Probabilistic Networks models for Statistical Seismology

Bala Rajaratnam

Stanford University

Joint work with Michael Tsiang & Doug Sparks (Stanford)

Questions

- Can we identify local spatial correlation structure in seismic catalog data?
- Given an earthquake, which are the areas where the aftershocks could be worse?
- Does knowing the seismic activity in a given location help to explain the seismic activity in nearby locations?
- Does seismic activity exhibit long-distance connections between regions that cannot be explained only by local behavior?

Modeling relationships in Seismology (and other fields)

- Correlation: basic measure of linear pairwise relationships
- Covariance matrix Σ: collection of relationships
- Estimates of Σ required in procedures such as PCA, CCA, MANOVA, etc.
- Estimating (functions of) Σ and $\Omega = \Sigma^{-1}$ are of statistical interest
- Estimating Σ is difficult in high dimensions

Sparse estimates

- Matrix Σ or Ω of size p-by-p has $O(p^2)$ elements
- Estimating $O(p^2)$ parameters with classical estimators is not viable, especially when $n \ll p$
- **Reliably estimate small number of parameters** in Σ
- Model selection: zero/non-zero structure recovery
- Gives rise to sparse estimates of Σ or Ω
- Sparsity pattern can be represented by graphs/networks

Sparse graphs/networks

- **Many physical networks are assumed to be sparse**
- Complex networks (internet, citation networks, social networks) tend to be sparse [Newman, 2003]
- Genetic networks are sparse [Gardner et al, 2003, Jeong et al, 2001]
- Model selection: recovery of edges in graphs

Gaussian Graphical Models (GGM)

- Assume $Y = (Y_1, \ldots, Y_p)'$ has distribution $N_p(0, \Sigma)$
- Denote $V = \{1, 2, \ldots, p\}$
- Covariance matrix $cov(Y) = \Sigma$ encodes marginal dependencies
$$Y_i \perp Y_j \iff cov(Y_i, Y_j) = [\Sigma]_{ij} = 0$$
- Inverse covariance matrix $\Omega = \Sigma^{-1}$ encodes conditional dependencies given the rest
$$\underbrace{\left(Y_i \perp Y_j \mid Y_{V \setminus \{i,j\}}\right)}_{\text{conditional independence}} \iff \underbrace{[\Omega]_{ij} = 0}_{\text{matrix element}}$$
- Also known as Markov Random Fields (MRF)

Gaussian Graphical Models (GGM)

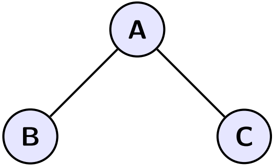
- Graph summarizes relationships with nodes $V = \{1, \dots, p\}$ and set E of edges

$$\underbrace{[\Omega]_{ij} = 0}_{\text{matrix element}} \iff \underbrace{i \not\sim j}_{\text{network/graph}}$$

- Build a graph from sparse Ω

$$\Omega = \begin{array}{c} \\ \left(\begin{array}{ccc} 1 & 0.2 & 0.3 \\ 0.2 & 2 & 0 \\ 0.3 & 0 & 1.2 \end{array}\right) \end{array} \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \end{array}$$

$$\begin{array}{cccc} \text{A} & \text{B} & \text{C} & \\ 1 & 0.2 & 0.3 & \text{A} \\ 0.2 & 2 & 0 & \text{B} \\ 0.3 & 0 & 1.2 & \text{C} \end{array}$$

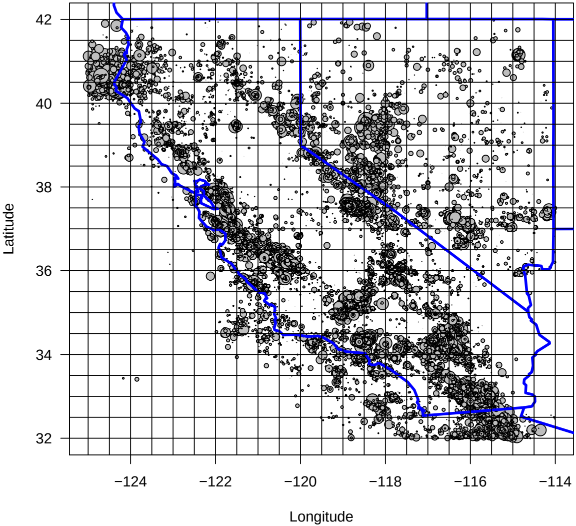


Description of the Data

We extracted data from the Advanced National Seismic System (ANSS) catalog.

