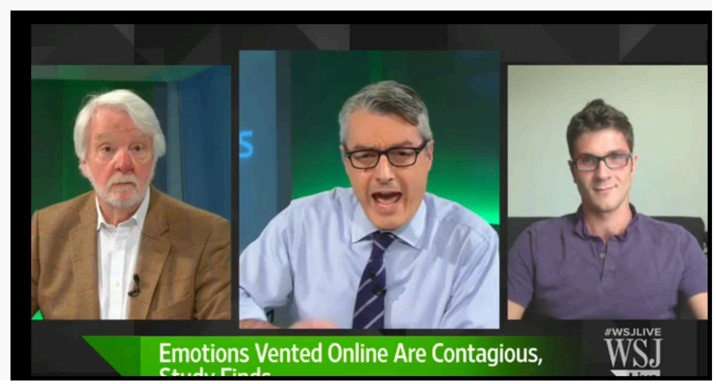
Non-experimental data analysis

Coviello, et al. (PLOS 2014, Proc-IEEE, 2015)

External variable → User's expression → Friends' expression

We use observational data only, without running an experiment

Instrumental variable regression, based on identifying an external variable that we cannot control but that we can observe performing a "natural" experiment



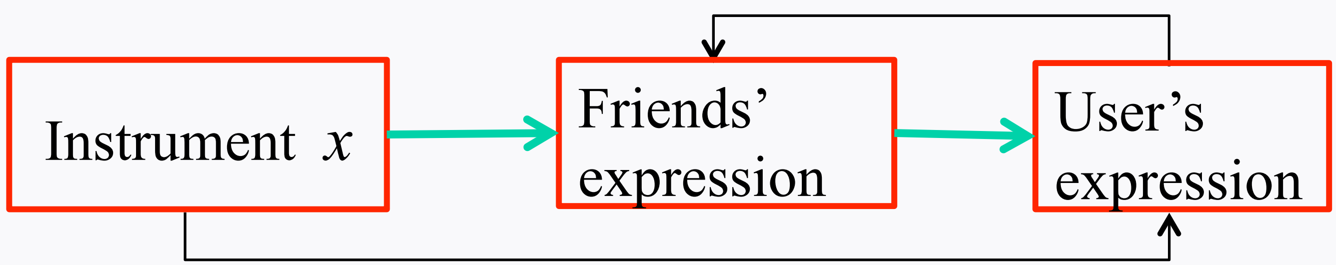
Statistical model of emotional contagion

$$y_i(t) = \theta(t) + f_i + \beta x_i(t) + \frac{\lambda}{\delta_i(t)} \sum_j a_{i,j}(t) y_{i,j}(t) + \epsilon_i(t)$$

Problem of identifying a valid external instrument

Problem of data reduction

Problem of causal dependencies yielding biased estimates (feedback)



Instrumental variable

$$y_i(t) = \theta(t) + f_i + \boxed{\beta x_i(t)} + \frac{\lambda}{\delta_i(t)} \sum_j a_{i,j}(t) y_{i,j}(t) + \epsilon_i(t)$$

Weather affects emotion

Use meteorological data for the 100 most populous US cities

US National climatic center (NCDC http://www.ncdc.noaa.gov)

Users were geo-located using IP addresses

Data aggregation

$$\boxed{\frac{1}{n_g}\sum_{i\in S_g} y_i(t)} = \frac{1}{n_g}\sum_{i\in S_g}\left(\theta(t) + f_i + \boxed{\beta x_i(t)} + \boxed{\frac{\lambda}{\delta_i(t)}\sum_j a_{i,j}(t) y_{i,j}(t)} + \epsilon_i(t)\right)$$

Need to aggregate data of hundred-millions users, billions friends, period of observation of 1180 days

