Applied Spatial Statistics in R, Section 3

Spatial Weights

Yuri M. Zhukov

IQSS, Harvard University

January 16, 2010

Outline

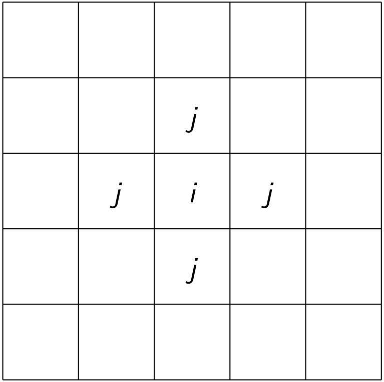
Choosing your neighbors?

- Most spatial weights matrices $\mathbf{W}$ are based on some version of a connectivity matrix $\mathbf{C}$.
- $\mathbf{C}$ is an $n \times n$ binary matrix, where $i = \{1, 2, \ldots, n\}$ and $j = \{1, 2, \ldots, n\}$ are the units in the system (for example, countries in the international system).
- Entry $c_{ij} = 1$ if two units $i \neq j$ are considered connected, and $c_{ij} = 0$ if they are not.
- The tricky part is how the word "connected" is defined.

Areal Contiguity I: Regular Grids

Rook's case

Cells sharing a common edge are considered contiguous



Areal Contiguity I: Regular Grids

Bishop's case

Cells sharing a <u>common vertex</u> are considered contiguous

	j		j	
		i		
	j		j	

Areal Contiguity I: Regular Grids

Queen's case

Cells sharing a common edge or common vertex are considered contiguous

	j	j	j	
	j	i	j	
	j	j	j	

Areal Contiguity I: Regular Grids

Second-order neighbors: (rook's case)

Cells sharing a common edge with first-order neighbors are considered contiguous

		k		
	k	j	k	
k	j	k	j	k
	k	j	k	
		k		

Areal Contiguity I: Regular Grids

- These conceptions of contiguity are useful when dealing with regular square grids or rectangular lattices, where the spatial structure can be easily summarized in elegant mathematical terms.
- But when spatial units consist of irregularly-shaped polygons, as is the case in most applied work (countries, census tracts, various administrative units), this simple characterization of contiguity breaks down...

Areal Contiguity II: Polygons ($\mathbf{W}_{CONT}$)

Two polygons x_i and x_j are neighbors if they share a common boundary.

Advantage

- Makes substantive sense

Disadvantage

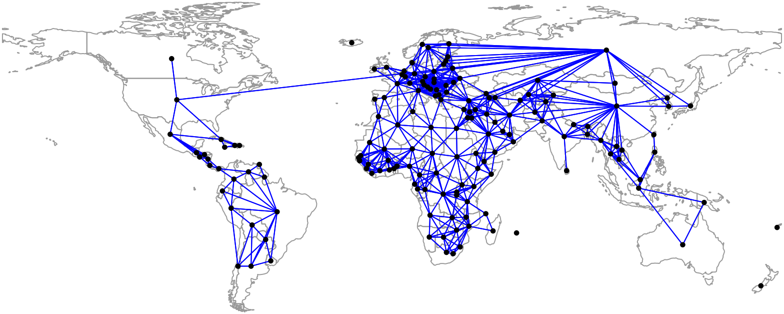
- Neighborless units

Lists with no-neighbor areas are problematic for the estimation of spatial weights.

- Should the weight representation of the empty set be a numeric zero or a missing value?
- This choice, and the resulting size of n, is highly consequential for tests of spatial autocorrelation.

Areal Contiguity II: Visualization of Connections

Figure: Contiguity neighbors with 500 km snap distance

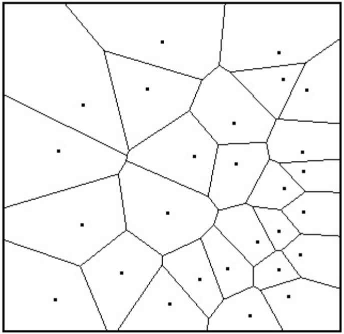


Areal Contiguity II: Polygons

- While intuitive, polygon contiguity is not always appropriate or feasible.
- When the spatial units consist of points, such as cities or event locations, aggregation into polygons is often undesirable due to the Modifiable Areal Unit Problem (MAUP) (Openshaw and Taylor 1979, 1981).
  - Aggregation of point data is only meaningful if the underlying phenomenon is homogeneous across space.
  - Otherwise, any aggregation scheme which does not account for heterogeneity and structural instability will be misleading.
  - Furthermore, the level of aggregation affects the magnitude of various measures of association, such as autocorrelation coefficients and estimated regression parameters.
  - The MAUP is closely conceptually similar to the ecological fallacy problem (King 1997).