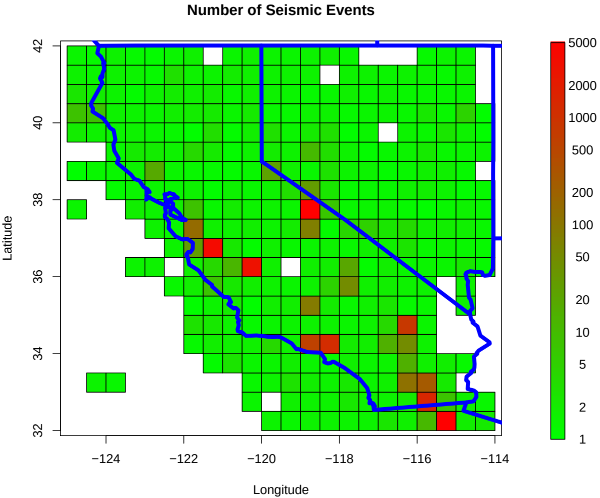
- The data contains all seismic events of magnitude at least 3.0 in an $11^\circ \times 10^\circ$ rectangle covering California and Nevada since January 1, 1950.
- We grouped the data into $0.5^\circ \times 0.5^\circ$ grid boxes and computed (1) the cumulative magnitudes and (2) the number of all events in a six month period in each grid box.
- Regarding the gridded cumulative magnitudes (or the number of events) during a given six month period as an observation from a random vector, we have a random vector of length $p = 340$ and a total of $n = 127$ observations.

Plot of the Data



Larger points denote seismic events of higher magnitude.

Seismic Events in Each Geographic Bin



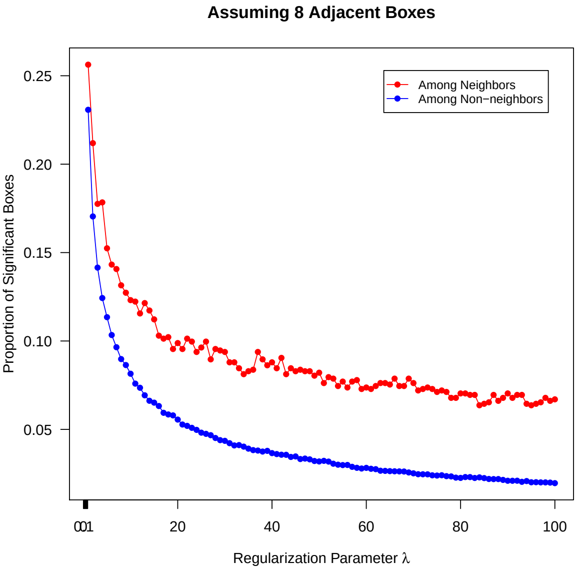
Methodology

- Neighboring grid boxes were defined to be those boxes directly adjacent to a given grid box.
- We computed the sample correlation matrix between grid boxes.
- Since $p > n$, the sample correlation matrix is a poor estimate of the true correlation matrix.
- Also, since $p > n$, the sample correlation matrix is singular. This means that its inverse (the sample partial correlation matrix) cannot be computed at all.
- Some form of regularization is required to obtain good estimates.

Regularization Approaches Considered

- **Correlation hard-thresholding** estimates the correlation matrix. We can consider either the Pearson correlation or the Spearman rank correlation.
- **Graphical LASSO (GLASSO)** estimates the partial correlation matrix. Again, we can consider either the Pearson correlation or the Spearman rank correlation.
- **SPACE** also estimates the partial correlation matrix. Here we can only consider the Pearson correlation.

