

Statistical Modeling of Social Groups on Communication Networks

Mark Goldberg (RPI)

Paul Horn (RPI)

Malik Magdon-Ismail (RPI)

Jessie Riposo (RPI)

David Siebecker (RPI)

William Wallace (RPI)

Bulent Yener (RPI)

June 25, 2003

Motivation

We would like to know how social groups behave.

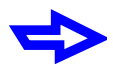
Why?

Curiosity?

Knowing how they behave might help in

- Resource allocation,
- Identifying new emerging groups early
- How communicable diseases spread.
- etc.

We can only observe communications.



INTRODUCTION

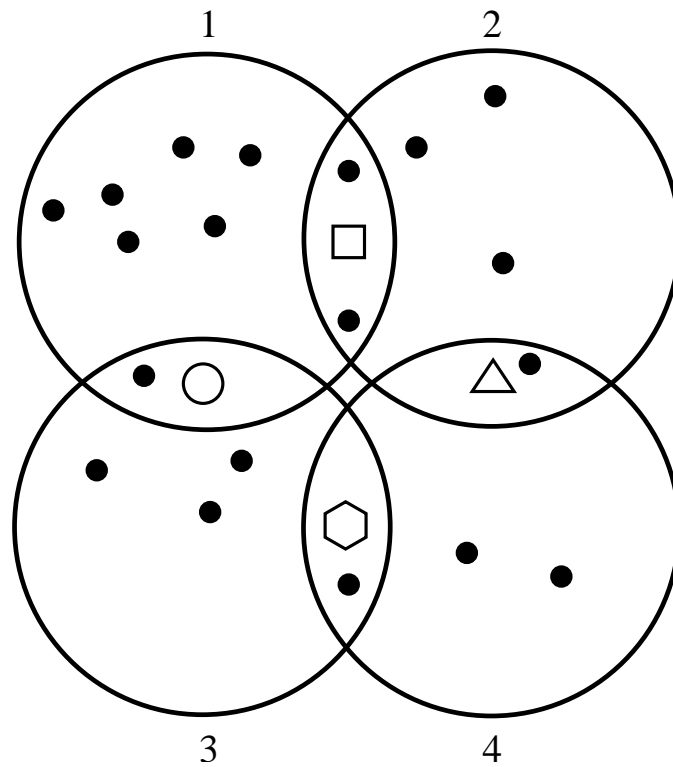
PROBABILISTIC SETUP

ALGORITHMS

EXPERIMENTS

DISCUSSION

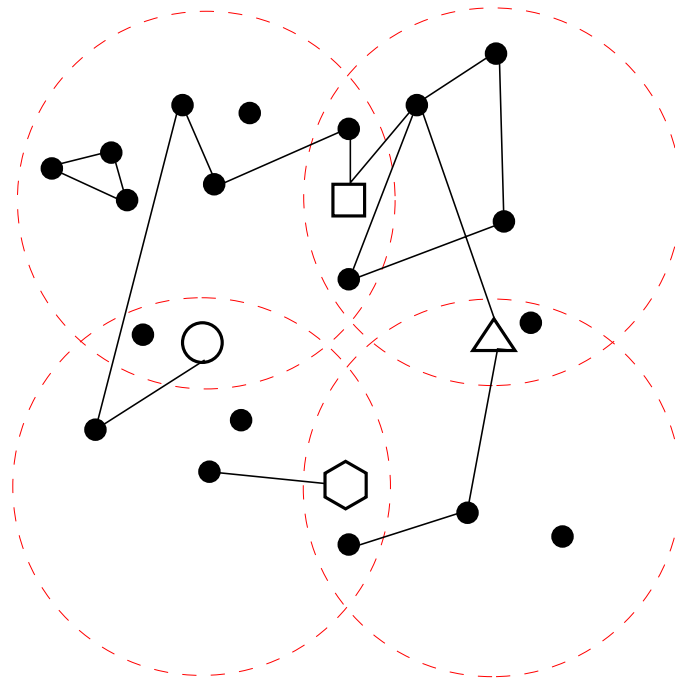
Communication Networks (eg. Newsgroups)



Individuals form groups.

