

---

# Método Bayesiano para Clusterização Espaço-Tempo

---

**Marcelo Azevedo Costa**

**Sérgio Henrique Rodrigues Ribeiro**

# Bayesian Detection of Clusters and Discontinuities in Disease Maps

Leonhard Knorr-Held and Günter Raßer  
Biometrics 56, 13-21 (2000)



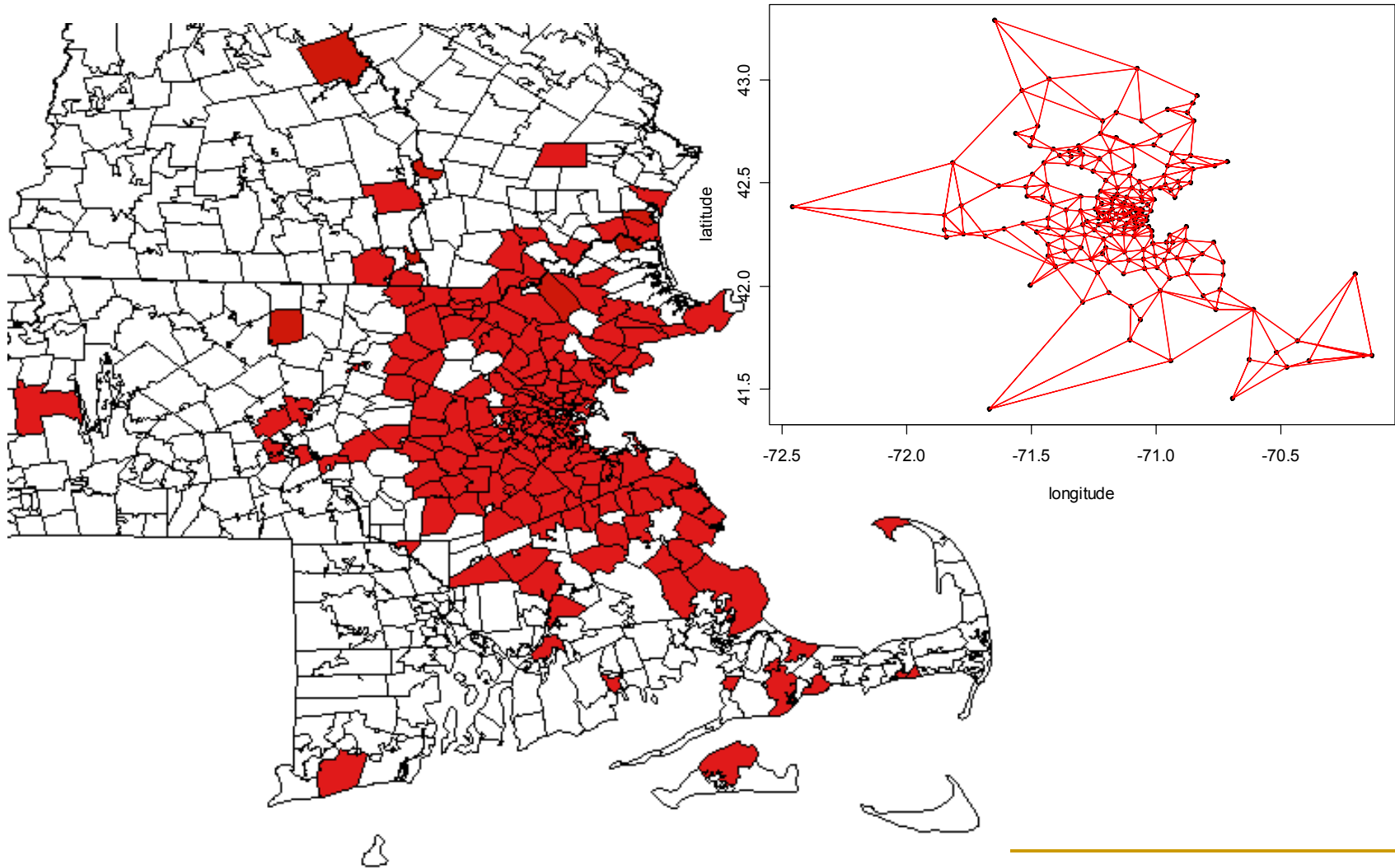
$y_i$  número de casos na  $i$ -ésima área,  $i = 1, \dots, n$ .

