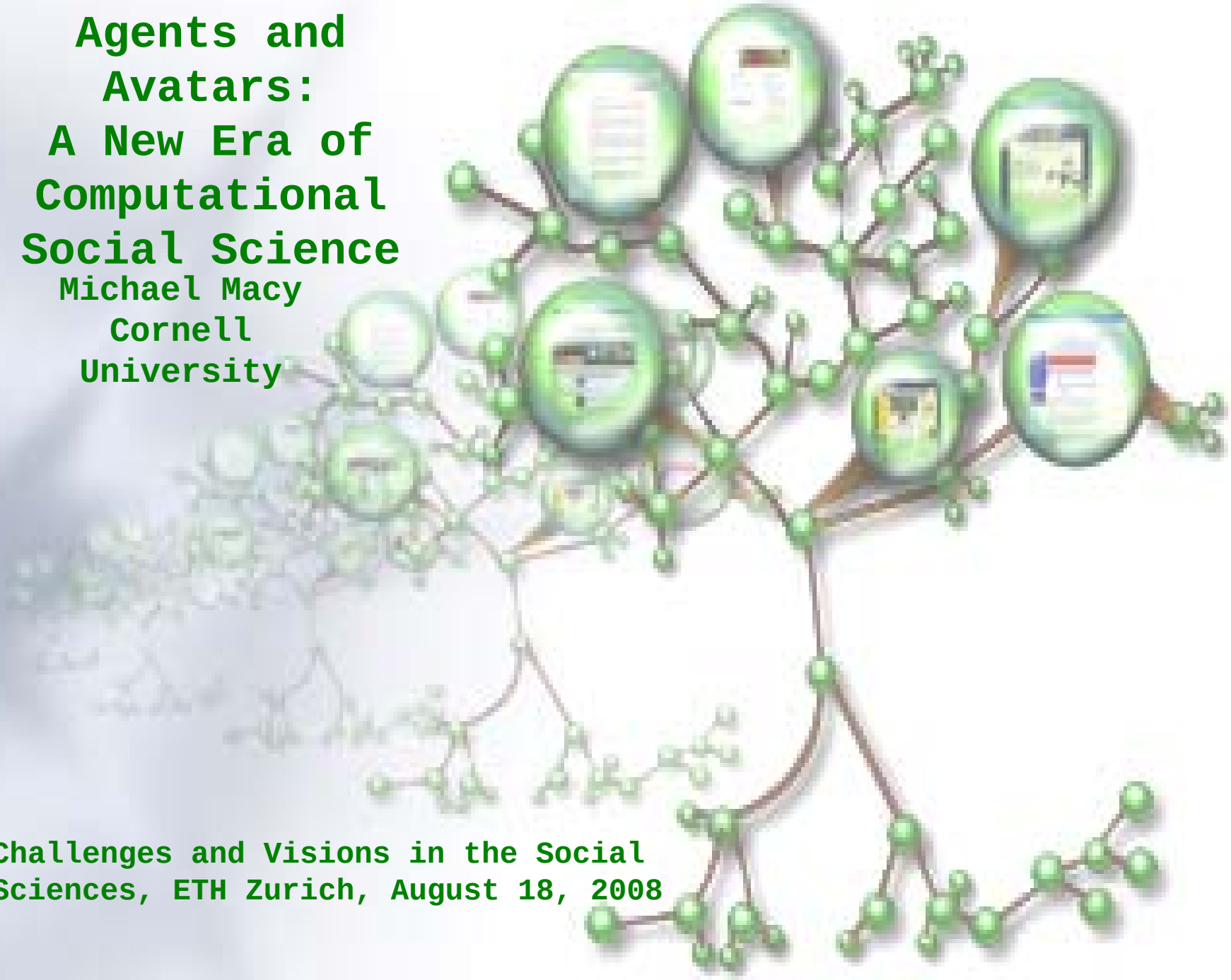
Agents and Avatars: A New Era of Computational Social Science

Michael Macy
Cornell University



Challenges and Visions in the Social Sciences, ETH Zurich, August 18, 2008

Acknowledgements

- Members of the Cornell cybertools team
- National Science Foundation
- Brewster Kahle, Internet Archive
- Tim Clark, DARPA
- Cornell Institute for the Social Sciences

- Social life is more like improv jazz than an orchestra
- How is this possible with
  - millions of players?
  - each aware only of local “neighbors”?

Outline

- Part 1: A Paradigm Shift in Social Science?
- Part 2: Research Projects
  - How does network structure affect social behavior?
  - What causes network ties to form and break?

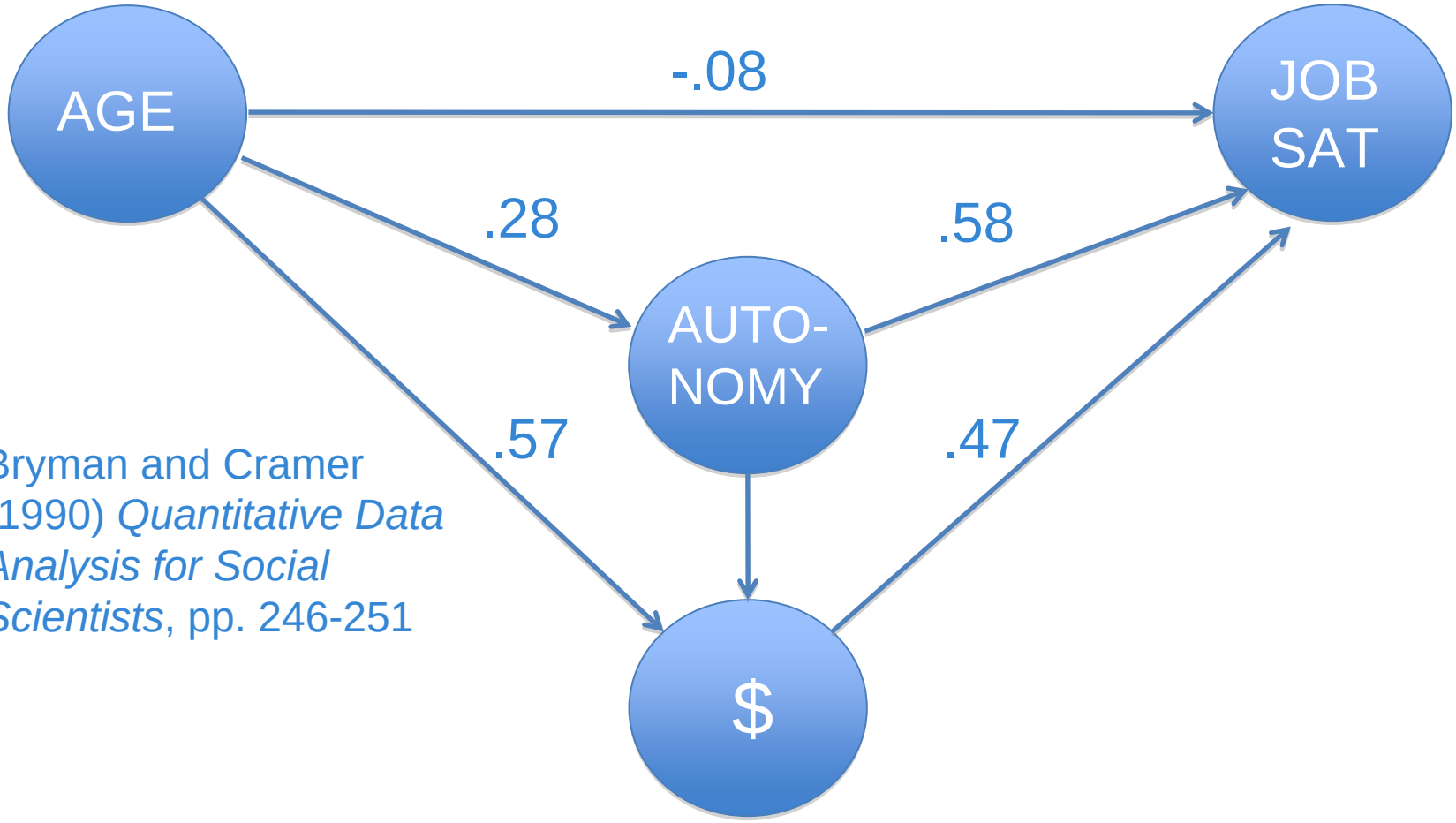
Part 1

How Digital Traces are Transforming Social Science

“Old School” Social Science

- Interactions among variables, not individuals
  - Correlations, no causal mechanisms
  - Population of individuals – no local structure
  - Observations are independent – no influence
- Problem is conceptual, not methodological
  - "These regression equations are the 'laws' of a science” (Blalock 1960:275)
  - Abbott: “General Linear Reality”

Variables Interact, Not People



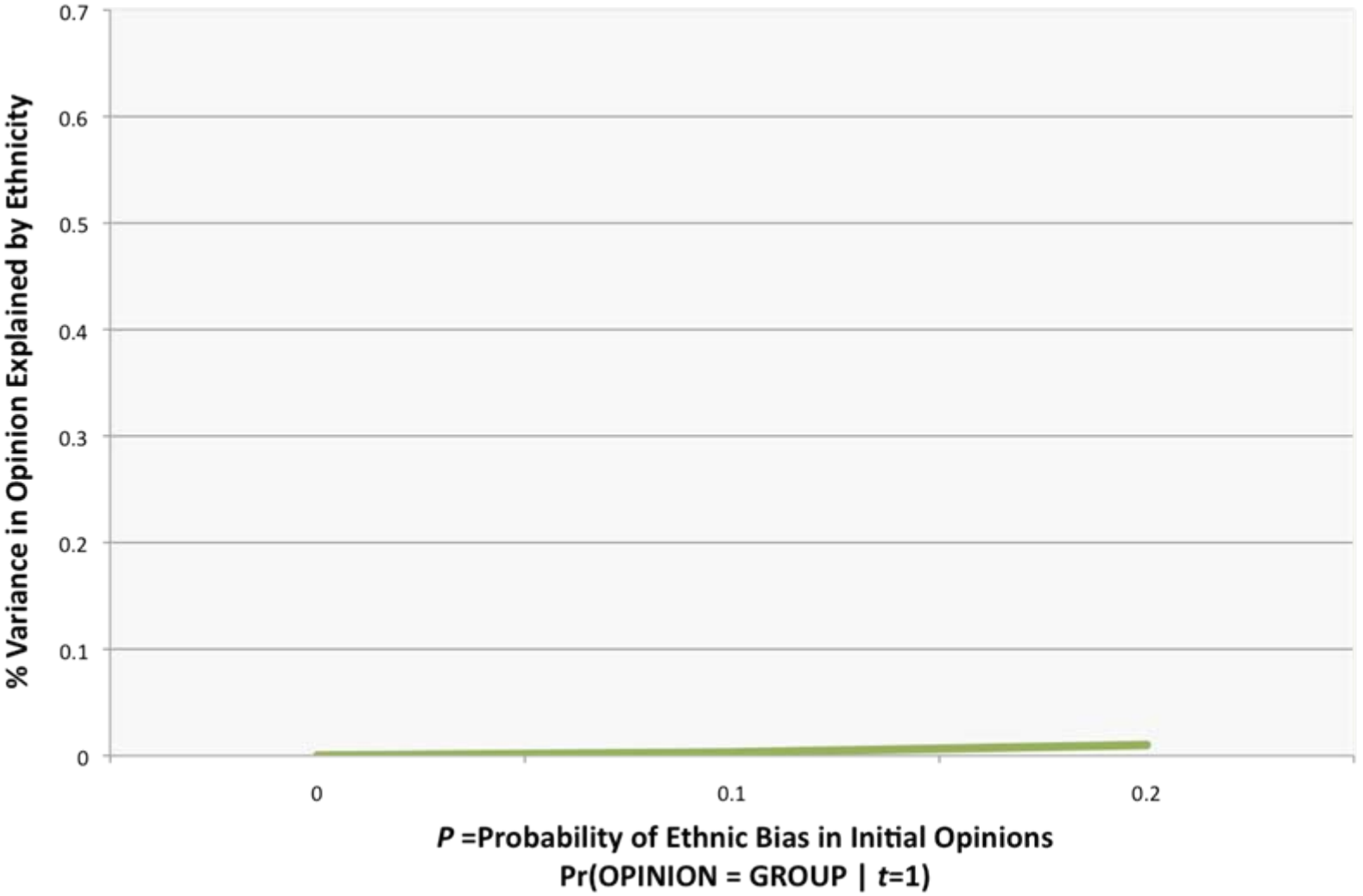
Bryman and Cramer (1990) *Quantitative Data Analysis for Social Scientists*, pp. 246-251