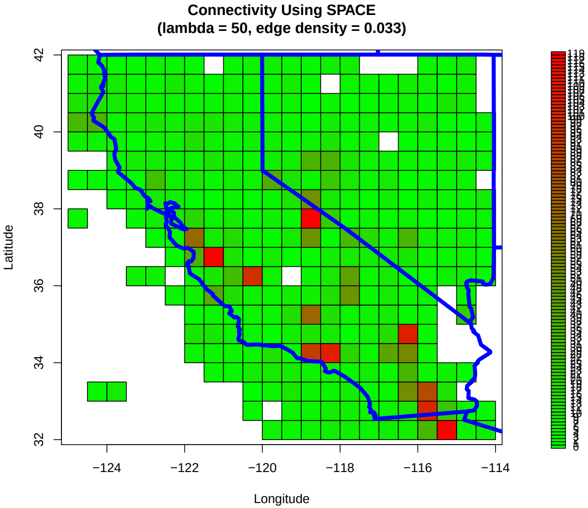
SPACE: Sum of Magnitudes of Seismic Events



Results

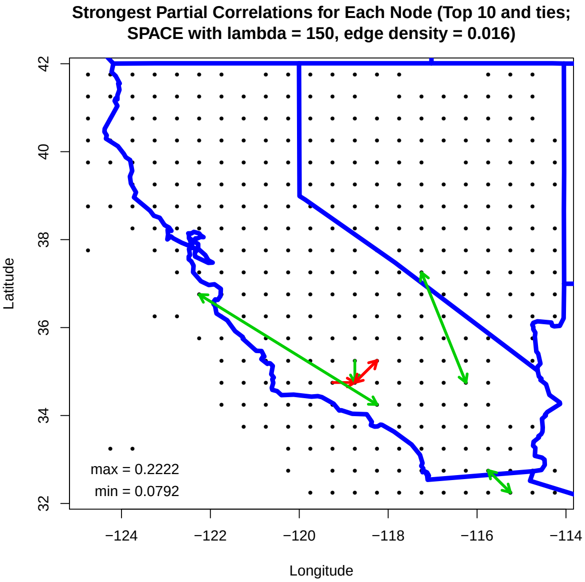
- Like the GLASSO, SPACE also estimates the partial correlation matrix, the inverse of the correlation matrix itself.
- Once again, we still see the same phenomenon: nodes that are geographic neighbors are much more likely than non-neighbors to have a nonzero element (here, a nonzero partial correlation).
- Again, the other plots also behave similarly (next few slides).

SPACE: Sum of Magnitudes of Seismic Events

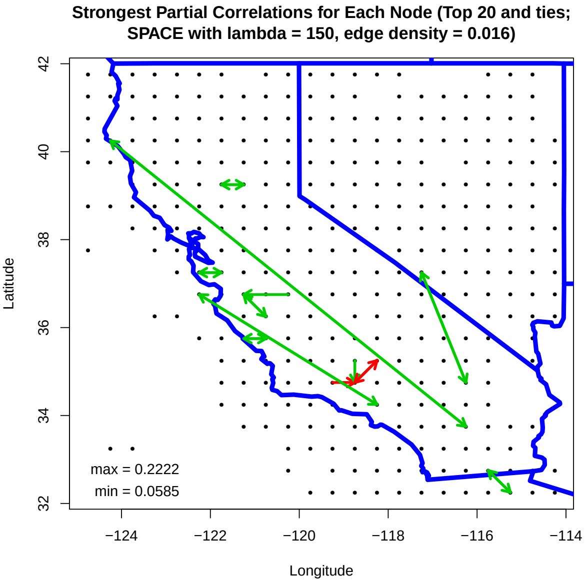


The **connectivity** of node i is the number of other nodes for which the estimated correlation with node i is nonzero.

