Locating Hidden Groups in Communication Networks Using Hidden Markov Models

Malik Magdon-Ismail (RPI)
Mark Goldberg (RPI)
William Wallace (RPI)
David Siebecker (RPI)

June 2 2003

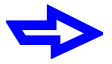
Why?

September 11, 2001

Need tools to detect groups that attempt to hide their communications within a communication network.

Identifying new emerging groups early

Resource allocation, etc.



INTRODUCTION

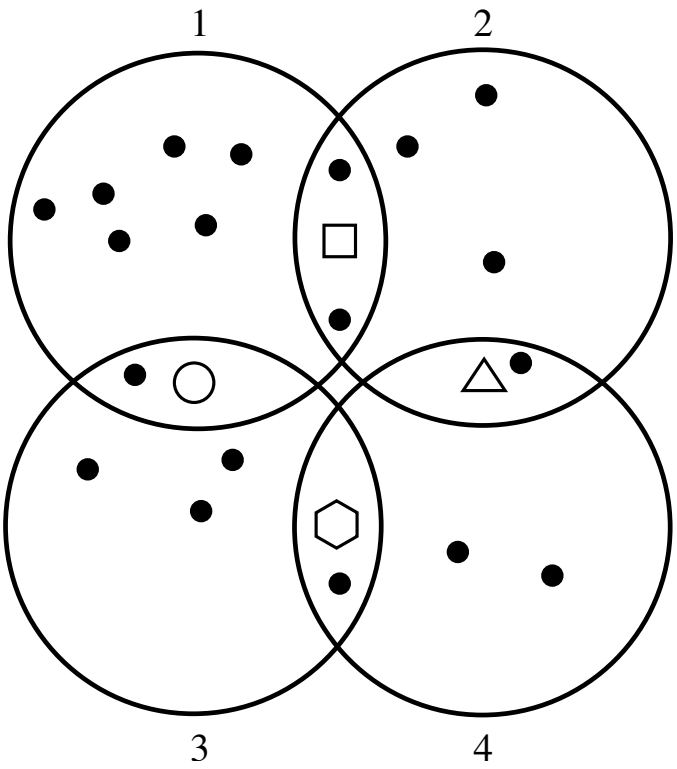
POBABILISTIC SETUP

ALGORITHMS

EXPERIMENTS

DISCUSSION

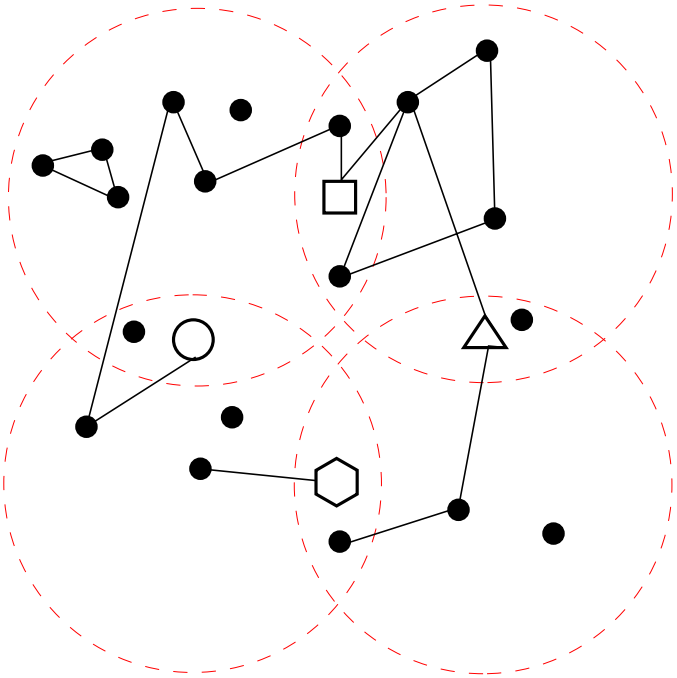
Communication Networks (eg. Newsgroups)



Individuals form groups.

Individuals communicate via an underlying communication medium.

