Modeling and Analysis of Dynamic Social Communication Networks

Malik Magdon-Ismail

CS, RPI.

www.cs.rpi.edu/~magdon

SNAG: **S**ocial **N**etwork **A**nalgorithms **G**roup

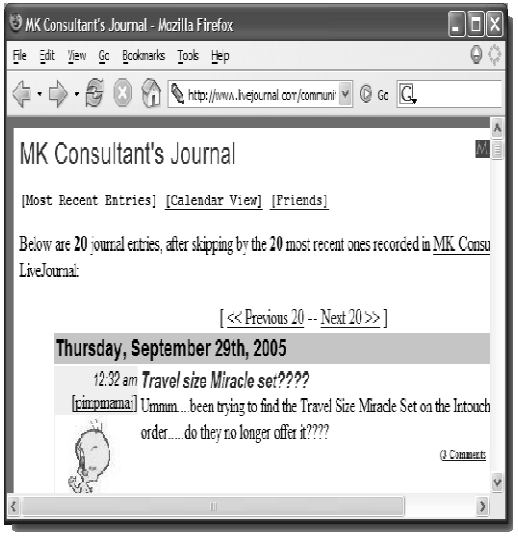
- Mark Goldberg
- M-I
- Al Wallace

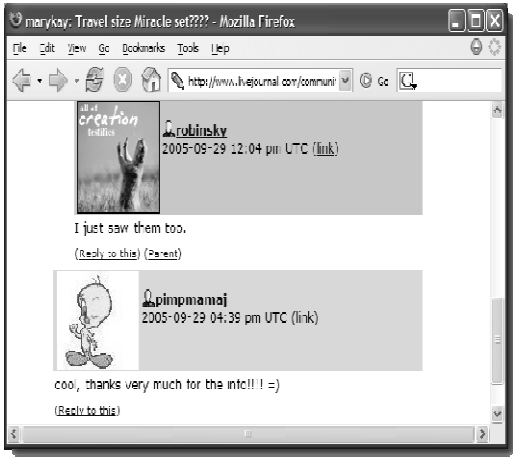
- Jeff Baumes
- Sean Barnes
- Justin Chen
- Matt Francisco
- Mykola Hayvanovich
- Konstantin Mertsalov
- Yingjie Zhou

Sponsors:

National Science Foundation
WHERE DISCOVERIES BEGIN

Communications





Time: January 12, 2005, 09:35

From: joe@xyz.com

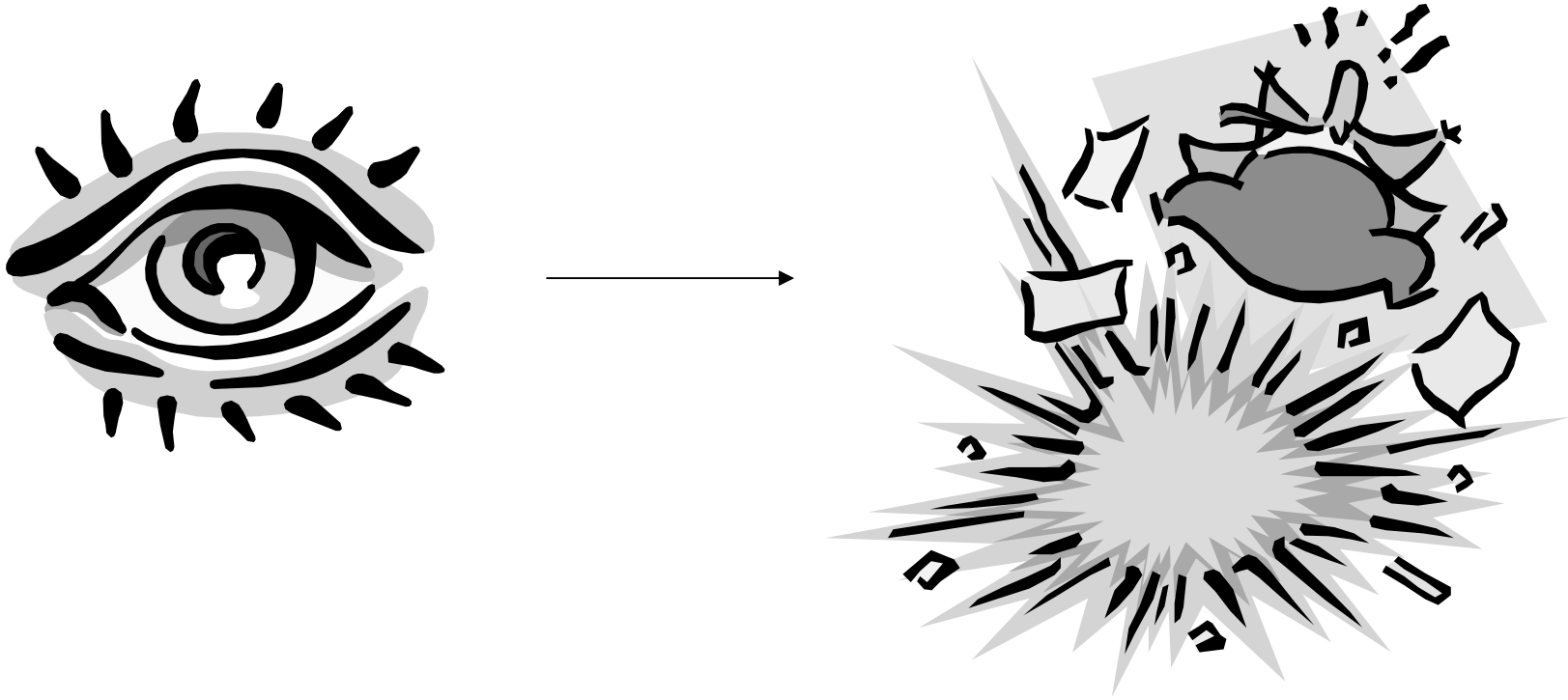
To: sue@abc.com

Subject: Hello

Message: Where have you been?

16:06:31] <FreeTrade> Republicans were the worst pacifists before ww1 and ww2
[16:06:43] <SweetLeaf> France Fries
[16:06:50] <FreeTrade> As a generality, of course their were Republican Hawks.
[16:07:13] <FreeTrade> Sweet, good pun but bad story!
[16:07:18] <SweetLeaf> yup
[16:07:23] <Lupine> anyways, he's perpetually tormented by presidential actions
[16:07:25] <SweetLeaf> it aint good for no one
[16:07:47] <SweetLeaf> I think they knew it was commiing
[16:07:51] <FreeTrade> Rossevelt met monthly in New York with mostly trusted Republicans to talk about how to get america into the war.
[16:08:10] <FreeTrade> and he spent 2 year with Churchill meeting him sometimes secretly in the ocean to discuss the same topic.
[16:08:22] <FreeTrade> Exchanging a lot of letters.
[16:08:25] <FreeTrade> telegrams
[16:08:28] <Lupine> There really is nothing like a shorn scrotum. It's breathtaking, I suggest you try it.
[16:08:55] <FreeTrade> Well they didnt literally meet in the ocean, they were on ships.

Minimal Intrusion



- Don’t use communication content.
  - Less intrusive
  - Easier

Overview

Part I:

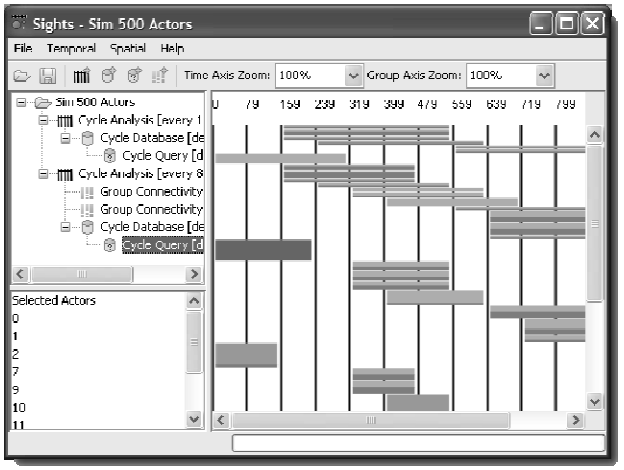
- Finding groups from communications.

Part II:

- Virtual Social Science Laboratory.

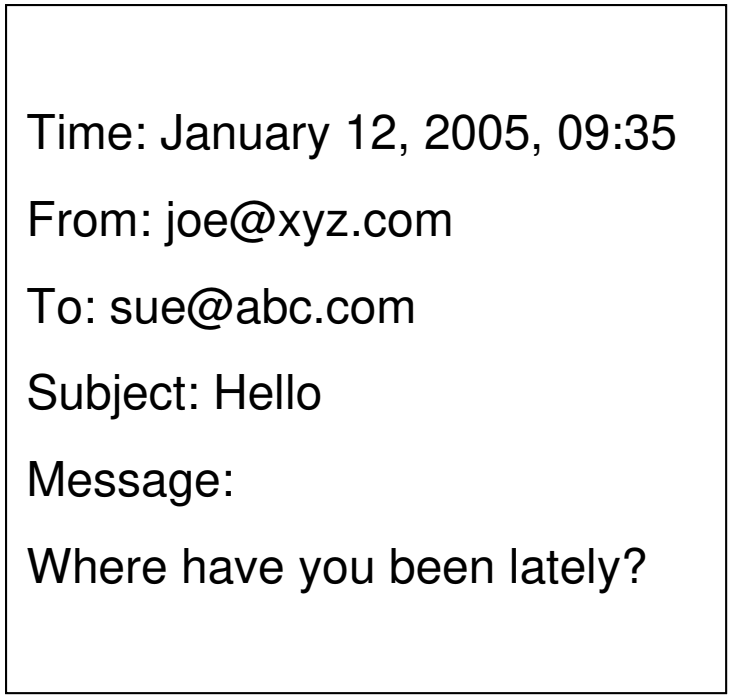
I: Groups from Communications

- Algorithms
  - Spatial algorithms (clustering)
  - Temporal hidden group algorithms
- Software tool SIGHTS
  - Statistical Identification of Groups Hidden in Time and Space
- Applications
  - Simulated datasets
  - Web logs
  - Enron email corpus