Areal Contiguity II: Polygons

- Another approach is to draw polygons around each point by spatial tessellation, as in Voronoi or Dirichlet diagrams.
- But here, notions of boundary locations, length and area are largely artificial constructs, determined by the particular tessellation algorithm used.



Interpoint Distance Neighbors I: Minimum Distance Neighbors ($\mathbf{W}_{MDN}$)

Neighbors of unit x_i defined by interpoint distance:

- Lower bound: 0
- Upper bound: $\max_{i=1}^{n}\left(\min_{j\neq i}^{n-1} d(x_i, x_j)\right)$

Advantage

- No neighborless units

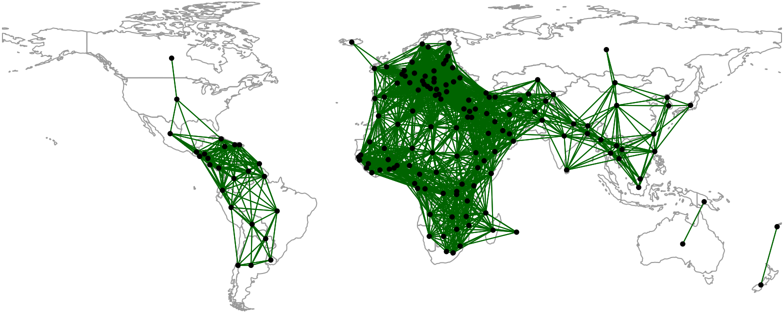
Disadvantage

- Inefficient for irregularly-spaced data
- Potentially high number of politically irrelevant connections

Choice of points (centroids vs. capital cities) is potentially quite significant and requires theoretical justification

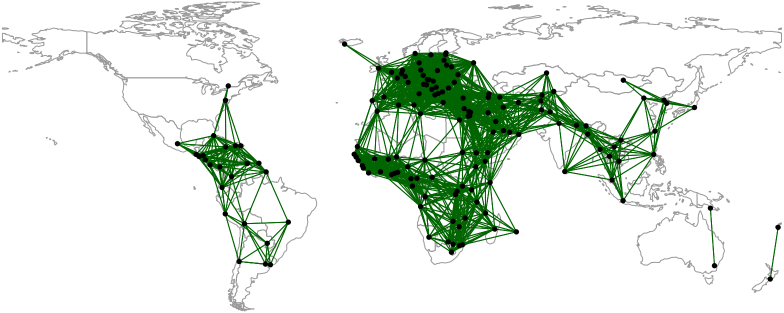
Interpoint Distance Neighbors I: Visualization of Connections

Figure: Minimum distance neighbors (polygon centroids)



Interpoint Distance Neighbors I: Visualization of Connections

Figure: Minimum distance neighbors (capital cities)



Interpoint Distance Neighbors II: k Nearest Neighbors ($\mathbf{W}_{KNN}$)

Neighbors of unit x_i defined by user-defined parameter k. x_j is a neighbor of x_i if $x_j \in N_k x_i$, where $N_k x_i$ are the k nearest neighbors of x_i.

Advantage

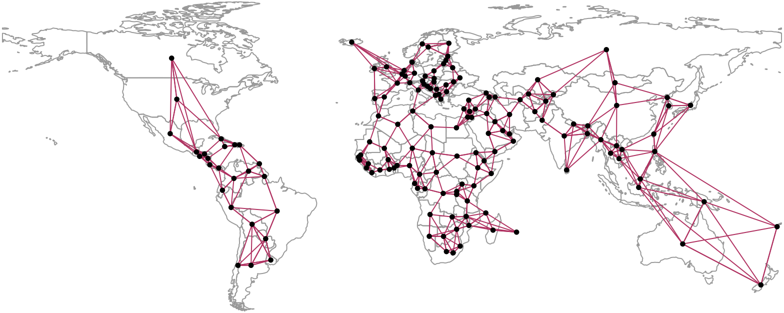
- No neighborless units
- Less noisy than $\mathbf{W}_{MDN}$