Individuals communicate via an underlying communication medium.

Communication



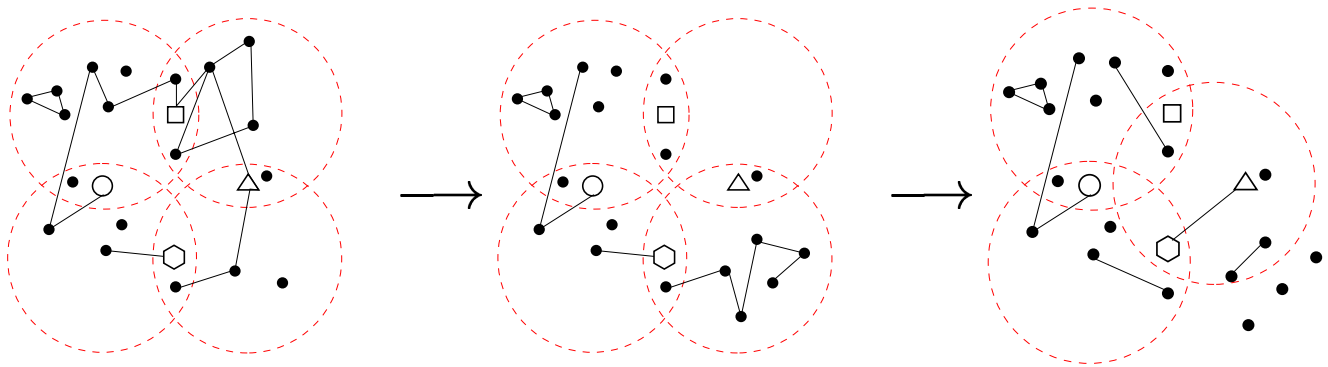
We **know** the groups.

We **do not know** group membership.

We **only observe the communications**.

Infer group membership? (For example members of the same group are more likely to communicate).

Group Dynamics

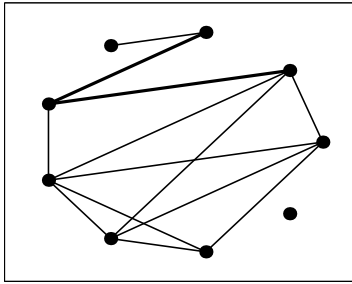


Group membership may evolve.

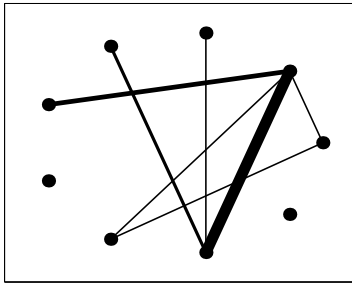
Groups may appear/disappear.

Communications evolve according to group structure.

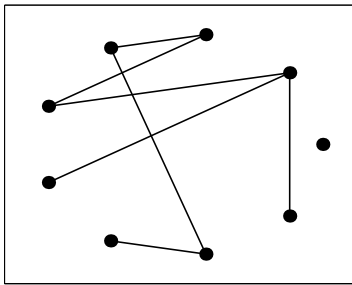
Communication Graph, t=1



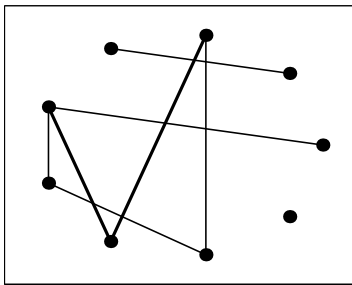
Communication Graph, t=2



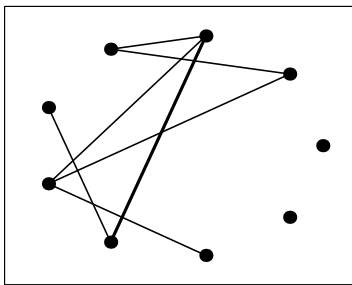
Communication Graph, t=3



Communication Graph, t=4



Communication Graph, t=5



Task

Determine how the social groups and group memberships are evolving.