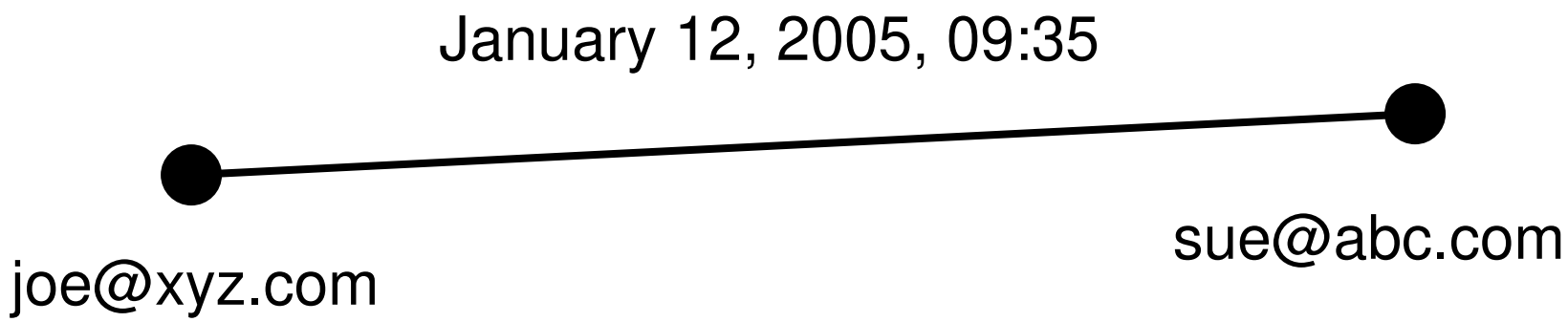


Communications Data

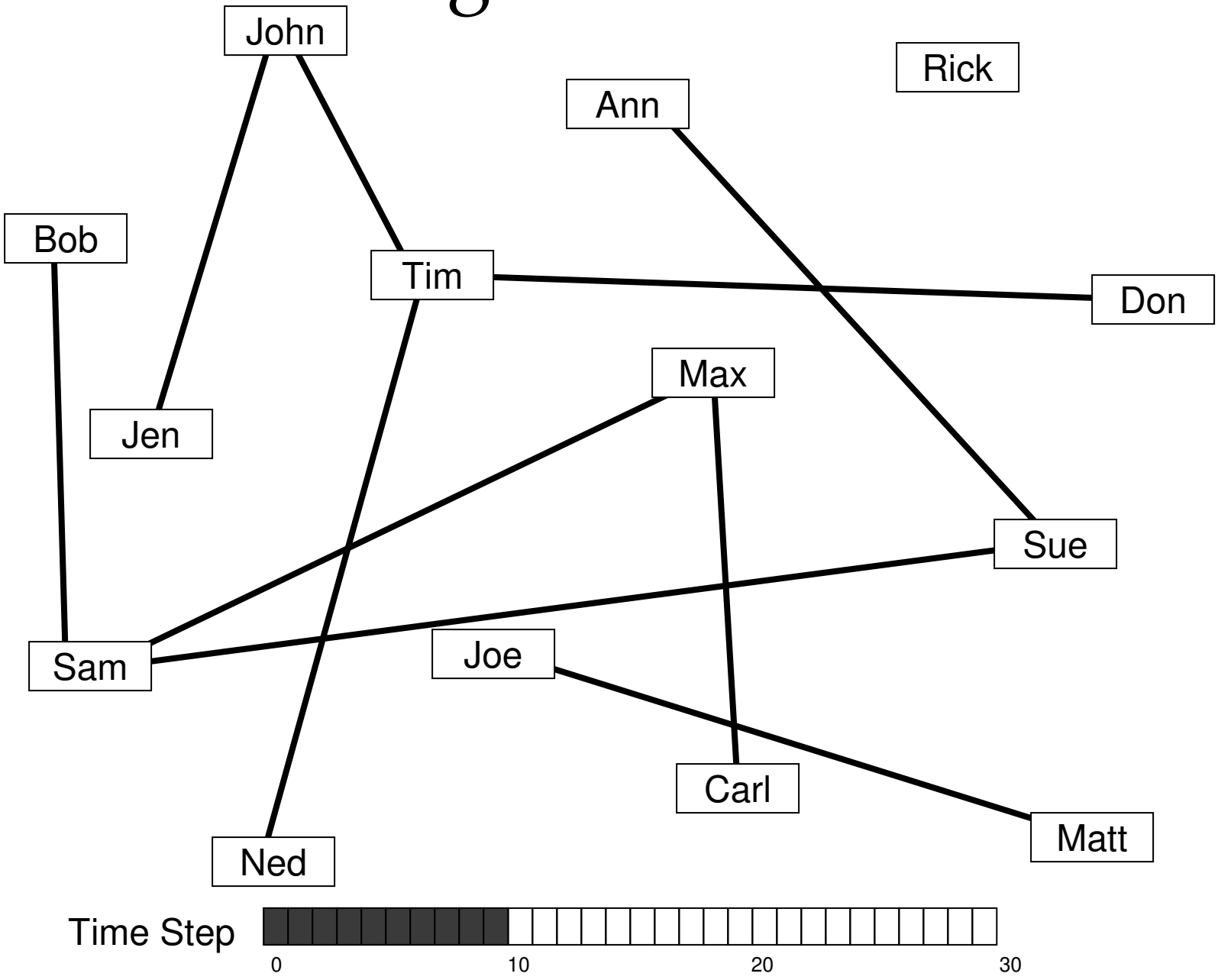
- Email, Telephone, Newsgroup, Weblog, Chatrooms, …



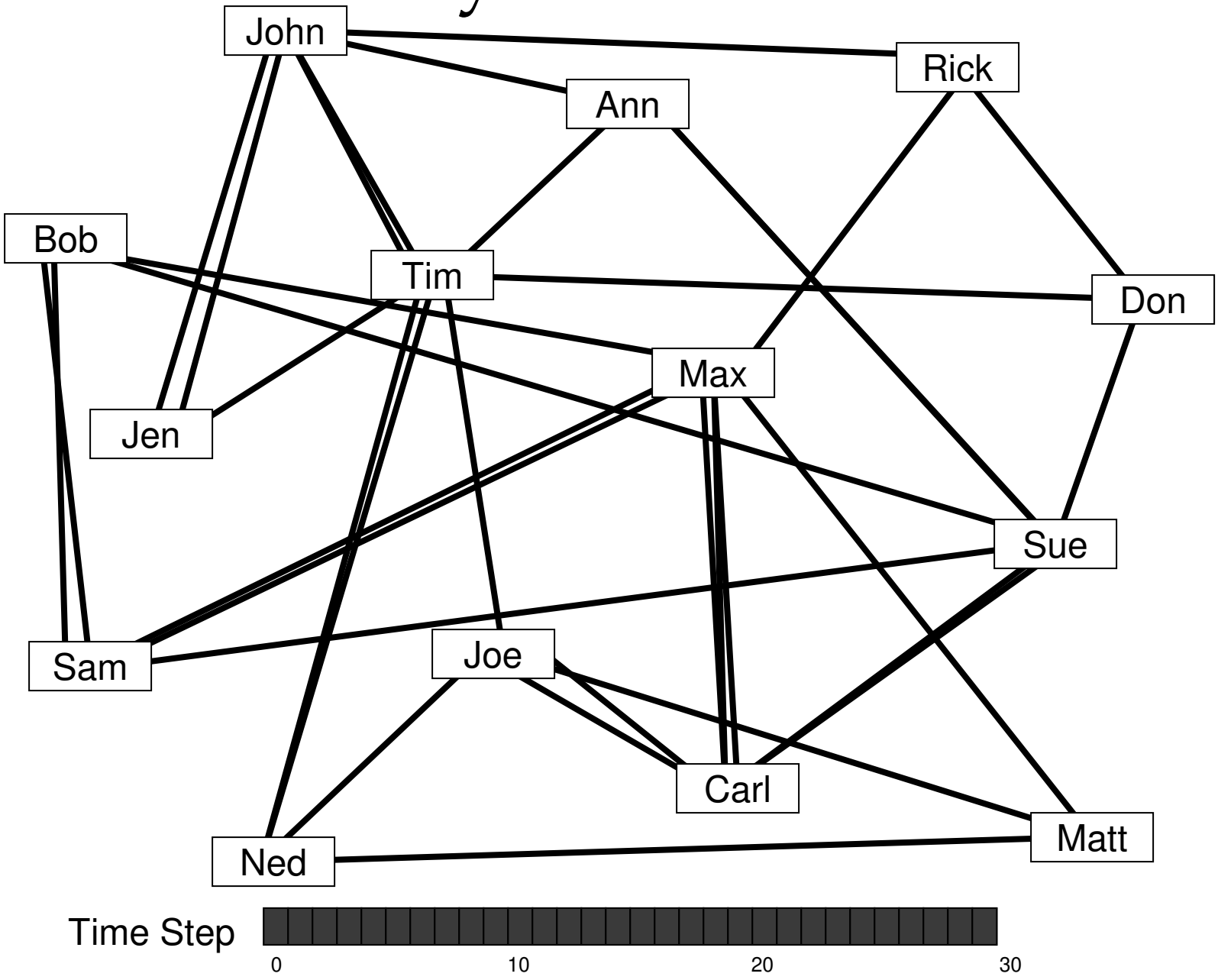
Communication Graph



Streaming Communications



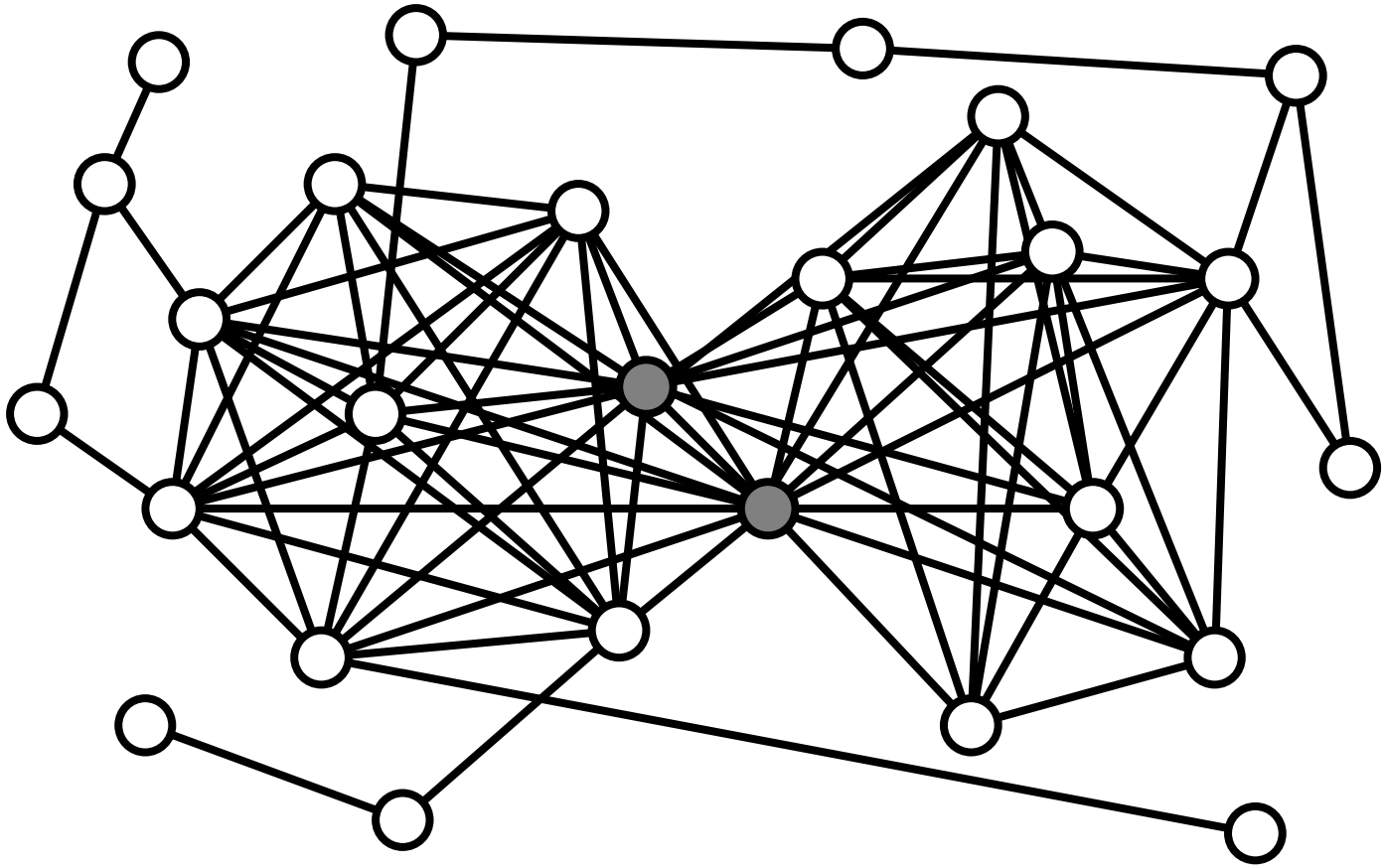
Cycle Model


John
Rick
Ann
Bob
Tim
Don
Max
Jen
Sue
Sam
Joe
Carl
Matt
Ned
Time Step
0
10
20
30

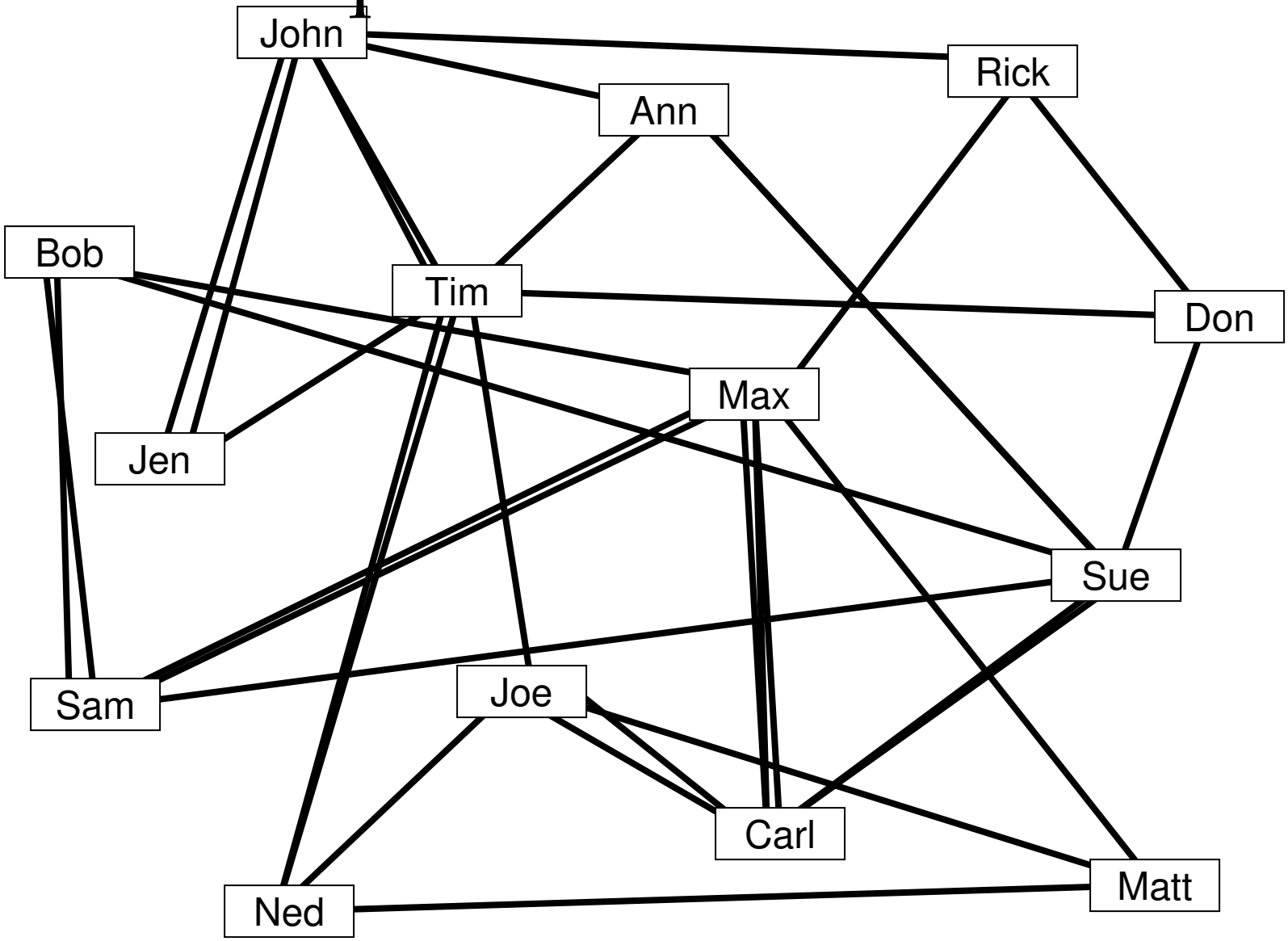
Types of Structure

- Spatial Correlation (spatial groups)
- Temporal Correlation (temporal or planning groups)