100 observations per day in different cities

Average emotion of user in city g at time t

Average emotional influence on user in city g at time t by all of her friends

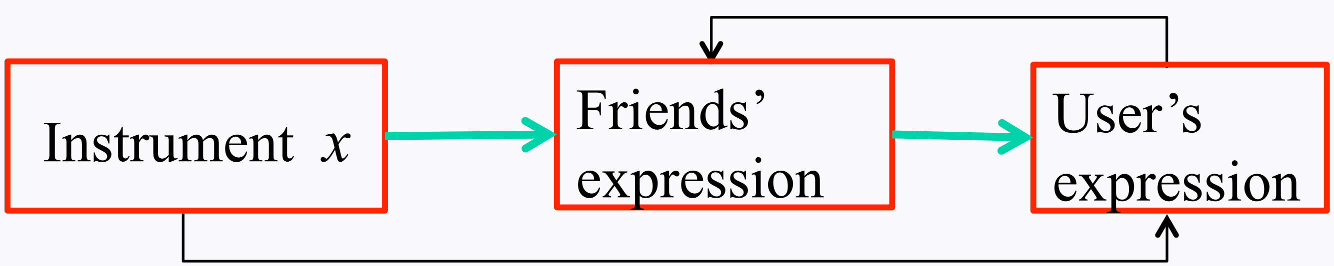
Average emotional influence on user in city g at time t by external variable

Dealing with causality

My friend's emotion is affected by her weather and by my weather (indirectly, through contagion)

My emotion is affected my weather and by the cumulative effect of my friends emotion (that could also be experiencing my same weather)

Need to separate effect of weather and effect of contagion to obtain unbiased estimates

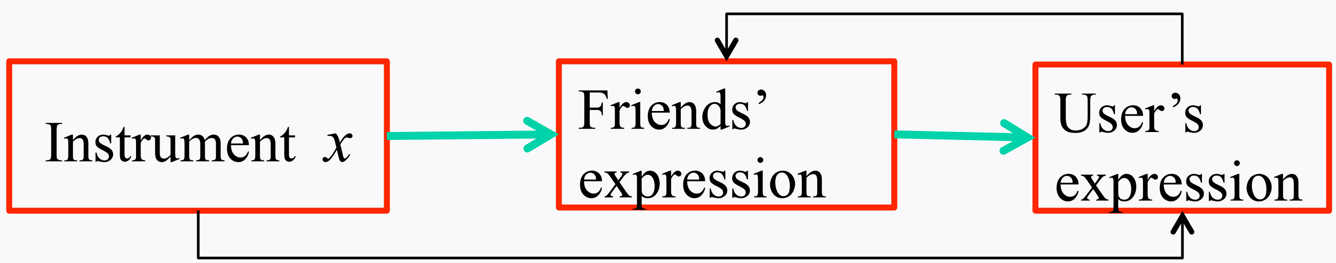


Dealing with causality

$$\bar{y}_g(t) = \theta(t) + \bar{f}_g + \cancel{\beta \bar{x}_g(t)} + \lambda \bar{Y}_g(t) + \bar{\epsilon}_g(t)$$