Disadvantage

- Parameter selection may not reflect 'true' level of connectedness or isolation

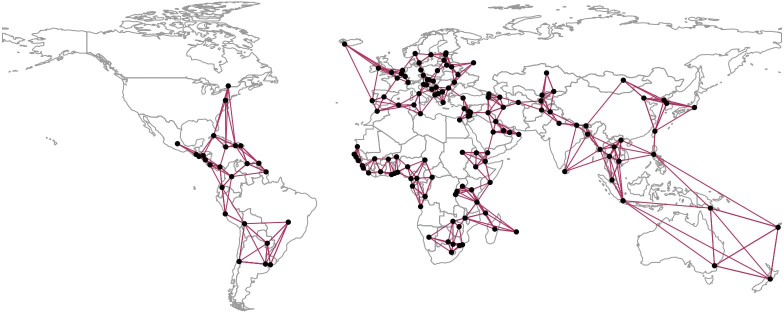
Interpoint Distance Neighbors II: Visualization of Connections

Figure: $k = 4$ Nearest Neighbors (polygon centroids)



Interpoint Distance Neighbors II: Visualization of Connections

Figure: $k = 4$ Nearest Neighbors (capital cities)



Graph-Based Neighbors: Sphere of Influence Neighbors ($\mathbf{W}_{SOI}$)

For each point $x \in S = \{x_1, \ldots, x_n\}$,

- Let $r_i = \min_{k \neq i} d(x_i, x_k)$.
- Let C_i be a circle of radius r_i, centered at x_i.

Points x_i and x_j are neighbors whenever C_i and C_j intersect in exactly two points.

Advantage

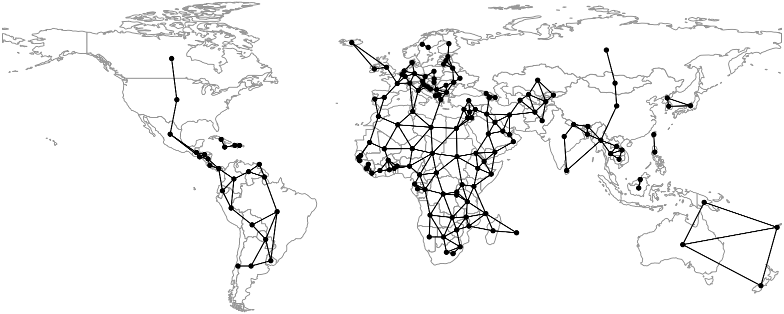
- No neighborless units
- Less noisy than $\mathbf{W}_{MDN}$
- Less arbitrary than $\mathbf{W}_{KNN}$

Disadvantage

- Uses Euclidean, not Great Circle Distances

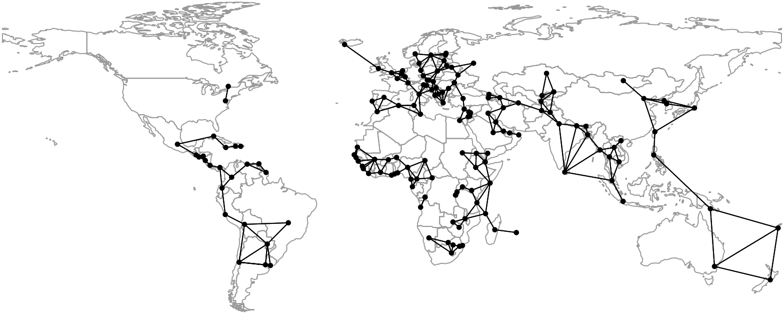
Graph-Based Neighbors: Visualization of Connections

Figure: Sphere of Influence Neighbors (polygon centroids)



Graph-Based Neighbors: Visualization of Connections

Figure: Sphere of Influence Neighbors (capital cities)



Application: Democratic Diffusion

Changes of political regime modeled as a first-order Markov chain process with the transition matrix

$$\mathbf{K} = \begin{bmatrix} Pr(y_{i,t} = 0 | y_{i,t-1} = 0) & Pr(y_{i,t} = 1 | y_{i,t-1} = 0) \\ Pr(y_{i,t} = 0 | y_{i,t-1} = 1) & Pr(y_{i,t} = 1 | y_{i,t-1} = 1) \end{bmatrix}$$

where $y_{i,t} = 1$ if an (A)utocratic regime exists in country i at time t, and $y_{i,t} = 0$ if the regime is (D)emocratic.
... in other words:

$$\mathbf{K} = \begin{bmatrix} Pr(D \rightarrow D) & Pr(D \rightarrow A) \\ Pr(A \rightarrow D) & Pr(A \rightarrow A) \end{bmatrix}$$