Micro-Laws

Using observed communications.

Macro-Laws

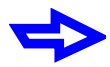
Intuition:

If the communications are not **consistent** with the model then need to change the model.

Consistent? } probabilistic interpretations...



INTRODUCTION



POBABILISTIC SETUP

ALGORITHMS

EXPERIMENTS

DISCUSSION

Micro/Macro-State

Groups F_1, \dots, F_{N_g} .

$\mathbf{F}(t)$, is the **micro-state** matrix.

$$\mathbf{F}_{ij}(t) = \begin{cases} 1 & \text{if node } x_i \text{ is in group } F_j, \\ 0 & \text{otherwise.} \end{cases}$$

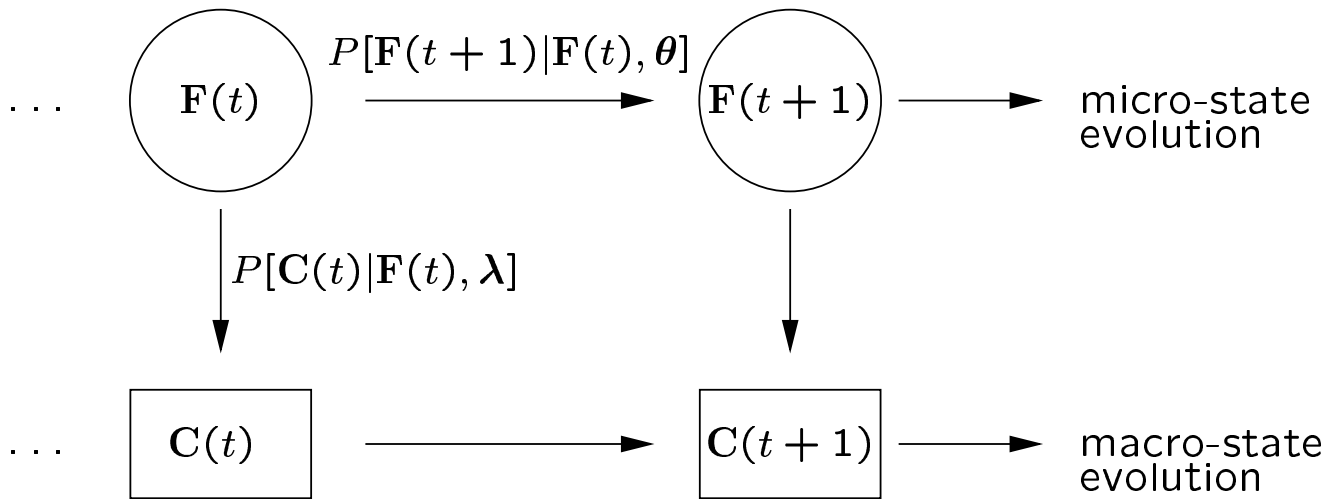
The micro state is **hidden**.

The **communication graph** $\mathbf{C}(t)$ is the **macro-state**,

$$\mathbf{C}_{ij}(t) = \text{Intensity of } i\text{-}j \text{ communication.}$$

The macro-state is observed.

Hidden Markov Model



Micro-state evolves according to Markov process.

Macro-state is determined by micro state
– **Hidden Markov Process.**

Example - Micro-Laws

- Nodes “prefer” to join smaller groups.
- Fixed energy (budget) for activities.
- Nodes join groups by reference.
- Some nodes are more aggressive than others.
- Communication between nodes in same group.

Example – Macro-Laws.

Communication graph obeys power law (Web graph).

Nodes tend to keep their communications balanced.

Task.

We would like to determine the micro-laws from the observed communications.

Focus on a simplified case, the models is general.

- Individual preference for group size.

Model.

Micro-Laws:

Group

Size preference parameter: $-1 \leq \rho \leq 1$.

$$\rho = \begin{cases} 1 & \text{large groups} \\ 0 & \text{medium sized groups} \\ -1 & \text{small groups} \end{cases}$$

Communication: Poisson communications.