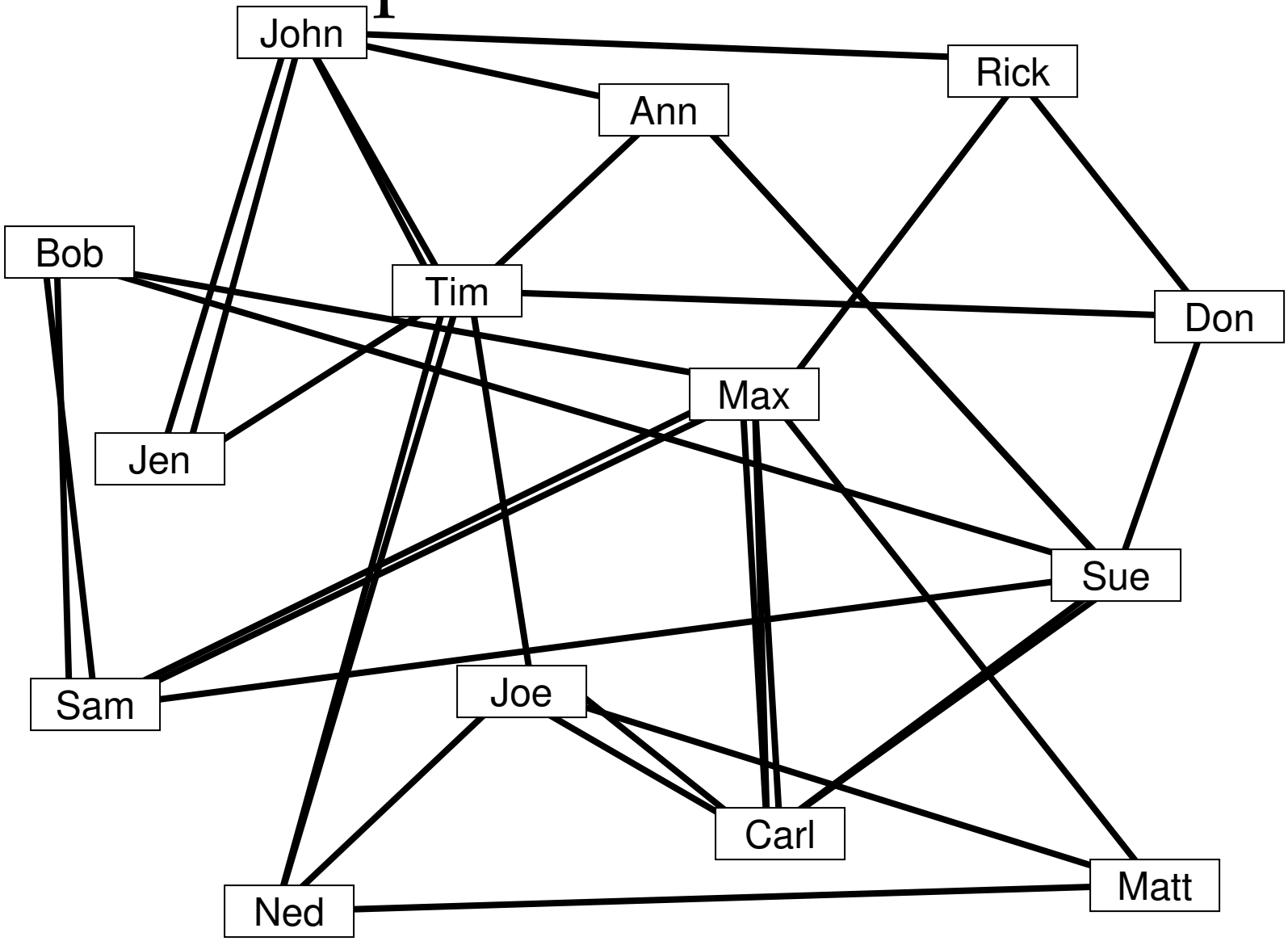
Groups Correlated in Space



Groups Correlated in Time



Groups correlated in time

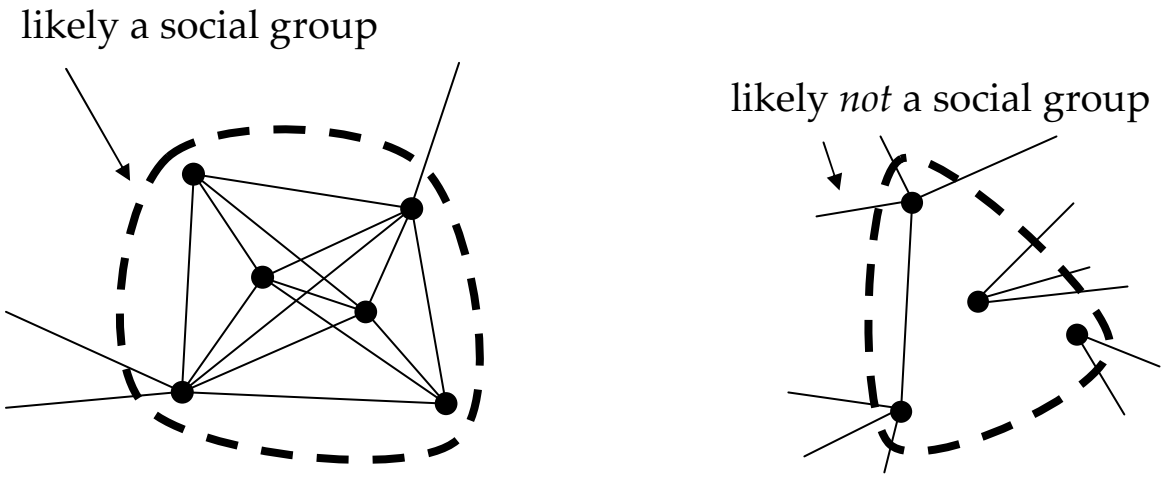


Spatial Correlation

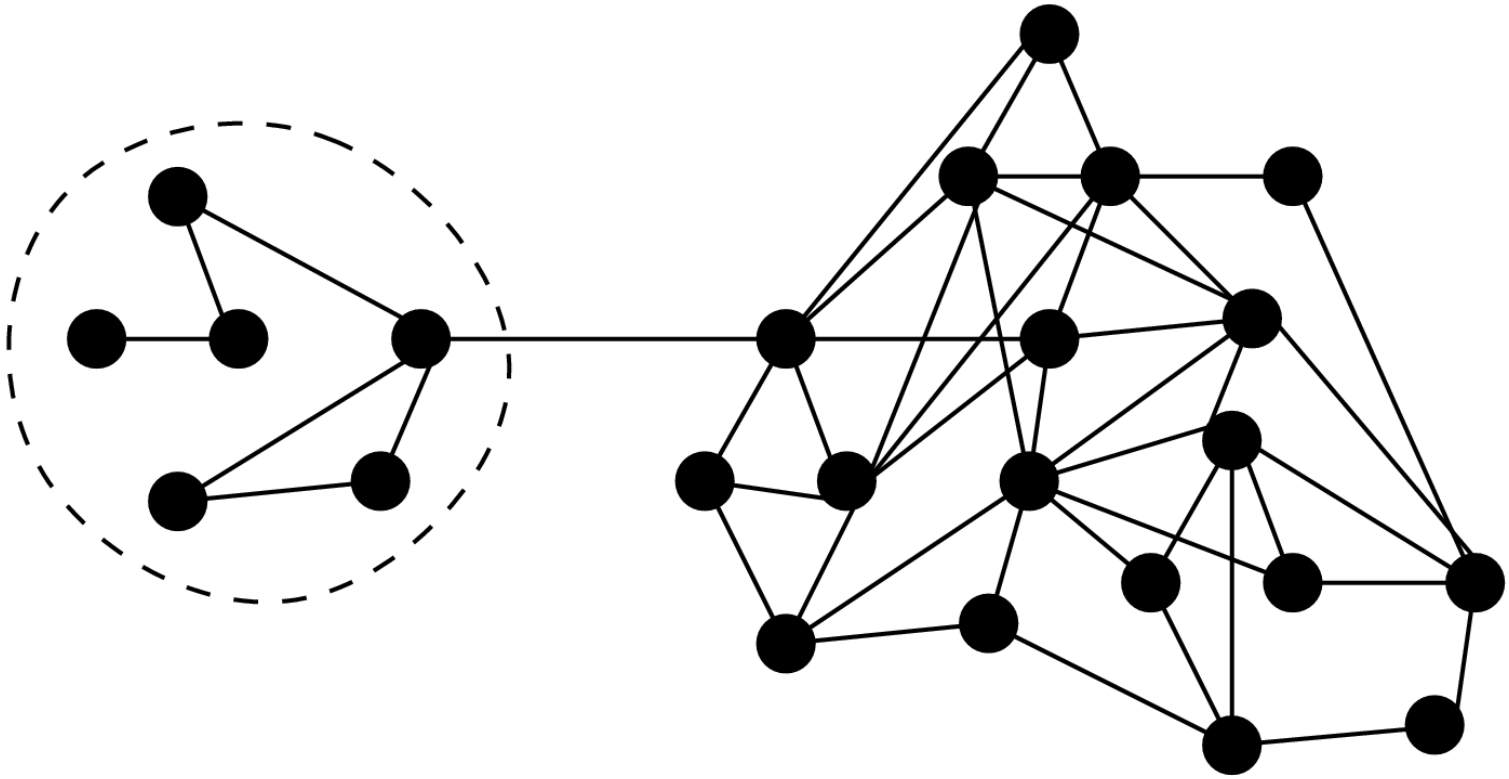
Clustering graphs into *overlapping* clusters

Groups as Clusters

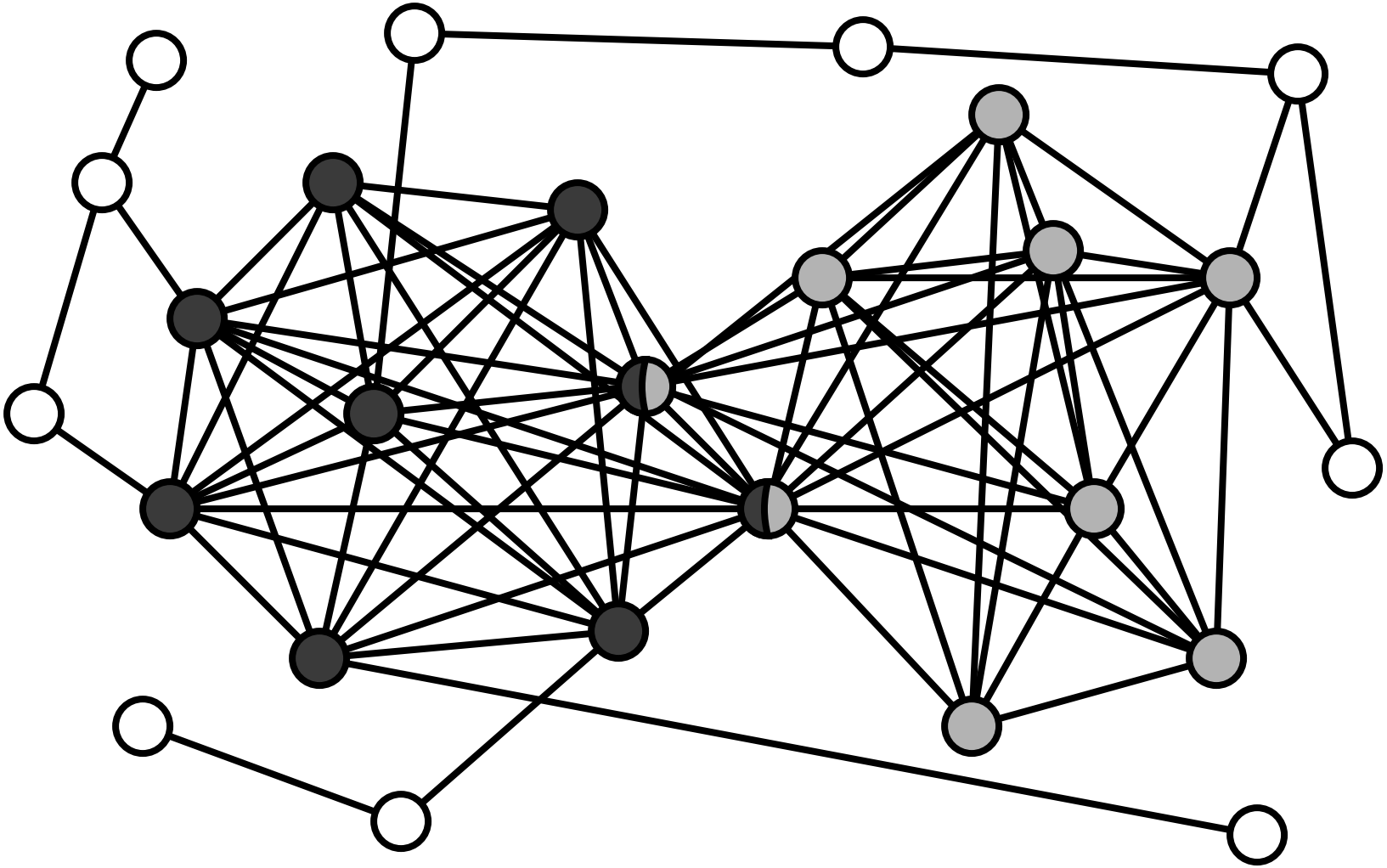
- Social groups tend to communicate with each other
- Find social groups by finding locally dense clusters