Estimation

Conditional transition probabilities are estimated by a probit link:

$$Pr(y_{i,t} = 1 | y_{i,t-1}, \mathbf{x}_{i,t}) = \Phi[\mathbf{x}_{i,t}^T \beta + y_{i,t-1} \mathbf{x}_{i,t}^T \alpha]$$

Previous uses:

- Takeshi Amemiya, *Advanced Econometrics* (Cambridge, MA: Harvard University Press, 1985)
- Adam Przeworski and Fernando Limongi, “Modernization: Theories and Facts,” *World Politics* 49 (1997): 155-83.
- Kristian S. Gleditsch and Michael D. Ward, “Diffusion and the International Context of Democratization,” *International Organization* 60 (2006): 911-33.

Equilibrium Effects of Democratic Transition

If a regime transition takes place in country i, what is the change in predicted probability of a regime transition in country j (country i's neighbor)?

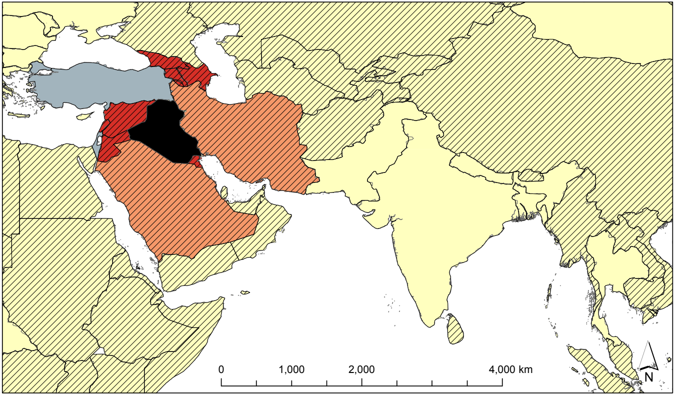
$$\text{QI} = Pr(y_{j,t}|y_{i,t} = y_{i,t-1}) - Pr(y_{j,t}|y_{i,t} \neq y_{i,t-1})$$

where $y_{i,t} = 0$ if country i is a democracy at time t and $y_{i,t} = 1$ if it is an autocracy. All other covariates are held constant.

Illustrative cases

- Iraq transitions from autocracy to democracy.
- Russia transitions from democracy to autocracy.

Iraq's democratization and regional regime stability



Contiguity + 500 km

Iraq transitions from autocracy to democracy (1998 data)

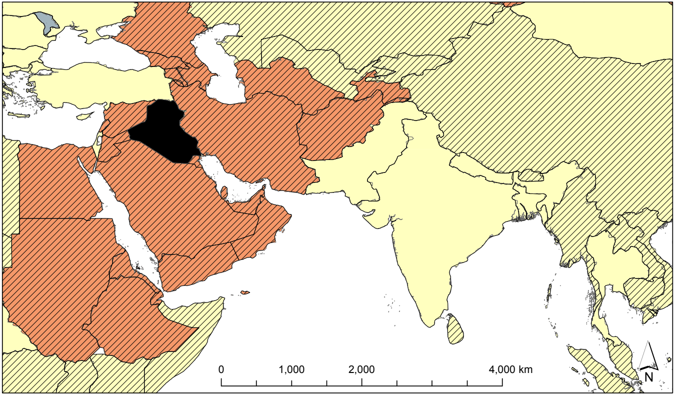
Monte Carlo simulation (1,000 runs)

Iraq	
Regime Type	
Democracy	
Autocracy	

Change in Transition Probability

- -0.05- -0.025
- -0.025 - -0.001
- 0
- 0.001 - 0.025
- 0.025 - 0.05

Iraq's democratization and regional regime stability



Minimum Distance

Iraq transitions from autocracy to democracy (1998 data)

Monte Carlo simulation (1,000 runs)