$e_i$  número esperado de casos na  $i$ -ésima área.

$k$  número de partições do mapa,  $k$  in  $(1, \dots, n)$ .

A cada partição está associada um conjunto de áreas contíguas que possuem um mesmo risco relativo,  $h_j$

# A Representação por Grafos



---

# O Modelo

$$y_i \sim \text{Poisson}(h_j \cdot e_i)$$

$$L(y|H_k) = \prod_{j=1}^k \prod_{i \in C_j} \frac{(e_i h_j)^{y_i}}{y_i!} \exp(-e_i h_j)$$

**Prioris:**

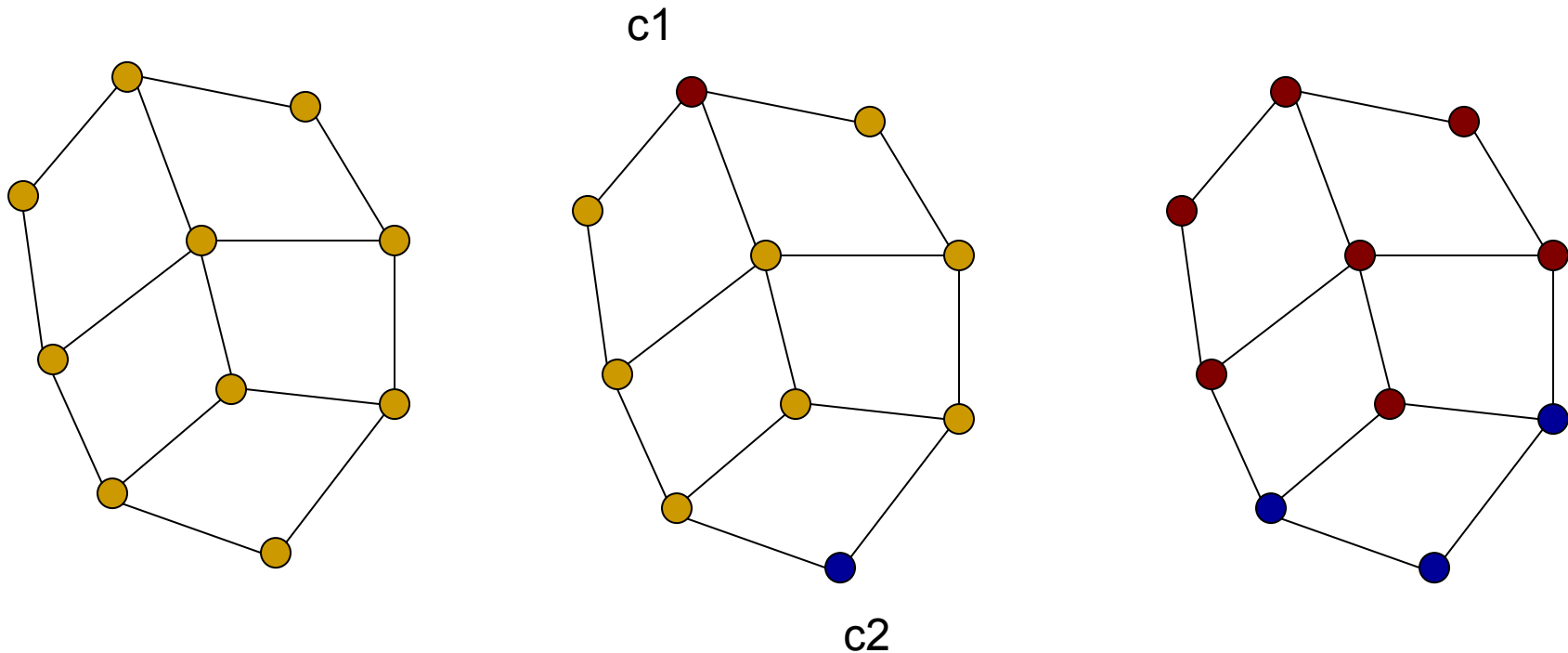
$$P(k) \propto (1-c)^k$$

$$h_j \sim \log \text{Normal}(\mu, \sigma^2)$$

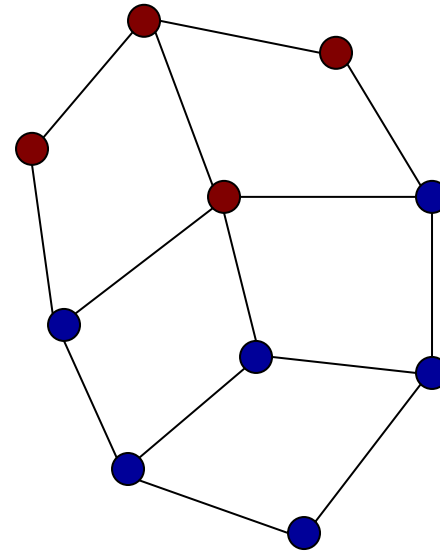
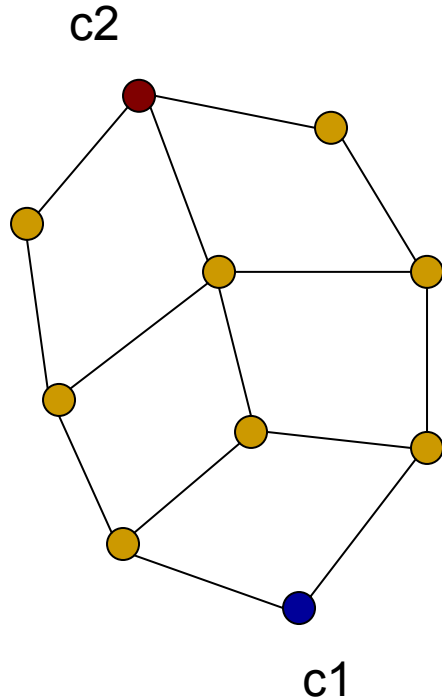
---

# Definindo as partições

- Uma partição é definida pelo seu centro e pela sua “prioridade” ou ordem em um vetor de centros.



# Definindo as partições (b)



Ou seja a dominância de uma partição depende da sua posição relativa no “vetor de centros”

# Gerando amostras das posteriores via Reversible Jump MCMC

- A cada iteração do algoritmo um dos seguintes movimentos é proposto:
  - **Birth**: Um novo centro é criado, aumentando em uma unidade o número de partições;
  - **Death**: Um centro é deletado, diminuindo de uma unidade o número de partições;
  - **Shift**: A posição de um centro de uma partição é alterada;
  - **Switch**: A posições relativa de dois centros é alternada;
  - **Height**: Os riscos relativos são atualizados (Metropolis-Hastings)
  - **Hyper**: Os hiperparâmetros são atualizados

**Cada movimento pode ou não ser aceito com uma certa probabilidade.**

---

# Algumas Alterações Propostas

- Trocamos a priori logNormal por uma Gama
- Retiramos os Hyper-Parâmetros;
- Somente dois movimentos são testados: **Birth** e **Death**;
- Sempre ao final de um movimento (**Birth** ou **Death**), é realizado um passo de Gibbs para a atualização de TODOS os riscos relativos.

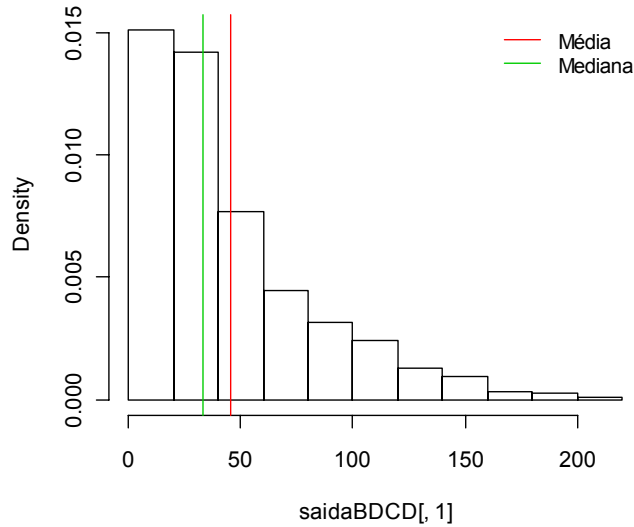
**Implementamos o código em C\C++ para garantir velocidade de execução.**