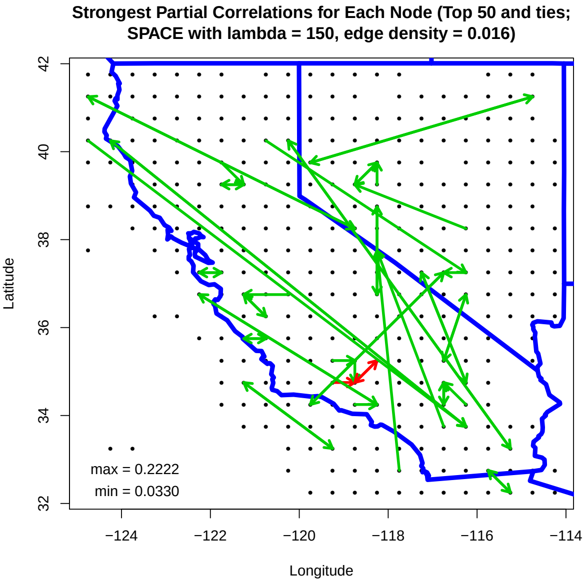
SPACE: Sum of Magnitudes of Seismic Events



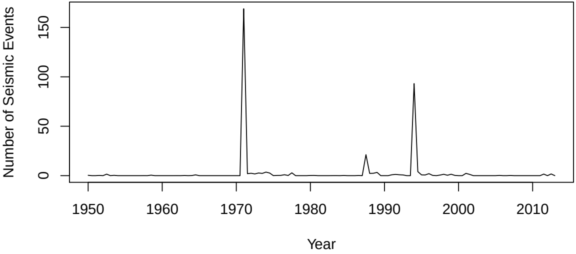
SPACE: Sum of Magnitudes of Seismic Events



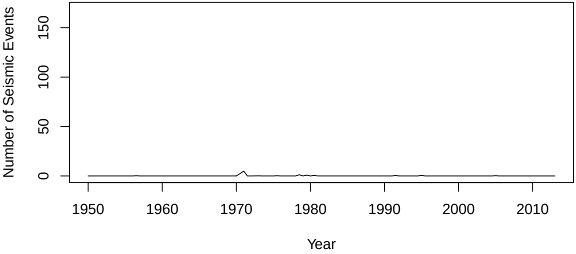
SPACE: Sum of Magnitudes of Seismic Events