A Hopfield Model of Opinion Dynamics*

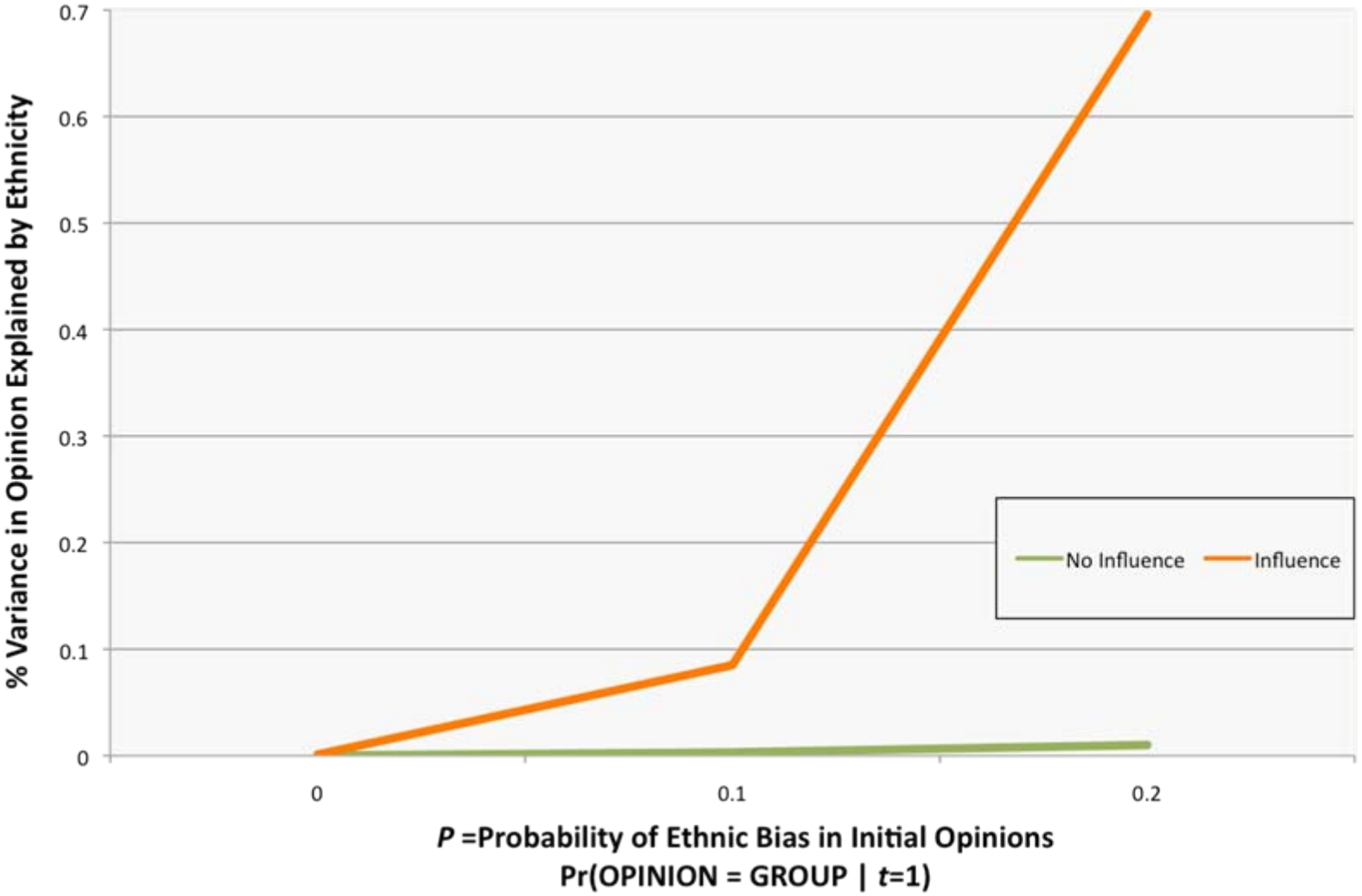
- Agents belong to one of two ethnic groups
- All opinions are binary (for or against)
- Homophily + Xenophobia
- Two experimental manipulations:
  1. Attraction increases influence vs. no influence
  2. Random start, except ethnicity biases initial opinions with probability p = [0, 0.1, 0.2]
- Run until equilibrium (population polarizes)
- Compare OLS estimates with “ground truth” across conditions.

*with Andreas Flache (Groningen)

True Results When Influence is Precluded



Spurious Results When Influence is Allowed



From Factors to Actors

- What we really want to know about are the interactions among people, not variables.
  - How we influence network neighbors in response to local influence received
  - How network structure affects propagation of influence
  - How attributes of actors affect tie formation

Problem: Social Life is Hard to See

- You can interview friends, but you cannot interview a friendship.
  - Fleeting interaction
  - In private
  - Tedious to record over time, especially in large groups

Why This is Changing

- Humans increasingly interact online
  - Publicly in Web pages, Facebook, blogs, news groups, wikis, MMOGs, Second Life, eBay, epinions, LastFM, flickr
  - Privately in email, SMS, Facebook, phones
- Computer-mediated interaction leaves digital traces

New Era of Computational Social Science

- How local micro interaction generates macrosocial patterns
  - Agent-based computational models
    - Relations among actors, not attributes
    - Out of equilibrium dynamics
    - Network evolves as it constrains
  - On-line controlled X-cultural experiments
  - Digital traces of on-line interactions

Cover Story in *Science*

"Online virtual worlds, electronic environments where people can work and interact in a somewhat realistic manner, have great potential as sites for research in the social, behavioral, and economic sciences … Second Life and World of Warcraft … foreshadow future developments, introducing a number of research methodologies that scientists are now exploring, including formal experimentation, observational ethnography, and quantitative analysis of economic markets or social networks."

--William Sims Bainbridge
"The Scientific Research Potential of Virtual Worlds"
Science, 27 July 2007

Cover Story in *Science*

"Online **virtual worlds**, electronic environments where people can work and interact in a somewhat realistic manner, **have great potential as sites for research in the social, behavioral, and economic sciences** … Second Life and World of Warcraft … foreshadow future developments, introducing a number of research methodologies that scientists are now exploring, including formal experimentation, observational ethnography, and quantitative analysis of economic markets or social networks."

--William Sims Bainbridge
"The Scientific Research Potential of Virtual Worlds"
Science, 27 July 2007

But is this the "Real World"?

- When humans interact via digital devices, do they enter a parallel universe?
- Historically unique
  - User-generated identities (aka "self-presentation")
  - Few spatial, cultural, or legal constraints on interaction (may be revealing as well as concealing)
  - Some community members are silicon life forms
  - Limited demographic data

The Web is All Too Real

- Three incompatible goals:
  1. Automatic data collection of large-scale computer-mediated human social interactions
  2. Widest possible access to these data by diverse research teams
  3. Protection of privacy

The Web as a Record of Social Interaction

- The spread of hoaxes, rumors, urban legends
- Diffusion of innovation (e.g. free/paid hotel wifi)
- Movement of personnel among organizations
- Political campaigns
- Status hierarchies (who links/refers to whom)
- Opinion dynamics in blogs, newsgroups, product ratings

The Internet Archive

- 10 years of bi-monthly crawls
  - Approaching 2 PB (compressed)
  - Page content plus metadata (format, links, anchor text, file type (pdf, jpg, flash)
- NSF: $2M Cybertools Grant
  - To copy the Archive to Cornell
  - Reconfigure it as a searchable database

ETH Website 2 March 1997

ETH Eidgenössische Technische Hochschule Zürich

The Swiss Federal Institute of Technology at Zürich, ETH-Zentrum, CH-8092 Zürich, Switzerland

---

Welcome to the WWW Server of the ETH Zürich !

Diese Seite gibt es auch in Deutsch

The ETH Zürich:

- An Overview, the Profile ("ETH-Leitbild")
- Administration and Operation
- ETHZ BULLETIN, the publication of the ETH Zürich
- Maps, lists of buildings, emergency phone numbers

- The Departments do research, see also Research Report
- The Faculties are responsible for teaching, in addition: Center for Continuing Education und Center for Teaching and Learning

- ETH library: (ETHICS and 3270 emulation
- Services at the ETH Zürich, ETH-related and external Information Services

Part 2

New Research Opportunities
(coming attractions)

I. How Do Ties Matter?

- How does “six degrees of separation” affect social influence?
- Is influence more effective in densely clustered neighborhoods?
- How important are strong vs. weak ties?

A Chance Encoun ads to Small Talk…

Six Degrees of Separation

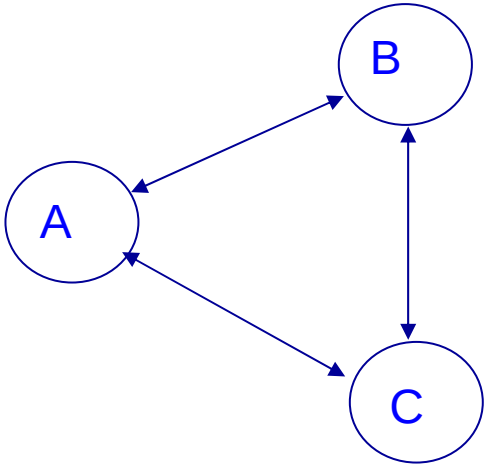
The planet is very large: 6.5*b*!