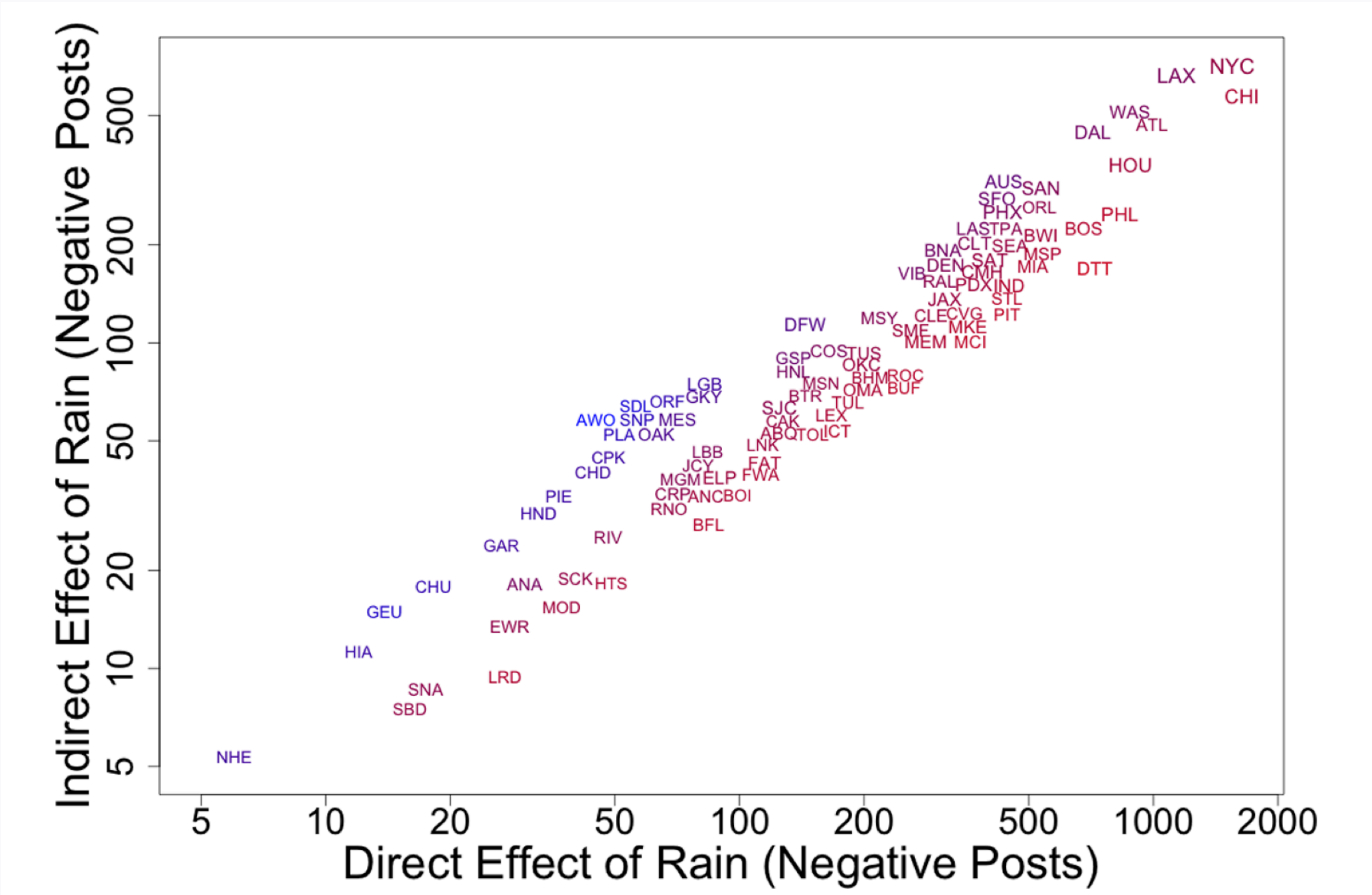
$$\bar{Y}_g(t) = \theta'(t) + \bar{f}'_g + \beta_1 \bar{X}_g(t) + \cancel{\beta_2 \bar{x}_g(t)} + \bar{\epsilon'}_g(t)$$

$$\bar{y}_g(t) = (\theta(t) + \lambda\theta'(t)) + (\bar{f}_g + \lambda \bar{f}'_g(t)) + \lambda\beta_1 \bar{X}_g(t) + \bar{\epsilon}''_g(t)$$