Locally vs. Globally Dense

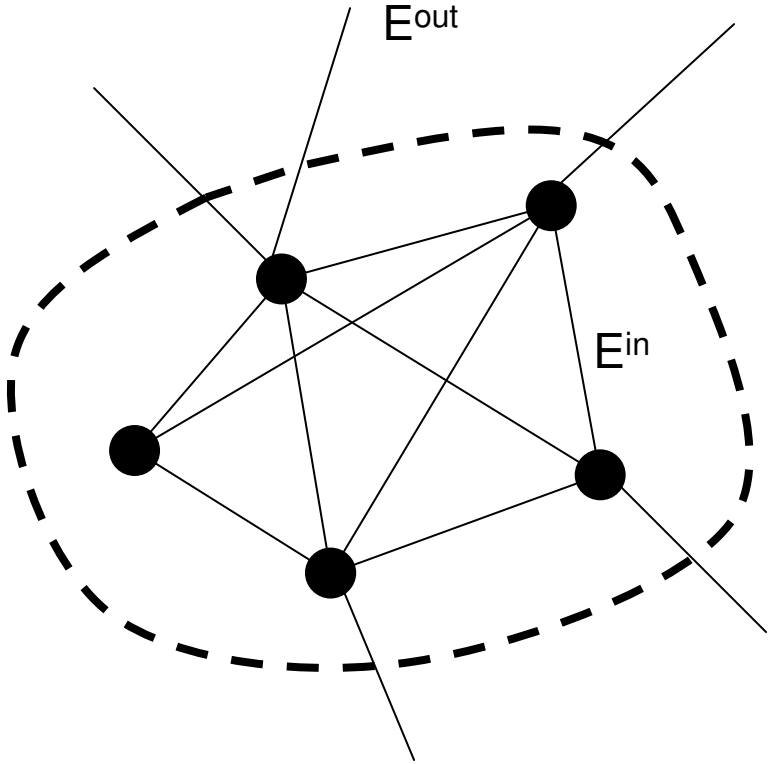


Clustering vs. Partitioning



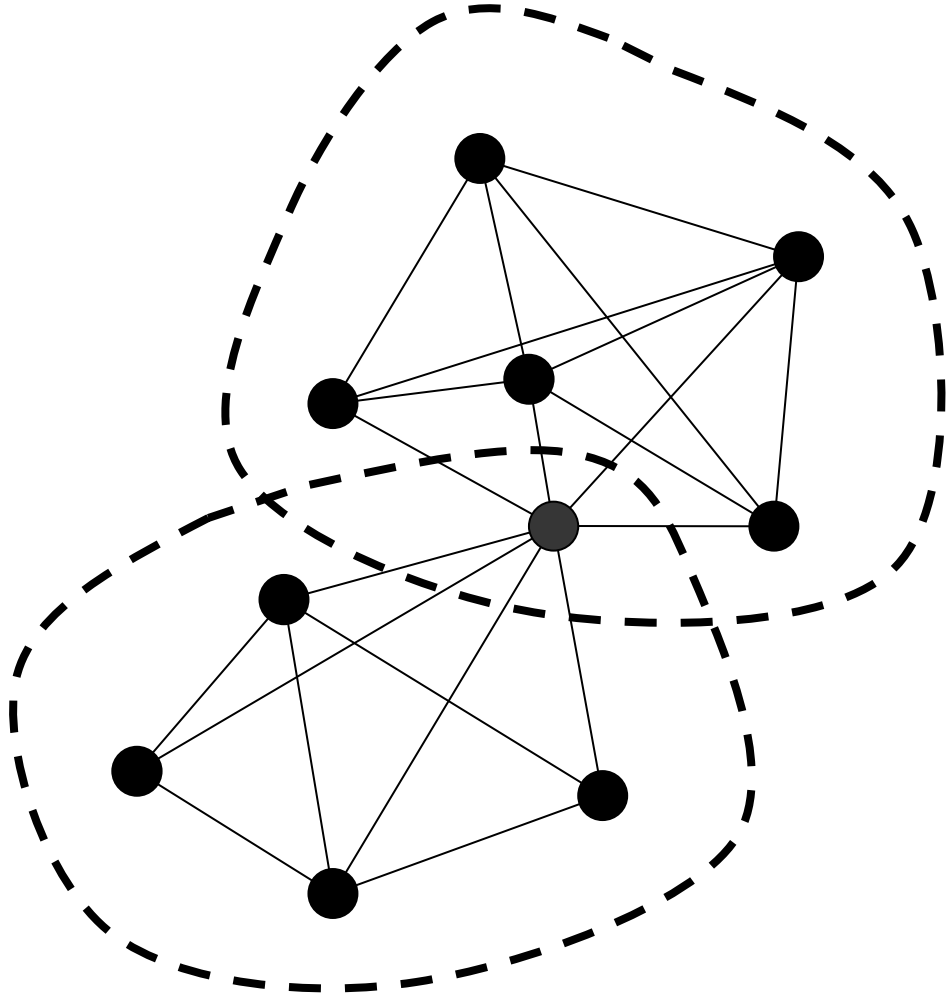
Clustering density metrics

- $P^{in}=E^{in}/E^{poss}$
- $E^{in}/(E^{in}+E^{out})$
- $P^{in}/(P^{in}+P^{out})$



Influential Nodes

- Page Rank
- Centrality
- …



Iterative Improvement

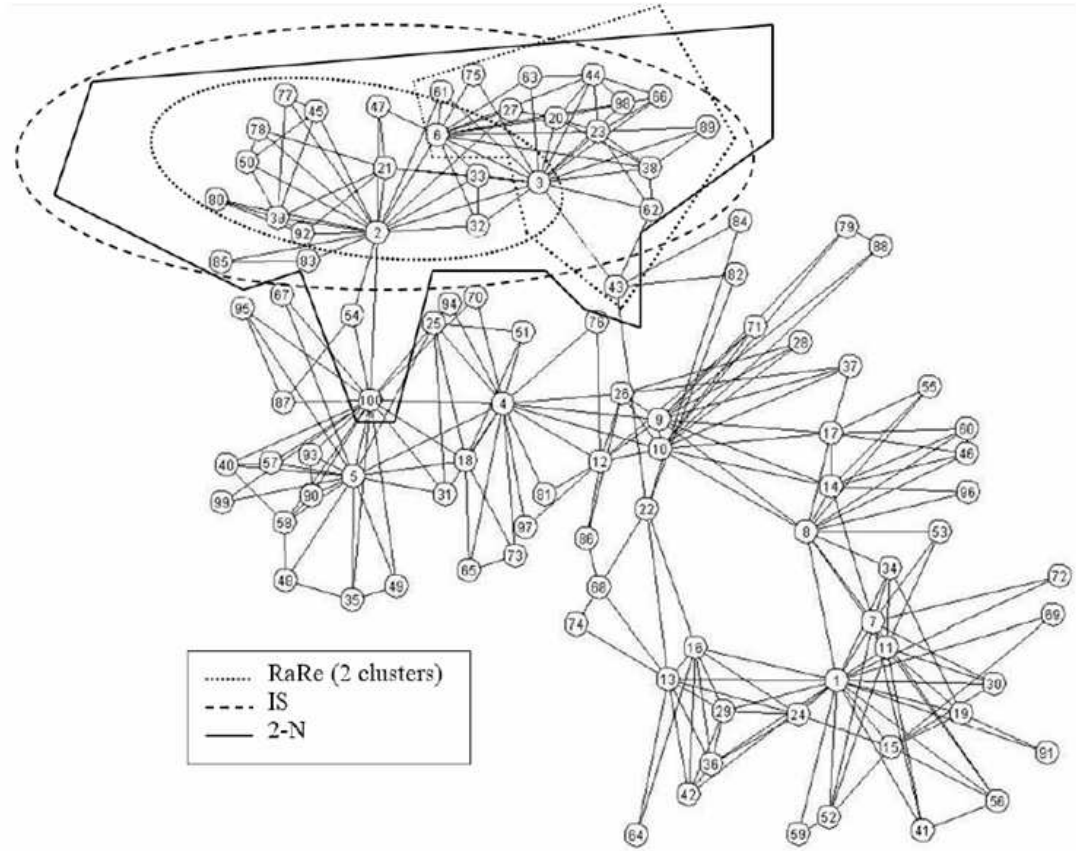
- Improve initial clusters using iterative local optimization.

Link Agregate (LA) *[B,G,M-I '05].*

RaRe & Iterative Scan (IS) *[B,G,K,M-I,P '05].*

Some Real Social Networks

- Semantic Web



Some Real Social Networks

- **CiteSeer** (co-authorship graph)

Example clusters:

Electric circuit design:

"An optimization strategy for reconfigurable control systems"

Optimization of Neural Networks:

"A new activation function in the Hopfield network for solving optimization problems"

Intersection:

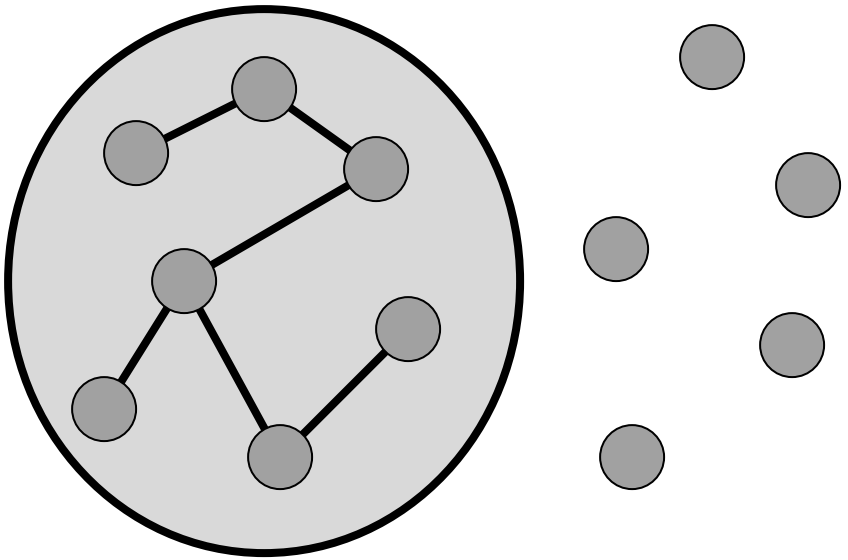
"Sensitivity analysis in degenerate quadratic programming"

Temporal Correlation

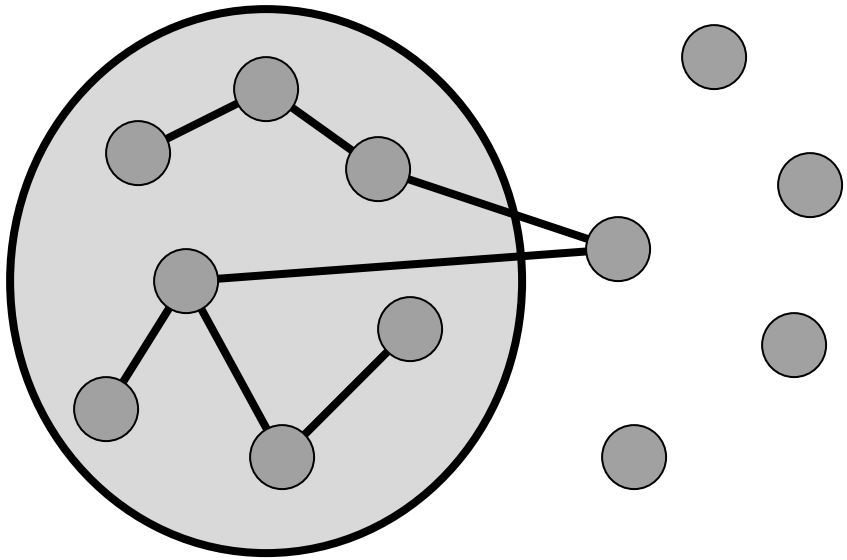
Finding hidden groups that are
planning over time

Connectivity and Planning

Internally connected



Externally connected



Persistence

- Group connected in successive time periods.

Persistence ⇔ planning over time.

Finding Temporal Hidden Groups

Given: communication graphs $G_1,\ldots,G_T$

- Is there a hidden group of size > K?
- Find all such hidden groups?
- Over what period is the hidden group active?