---

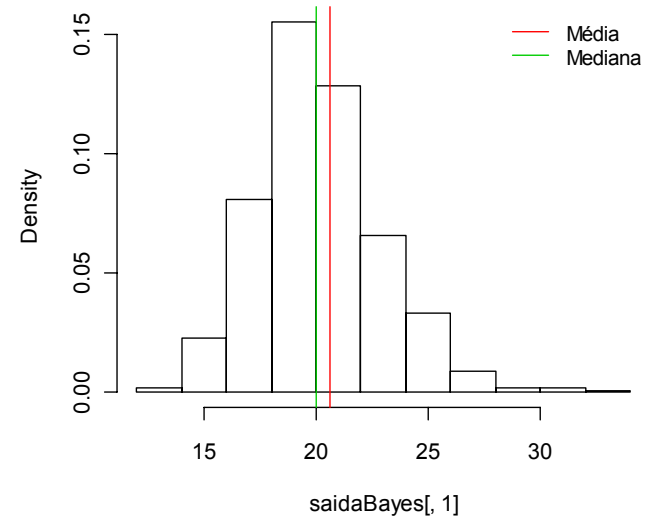


# Posteriori do Número de Partições

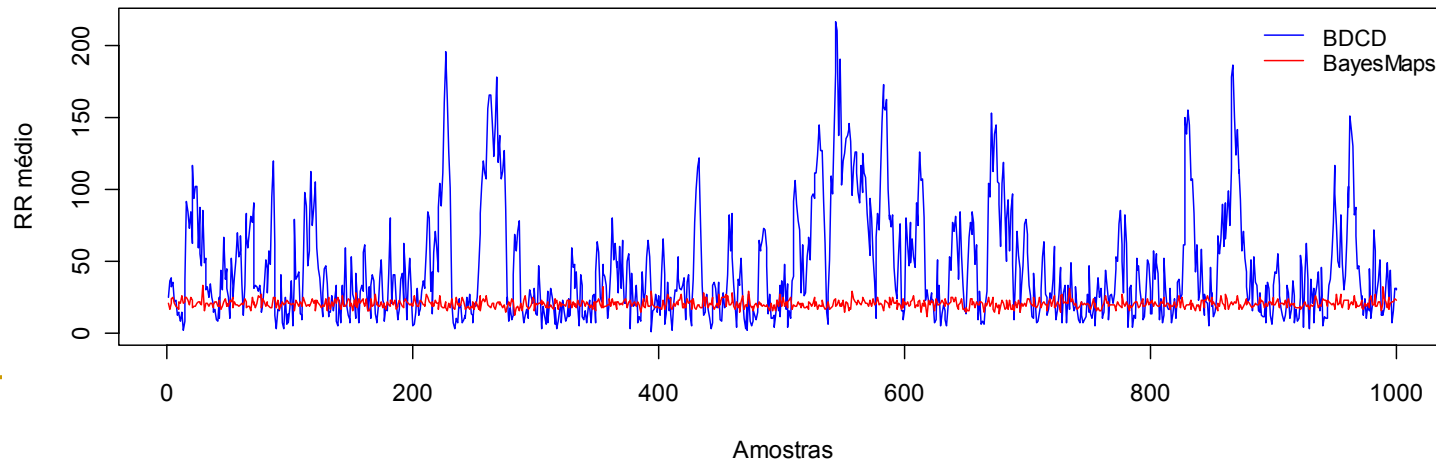
Distribuição a posteriori de K (BDCD)



Distribuição a posteriori de K (BayesMaps)