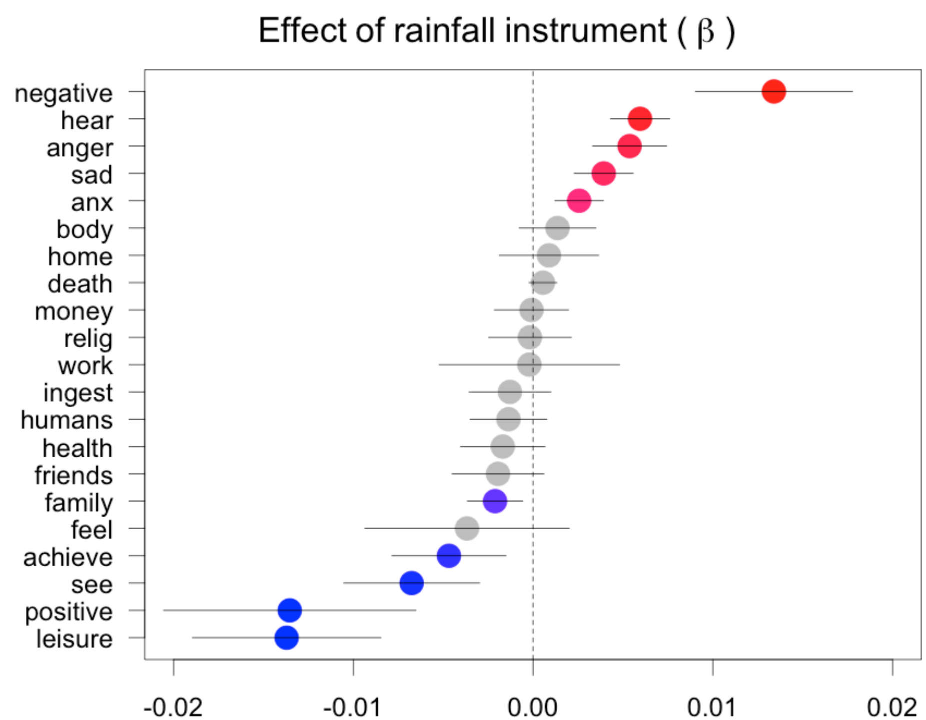
Only consider observations for city/day pairs that experience different weather

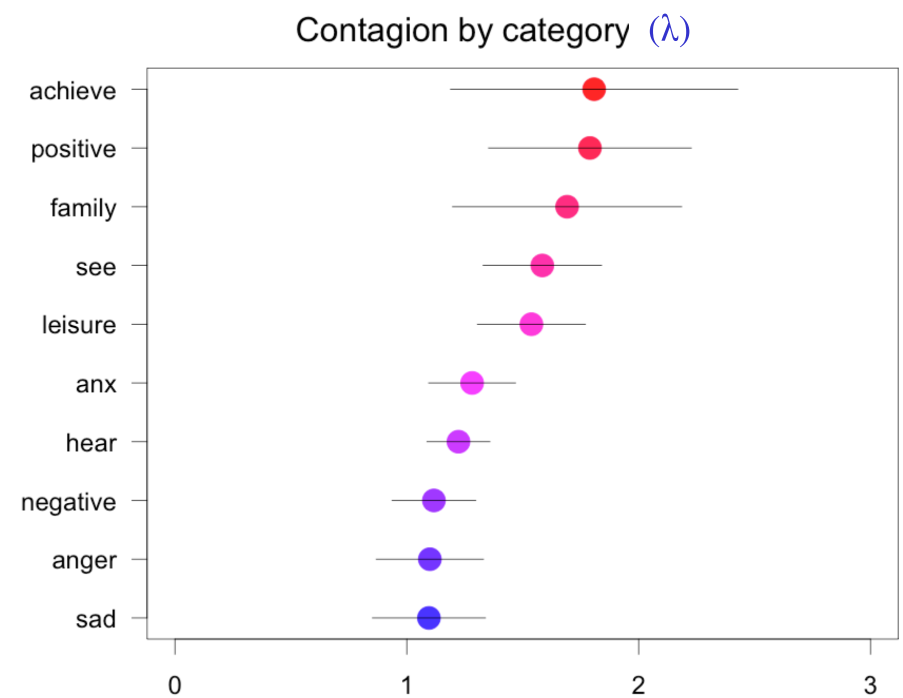
Results



Indirect Effect of Rain (Negative Posts)
Direct Effect of Rain (Negative Posts)
5
10
20
50
100
200
500
1000
2000
NYC
LAX
CHI
WAS
ATL
DAL
HOU
AUS
SAN
SFO
PHX
ORL
PHL
LAS
TPA
BWI
BOS
BNA
CLT
SEA
MSP
DEN
SAT
MIA
DTT
VIB
RAL
CMH
PDX
IND
JAX
STL
MSY
CLE
CVG
PIT
DFW
MKE
MEM
MCI
GSP
COS
TUS
HNL
OKC
MSN
BHM
ROC
LGB
BTR
OMA
BUF
SDL
ORF
GKY
TUL
SJC
CAK
LEX
AWO
SNP
MES
ABQ
TOL
ICT
PLA
OAK
LNK
LBB
CPK
JCY
FAT
CHD
MGM
ELP
FWA
PIE
CRP
ANC
BOI
HND
RNO
BFL
RIV
GAR
CHU
ANA
SCK
HTS
MOD
GEU
EWR
HIA
LRD
SNA
SBD
NHE