Communication



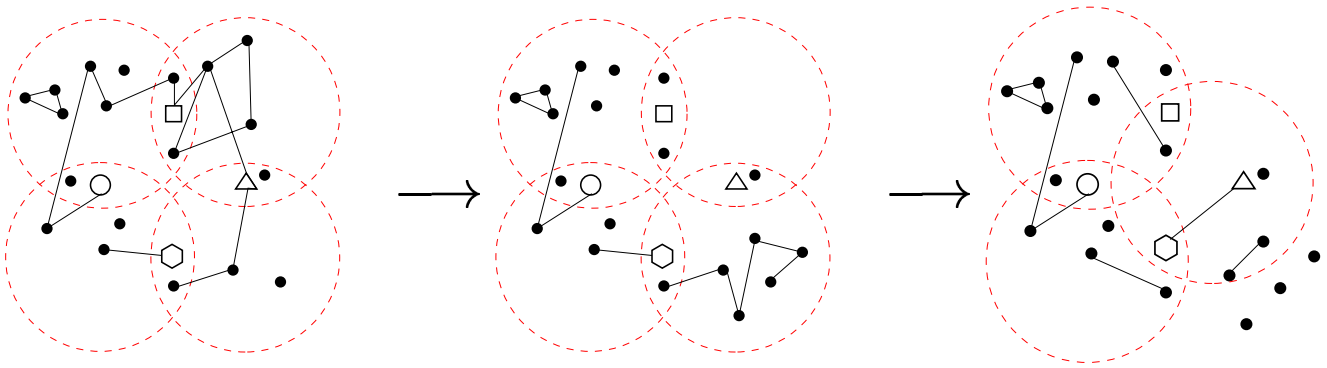
We **know** the groups.

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

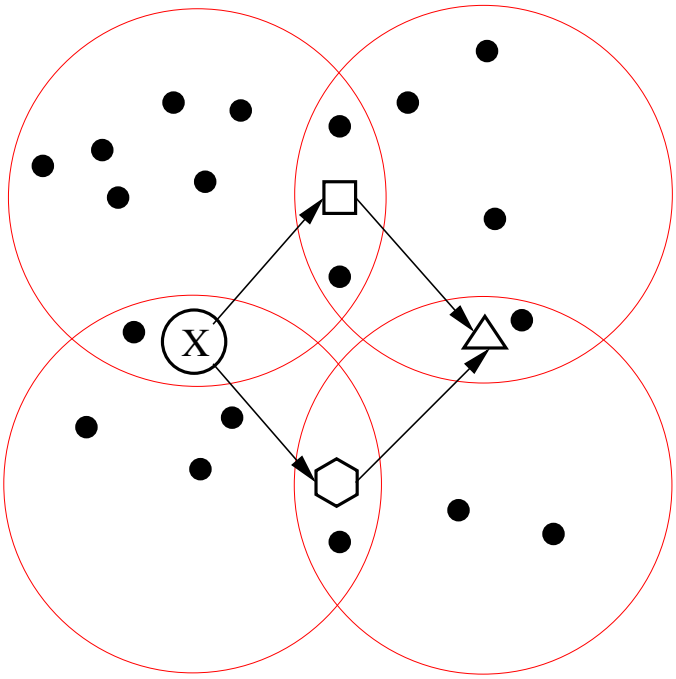


Group membership may evolve.

Groups may appear/disappear.

Communications evolve according to group structure.

Hidden Groups



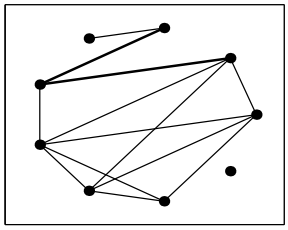
Above is an example of a **Broadcast** hidden group.

Hidden and regular communications are simultaneous.