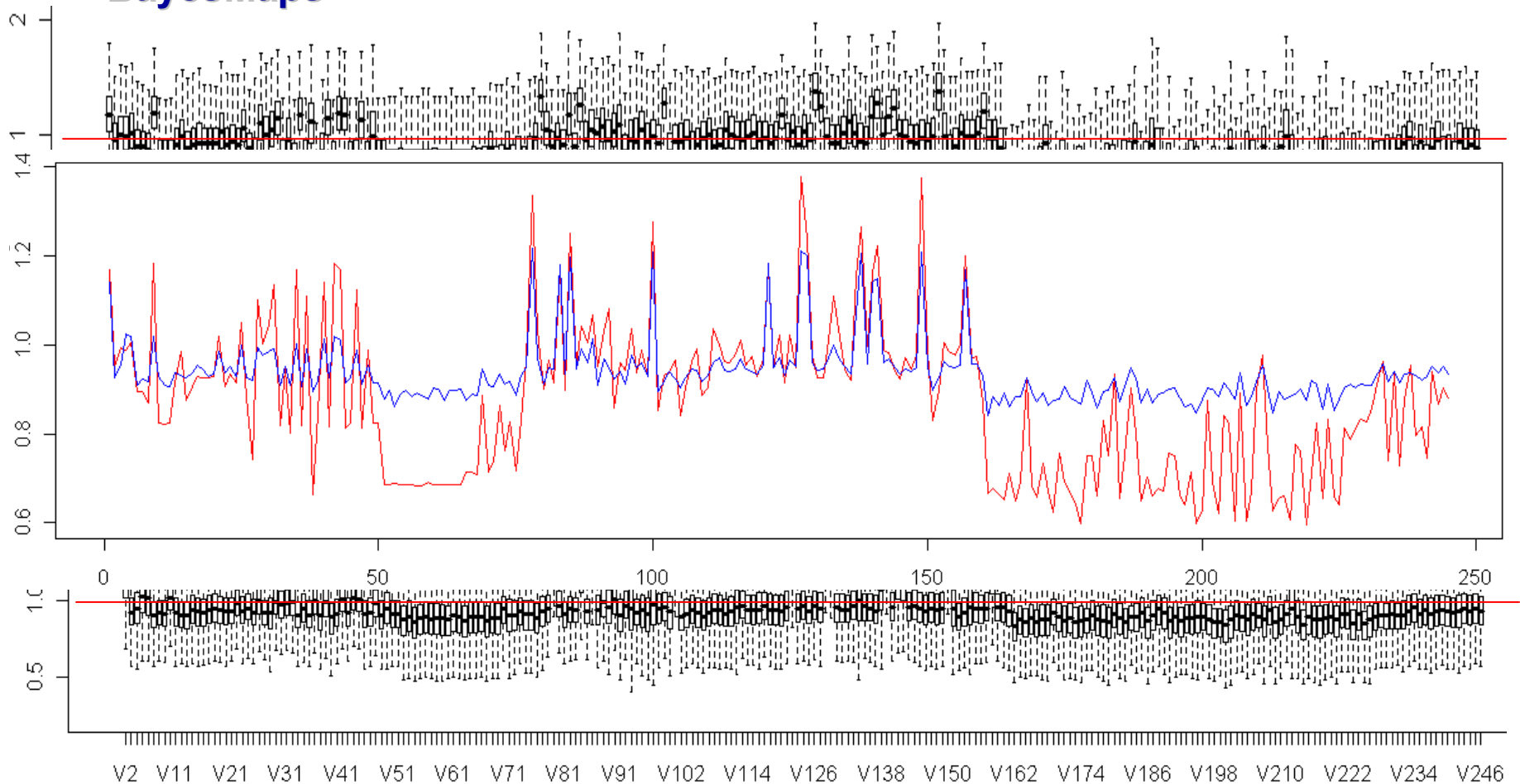


Número de clusters por amostra

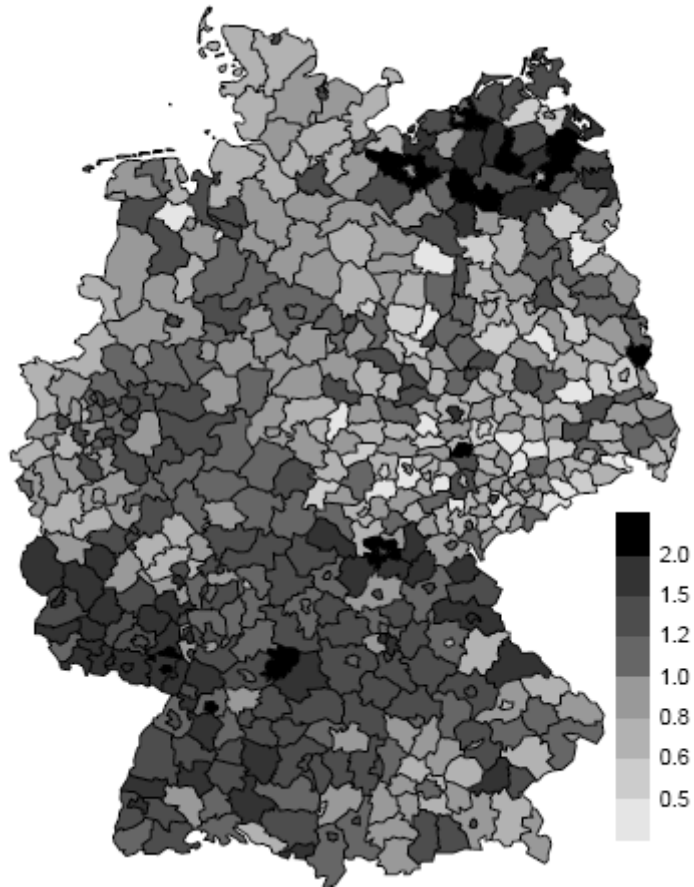


