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

POBABLISTIC SETUP

ALGORITHMS

EXPERIMENTS

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

$F(t)$, is the **micro-state** matrix.

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

The micro state is **hidden**.

The **communication graph** $C(t)$ is the **macro-state**,

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

The macro-state is observed.
Hidden Markov Model

Micro-state evolves according to Markov process.

Macro-state is determined by micro state
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.
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 ⇒ macro-state evolution.

Model Validation:
- Compare model macro-evolution with observed.

Comparison Metric?
- Probability of observed given model.
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:

![Graph showing the probability over time for movies newsgroups]

Revisionism:

![Graph showing the probability over time for revisionism newsgroups]
Better Models?

Better model ($\rho$)?

Different nodes have different $\rho$?

Current approach:
Hypothesize model, see if data agrees.
Learn the model from the data
Find the set of parameters for the model that makes the observed data the most likely.
✓ INTRODUCTION
✓ POBABLISTIC 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:

% of Nodes Identified Correctly vs. Time

$p$-error:

Prediction Error of $p$ vs. Time
Communication is a reply to a posting.

Use some data to determine the model parameters.

Use these parameters to predict.

Predict what?
  future communications
Learning the Size Preference

Time

Probability

Leader
Socialite
Learned

Student Version of MATLAB
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