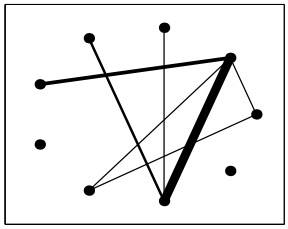
Example

No Hidden Group

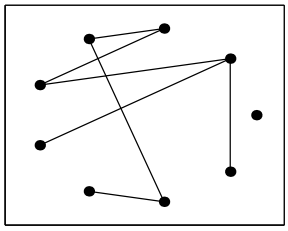
Communication Graph, t=1



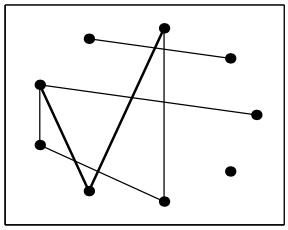
Communication Graph, t=2



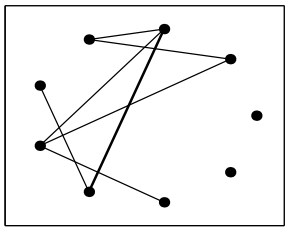
Communication Graph, t=3



Communication Graph, t=4

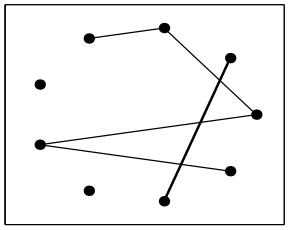


Communication Graph, t=5

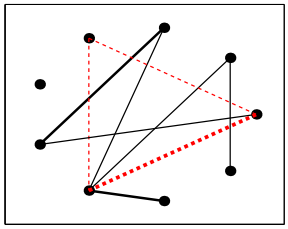


Hidden Group

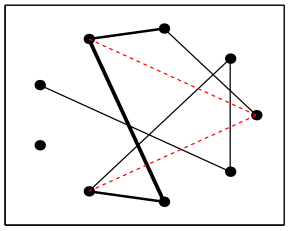
Communication Graph, t=1



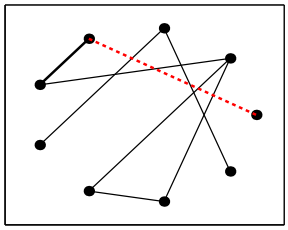
Communication Graph, t=2



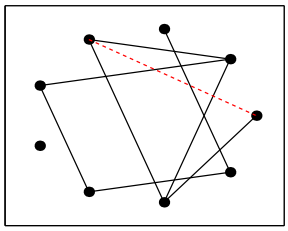
Communication Graph, t=3



Communication Graph, t=4



Communication Graph, t=5



Task

1. Determine whether there is a hidden group.

2. Who the hidden group members are.

Only using observed communications?

Intuition:

If the communications are not **consistent** with the observed groups (for example slightly more intense) then there may be a hidden group.

Consistent?
Slightly more intense? } probabilistic interpretations. . .

Formal Problem Statement

Given:

Communications data.

Model for observed group dynamics and communication.

Model for hidden group dynamics and communication.

Determine:

Whether a hidden group is present.

Who the hidden group members are.

Simplified Example

Observed Group Model:

Dynamics:	**Static**
Communication:	**mostly within group.**

Hidden Group Model:

Dynamics:	**Static**
Communication:	**mostly within group.**
	(emerging groups)
	Broadcast
	(abberent hidden group)

✓ INTRODUCTION

⇨ **POBABILISTIC SETUP**

ALGORITHMS

EXPERIMENTS

DISCUSSION

Micro/Macro-State

Groups $F_1, \ldots, F_{N_g}$.

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

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

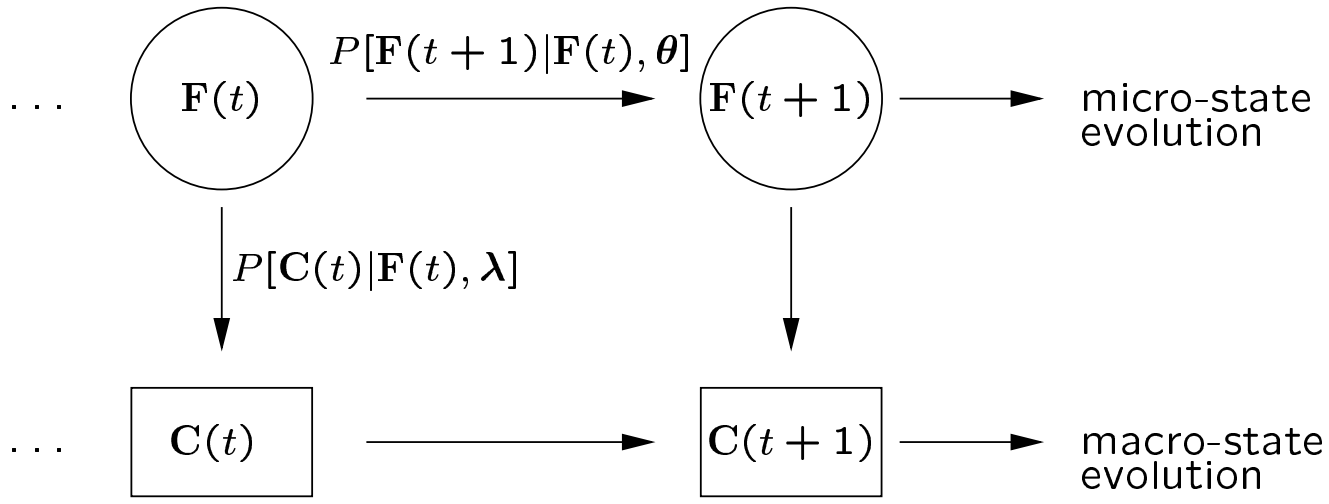
The micro state is **hidden**.