# Posteriori dos Riscos Relativos

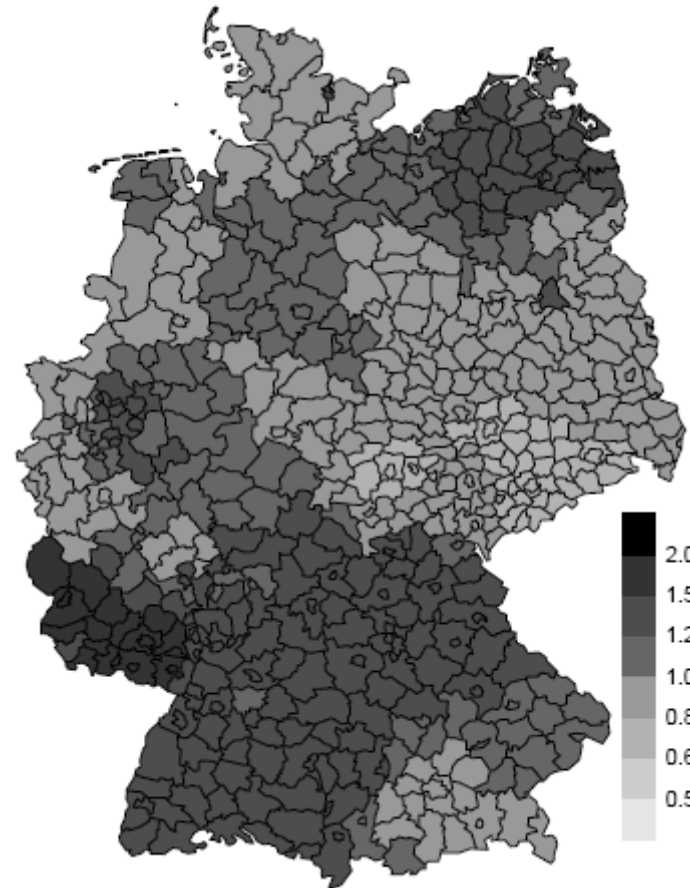
BayesMaps



# Figuras do Artigo