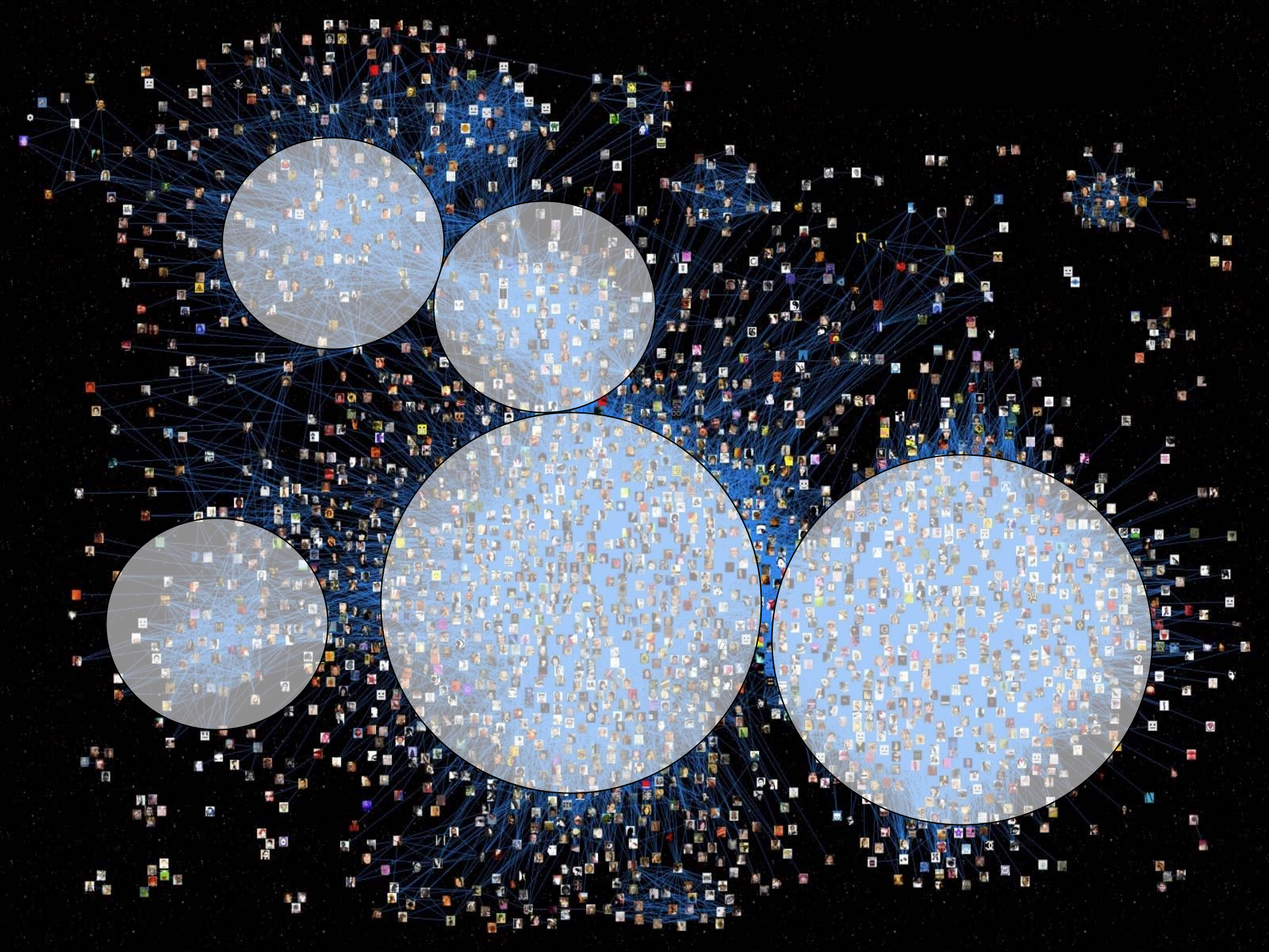
Yet the world is small: 6°

How is this possible?

Adding to the Mystery…

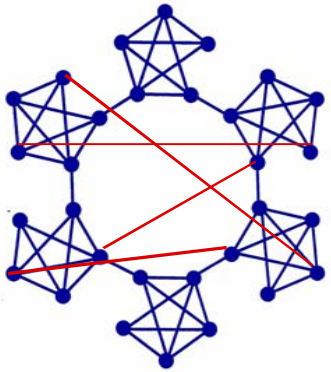
- Easy to explain if the social ties were random
- But friendships tend to be highly clustered





Solved by Watts & Strogatz

- A few long-range ties
  - Create “shortcuts” between otherwise distant nodes
  - While preserving the clustering of a social network

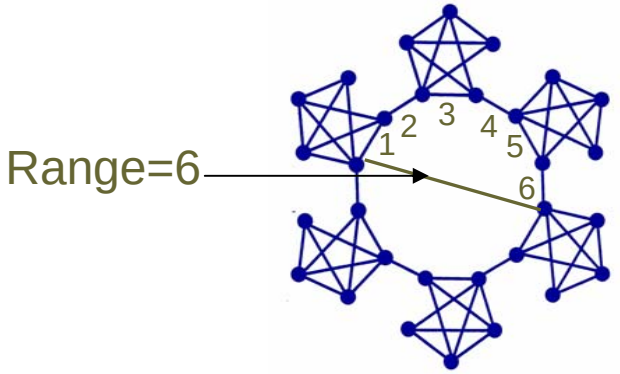


Granovetter: Strength of Weak Ties

- Social ties vary in strength
  - Frequency of interaction
  - Trust, commitment, attachment
- Strong ties form short cycles
- Strength of weak ties is their range

Does Tie Strength Decline with Range?*

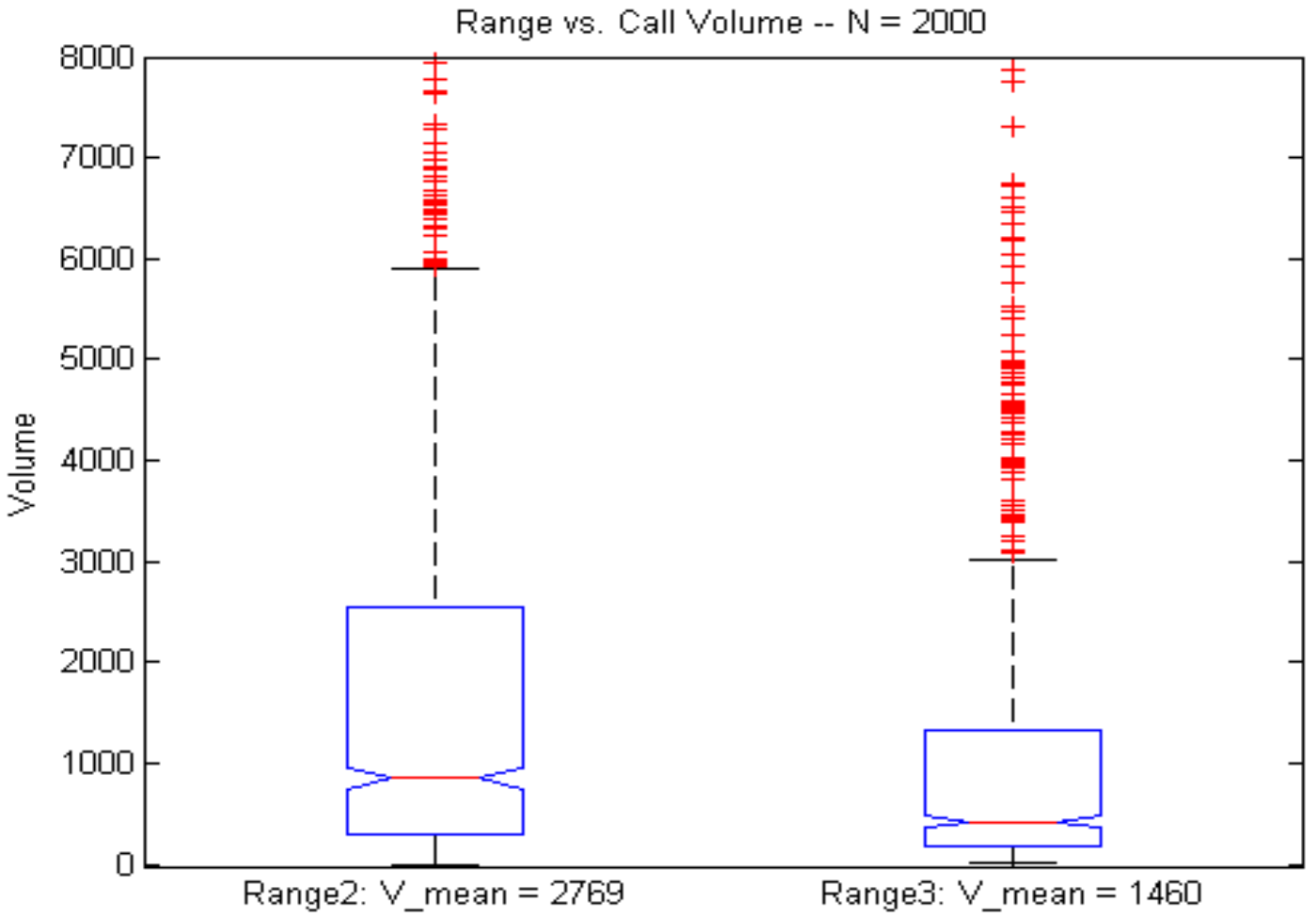
- Range is the path length traversed



- Tie strength
  - Number of emails, phone calls
  - Email lag times
  - Duration of phone calls

*with Gueorgi Kossinets (Cornell), Nathan Eagle (MIT/SFI)

Call Volume Declines with Range



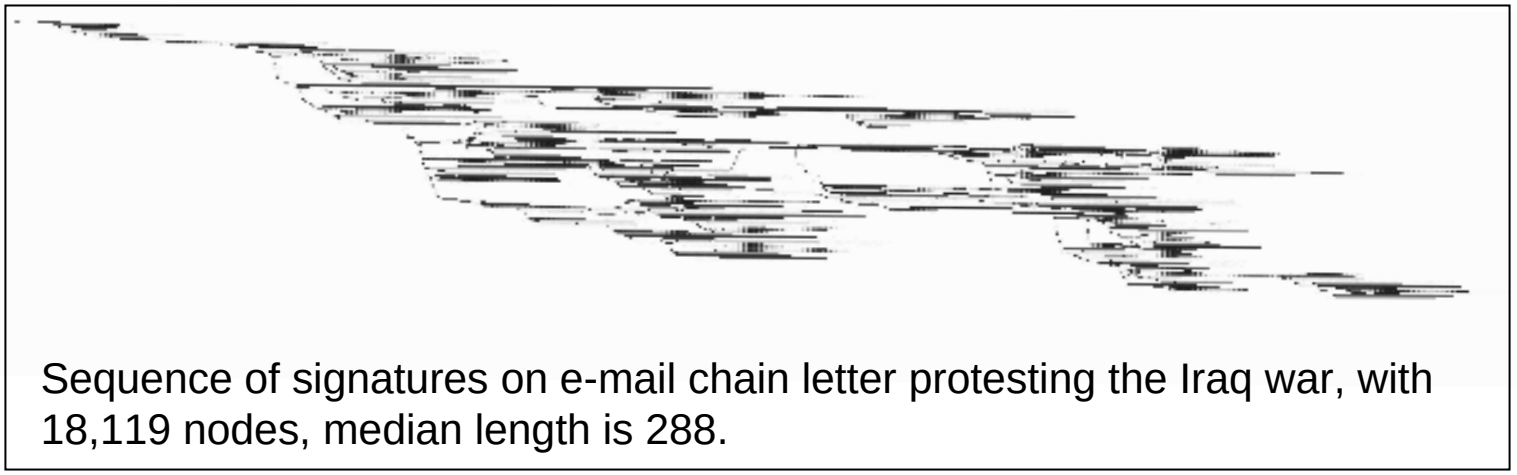
Range Compensates for Weakness

"Whatever is to be diffused can reach a larger number of people, and traverse a greater social distance, when passed through weak ties rather than strong." -- Mark Granovetter, 1973

- Access/exposure to new ideas and information
- A truism across the social & information sciences
- But there are some intriguing anomalies...

