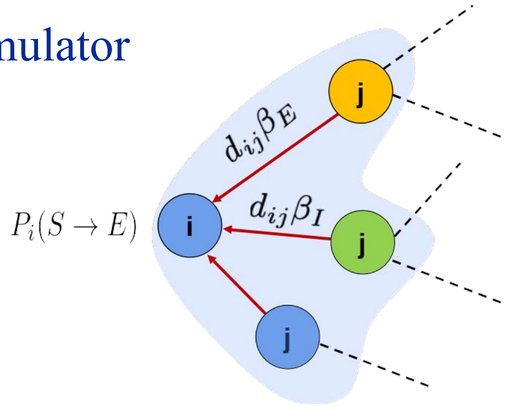
Inference in Network-based Epidemiological Simulations with Probabilistic Programming

Niklas Smedemark-Margulies[1], Robin Walters[1], Heiko Zimmermann[1], Lucas Laird, Neela Kaushik, Rajmonda Caceres and Jan-Willem van de Meent

AI for Public Health
ICLR 2021 Workshop

Introduction

- ➢ Designing interventions for pandemics requires a realistic simulator
  - ○ Complex enough to fit real-world dynamics
  - ○ Accurately fit to real-world data
  - ○ *Hard to do both at once.*

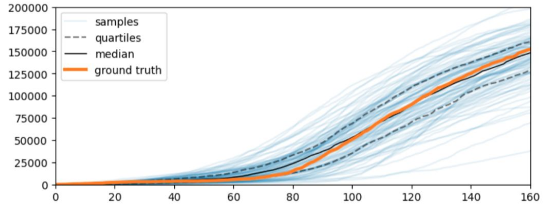


- ➢ Our disease model: Network-SEIR
  - ○ Mobility network derived from cell phone data
  - ○ Agent-based transmission

- ➢ Parameter estimation by probabilistic programming
  - ○ Black-box variational inference
  - ○ Disease parameters, initial exposure patterns

$$p(\beta \mid \text{data}) = \frac{p(\text{data} \mid \beta)\, p(\beta)}{p(\text{data})}$$

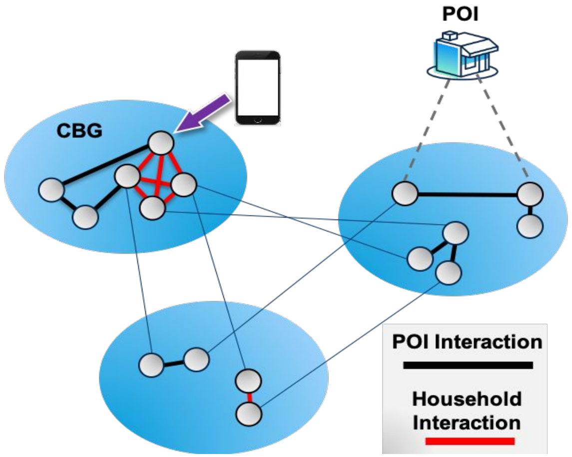
- ➢ Better fit to real data
  - ○ Model multi-peak dynamics
  - ○ Replicates actual statistics for different regions

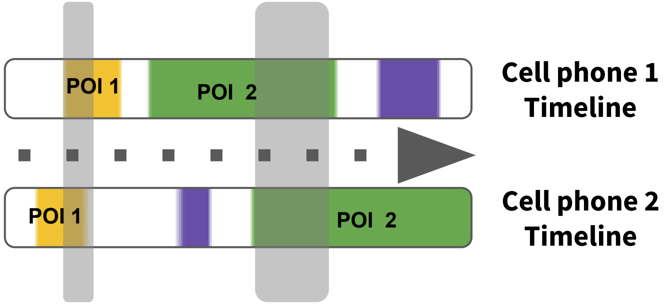


Mobility Networks as Stochastic Block Models

SAFEGRAPH

- ➢ Smartphone location data at points-of-interests (**POI**)[1]
- ➢ Each phone mapped to "home" Census Block Group (**CBG**)
- ➢ Simulate geographic contact patterns using **DCSBM**[2]
    - ○ Nodes are individuals; stochastic blocks are **CBGs**
    - ○ **POI** interaction probability based on shared **POI** visit counts
    - ○ Weight of interactions captures the duration of co-location