SPACE: Sum of Magnitudes of Seismic Events

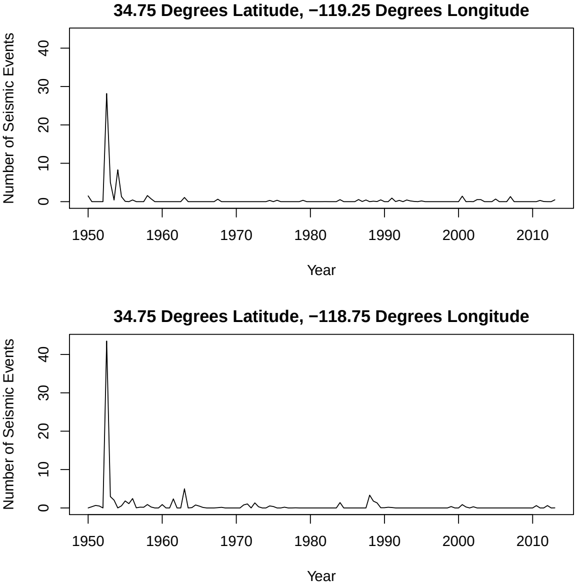


34.25 Degrees Latitude, −118.25 Degrees Longitude

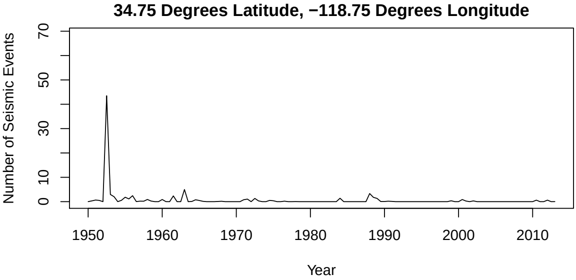


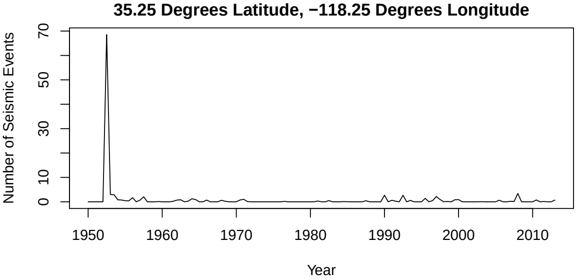
36.75 Degrees Latitude, −122.25 Degrees Longitude

SPACE: Sum of Magnitudes of Seismic Events



SPACE: Sum of Magnitudes of Seismic Events

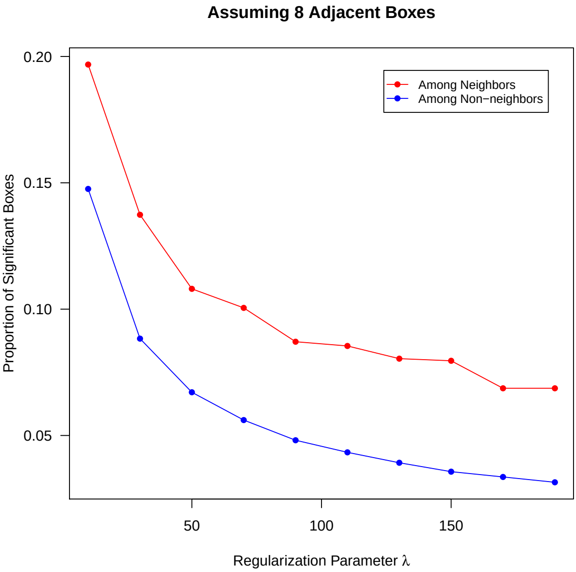




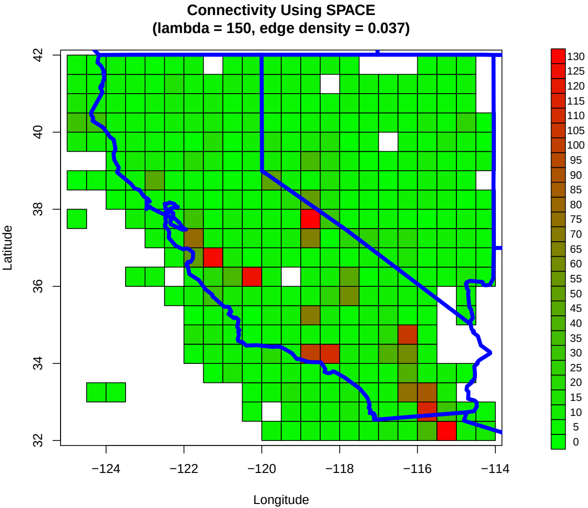
SPACE: Number of Seismic Events

- The results are fairly similar if we consider simply the number of seismic events in each spatial and temporal bin, rather than the sum of the magnitudes.
- This alternative approach yields the plots shown in the following slides.