The Chain-Letter Paradox*

If most people are separated by only six degrees, why are e-mail chain letters hundreds of links long?



Sequence of signatures on e-mail chain letter protesting the Iraq war, with 18,119 nodes, median length is 288.

*Liben-Nowell & Kleinberg 2008, "Tracing information flow on a global scale using Internet chain-letter data," *PNAS* 105:4633-38.

The Triangle Paradox

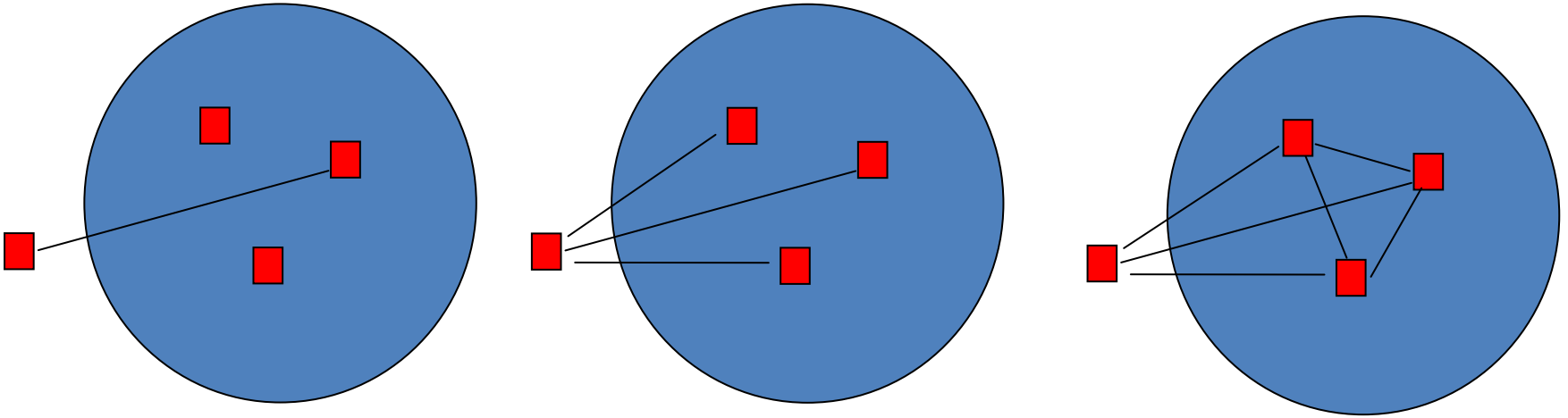
- McAdams: People recruit their friends to social movements and communities
- Well documented that the probability of recruitment increase with the number of friends who are already members.
- A new study discovered something odd…

Friendship and Diffusion*

- All BT residential landlines in the UK
- Individuals one degree removed (non-adopters with at least one adopter tie)
- What is the probability of adopting BT Voice Mail as a function of
  - Number of friends who already adopted
  - Clustering among friends

*With Nathan Eagle (MIT, SFI).