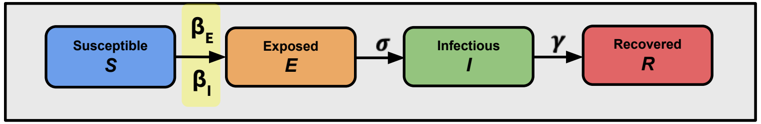
[1]https://www.safegraph.com/

[2] DCSBM: Degree Corrected Stochastic Block Model: https://en.wikipedia.org/wiki/Stochastic_block_model

Agent-Based Disease Simulator

- Compartmental models

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \qquad \frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$



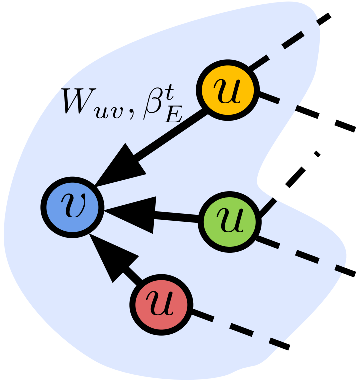
- Our approach: Network-based SEIR Model
  - Poisson modeling of exposures:

$$P(v \in E_{t+1} | v \in S_t, u \in E_t) = 1 - \exp(-W_{uv}\beta_E^t)$$

  - Model local network effects

$$P(v \in E_{t+1} | v \in S_t) = 1 - \exp\left(-\left[\sum_{u \in N_E(v)} W_{uv}\beta_E^t + \sum_{u \in N_I(v)} W_{uv}\beta_I^t\right]\right)$$

$$\approx \min\left\{1, \left[\sum_{u \in N_E(v)} W_{uv}\beta_E^t + \sum_{u \in N_I(v)} W_{uv}\beta_I^t\right]\right\}$$



Agent-Based Disease Simulator

Function $f_{\mathrm{SEIR}}\,(G,\, \alpha^{1:C},\, \{\beta_E^{t_1} \ldots \beta_E^{t_K}\},\, \{\beta_I^{t_1} \ldots \beta_I^{t_K}\}, \gamma,\, \lambda,\, T)$ **:**

Agent-Based Disease Simulator

Function f_{SEIR} $(G, \alpha^{1:C}, \{\beta_E^{t_1} \dots \beta_E^{t_K}\}, \{\beta_I^{t_1} \dots \beta_I^{t_K}\}, \gamma, \lambda, T)$ **:**

    for $c \leftarrow 1$ **to** C **do** `// Initial Exposure`

        for $v \in \mathcal{V}^c$ **do** **if** $\text{Unif}(0,1) < \alpha^c$ **then** $v \rightarrow E^1$ **else** $v \rightarrow S^1$

Agent-Based Disease Simulator

Function f_{SEIR} $(G, \alpha^{1:C}, \{\beta_E^{t_1} \ldots \beta_E^{t_K}\}, \{\beta_I^{t_1} \ldots \beta_I^{t_K}\}, \gamma, \lambda, T)$**:**

```
for c ← 1 to C do                                                    // Initial Exposure
    for v ∈ V^c do  if Unif(0,1) < α^c then v → E^1 else v → S^1
for t ← 1 to T−1 do                                                  // Simulate T days
    β_E^t ← INTERPOLATE(β_E^{t_1}, ..., β_E^{t_K}) ;   β_I^t ← INTERPOLATE(β_I^{t_1}, ..., β_I^{t_K})
```

for $c \leftarrow 1$ **to** C **do** // Initial Exposure

for $v \in \mathcal{V}^c$ **do** **if** $\mathrm{Unif}(0,1) < \alpha^c$ **then** $v \rightarrow E^1$ **else** $v \rightarrow S^1$

for $t \leftarrow 1$ **to** $T-1$ **do** // Simulate T days

$\beta_E^t \leftarrow \textsc{Interpolate}(\beta_E^{t_1}, \ldots, \beta_E^{t_K})$; $\beta_I^t \leftarrow \textsc{Interpolate}(\beta_I^{t_1}, \ldots, \beta_I^{t_K})$