Results


Effect of rainfall instrument (β)
negative
hear
anger
sad
anx
body
home
death
money
relig
work
ingest
humans
health
friends
family
feel
achieve
see
positive
leisure
-0.02
-0.01
0.00
0.01
0.02


Contagion by category (λ)
achieve
positive
family
see
leisure
anx
hear
negative
anger
sad
0
1
2
3

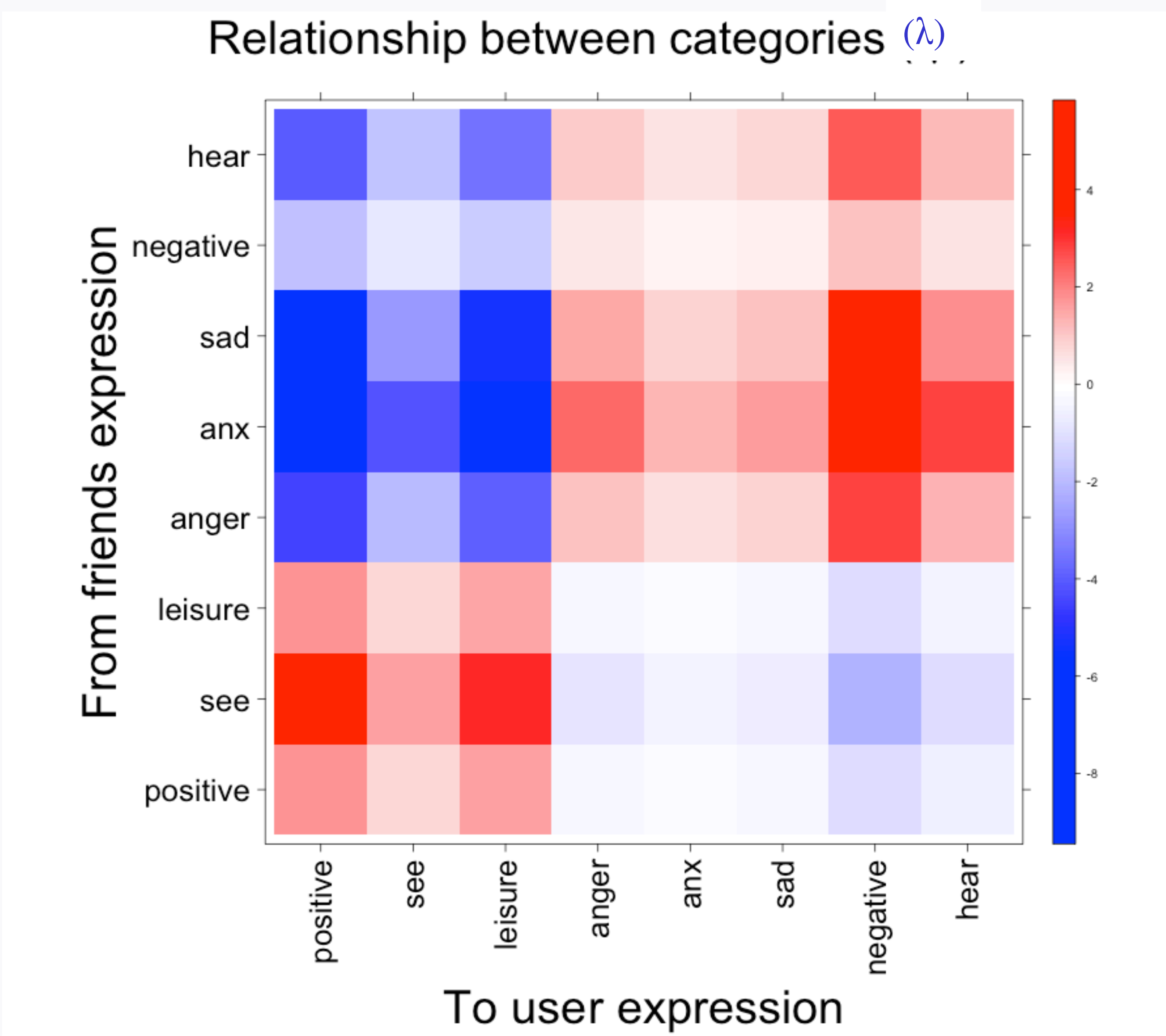
Results

Global emotional synchrony

Emotional contagion: We tend to mirror the semantic categories of our friends

Each post in a semantic category causes friends who live in other cities to make about 1 to 2 posts in the same category

Results



Relationship between categories (λ)
From friends expression
hear
negative
sad
anx
anger
leisure
see
positive
To user expression
positive
see
leisure
anger
anx
sad
negative
hear
4
2
0
-2
-4
-6
-8

Summary

The use of semantic expression spreads from person to person