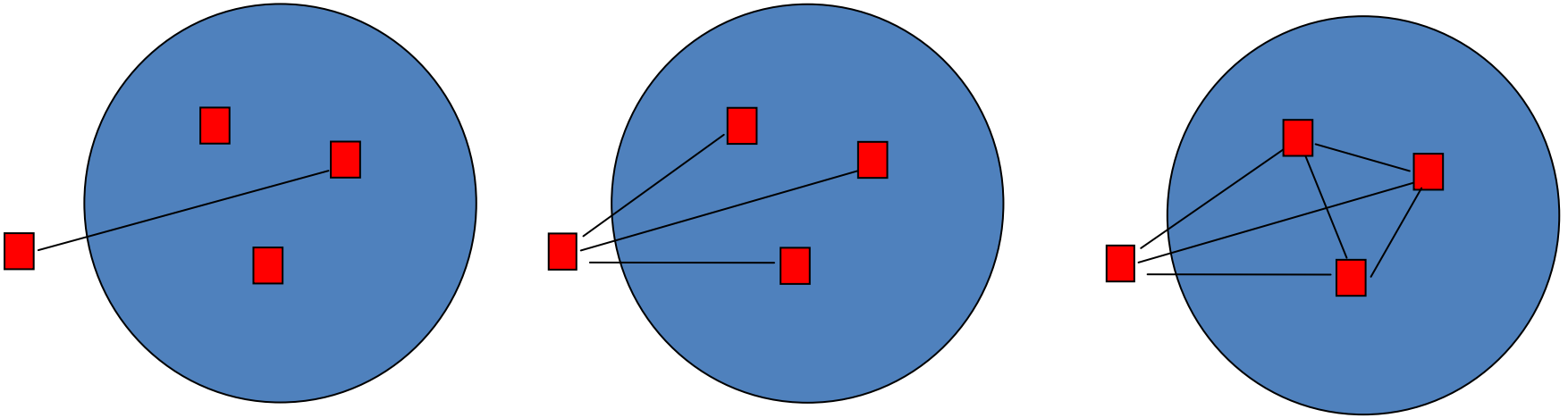
Number and clustering of friends



A B C

Time 1

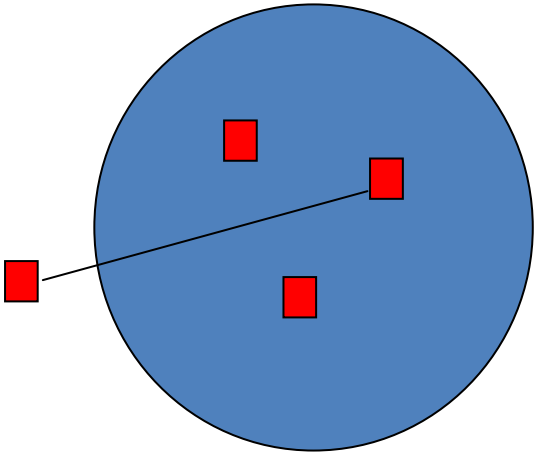
Number and clustering of friends



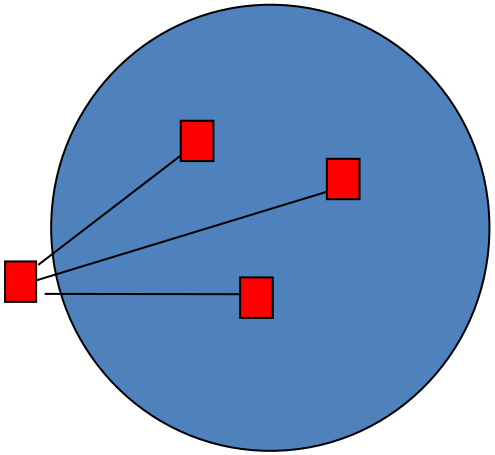
A B C

Time 2

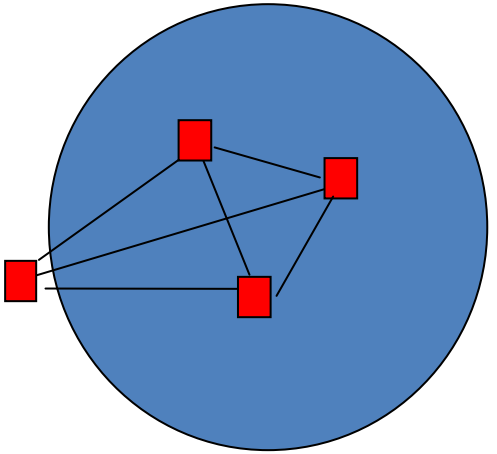
Number and clustering of friends



A



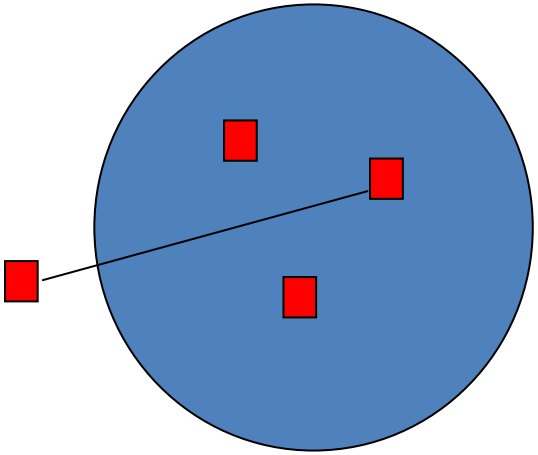
B



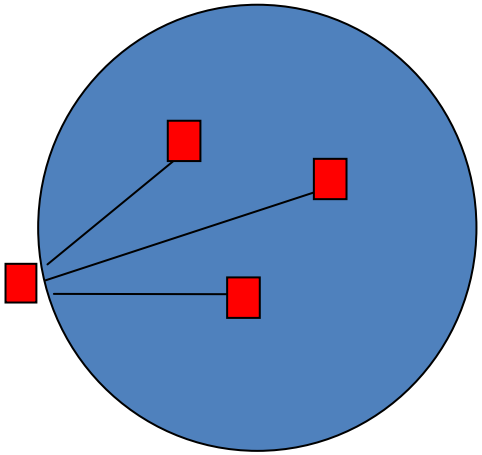
C

Time 3

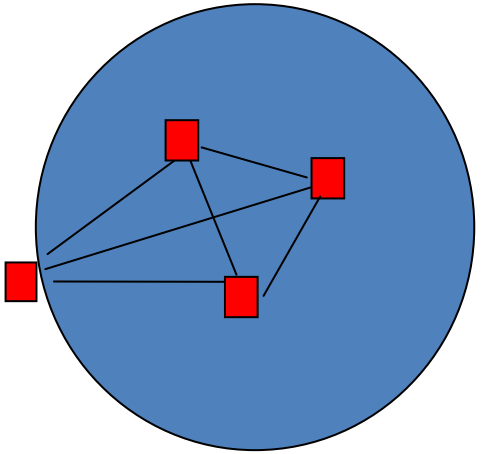
Number and clustering of friends



A



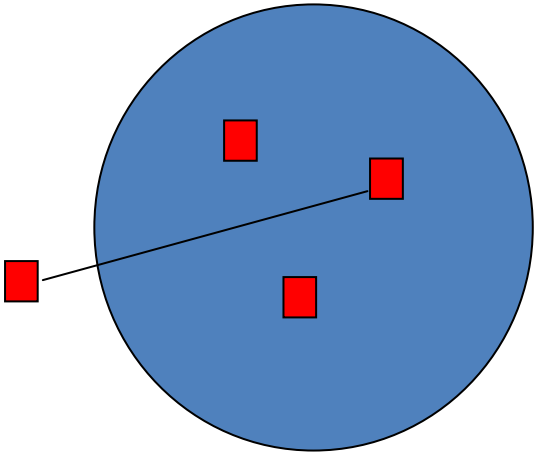
B



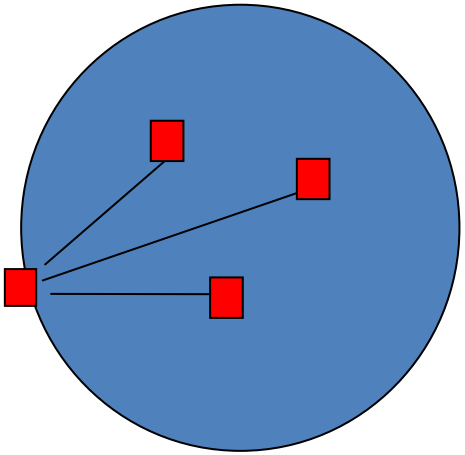
C

Time 4

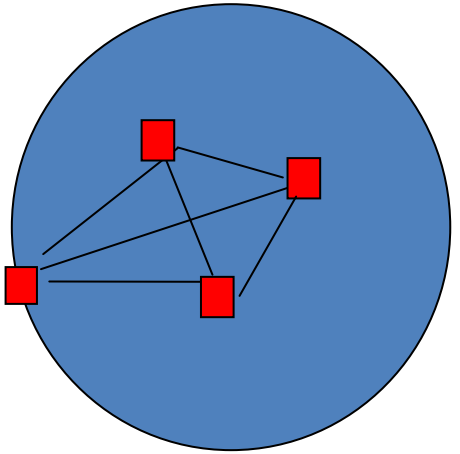
Number and clustering of friends



A



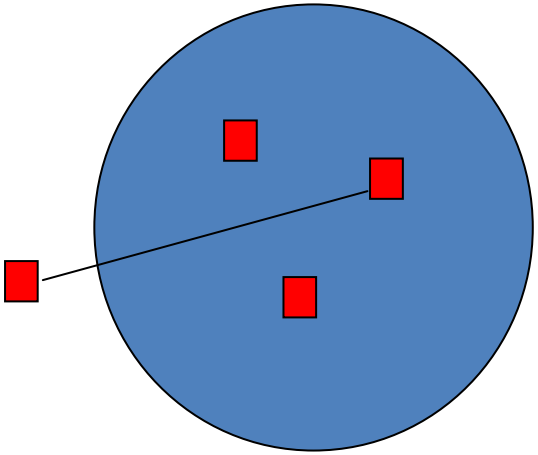
B



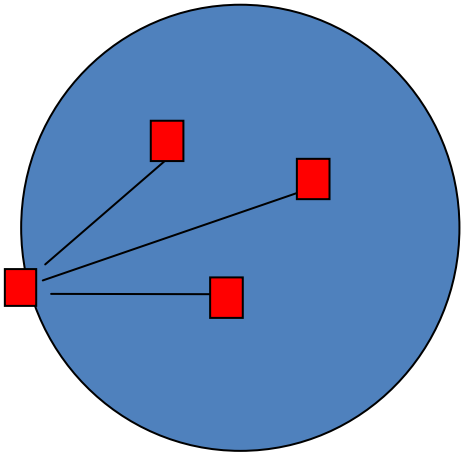
C

Time 5

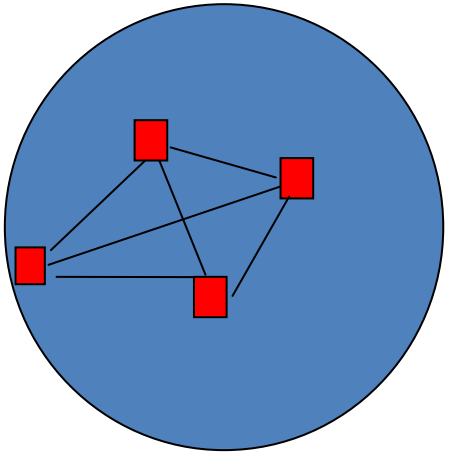
Number and clustering of friends



A



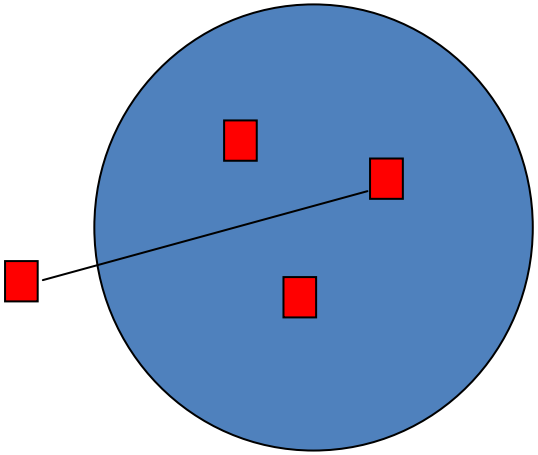
B



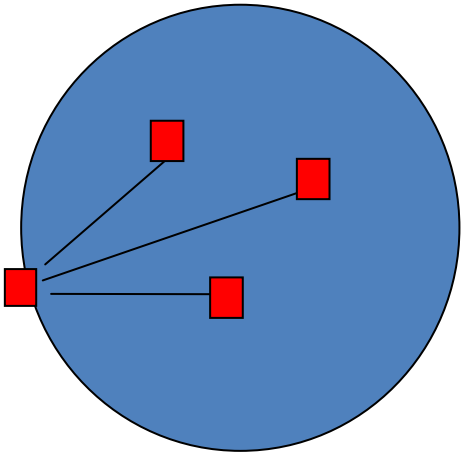
C

Time 6

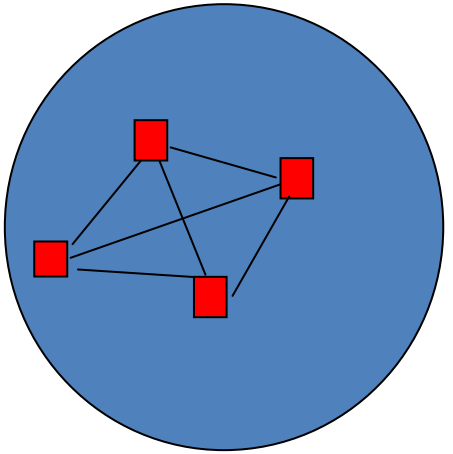
Number and clustering of friends



A



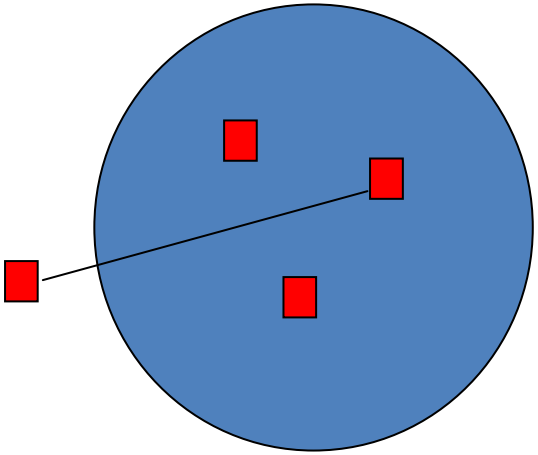
B



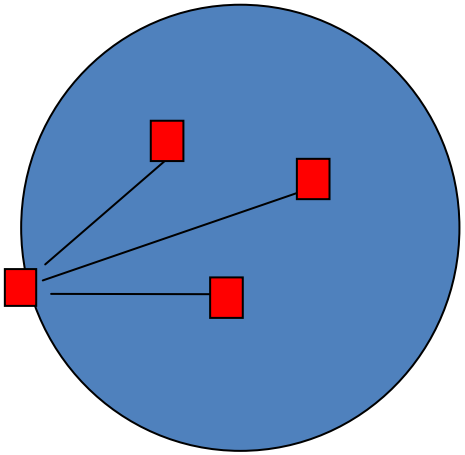
C