Agent-Based Disease Simulator

```
Function f_SEIR (G, α^{1:C}, {β_E^{t_1} … β_E^{t_K}}, {β_I^{t_1} … β_I^{t_K}}, γ, λ, T):
    for c ← 1 to C do                                           // Initial Exposure
        for v ∈ V^c do  if Unif(0,1) < α^c then v → E^1 else v → S^1
    for t ← 1 to T−1 do                                         // Simulate T days
        β_E^t ← INTERPOLATE(β_E^{t_1}, …, β_E^{t_K}) ;   β_I^t ← INTERPOLATE(β_I^{t_1}, …, β_I^{t_K})
        for v ∈ S_t do
            E_pressure ← Σ_{u∈N_E^t(v)} W_uv β_E^t ;   I_pressure ← Σ_{u∈N_I^t(v)} W_uv β_I^t
            if Unif(0,1) < (E_pressure + I_pressure) then v → E^{t+1}
```

Agent-Based Disease Simulator

Function $f_{\mathrm{SEIR}}\,(G,\, \alpha^{1:C},\, \{\beta_E^{t_1} \ldots \beta_E^{t_K}\},\, \{\beta_I^{t_1} \ldots \beta_I^{t_K}\}, \gamma,\, \lambda,\, T)$:

```
    for c ← 1 to C do                                              // Initial Exposure
        for v ∈ V^c do  if Unif(0,1) < α^c then v → E^1 else v → S^1
    for t ← 1 to T−1 do                                            // Simulate T days
        β_E^t ← INTERPOLATE(β_E^{t_1}, ..., β_E^{t_K});   β_I^t ← INTERPOLATE(β_I^{t_1}, ..., β_I^{t_K})
        for v ∈ S_t do
            E_pressure ← Σ_{u∈N_E^t(v)} W_uv β_E^t;   I_pressure ← Σ_{u∈N_I^t(v)} W_uv β_I^t
            if Unif(0,1) < (E_pressure + I_pressure) then v → E^{t+1}
        for v ∈ E^t do if Unif(0,1) < γ then v → I^{t+1}
        for v ∈ I^t do if Unif(0,1) < λ then v → R^{t+1}
```

Agent-Based Disease Simulator

Function $f_{\text{SEIR}}\,(G,\,\alpha^{1:C},\,\{\beta_E^{t_1}\ldots\beta_E^{t_K}\},\,\{\beta_I^{t_1}\ldots\beta_I^{t_K}\},\gamma,\,\lambda,\,T)$ **:**

```
for c ← 1 to C do                                                   // Initial Exposure
    for v ∈ 𝒱^c do if Unif(0,1) < α^c then v → E^1 else v → S^1
for t ← 1 to T−1 do                                                 // Simulate T days
    β_E^t ← INTERPOLATE(β_E^{t_1}, …, β_E^{t_K});   β_I^t ← INTERPOLATE(β_I^{t_1}, …, β_I^{t_K})
    for v ∈ S_t do
        E_pressure ← Σ_{u∈N_E^t(v)} W_uv β_E^t;   I_pressure ← Σ_{u∈N_I^t(v)} W_uv β_I^t
        if Unif(0,1) < (E_pressure + I_pressure) then v → E^{t+1}
    for v ∈ E^t do if Unif(0,1) < γ then v → I^{t+1}
    for v ∈ I^t do if Unif(0,1) < λ then v → R^{t+1}
return (Σ_{t=1}^{j} I^t)_{j=1}^{T}                                  // List of Cumulative Infections
```