INTRODUCTION



POBABILISTIC SETUP

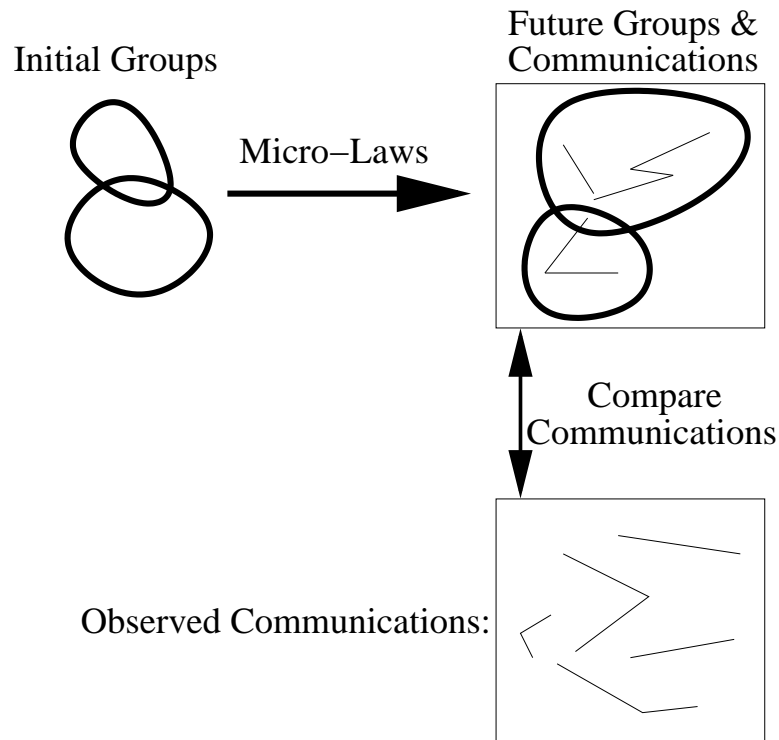


ALGORITHMS

EXPERIMENTS

DISCUSSION

What can we do given a Model



Forward Simulation:

- Micro-laws, initial groups \Rightarrow macro-state evolution.

Model Validation:

- Compare model macro-evolution with observed.

Comparison Metric?

- Probability of observed given model.

Newsgroup Data

Movies and Revisionism Newsgroups.

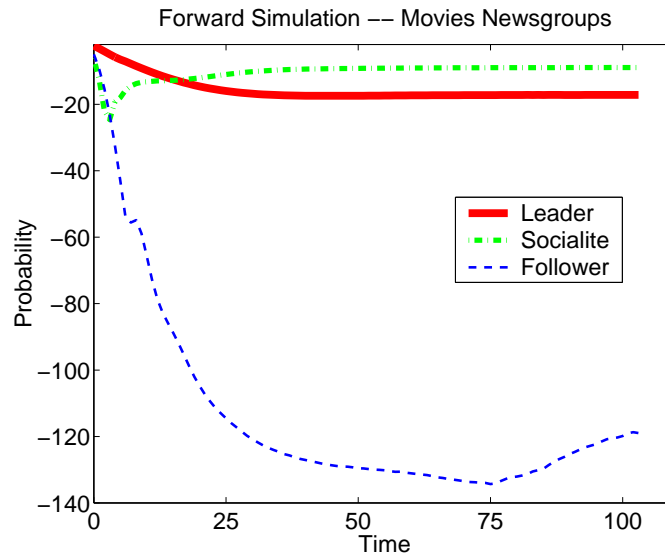
A reply to a posting is a communication.