Time 7

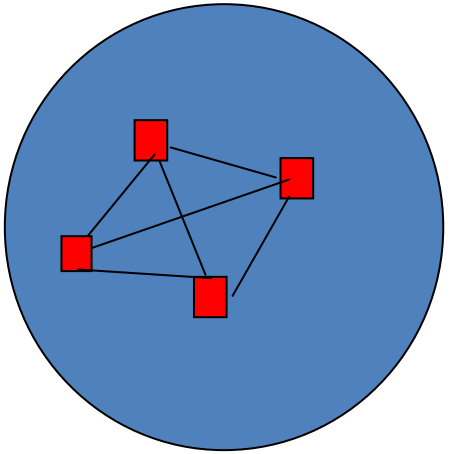
Number and clustering of friends



A

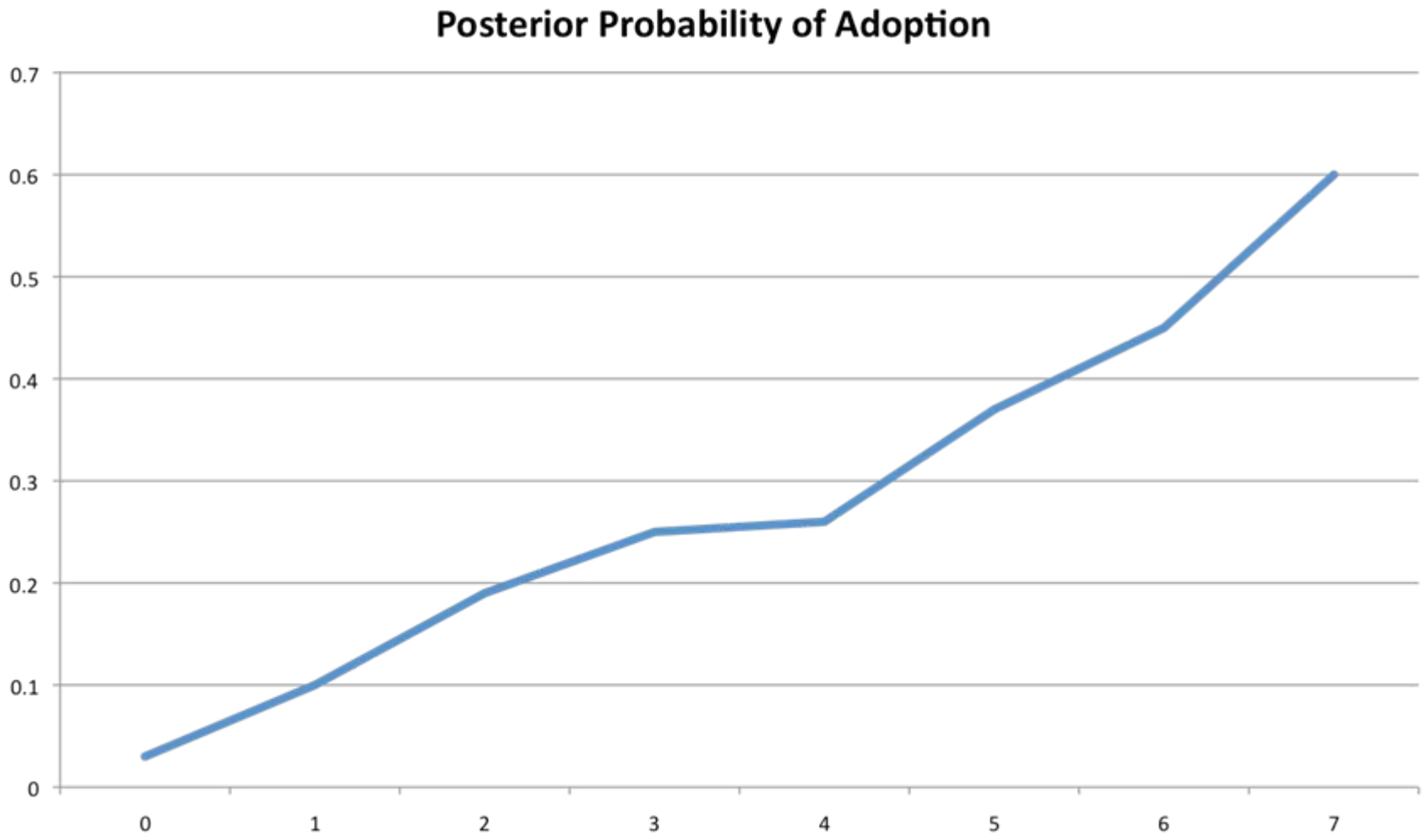


B



C

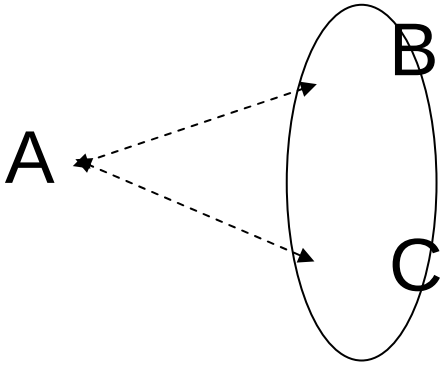
Time 8

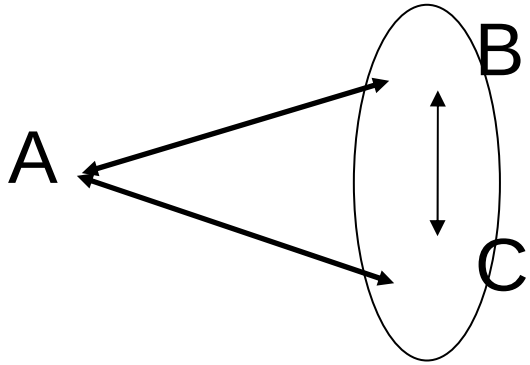


Number of Triangles among those with 7 Adopter Ties

Why Do Triangles Matter?

- A proxy for tie strength?





- Or effect of coordinated influence?

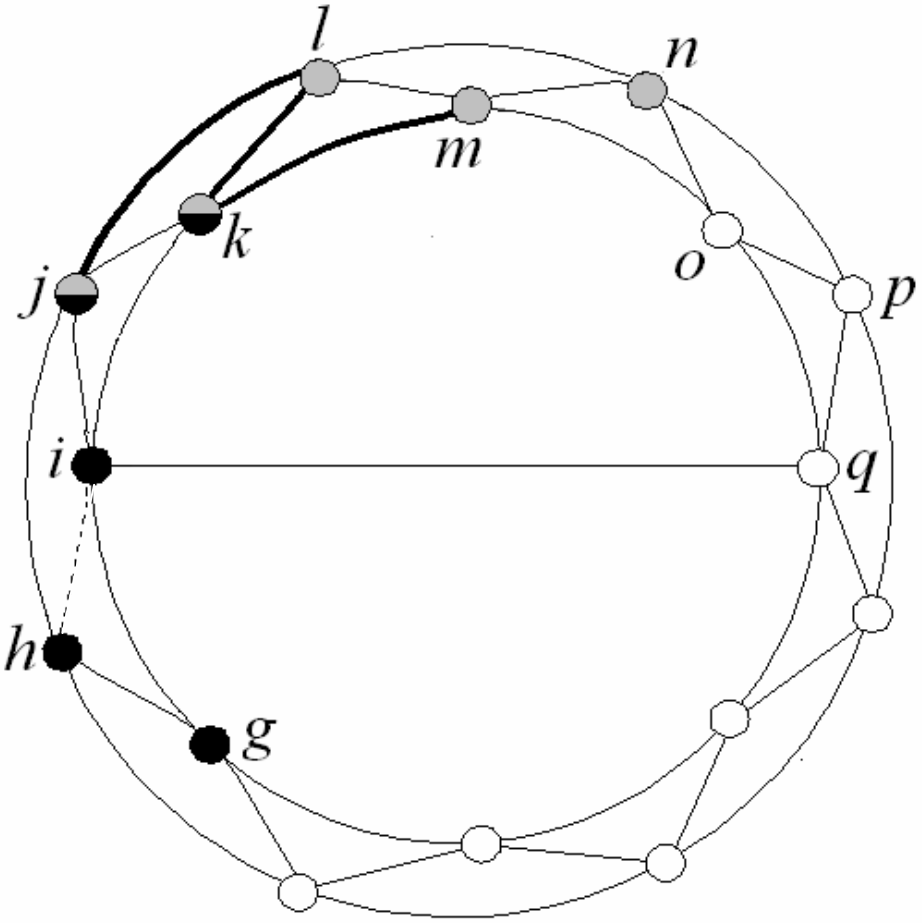
Why is Clustering Important?

- Why do chain-letters seem to avoid taking "shortcuts" across the network?
- Why is it the mutual friends that diffuse an innovation?

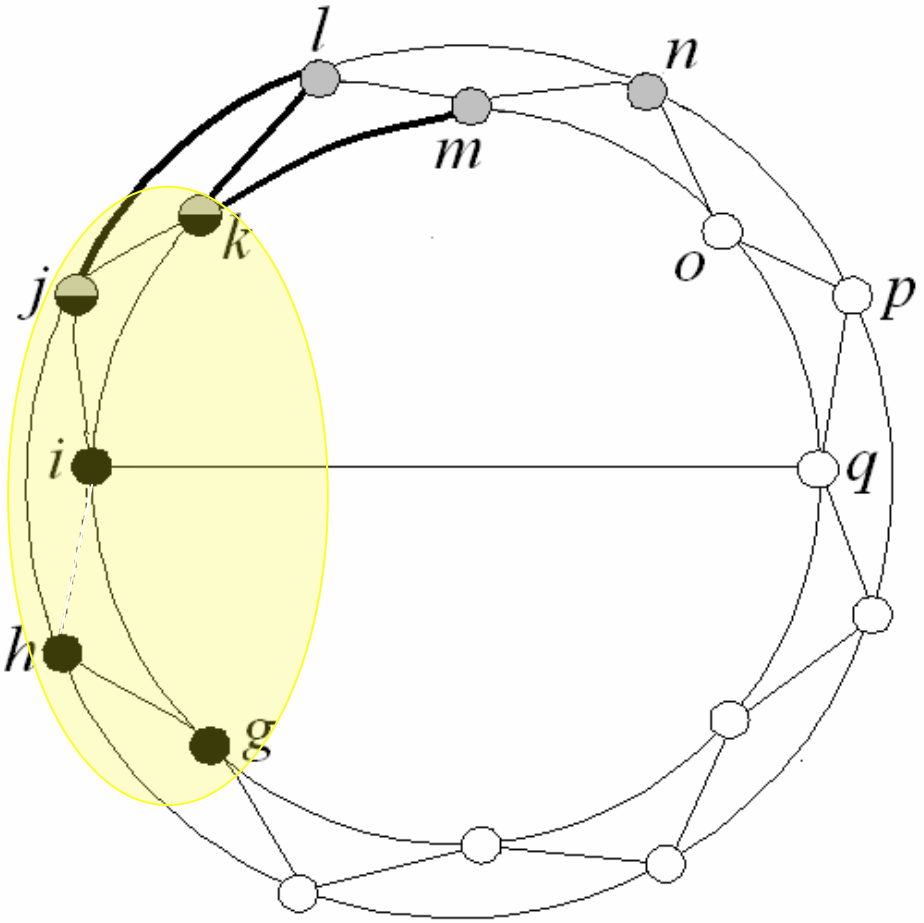
A Simple Explanation*

- Long-range ties inform but do not persuade
  - Acquiring information ≠ acting on it
  - Credibility, legitimacy increase with prior adopters
    - The same information from two friends is redundant
    - The same advice from two friends is not
- Maybe it's not such a small world after all?

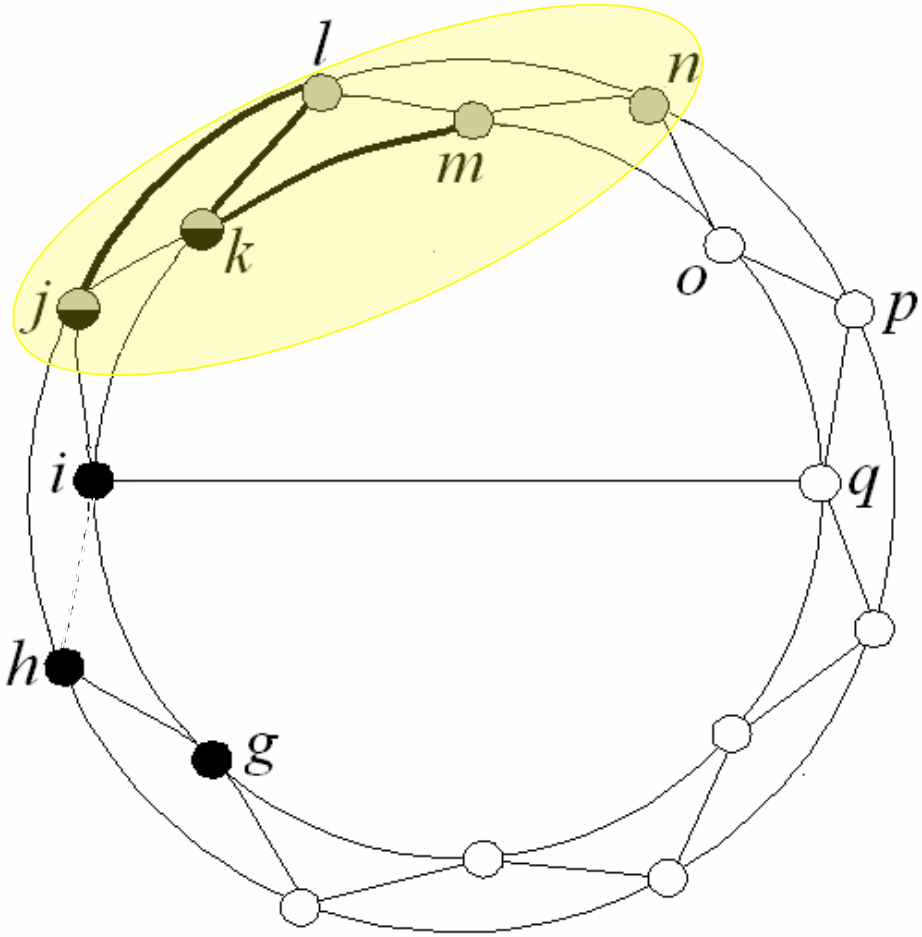
*Centola, D. and M. Macy. 2007. "Complex Contagions & the Weakness of Long Ties." *American Journal of Sociology* 113:702-34