SPACE: Number of Seismic Events

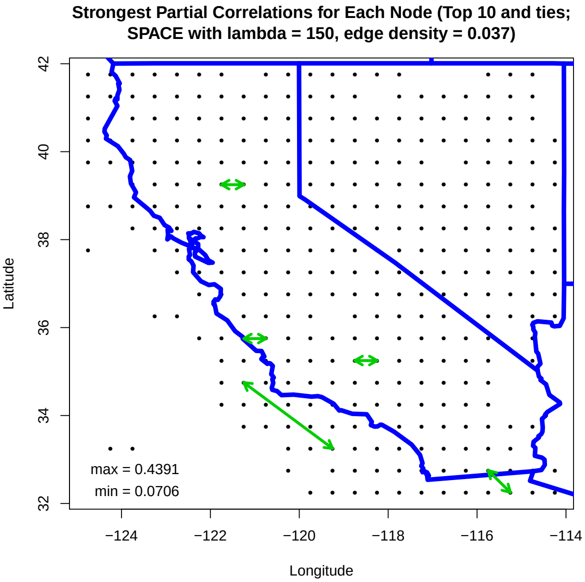


SPACE: Number of Seismic Events

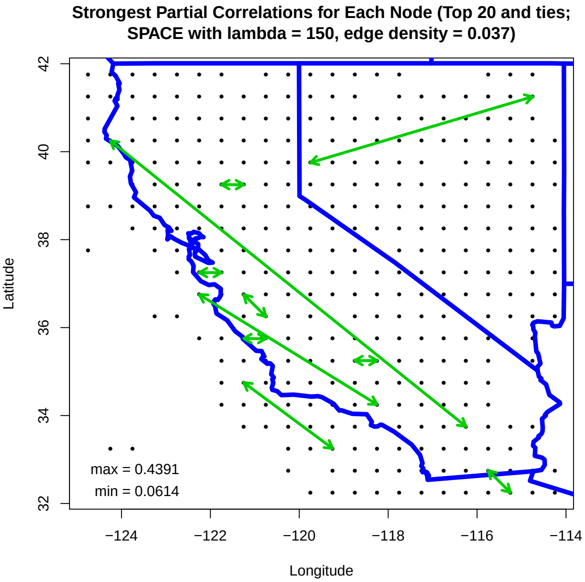


The **connectivity** of node i is the number of other nodes for which the estimated correlation with node i is nonzero.