Algorithms

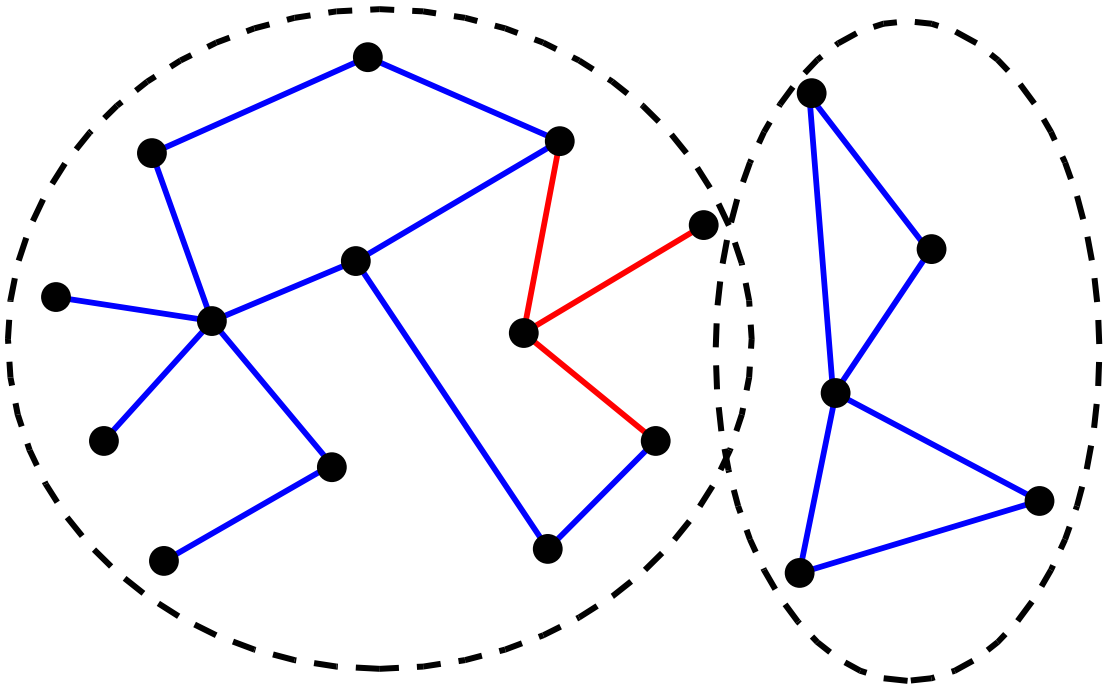
Low order poly-time algorithms:

[B,G,M-I,W '05]

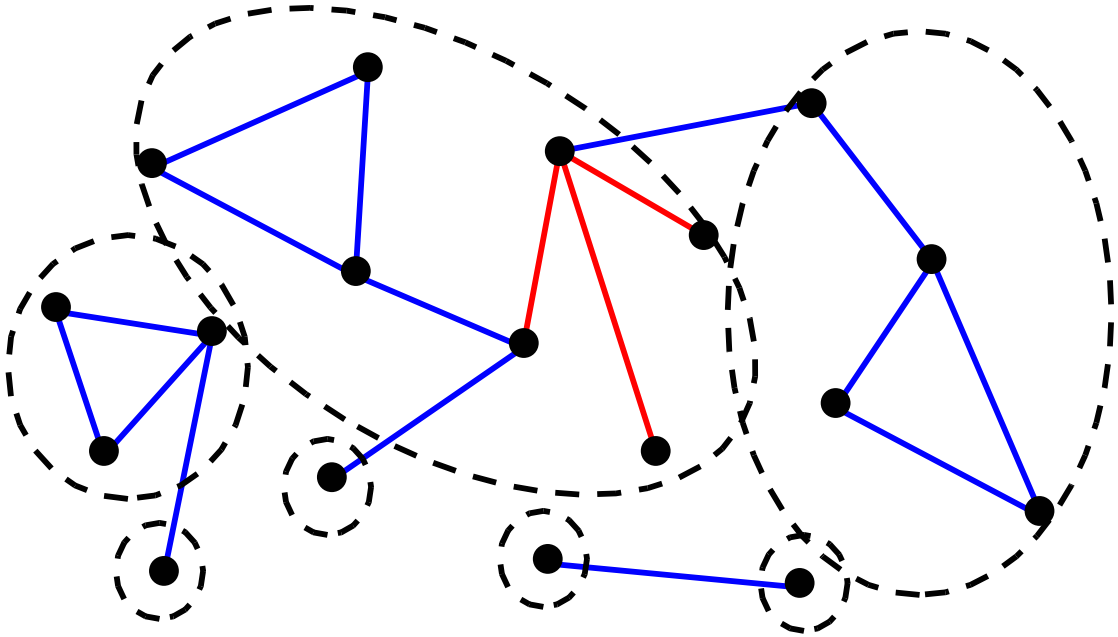
- Not all members connected in every time period?
- Connected in most time periods?

NP-Hard

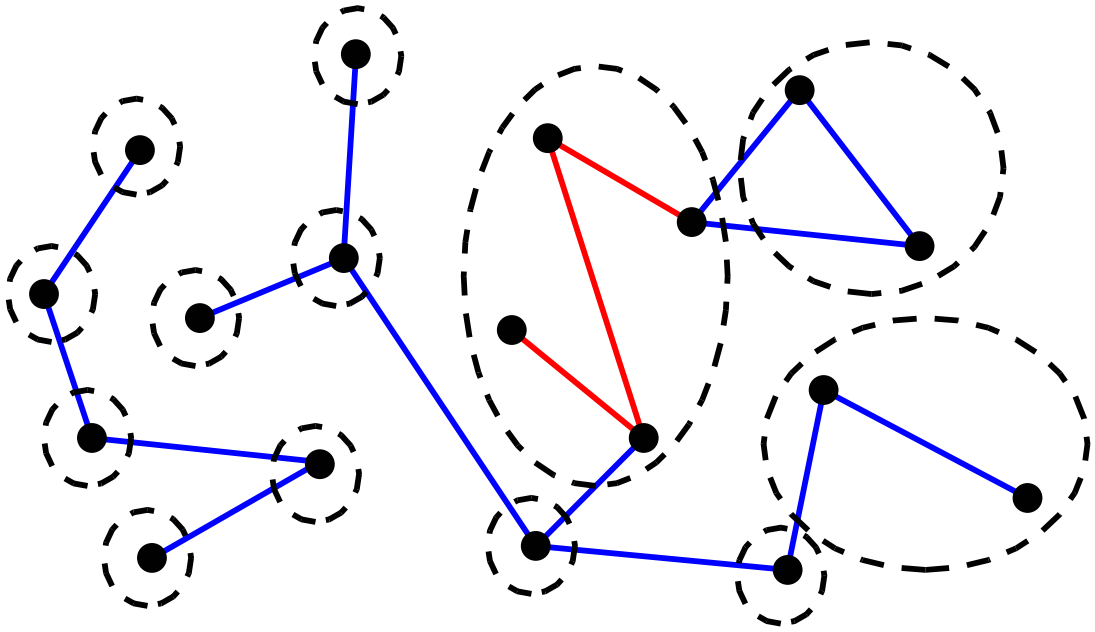
Example



Example

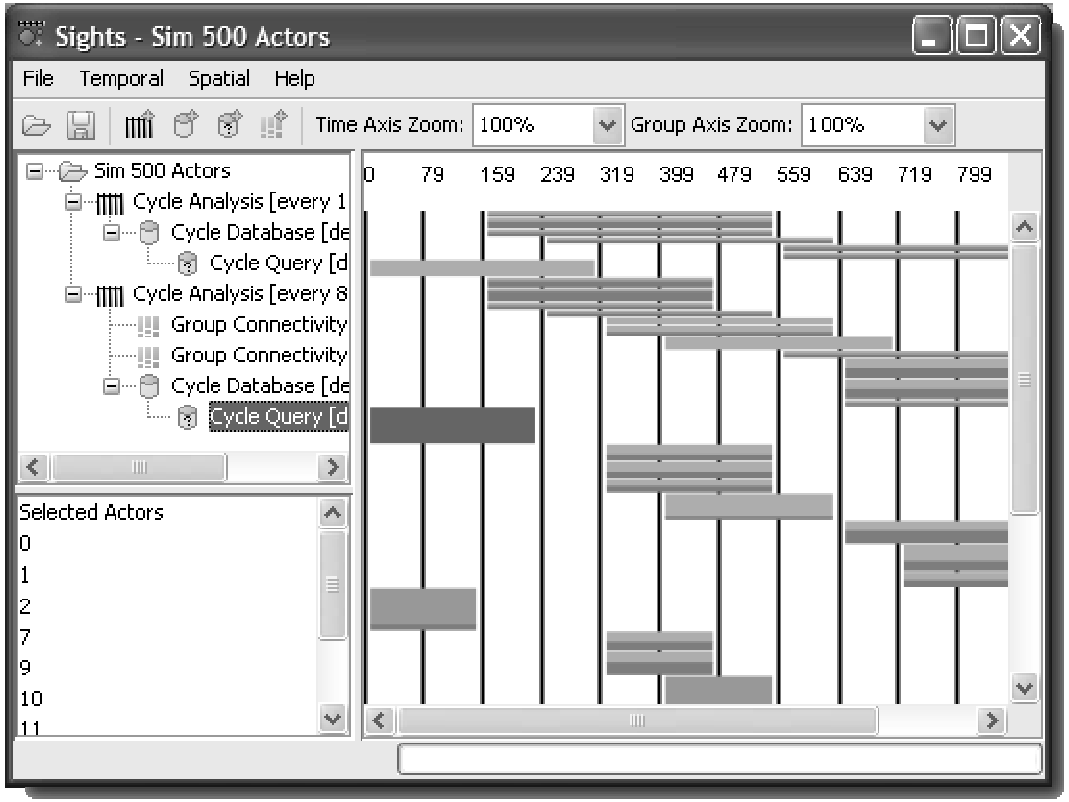


Example



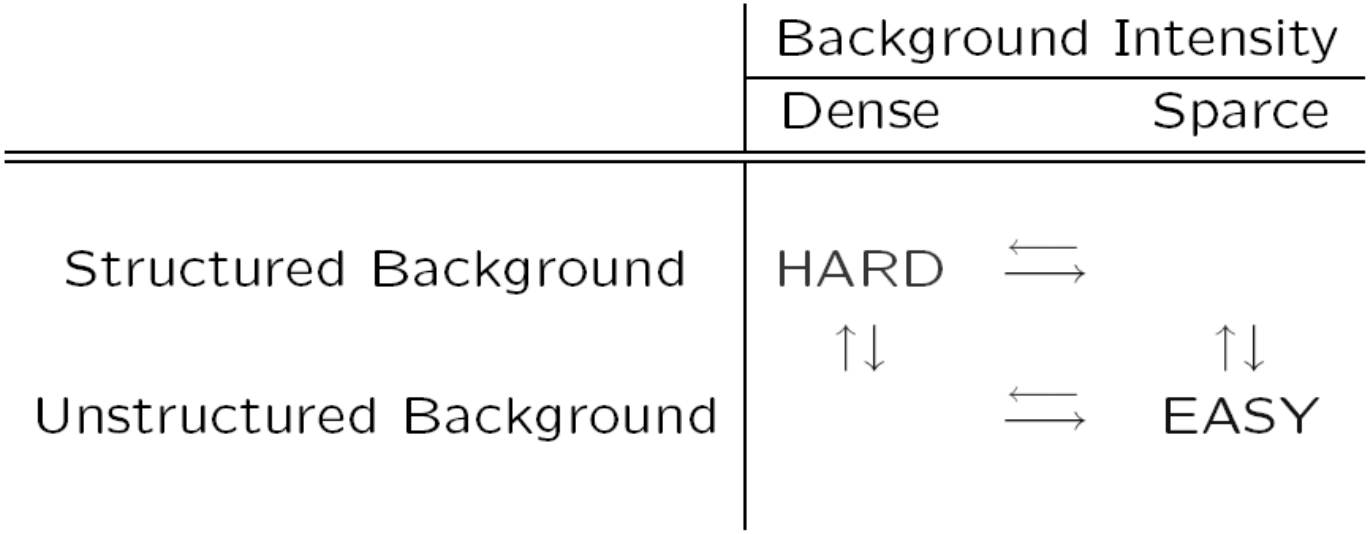
SIGHTS

Statistical Identification of Groups Hidden in Time and Space



Statistical Significance

- Background communications



- Nature of hidden group
  - Detecting non-trusting hidden groups is easier

Ali Baba dataset

- Unclassified synthesized data for the Department of Defense
- Used for specific case studies for initial validation of research
- Nine embedded hidden groups

Message content not used

Ali Baba initial results

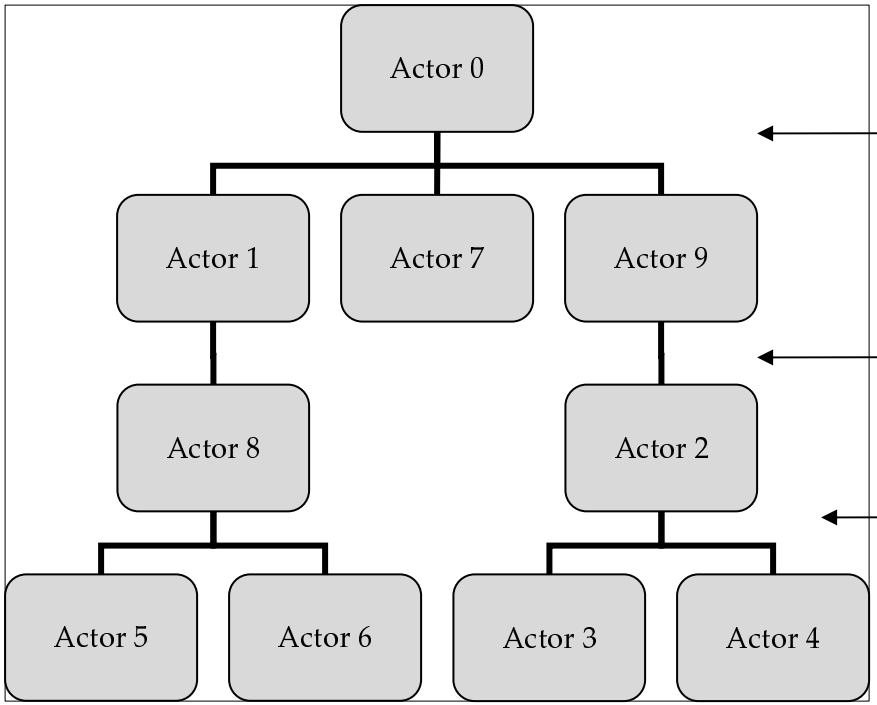
Ground Truth

- Group A
  - Dog
  - Vulture
  - Camel
  - Yassir Hussein
  - Bird
  - (6 others)
- Group B
  - Ahmet
  - Saleh Sarwuk
  - Shaid
  - Pavlammed Pavlah
  - Osan Domenik