Iraq

Regime Type

Democracy

Autocracy

Change in Transition Probability

-0.05- -0.025

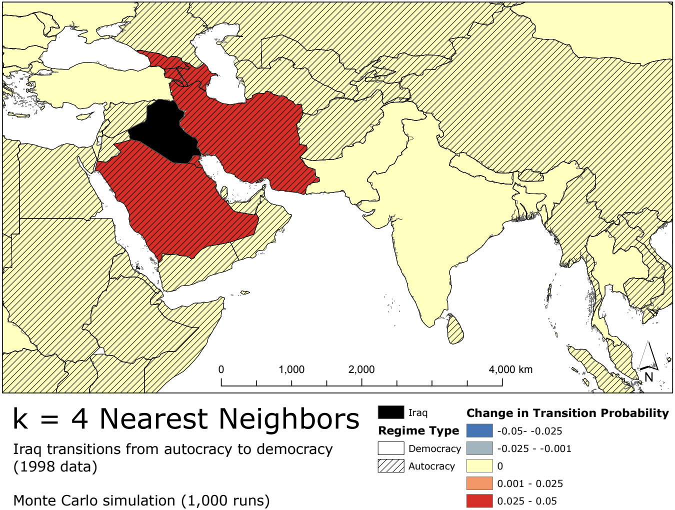
-0.025 - -0.001

0

0.001 - 0.025

0.025 - 0.05

Iraq's democratization and regional regime stability

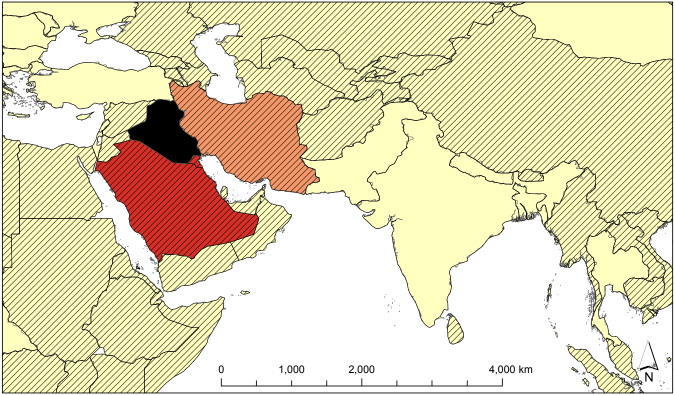


k = 4 Nearest Neighbors

Iraq transitions from autocracy to democracy (1998 data)

Monte Carlo simulation (1,000 runs)

Iraq's democratization and regional regime stability



Sphere of Influence

Iraq transitions from autocracy to democracy (1998 data)

Monte Carlo simulation (1,000 runs)

Iraq

Regime Type

Democracy

Autocracy

Change in Transition Probability

-0.05- -0.025

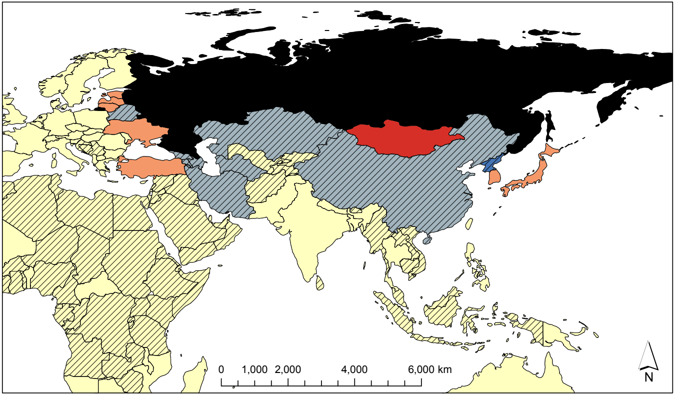
-0.025 - -0.001

0

0.001 - 0.025

0.025 - 0.05

Russia's autocratization and regional regime stability



Contiguity + 500 km

Russia transitions from democracy to autocracy (1998 data)

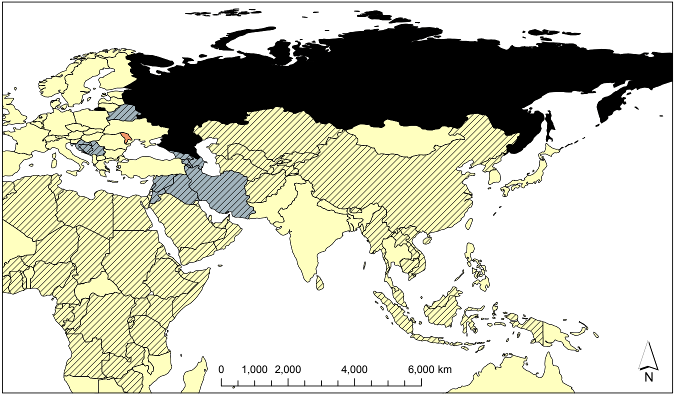
Monte Carlo simulation (1,000 runs)