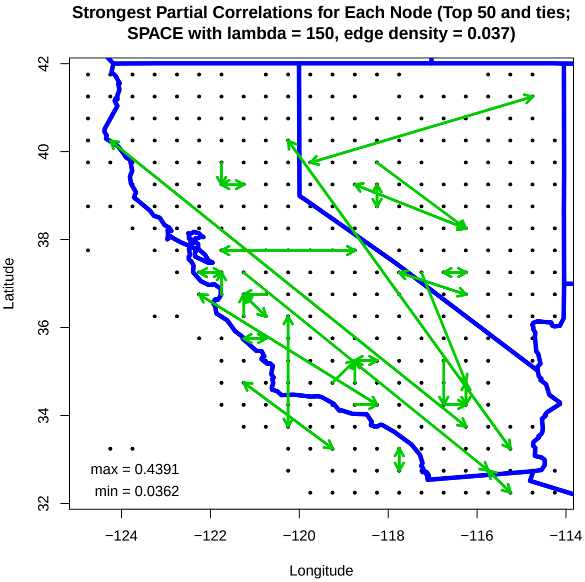
SPACE: Number of Seismic Events



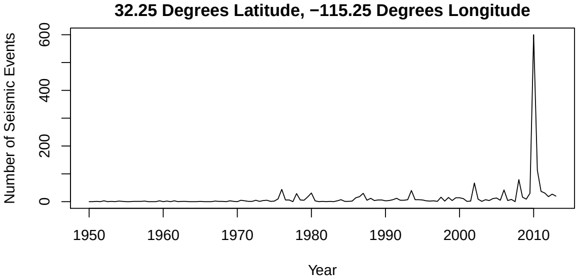
SPACE: Number of Seismic Events



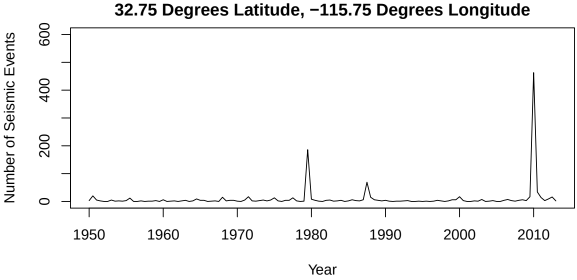
SPACE: Number of Seismic Events



SPACE: Number of Seismic Events

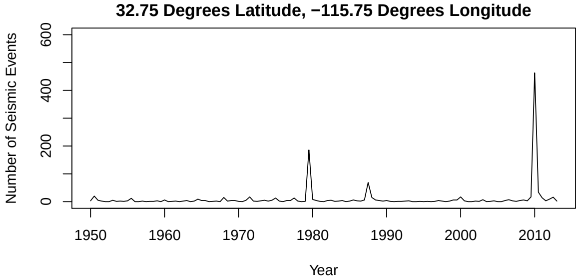


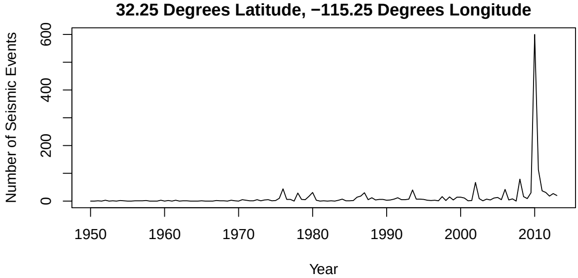
32.25 Degrees Latitude, −115.25 Degrees Longitude



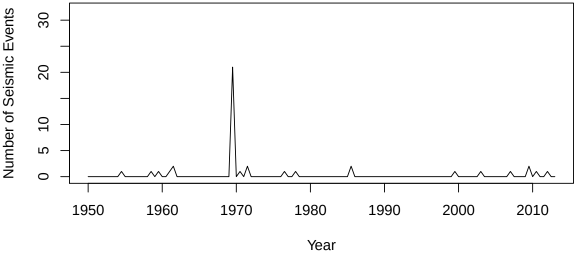
32.75 Degrees Latitude, −115.75 Degrees Longitude

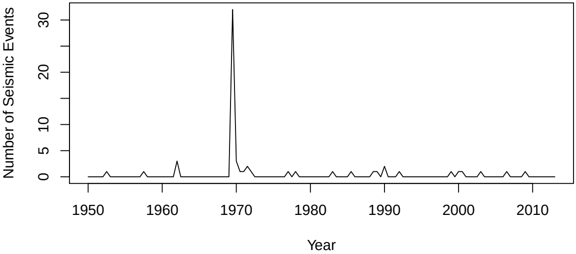
SPACE: Number of Seismic Events


32.75 Degrees Latitude, −115.75 Degrees Longitude


32.25 Degrees Latitude, −115.25 Degrees Longitude

SPACE: Number of Seismic Events

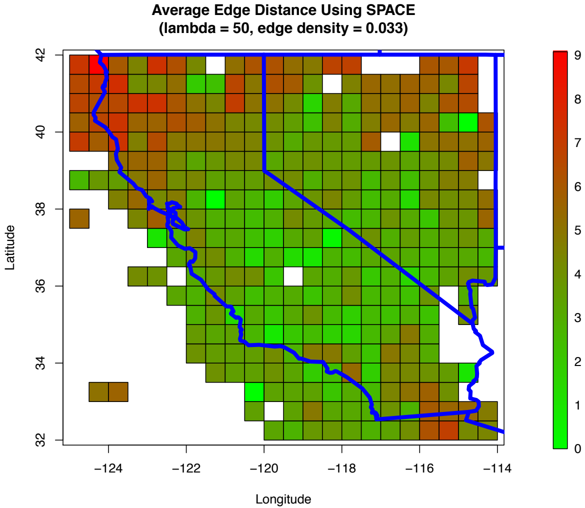

33.25 Degrees Latitude, −119.25 Degrees Longitude


34.75 Degrees Latitude, −121.25 Degrees Longitude

Conclusions

- Data suggests that Markov random fields can be useful for describing seismic activity.
- Seismic activity in a target area is more related to activity in geographically neighboring areas than to activity in non-neighboring areas.
- However, long-distance connections can still have substantial explanatory value when they are present.

SPACE



The **average edge distance** for node i is its average distance from the other nodes with which it has a nonzero estimated correlation.