We compare 3 models:

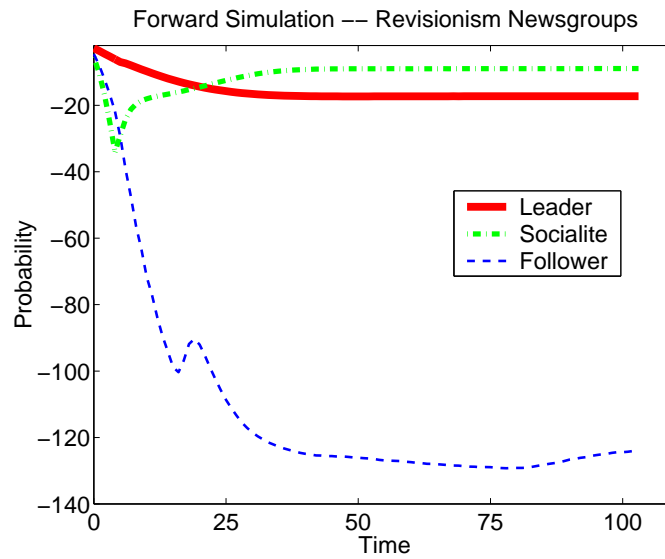
$$\rho = \begin{cases} 1 & \text{"Leaders"} \\ 0 & \text{"Socialites"} \\ -1 & \text{"Followers"} \end{cases}$$

Results

Movies:



Revisionism:



Better Models?

Better model (ρ)?

Different nodes have different ρ ?

Current approach:

Hypothesize model, see if data agrees.

Learning

Learn the model from the data

Reverse Engineering

Find the set of parameters for the model that makes the observed data the most likely.



INTRODUCTION



POBABILISTIC SETUP



ALGORITHMS



EXPERIMENTS

DISCUSSION

Illustrative Example

Two groups.

Two types of nodes p, q .

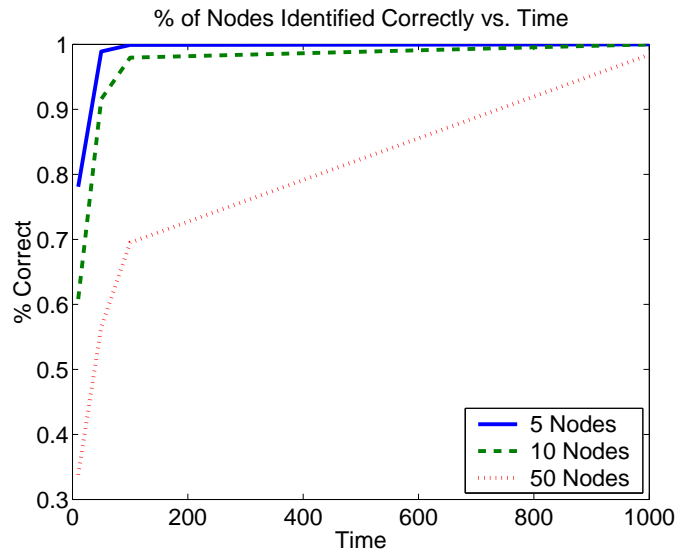
Those who switch frequently p .

Those who switch rarely q .

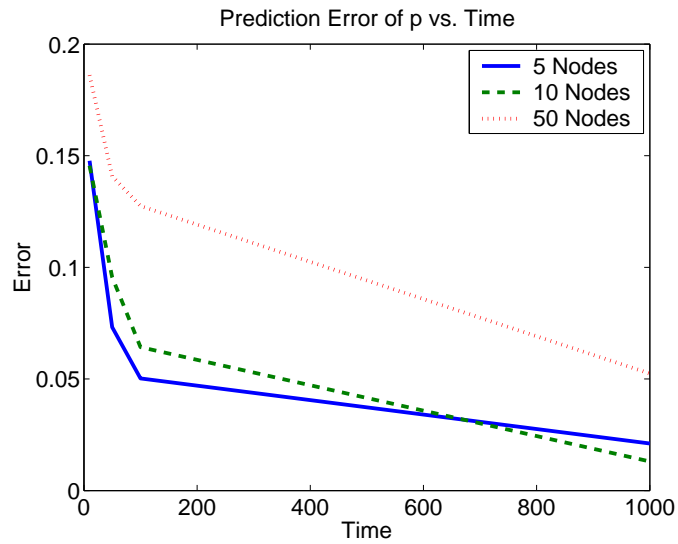
Observe behavior and determine who is p, q .

Learning Accuracy

% correct nodes:



p -error:



Newsgroups Data

Communication is a reply to a posting.