SIGHTS

- Group A
  - Dog
  - Vulture
  - Camel
  - Gopher
- Group B
  - Ahmet
  - Saleh Sarwuk
  - Shaid
  - Ahmett
  - Dajik

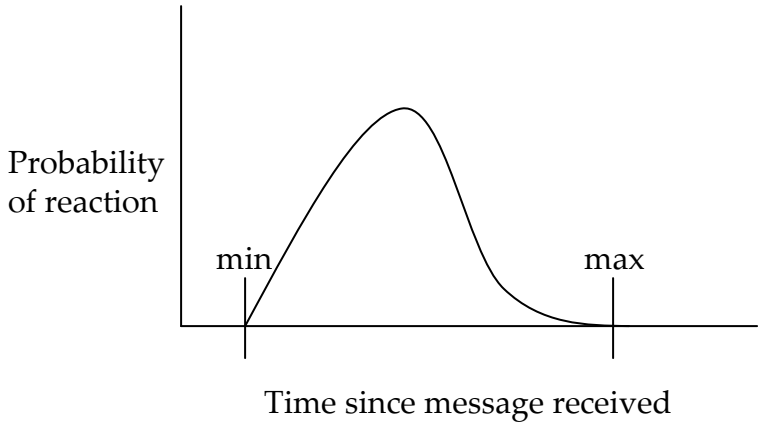
Cycle vs. Stream Model



Sent at time B

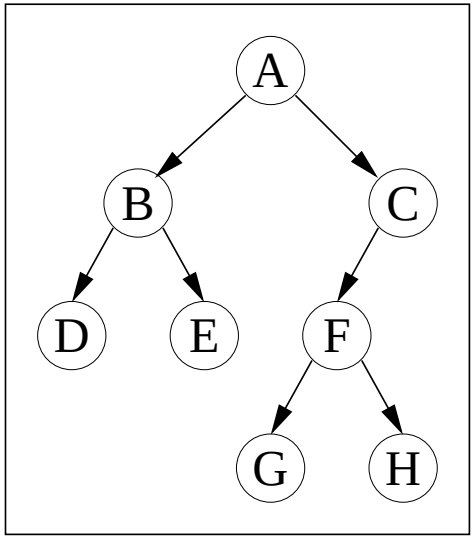
Sent at time B + 20

Sent at time B + 40



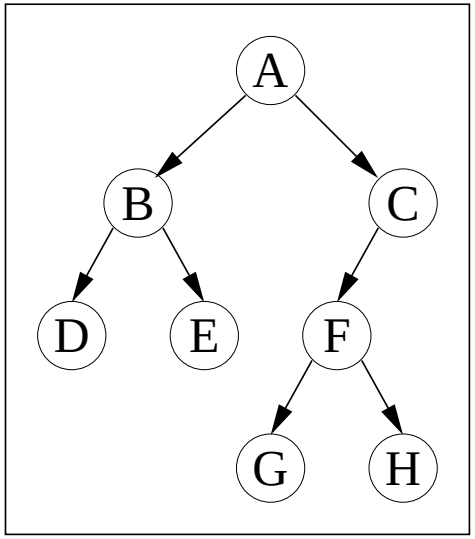
Stream Example

Time	From	To	Message
10:00	Alice	Charlie	Golf tomorrow? Tell everyone.
10:05	Charlie	Felix	Alice mentioned golf tomorrow.
10:06	Alice	Bob	Hey, golf tomorrow. Spread the word.
10:12	Alice	Bob	Tee off: 8am at Pinehurst.
10:13	Felix	Grace	Hey guys, golf tomorrow.
10:13	Felix	Harry	Hey guys, golf tomorrow.
10:15	Alice	Charlie	Pinehurst Tee time: 8am.
10:20	Bob	Elizabeth	We're playing golf tomorrow.
10:20	Bob	Dave	We're playing golf tomorrow.
10:22	Charlie	Felix	Tee time 8am at Pinehurst
10:25	Bob	Elizabeth	We tee off 8am at Pinehurst.
10:25	Bob	Dave	We tee off 8am at Pinehurst.
10:31	Felix	Grace	Tee time 8am, Pinehurst.
10:31	Felix	Harry	Tee time 8am, Pinehurst.



Stream Example

Time	From	To
10:00	Alice	Charlie
10:05	Charlie	Felix
10:06	Alice	Bob
10:12	Alice	Bob
10:13	Felix	Grace
10:13	Felix	Harry
10:15	Alice	Charlie
10:20	Bob	Elizabeth
10:20	Bob	Dave
10:22	Charlie	Felix
10:25	Bob	Elizabeth
10:25	Bob	Dave
10:31	Felix	Grace
10:31	Felix	Harry



Streams vs. Cycles

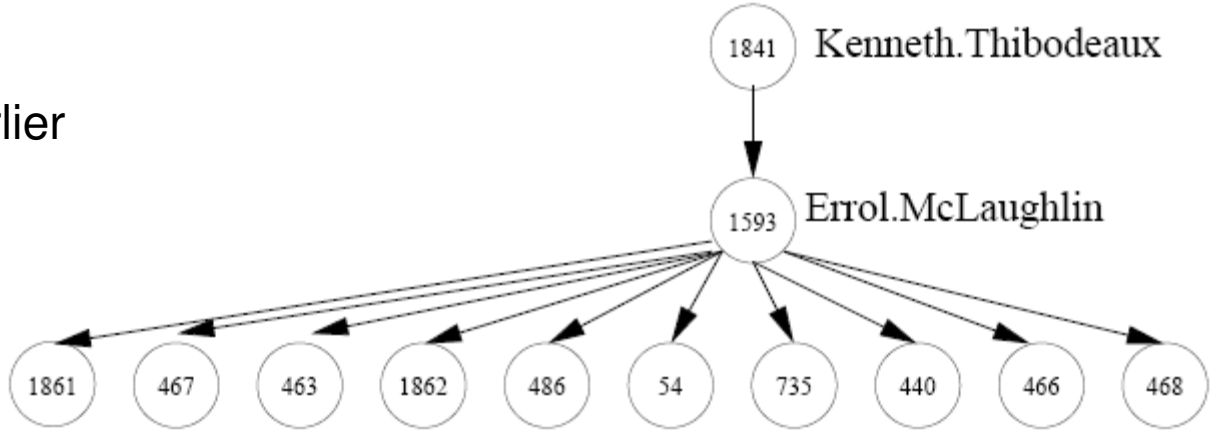
- Tree threads may overlap.
- Some may be short, some long.

Stream Algorithms

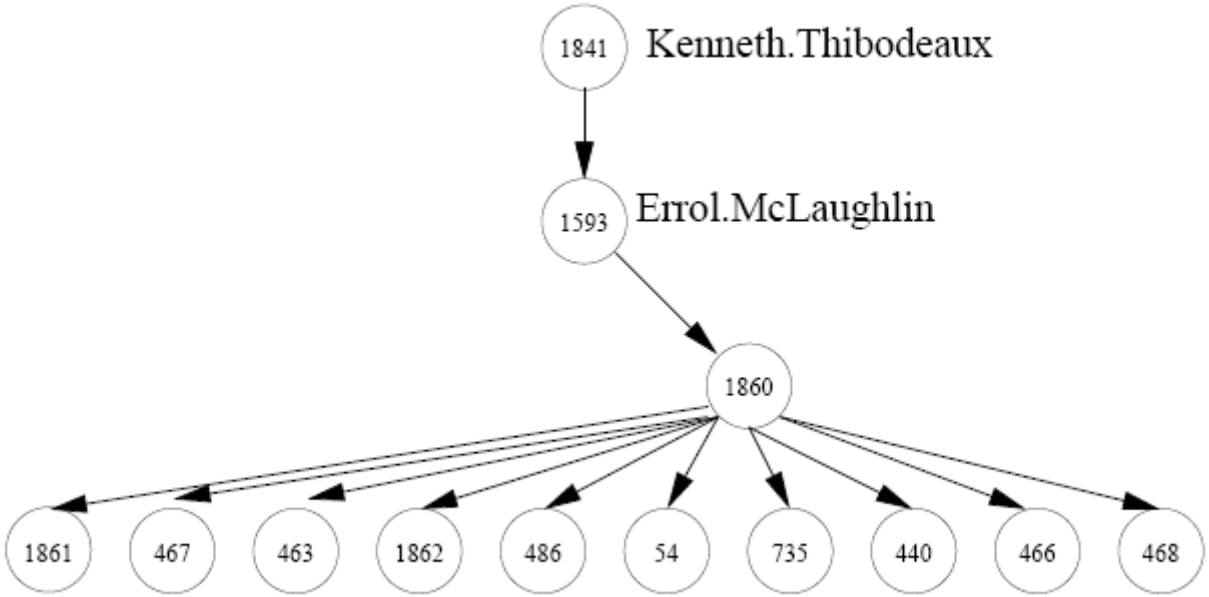
- Efficient algorithms for small trees (triples, chains).
- Build larger frequent trees from smaller.
- What size tree is statistically significant?

Enron data in stream model

Earlier



Later



II: Virtual Social Science Laboratory

- A general HMM model.
- Simulation
  - social science experiments.
- Reverse engineering
  - what makes a society tick?

Goal

Given a society's communication history,

1. Can we predict the society's future:

   eg: number of groups after 3 months?

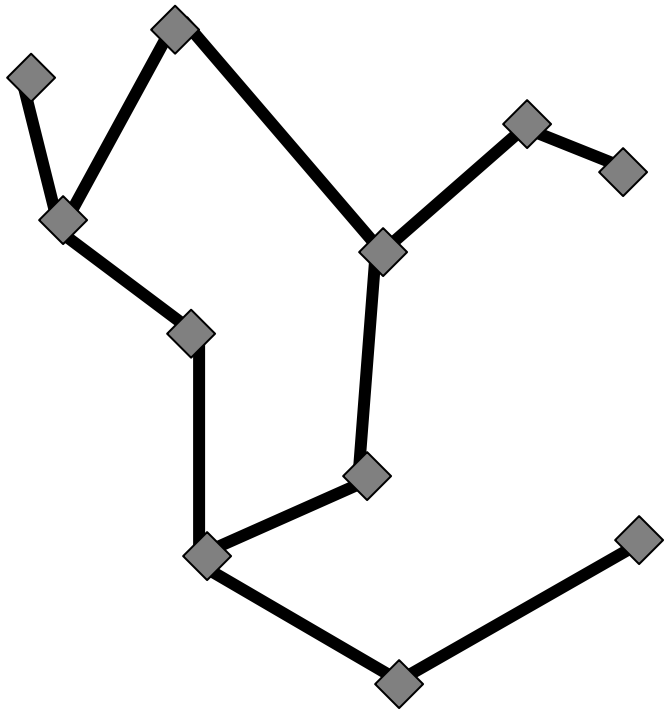
   average group size after 3 months?

2. Can we deduce something about the "nature" of the society:

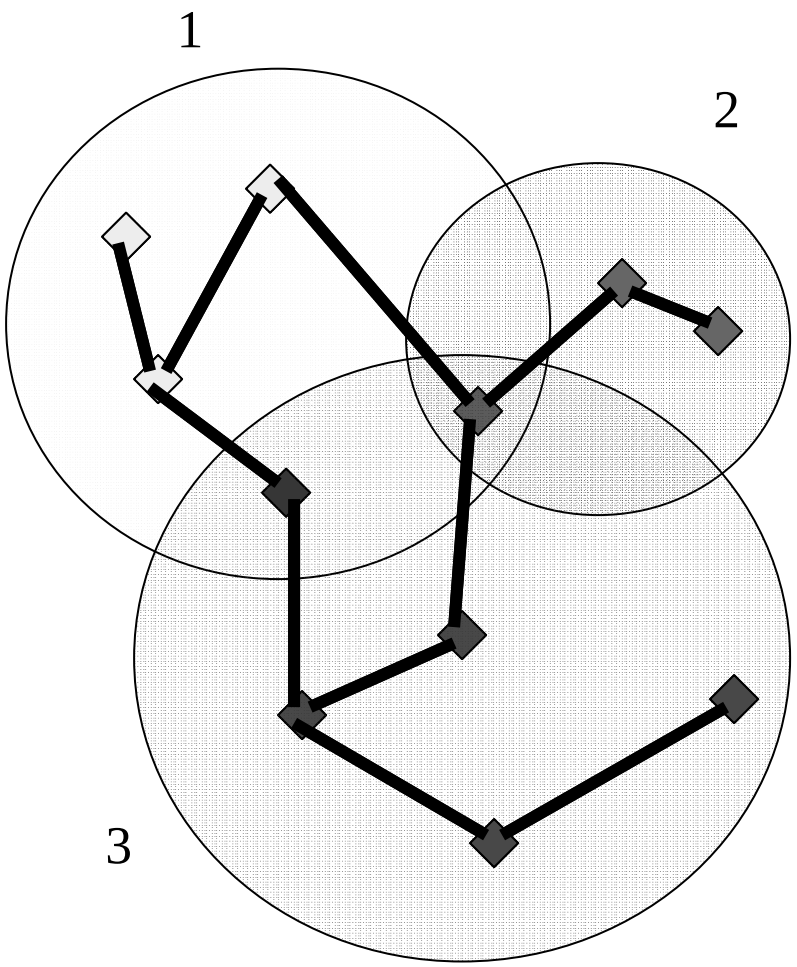
   eg: actors have a propensity to join small groups?