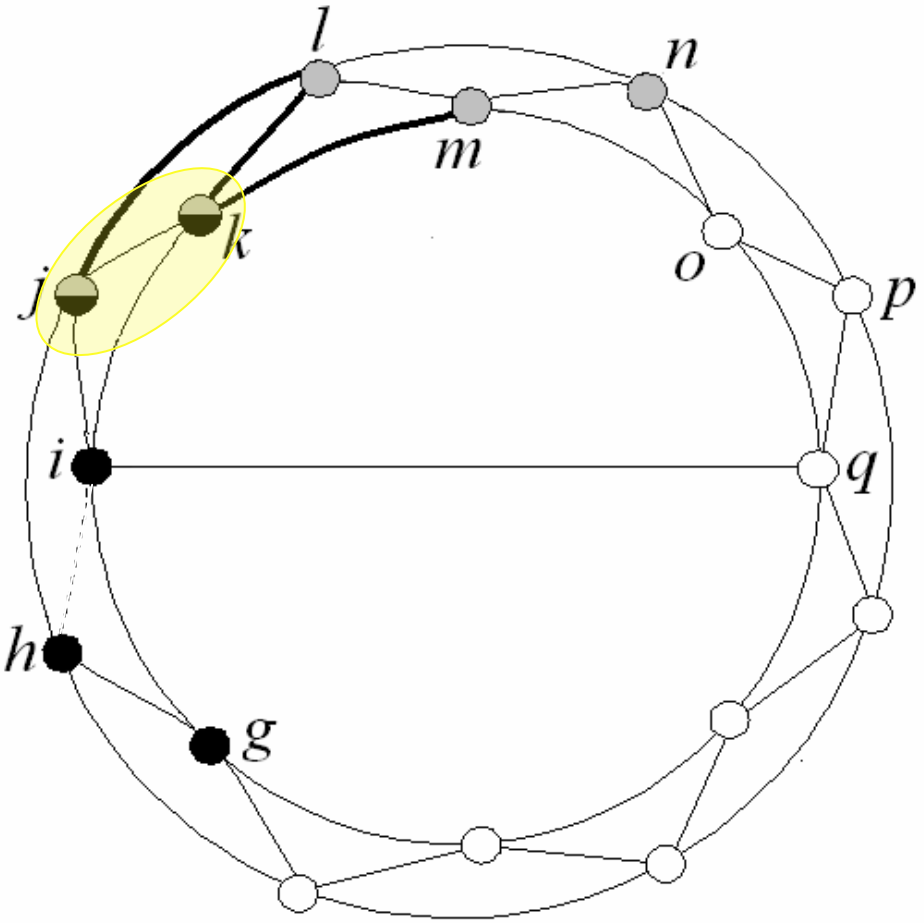
Illustrates the width of the bridge between the neighborhoods of ***i*** (black and gray/black nodes) and ***l*** (gray and gray/black nodes), showing the two common members (gray/black nodes). The bridge between these two neighborhoods consists of the three ties ***jl***, ***kl***, and ***km*** (shown as bold lines).



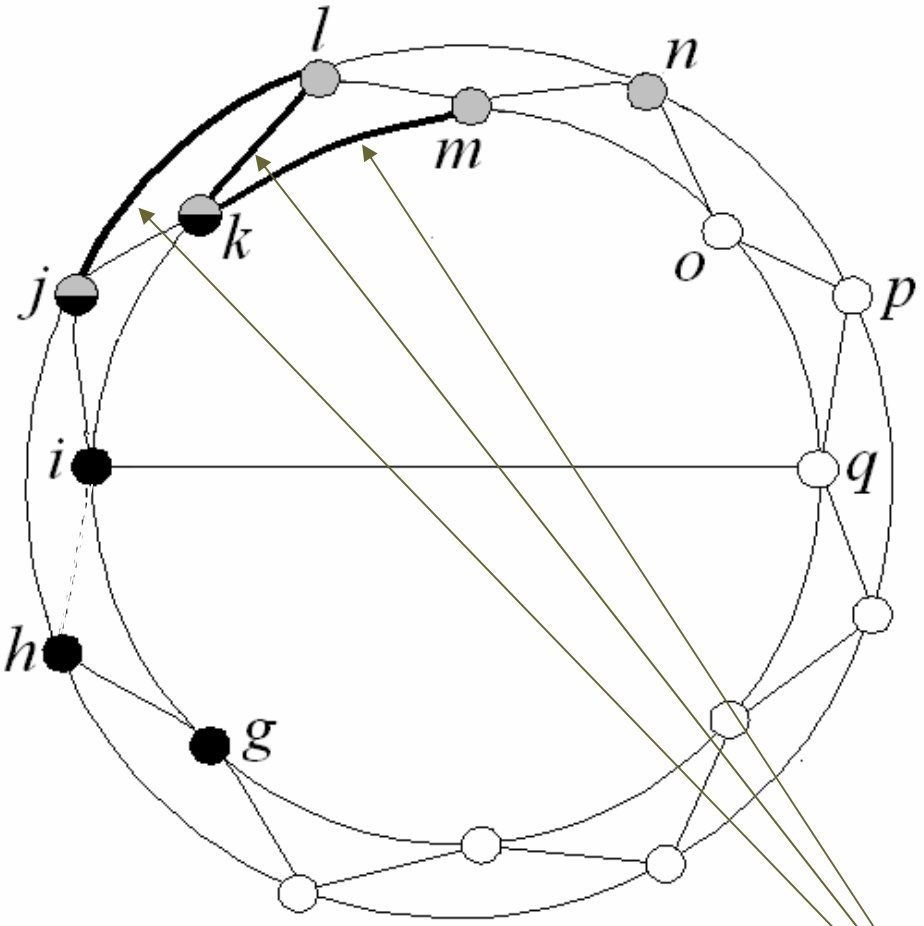
Illustrates the width of the bridge between the neighborhoods of ***i*** (black and gray/black nodes) and ***l*** (gray and gray/black nodes), showing the two common members (gray/black nodes). The bridge between these two neighborhoods consists of the three ties ***jl***, ***kl***, and ***km*** (shown as bold lines).



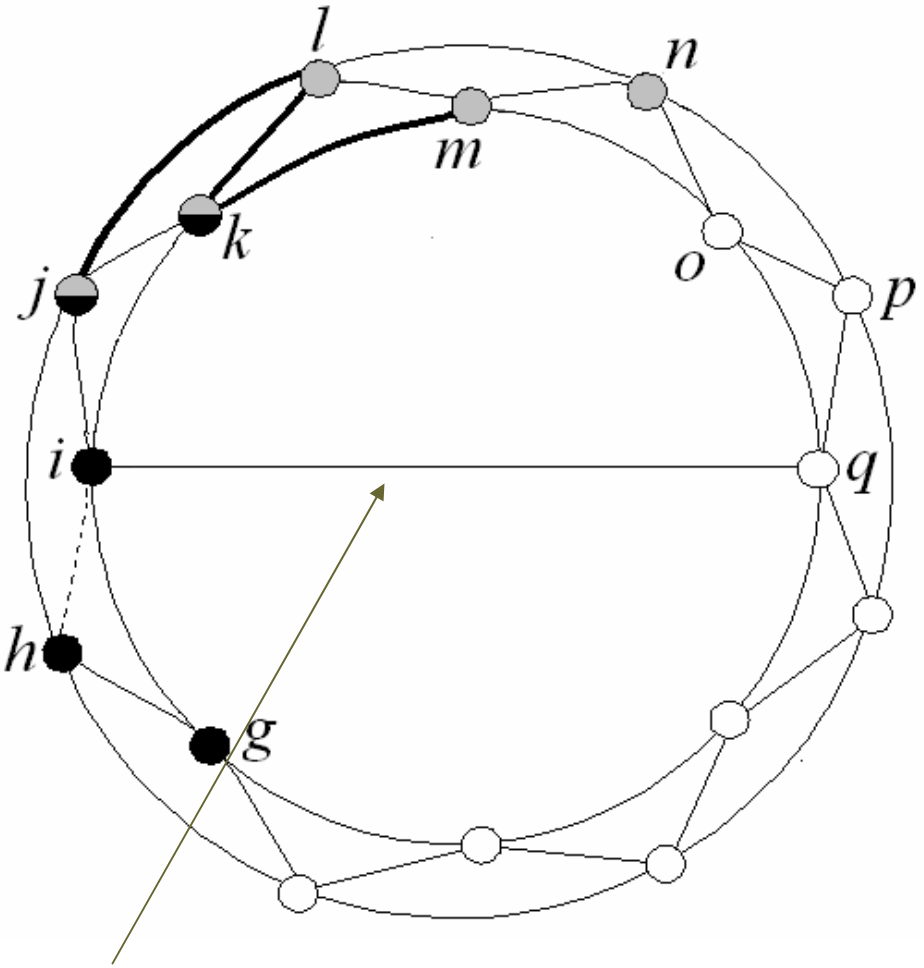
Illustrates the width of the bridge between the neighborhoods of ***i*** (black and gray/black nodes) and ***l*** (gray and gray/black nodes), showing the two common members (gray/black nodes). The bridge between these two neighborhoods consists of the three ties ***jl***, ***kl***, and ***km*** (shown as bold lines).



Illustrates the width of the bridge between the neighborhoods of ***i*** (black and gray/black nodes) and ***I*** (gray and gray/black nodes), showing the two common members (gray/black nodes). The bridge between these two neighborhoods consists of the three ties ***jl***, ***kl***, and ***km*** (shown as bold lines).



Illustrates the width of the bridge between the neighborhoods of ***i*** (black and gray/black nodes) and ***I*** (gray and gray/black nodes), showing the two common members (gray/black nodes). The bridge between these two neighborhoods consists of the three ties ***jl***, ***kl***, and ***km*** (shown as bold lines).