Emotional contagion can be detected and measured in online social networks from observational data, using a non-invasive method

Even a weak instrument (rainfall) is sufficient for large data sets

Summary

Simple models of distributed computation can predict the performance of real populations solving computational problems over networks

Global dynamics of complex agents with possibly diverse strategies can be well described by simple synthetic agents with uniform strategies

Predicting epidemic risk

Predicting epidemic risk

The Challenge Dataset:

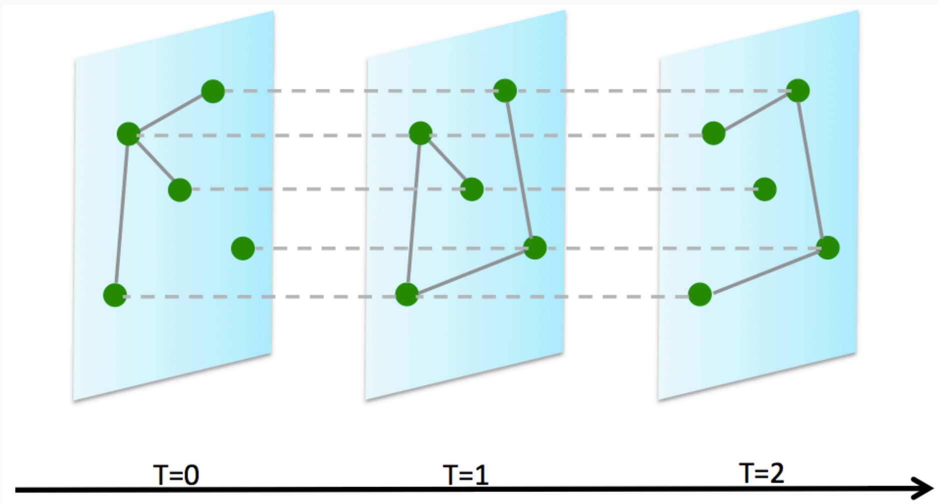
- **2.2M** reviews and **591K** tips by **552K** users for **77K** businesses
- **566K** business attributes, e.g., hours, parking availability, ambience.
- Social network of **552K** users for a total of **3.5M** social edges.
- Aggregated check-ins over time for each of the **77K** businesses
- **200,000** pictures from the included businesses