Social Networks

- Actors

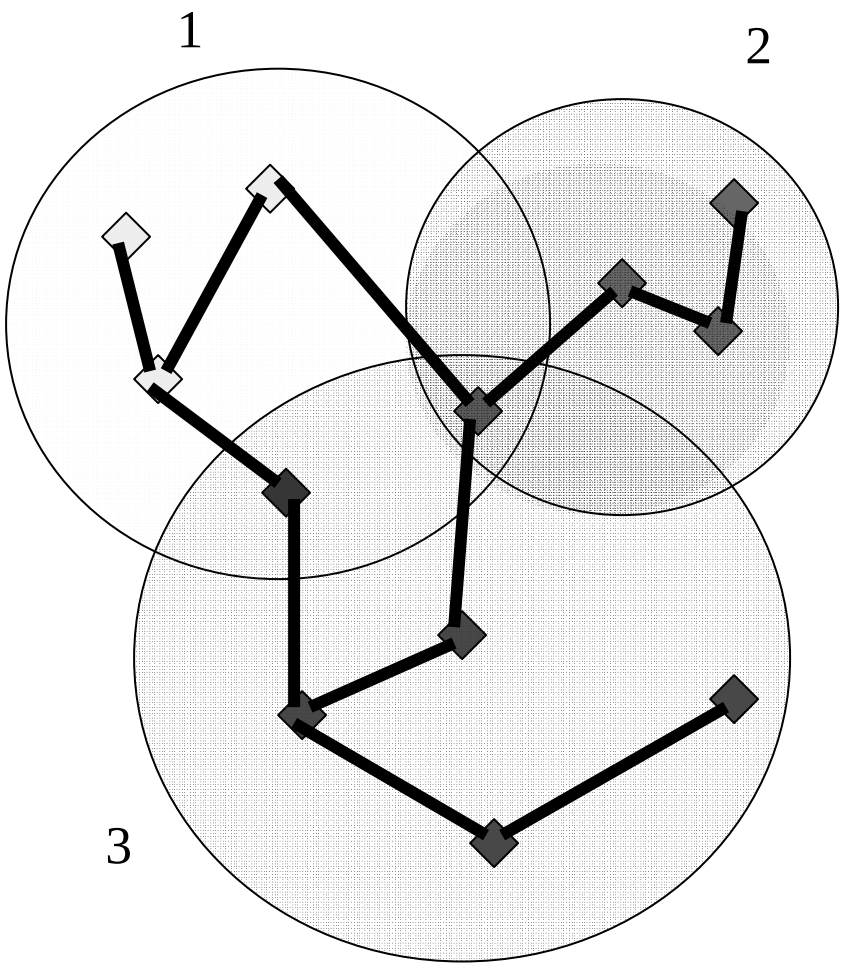


Social Networks



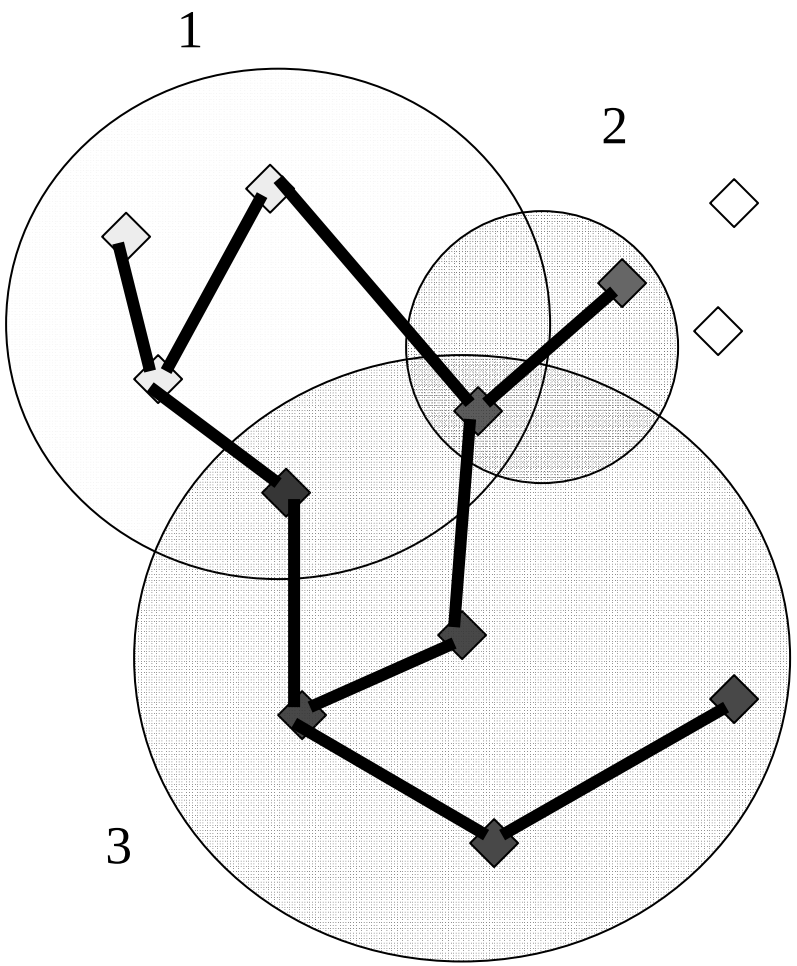
- Actors
- Groups

Social Networks



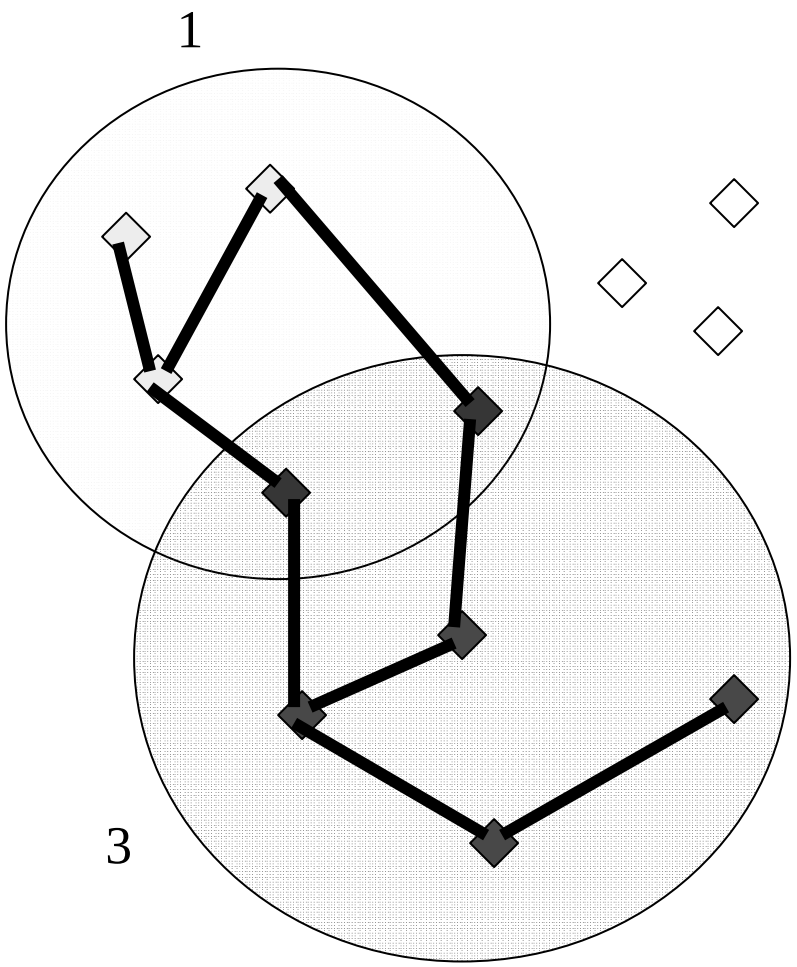
- Actors
  - Join
- Groups

Social Networks



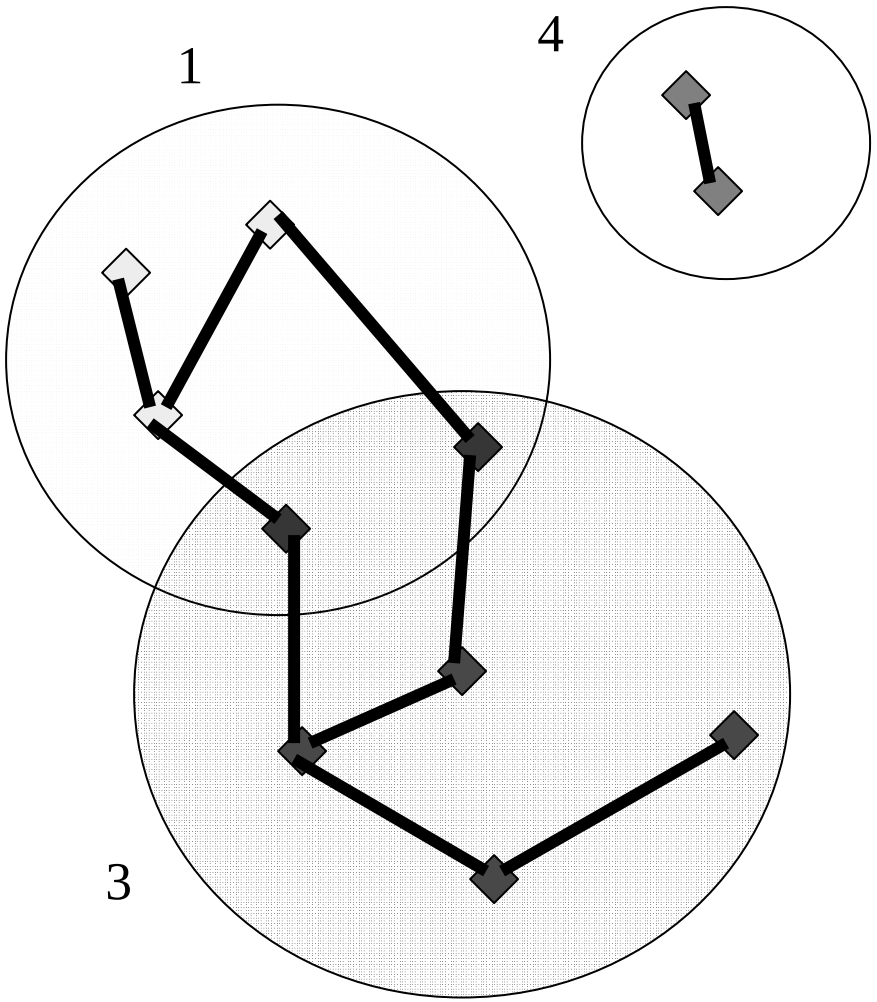
- Actors
  - Join
  - Leave
- Groups

Social Networks



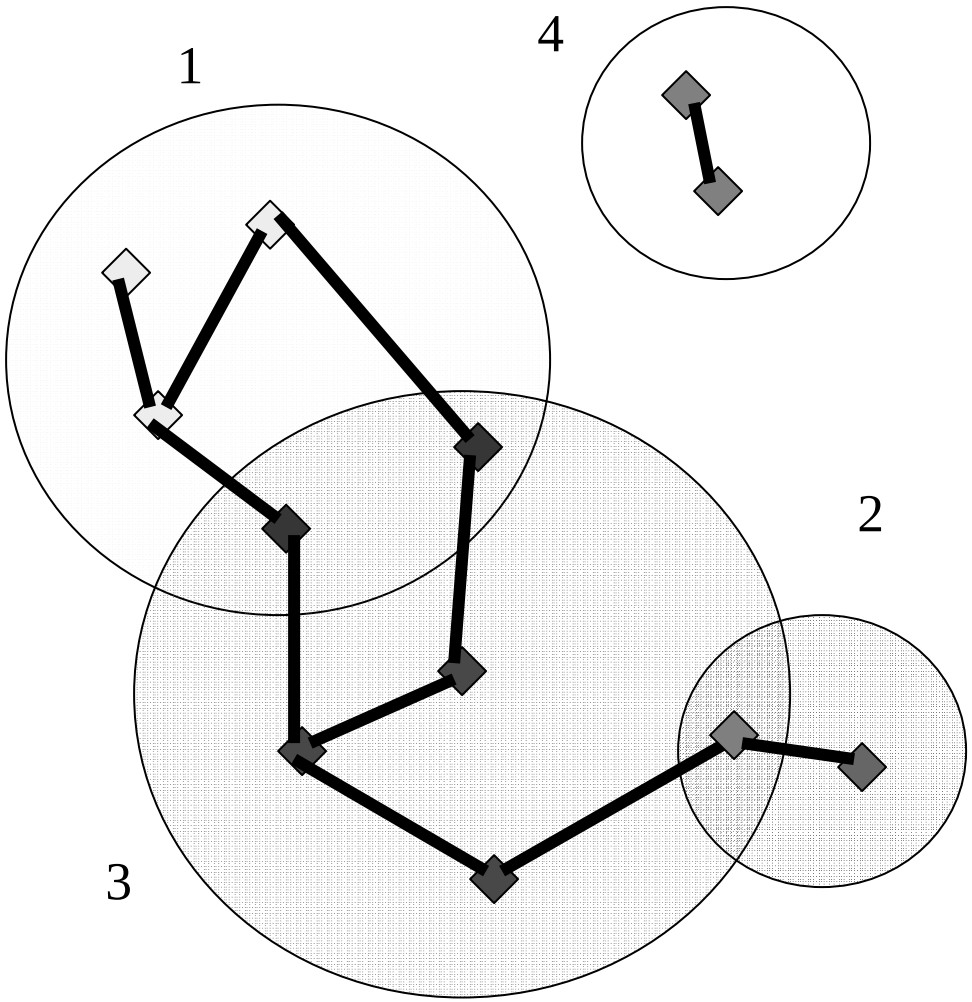
- Actors
  - Join
  - Leave
- Groups
  - Disappear