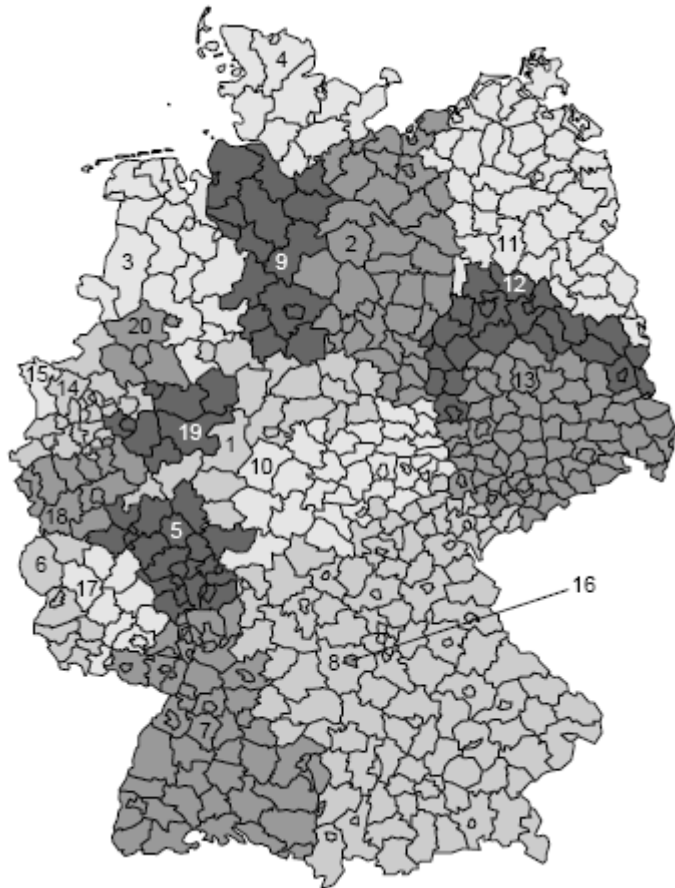


Taxas Brutas

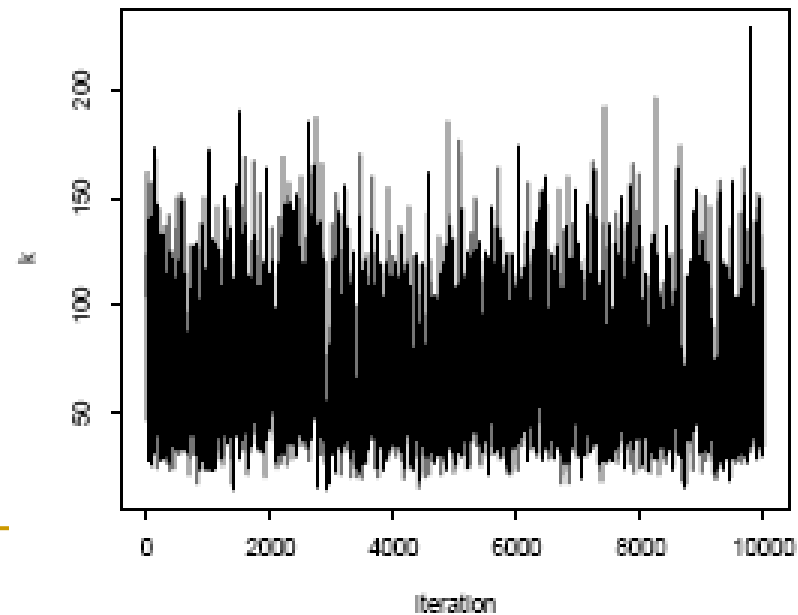
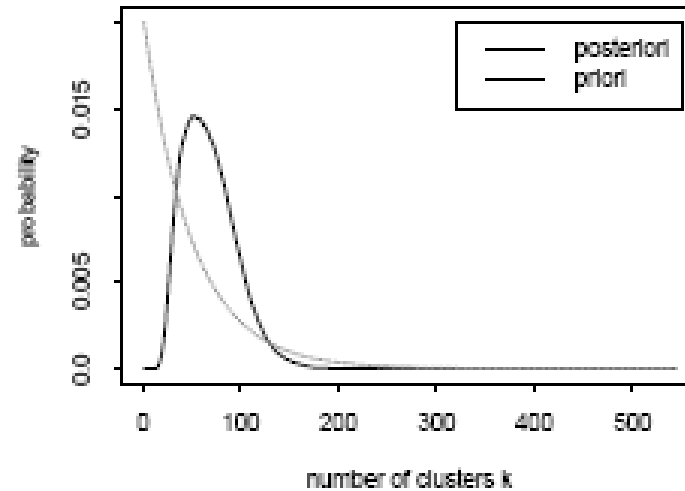