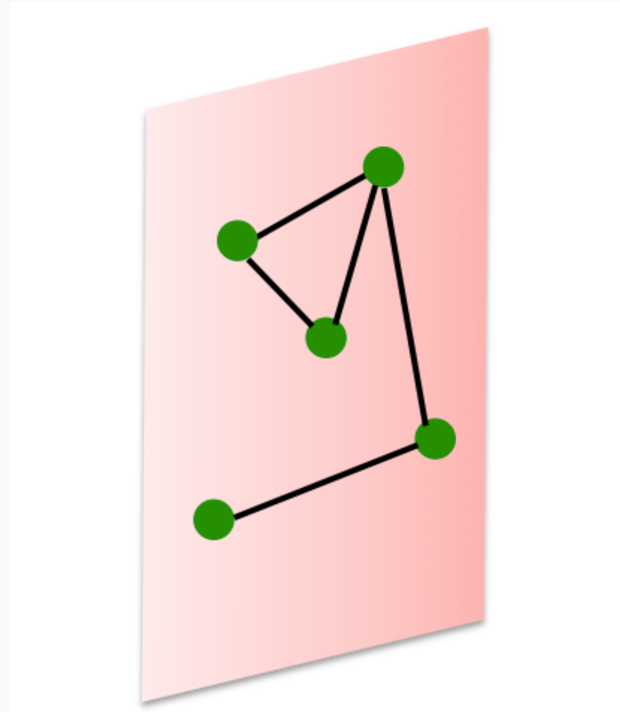
Get the Data

Cities:

- U.K.: Edinburgh
- Germany: Karlsruhe
- Canada: Montreal and Waterloo
- U.S.: Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison

Encounter Network vs. Friendship Network



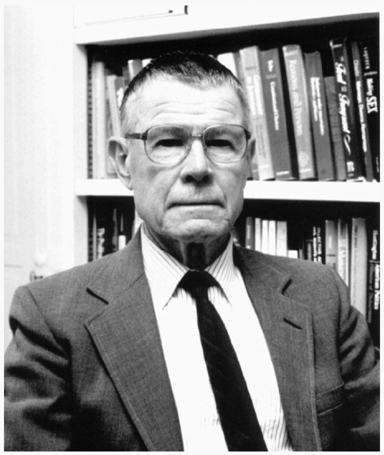


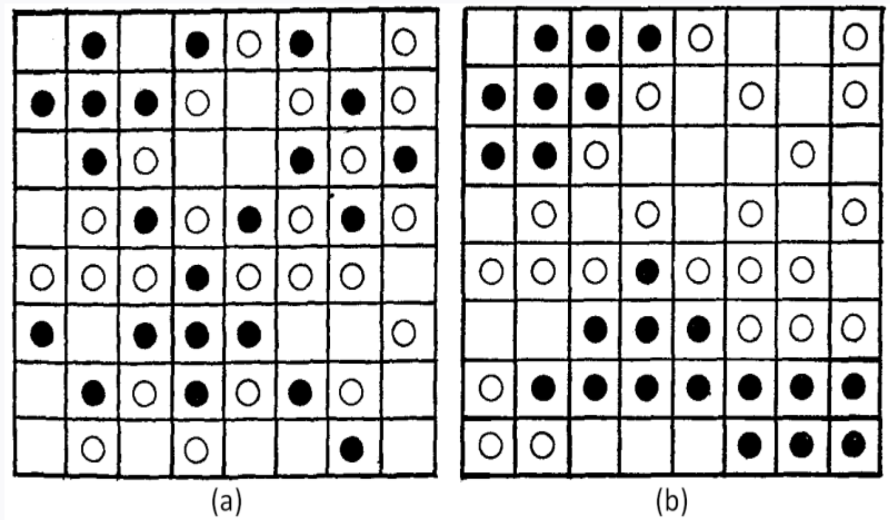
Predict risk of contagion

Contain epidemic spread

Using only knowledge of static friendship network

Residential segregation model





Thomas Schelling studied residential segregation in the US in the 70's using a simple probabilistic dynamical model

Dynamical system

Network: n by n torus

Agents: Type of agent is random iid Bernoulli: +1 or -1 spin

Neighborhood: Each agent considers the agents within Manhattan distance w as its "neighborhood"

Initialization: On each location of the grid there is an agent

State: If the fraction of agents in my neighborhood of my same type is larger than a threshold then I am happy.