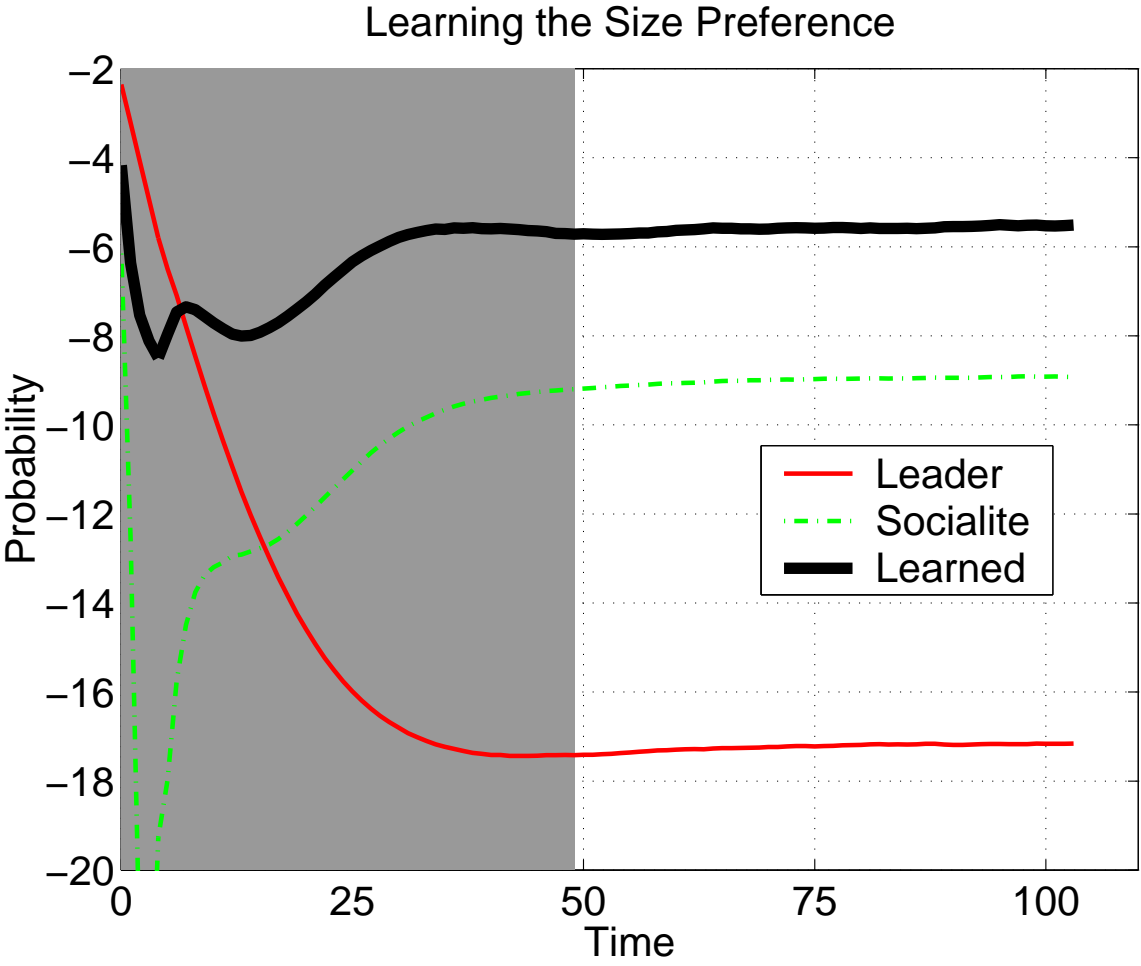
Use some data to determine the model parameters.

Use these parameters to **predict**.

Predict what?

future communications

Newsgroups



Discussion

1. Methodology is general

- given the model (HMM), can simulate.
- Can incorporate a wide range of behavior.
- Model validation.

2. Reverse Engineering / Learning from data

- data determine the model
- optimization heuristics.
- real societies (eg. newsgroups).

3. Parameterized HMM; datas pick the parameters.

Thank You!

Advertisement:

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