Estimated median relative risks

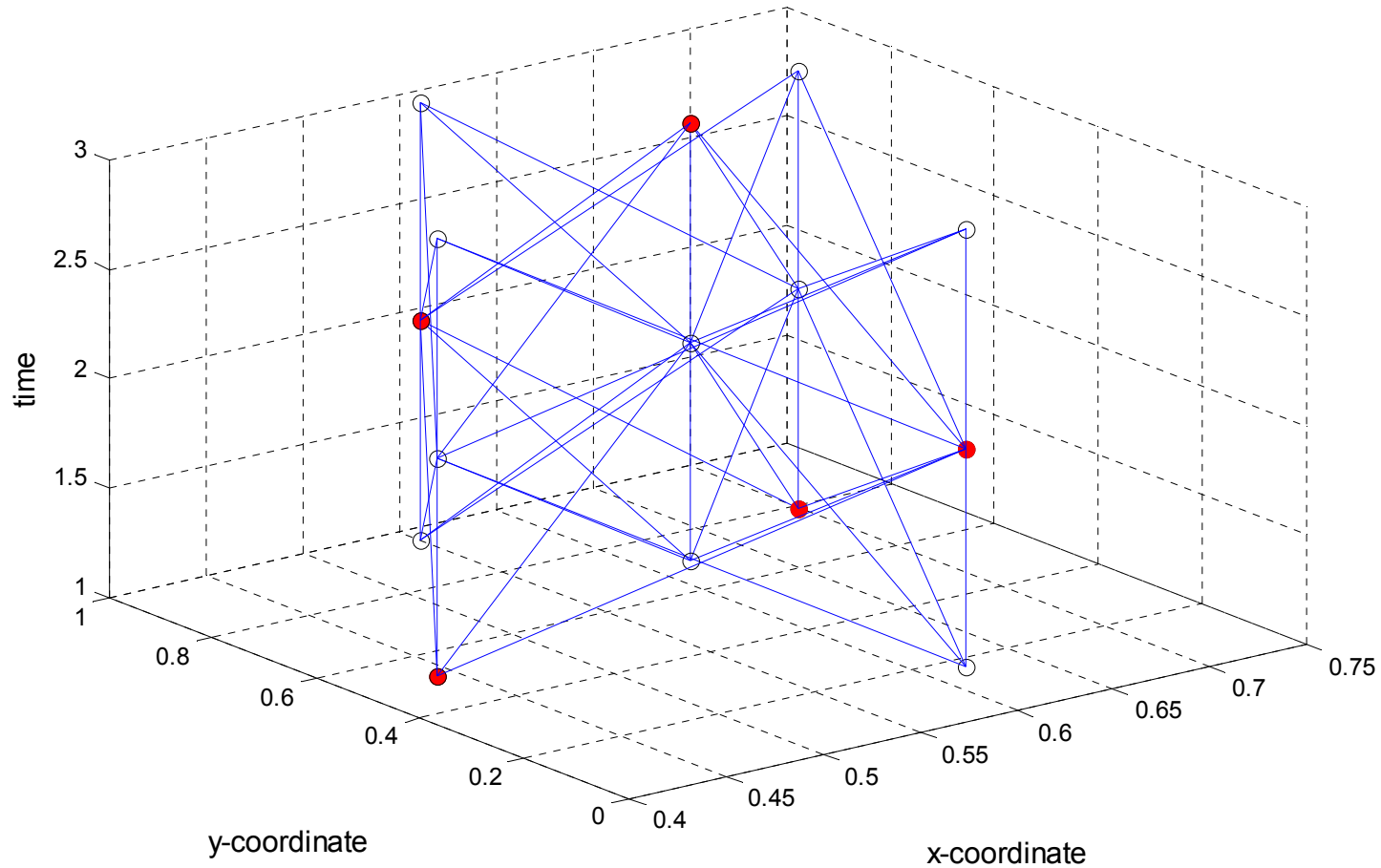
# Figuras do Artigo (b)



A cluster configuration with  $k=20$



# A Proposta Espaço-Tempo



---

# Modificações no Modelo para os dados de Produtividade

$y_i$  produtividade da  $i$ -ésima área,  $i = 1, \dots, n$ .

$$y_i \sim \text{Normal}(\mu_j, \sigma_j^2)$$



---

# Considerações:

- Os objetivos deste trabalho são:
    - ❑ Entender o princípio de funcionamento do método de clusterização bayesiano;
    - ❑ Entender e reproduzir o método de “Saltos Reversíveis”, propondo e testando modificações;
    - ❑ Implementar o código em C/C++;
    - ❑ Modificar o modelo para tratar dados espaço-temporais de produtividade;
    - ❑ Estimar as partições para os dados de produtividade.
-