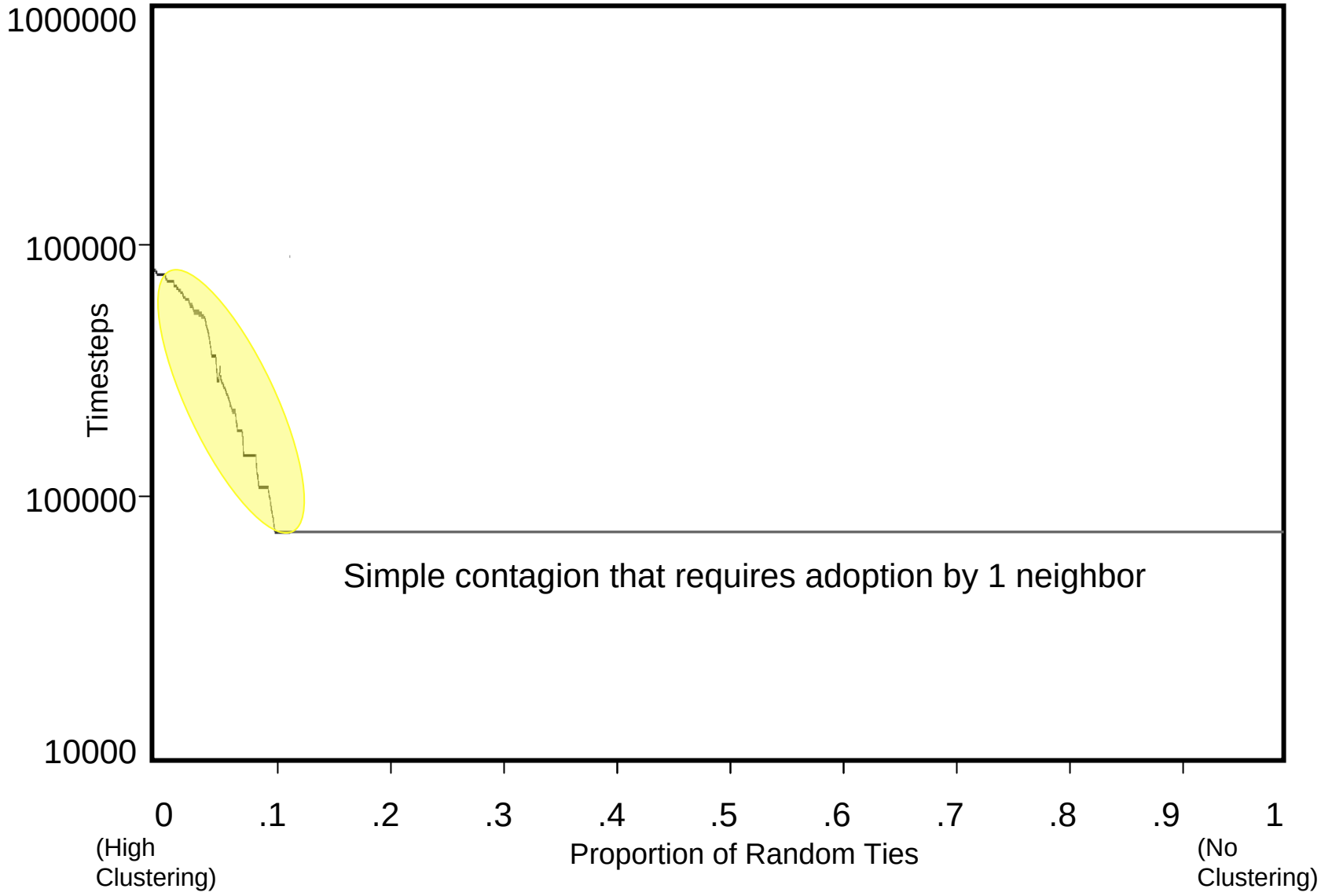


The long range tie from ***i*** to ***q*** provides a shortcut for information or disease but not the spread of social contagions.

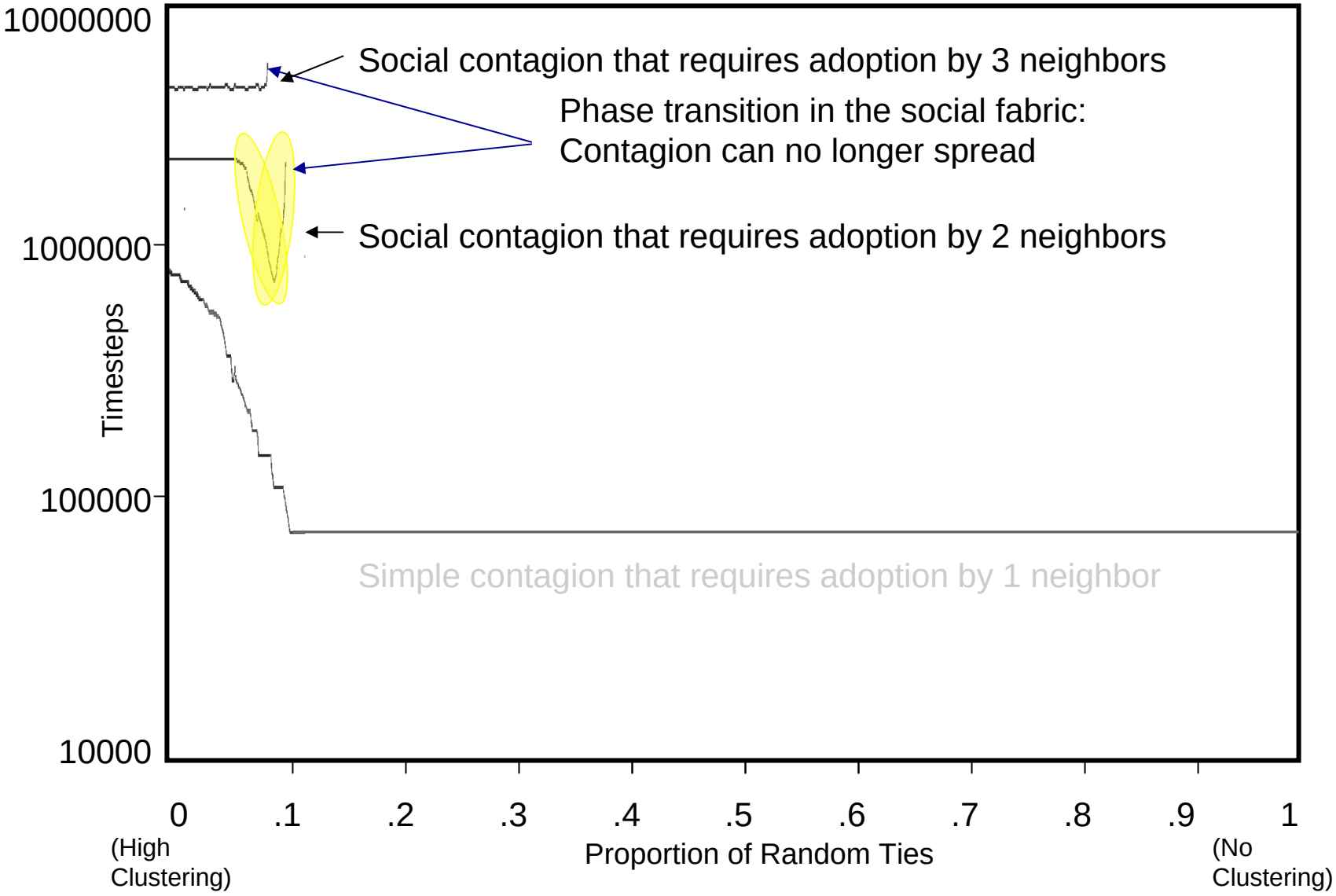
An Agent-Based Experiment

- Replicated Watts-Strogatz' original experiment
- Experimental condition: Increased activation threshold from 1 to 2+
- Observe effect of proportion of long-range ties on the rate and robustness of propagation

*Centola, D. and M. Macy. 2007. "Complex Contagions & the Weakness of Long Ties." *American Journal of Sociology* 113:702-34



Random ties promote the spread of information (lower is faster)



But not the spread of social contagions

Opinion Dynamics in US & JP*

- Are collectivist cultures more prone to forming a consensus?
- Tested using matched pairs of Amazon book reviews in the US and Japan
- Findings:
  - Collectivism promotes consensus
  - Consensus promotes consensus
  - Especially in Japan

*with Gueorgi Kossinets and Yongren Shi, Cornell

II. Why Do Ties Form?

- Homophily: likes attract
  - Attraction (shared identity, relevance)
  - Selection (opportunity to interact)
- Diversity: opposites attract
  - Complementarity and exchange
  - Avoid competitors
- Trust
  - Embeddedness (reputation, sanctioning)
  - Signal detection (“telltale signs of character”)

Empirical Tests

- UK business phone logs
- March 04 web crawl (95m pages)
- Web pages for 1081 US colleges & universities
- Similarity: in-link structural equivalence
- Results coming soon, initial indication is non-monotonicity

Why is Trust Lower in Japan?*

- Collectivist societies: risk-averse equilibrium
  - Risk of betrayal reduced by on-going relations
  - Provincialism precludes skills for judging strangers
- Individualist societies: risk-seeking equilibrium
  - Risk of betrayal reduced by detection skills
  - Better opportunities with long-range ("weak") ties
  - Experience with strangers improves detection skill

*with Ko Kuwabara, Robb Willer, Rie Mashima, Toshio Yamagishi

Web-Based Laboratory

- First XC trust experiment to independently vary both trustor and trustee nationality
- Designed to test effects of reciprocity (trust those who trust you)
- Findings
  - Japanese built more durable in-group relationships
  - Americans were more likely to trust strangers (US or JP)

Why Does Conflict Management Rarely Succeed?*

- Conventional strategy: promote out-group tolerance
- Agent-based model shows why this strategy fails:
  - Spiraling conflict intensifies in-group social control
  - An enemy without creates an enemy within
- An alternative strategy – promote tolerance *within* groups:
  - Protect "doves" from "hawk" reprisals
  - Promoting tolerance of dissent
  - Winning over opinion leaders (e.g., Grand Ayatollah Ali al-Sistani)

*with Steve Benard (Cornell), Lisa Troyer (UConn), & Elisa Bienenstock (Bose Allen Hamilton), funded by DARPA and NSF

Intergroup Conflict: The Enemy Within

- In 19th C. Corsica, those who failed to honor a vendetta faced reprisals by family members (Gould 2000).
- In 1995, a right-wing Israeli assassinated Yitzhak Rabin, even as the PLO renounced violence and recognized Israel.
- In WWI trench warfare, “cowards” were exiled to “no man’s land”

Sunni Militants Execute Moderate Sunnis Who Question Sectarian Violence

An Online Experiment

- Start with agent-based model of spiraling inter-group conflict
  - Spoils of war motivate conflict
    - A public good to winning group
    - But one contributor can tip a close conflict
  - Destruction from conflict (to both sides)
  - Individual cost of participating (private)
  - Costs of sanctioning and being sanctioned
- Replace agents with 1 or more avatars in large multiplayer virtual worlds

Test Strategies Online

- Phase 1: Generate an online inter-group conflict that spirals out of control.
- Phase 2: Test the effectiveness of between-group and within-group conflict management strategies.
- Pre-testing shows that incentives can reverse the spiral.

A Parting Thought …

- Our immediate goal is to collect and analyze digital traces.
- The hidden agenda
  - To advance a new way of doing social science
  - Dynamics, not comparative statics
  - How global patterns emerge from local interactions

And a Word of Caution …

- Unprecedented methodological challenges:
  - How to structure and search Web-scale data?
  - Billions of nodes at multiple time points
  - New tools needed to parse data into meaningful structures.
  - Manual coding is beyond human capabilities
  - The "New Social Science" must link the talents & tools of social, computer, and information scientists

The End