

# Population Sized Graphical Record Linkage

*Andee Kaplan and Rebecca C. Steorts*

## Introduction

### Prior work

## Bayesian Record Linkage

### Notation

### Generalized Model

### Empirically Motivated Priors

## Post-linkage Analysis

### General Class of Model

### Application: Population Estimation

## Experiments

On all datasets, pair results from each record linkage method with each CRC method.

Potential additions: blocking, bootstrap?

### Data sets

- NLTKS
- cora
- cd
- voter

### Record linkage methods

- unsupervised: Fellegi-Sunter, blink, fastLink, Sadinle
- supervised: logistic regression, random forests, svm, tree, adaboost, bagging, nnet, bumping

### Capture-recapture methods

- Johndrow
- Bayesian model averaging (Madigan and York (1997))

## Discussion