Dynamics: Choose two unhappy agents of opposite type at each iteration and swap their locations if this makes both happy

Dynamical system

Based on paper simulation segregation occurs
even for high tolerance level

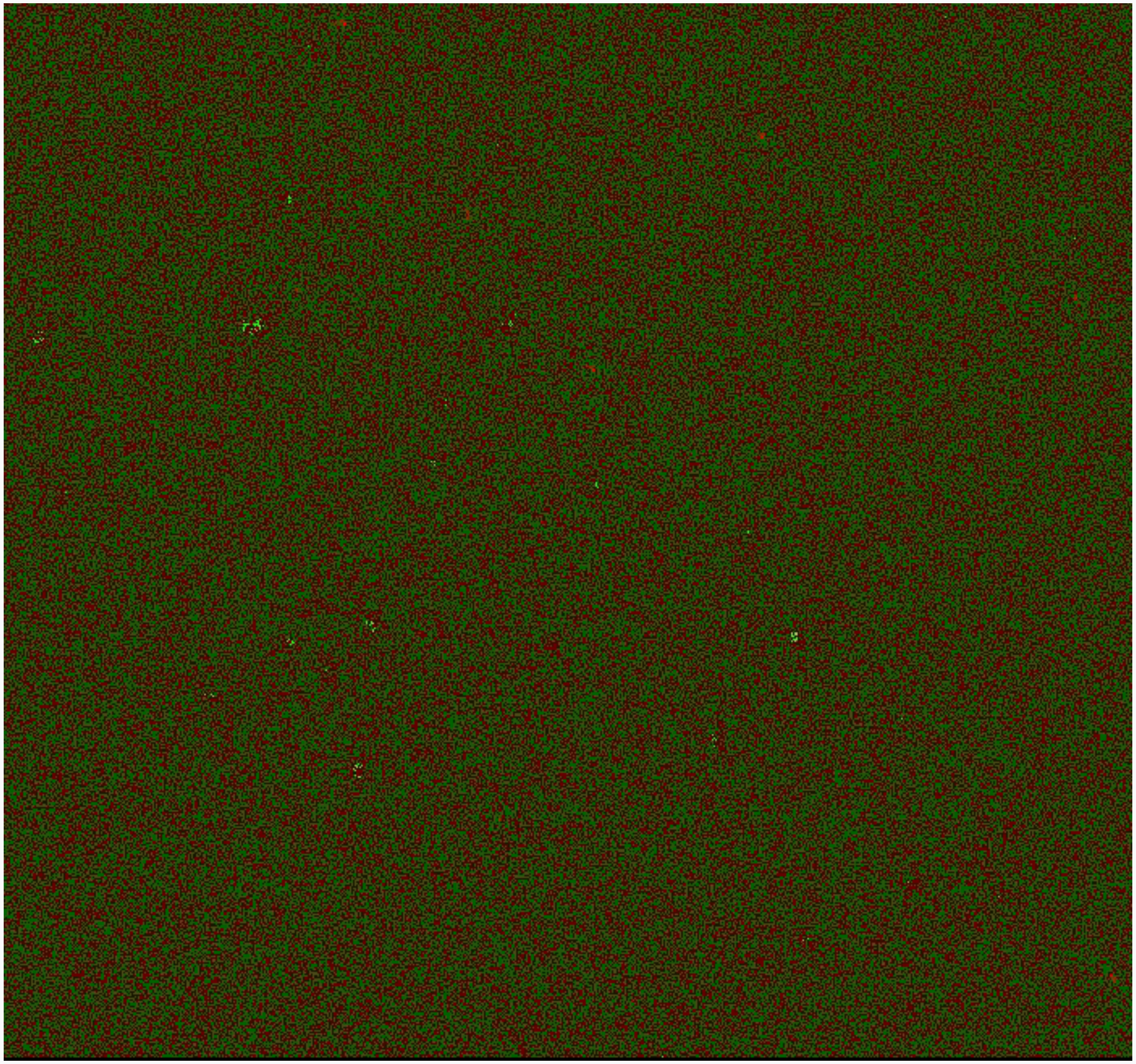
Local decisions can have global consequences

This simple model largely resisted rigorous analysis

Questions

- Under what conditions the system evolves into large segregated areas?
- How large will the segregated area be?
- How fast is the segregation process?
- How can we extend the model to more sophisticated settings?

Example [Hamed Omidvar]



Advocate for Aggregation not Segregation