Agent-Based Disease Simulator

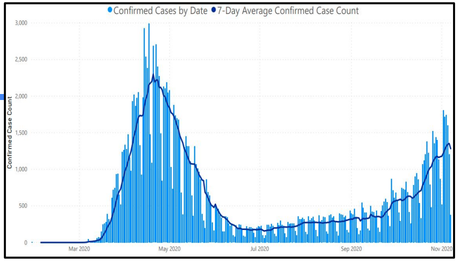
```
Function f_SEIR (G, α^{1:C}, {β_E^{t_1} ... β_E^{t_K}}, {β_I^{t_1} ... β_I^{t_K}}, γ, λ, T):
    for c ← 1 to C do                                        // Initial Exposure
        for v ∈ V^c do  if Unif(0,1) < α^c then v → E^1 else v → S^1
    for t ← 1 to T−1 do                                      // Simulate T days
        β_E^t ← INTERPOLATE(β_E^{t_1}, ..., β_E^{t_K}) ;   β_I^t ← INTERPOLATE(β_I^{t_1}, ..., β_I^{t_K})
        for v ∈ S_t do
            E_pressure ← Σ_{u∈N_E^t(v)} W_uv β_E^t ;   I_pressure ← Σ_{u∈N_I^t(v)} W_uv β_I^t
            if Unif(0,1) < (E_pressure + I_pressure) then v → E^{t+1}
        for v ∈ E^t do if Unif(0,1) < γ then v → I^{t+1}
        for v ∈ I^t do if Unif(0,1) < λ then v → R^{t+1}
    return (Σ_{t=1}^{j} I^t)_{j=1}^{T}                      // List of Cumulative Infections
```

Bayesian Inference

$$p(\beta \mid data) = \frac{p(data \mid \beta)\, p(\beta)}{\underbrace{\int p(data \mid \beta)\, p(\beta)\, d\beta}_{\textbf{intractable}}}$$

data

JHU Reported Infection Counts



$p(\beta)$ - prior belief about disease parameters

$p(data \mid \beta)$ - probability of observed data given disease parameters

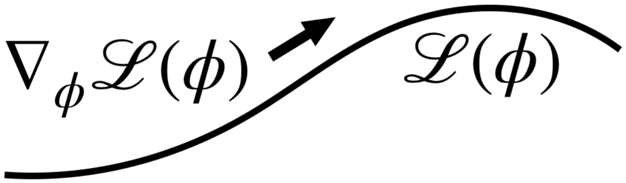
Stochastic Variational Inference

➢ Approximate intractable posterior with variational distribution q_ϕ

$$\phi^* = \arg\min_\phi \; \mathrm{KL}(q_\phi(\cdot) \,||\, \underbrace{p(\cdot \mid data)}_{\textbf{intractable}})$$

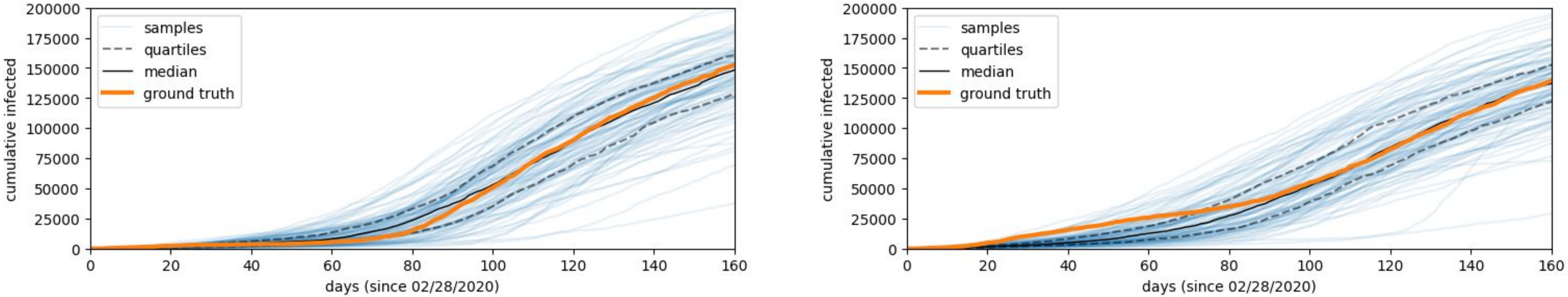
$$= \arg\max_\phi \; \mathscr{L}(\phi) \leftarrow \textbf{tractable surrogate objective (ELBO)}$$