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

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

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

$\mathbf{F}$ is **static**.

(Special case of Markov process).

$\mathbf{C}(t)$ is determined by a **Poisson process**.

$$P[\mathbf{C}_{ij} = k]$$

has Poisson distribution with Poisson parameter proportional to number of groups i, j have in common.

Example – continued.

If there is a hidden group:

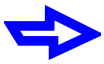
$$\boxed{\mathbf{C}(t) = \mathbf{R}(t) + \mathbf{H}(t)}$$

$\mathbf{R}(t)$ are regular Poisson communications.

$\mathbf{H}(t)$ is the hidden group broadcast communication.

INTRODUCTION

POBABILISTIC SETUP



ALGORITHMS

EXPERIMENTS

DISCUSSION

Maximum Likelihood Approach
(No Hidden Group)

Given $\mathbf{F}$ and the communication model,

$$l(\mathbf{F}) = \textbf{likelihood} = P[\mathbf{C}|\mathbf{F}, model]$$

Pick $\mathbf{F}$ to **maximize** $l(\mathbf{F})$.

$$\begin{aligned} l_0 &= \max_{\mathbf{F}} \ l(\mathbf{F}) \\ \mathbf{F}_0 &= \operatorname*{argmax}_{\mathbf{F}} \ l(\mathbf{F}) \end{aligned}$$

(can also maximize w.r.t. model choice)

Maximum Likelihood Approach (Hidden Group)

Given $\mathbf{F}$ **and the hidden nodes** ($\mathbf{v}$) and the model,

$$l(\mathbf{F}, \mathbf{v}) = \textbf{likelihood} = P[\mathbf{C}|\mathbf{F}, \mathbf{v}, model]$$

(Need to consider all possible decompositions of $\mathbf{C}$.)

Pick $\mathbf{F}_1, \mathbf{v}_1$ to **maximize** l.

$$\begin{aligned} l_1 &= \max_{\mathbf{F},\mathbf{v}} l(\mathbf{F}, \mathbf{v}) \\ \{\mathbf{F}_1, \mathbf{v}_1\} &= \operatorname*{argmax}_{\mathbf{F},\mathbf{v}} l(\mathbf{F}, \mathbf{v}) \end{aligned}$$

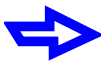
(can also maximize w.r.t. model choice)

$$\boxed{l_0 > l_1 \implies \text{no hidden group}}$$

INTRODUCTION

POBABILISTIC SETUP

ALGORITHMS



EXPERIMENTS

DISCUSSION

Experimental Setup

Small 9 node society.

Generate **many societies**
(with and without hidden group).

Montecarlo optimization with random perturbation

Obtain $l_0, l_1, \mathbf{F}_0, \mathbf{F}_1, \mathbf{v}_1$.

$$\boxed{l_0 > l_1 \implies \text{no hidden group}}$$

(identify hidden nodes, and group structure.)

Results

20 time steps

True H	Predicted H 1	0
1	0.78	**0.28**
0	**0.04**	0.96

% correct =**89%**

50 time steps

True H	Predicted H 1	0
1	0.88	**0.12**
0	**0.03**	0.97

% correct=**94%**

% correct=% nodes identified correctly when H=1.

false negative is generally higher than false positive:

- hidden group is small,
- communications hard to detect.

Discussion

1. **Methodology is general**
   - given the model, can detect hidden group.
   - focused on simplified example.

2. **Proof of concept.**
   - develop optimization heuristics.
   - larger societies
   - real societies (eg. newsgroups).

3. **Where to get model from?**
   - social sciences.
   - **learn it**!

Thank You!

Advertisement:

http://www.cs.rpi.edu/∼magdon