Social Networks



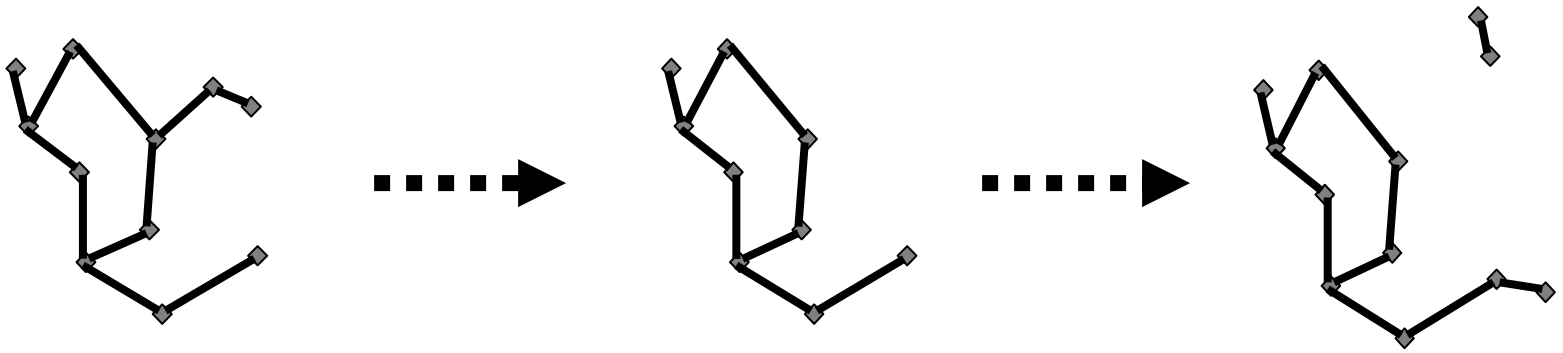
- Actors
  - Join
  - Leave
- Groups
  - Disappear
  - Appear

Social Networks

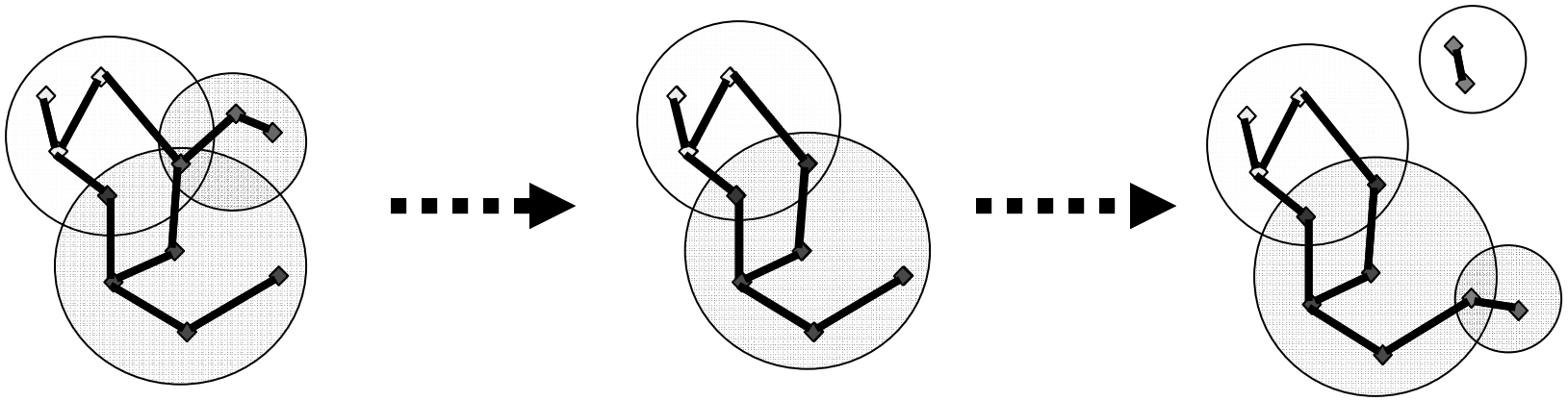


- Actors
  - Join
  - Leave
- Groups
  - Disappear
  - Appear
  - Re-appear

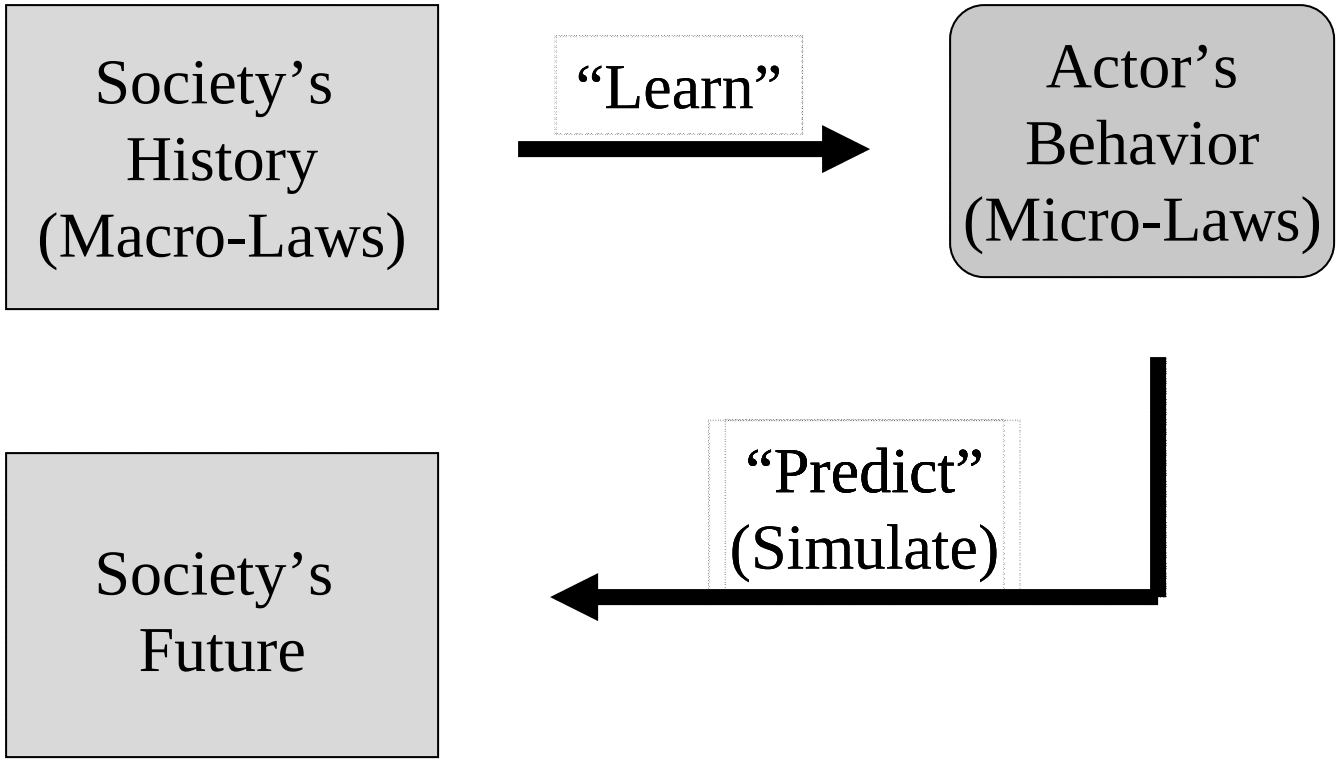
Communication History



Social Group History



Learning and Predicting



Example of Micro-Law

Actor X has a propensity to join *groups.*

Parameter

Micro-Laws

- Actor micro-laws:
  - Probabilistically specify actor decisions.
- Group micro-laws:
  - Probabilistically specify group decisions.

Hidden Markov Model

Society is a probabilistically driven complex system.

$$P(S_{T+1}|\text{micro-laws};S_0,\ldots,S_T)$$

Functions
Parameters

History

Social Capital Theory

Simulation

$$P(S_{T+1} | \text{micro-laws}; S_0, \ldots, S_T)$$

Reverse Engineering

$$P(S_{T+1} | \text{micro-laws}; S_0, \ldots, S_T)$$

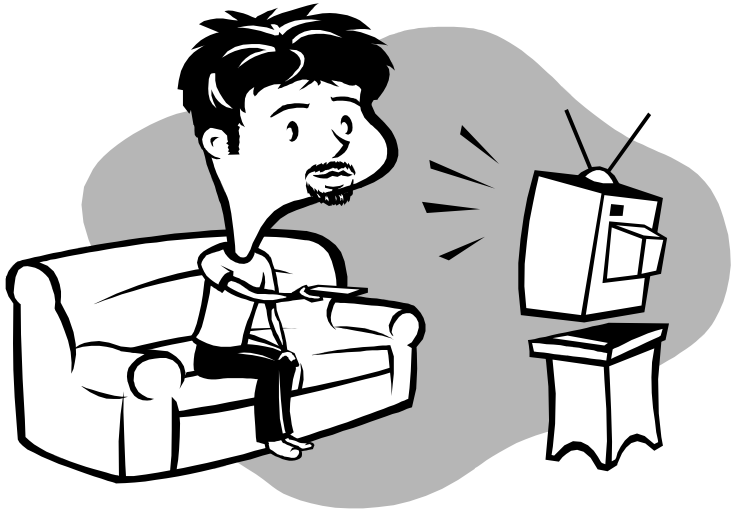
Putnam on Social Capital

"Collapse of social capital in United States communities"

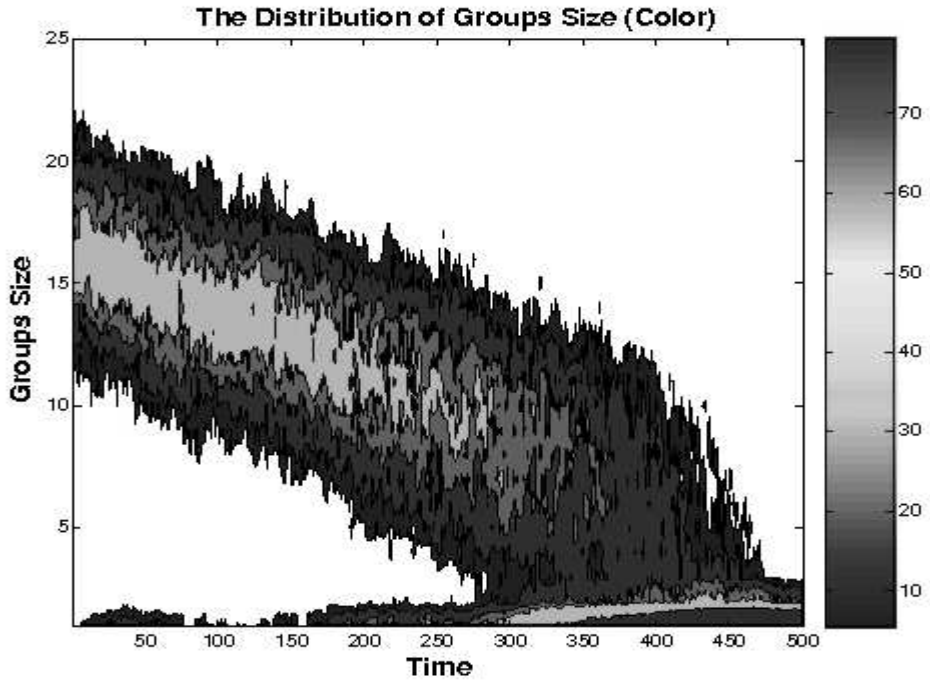
- Actors build social capital by belonging to social groups.

Why?

- Technological innovation?
- Cultural change?
- Demographics change?



Test Such Hypotheses in VSSL



Reverse Engineering

Simulated data – proof of concept.

	Small	Medium	Large
Small	49.2	0.8	0.0
Medium	0.3	73.3	1.5
Large	0.0	3.8	371.2

Newsgroups – actors prefer small groups

[Butler 1999]

Reverse Engineering can…

- Obtain actor preferences (eg. size).
- Determine society reward structure.
- Probabilistic micro-laws governing actor and group dynamics.
- ...

Summary

- Discovering groups in space and time
  - Society's social group history.

- VSSL: Virtual Social Science Lab
  - Simulation: social science experiments.
  - Reverse engineering: learn behavior.

Algorithms, tools, applications (data).

Ongoing Work

- Data
  - Weblogs, Chatrooms, Email (eg. Enron) …
- Finding hidden groups
  - Stream, cycle (“NP-hard”)
- Modeling and reverse engineering
- Visualization
  - Dynamic networks
  - Information visualization (*Knowledgization*)

Thank You

http://www.cs.rpi.edu/~magdon