➢ Optimization via stochastic gradient ascent



$$\phi_{t+1} \leftarrow \phi_t + \alpha_t \nabla_\phi \mathscr{L}(\phi_t)$$

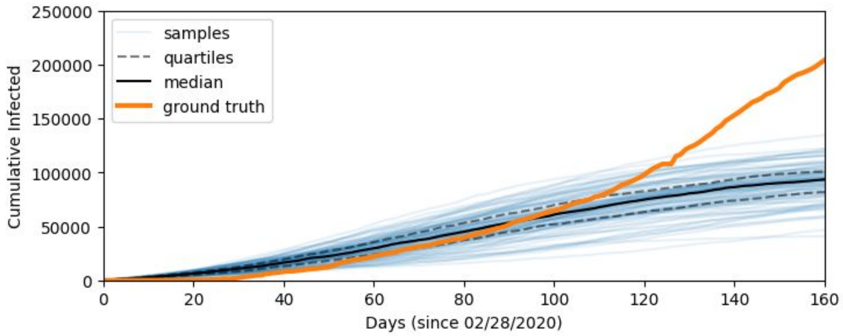
Fit Time-Varying Infection Rates



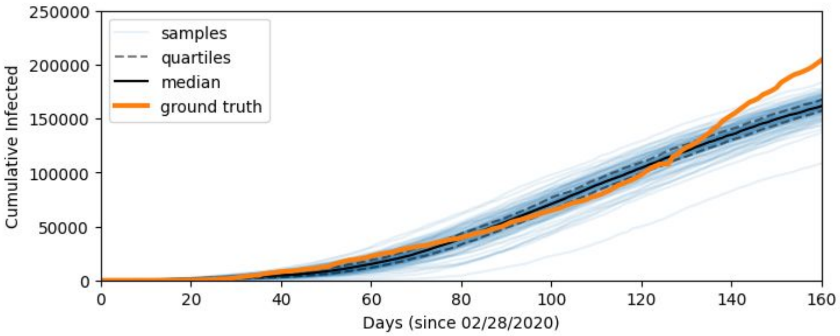
County	low	high	low-high	high-low	low-high-low	high-low-high
Miami-Dade	0.0052	0.0046	0.0042	0.0051	0.0043	0.0050
Los Angeles	0.0037	0.0046	0.0050	0.0044	0.0048	0.0047

$$\text{MDAE} \equiv \mathop{\mathbb{E}}_{q_\phi(z)}\left[\frac{\|f_{\text{SEIR}}(z) - x\|_1}{TN}\right] \approx \frac{1}{N}\frac{1}{ST}\sum_s \sum_t \left|f_{\text{SEIR}}(z_s)^t - x^t\right|$$

