Application of Spatial Mixing Models to Sampled Data from Facebook

Emma S. Spiro¹ Zack W. Almquist¹ Carter T. Butts^{1,2}

¹Department of Sociology ²Institute for Mathematical Behavioral Sciences University of California – Irvine

Presented at MURI Meeting June 3, 2011

This material is based on research supported by the Office of Naval Research under award N00014-08-1-1015.

University of California, Irvine

MURI Themes

- Theoretical foundation and substantive problems
- Statistical methods
- Fast algorithms and new data structures
- Rich models of large-scale data with complex covariates

▲□▶ ▲□▶ ▲臣▶ ★臣▶ 三臣 - のへで

Spatial Bernoulli Graphs, (Butts 2002)

 A simple family of models for spatially embedded social networks

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{D}) = \prod_{\{i,j\}} B(Y_{ij} = y_{ij} | \mathcal{F}_d(D_{ij}))$$
(1)

►
$$\mathbf{Y} \in \{0, 1\}^{N \times N}$$

► $\mathbf{D} \in [0, \infty)^{N \times N}$

- $\mathcal{F}_d: [0,\infty) \mapsto [0,1]$
- Assumes that dependence among edges is absorbed by the distance structure – edges conditionally independent.
- This model can be expressed in ERG form.

Spatial Interaction Function

Decay as a power law in distance

$$\mathcal{F}_d(x) = \frac{p_b}{(1+\alpha x)^{\gamma}}$$

where $0 \le p_b \le 1$ is a baseline tie probability, $\alpha \ge 0$ is a scaling parameter, and $\gamma > 0$ is the exponent which controls the distance effect

Attenuated power law, arctangent decay, etc.



Spatial Bernoulli Models with Covariates

- We can extend the model in a simple way to include tie covariates
- Add GLM structure to the parameters of the SIF, \mathcal{F}_d

$$\mathsf{Pr}(Y_{ij}=1) = rac{oldsymbol{p}_{bij}}{(1+lpha_{ij}oldsymbol{d}_{ij})^{\gamma_{ij}}}$$

where

$$p_{b_{ij}} = ilogit(\theta * X_{ij})$$
$$\alpha_{ij} = exp(\psi * W_{ij})$$
$$\gamma_{ij} = exp(\phi * U_{ij})$$

and where θ , ψ , and ϕ are parameter vectors, and **X**, **W**, and **U** are covariate matrices.

Application: Selective Mixing on Facebook

- Facebook is an extremely large online social network
- Data: sample of almost 1 million egocentric networks (Gjoka et al. 2010)
- Each Facebook user may indicate a university affiliation,
 < 4% actually do
- Rich set of covariates at the institution level
- Online context is a best case scenario for equal mixing and "weak" distance effects

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

Model Fitting and Selection

SIF:
$$\mathcal{F}_d(x) = \frac{p_b}{(1+\alpha x)^{\gamma}}$$



◆□> ◆□> ◆臣> ◆臣> ─臣 ─ のへで



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 三臣 - のへ⊙



E. Spiro espiro@uci.edu

University of California, Irvine

June 3, 2011

A Model of Facebook Friendship

Parameter	Covariate	Estimate	s.e.	
p _b	Intercept	-6.214	0.0048	**
	Public-Private	0.080	0.0325	**
	Public-Public	-0.610	0.0050	**
	Prestige Difference	-0.019	0.0001	**
α	Intercept	1.224	0.0248	**
	Public-Private	-0.376	0.0753	**
	Public-Public	-3.435	0.0260	**
	Prestige Difference	-0.02	0.0001	**
γ	Intercept	-1.006	0.0020	**
	Public-Private	0.358	0.0029	**
	Public-Public	0.901	0.0026	**

A Model of Facebook Friendship



E. Spiro espiro@uci.edu

University of California, Irvine

June 3, 2011

Behavior at the Origin



University of California, Irvine

June 3, 2011

Discussion

- Spatial mixing models to sampled data
- Model extension to include covariates
- Non-trivial model fitting procedure
- Inhomogeneous relationship between distance and tie probability
- Educational inequality and social stratification



E. Spiro espiro@uci.edu

University of California, Irvine

June 3, 2011

э

イロン 不良と 不良とう



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 三臣 - のへ⊙



E. Spiro espiro@uci.edu

University of California, Irvine

June 3, 2011

э

イロト イヨト イヨト イヨト