Better than Baselines



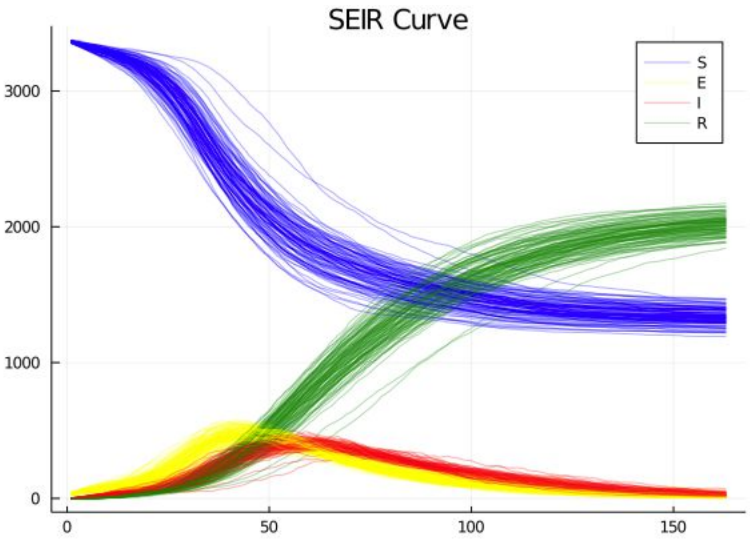
R_t-Analytic Baseline



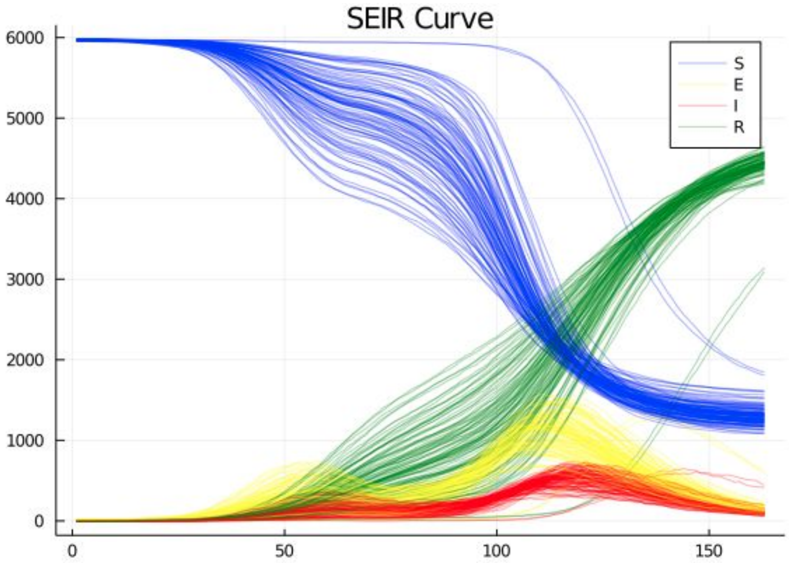
Our Method

Disease Model	Fitting Method	LA-MDAE	Miami-Dade-MDAE
Compartmental	CE-EM	0.0251	0.0161
Network	R_t-Analytic	0.0075	0.0086
Network	BBVI	**0.0029**	**0.0053**

Fit Different Regions



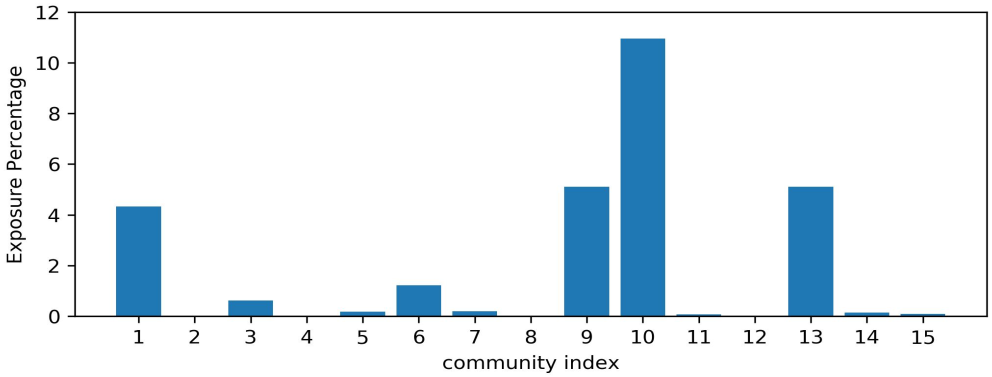
Middlesex



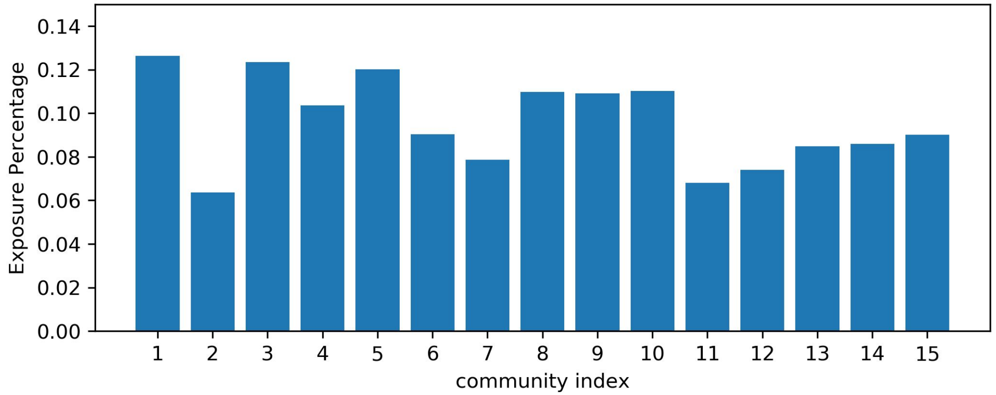
Los Angeles

Infer Starting Communities

Miami - tight noise



Miami - loose noise



Thank you!

- Paper ID: 36
- Email: smedemark-margulie.n@northeastern.edu
- Acknowledgements:

MIT Supercloud
Education and Outreach
DARPA
Northeastern University
Research Computing
intel
MIT
LINCOLN
LABORATORY