Russia

Regime Type
- Democracy
- Autocracy

Change in Transition Probability
- -0.05- -0.025
- -0.025 - -0.001
- 0
- 0.001 - 0.025
- 0.025 - 0.05

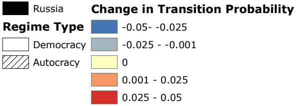
Russia's autocratization and regional regime stability



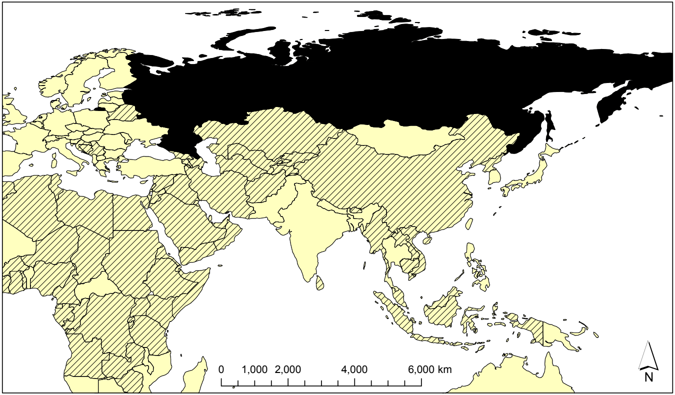
Minimum Distance

Russia transitions from democracy to autocracy (1998 data)

Monte Carlo simulation (1,000 runs)



Russia's autocratization and regional regime stability



k = 4 Nearest Neighbors

Russia transitions from democracy to autocracy (1998 data)

Monte Carlo simulation (1,000 runs)

Russia

Regime Type

Democracy

Autocracy

Change in Transition Probability

-0.05- -0.025

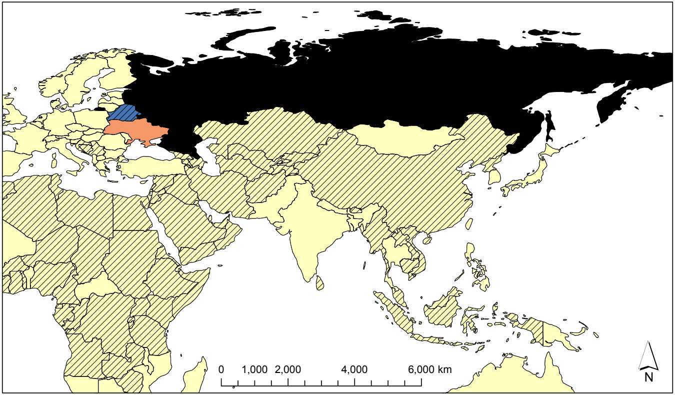
-0.025 - -0.001

0

0.001 - 0.025

0.025 - 0.05

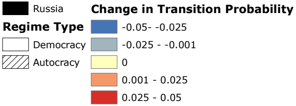
Russia's autocratization and regional regime stability



Sphere of Influence

Russia transitions from democracy to autocracy (1998 data)

Monte Carlo simulation (1,000 runs)



Network neighbors

The structure of spatial dependence can be non-geographic. Any theoretically-relevant dyadic relationship can form the basis of connectivity.

- **Individual level**: friendship, frequency of communication, citations, kinship.
- **Organizational level**: market competition, joint enterprises, personnel exchanges.
- **International level**: alliance relationship, trade flows, joint organizational membership, diplomatic contacts, cultural exchanges, migration flows.

From Connections to Weights

- Once a definition of connectivity is made, one must translate binary indicators into weights, which will form the elements w_{ij} of matrix $\mathbf{W}$.
- A plethora of options exist: inverse distance (IDW), negative exponentials of distance, length of shared boundary, relative area, accessibility...
- The resulting weights will often be asymmetric, unless the study region is a regular lattice.

From Connections to Weights

- The rows of $\mathbf{W}$ are often row-standardized, so that $\sum_{j=1}^{n} w_{ij} = 1$
- There is no mathematical or statistical requirement for this, but row standardization facilitates interpretation of lagged variables as a weighted average of neighboring values of some variable.
- This is not always desirable, however.
- Row-standardization also implies competition among neighbors: the fewer the neighbors, the stronger their individual influence on i.
- Further, when weights are based on some measure of distance decay (ie: IDW), scaling the rows to sum to one results in a loss of that interpretation.
- **Bottom line**: the weights should bear a direct relation to one's theoretical conceptualization of the structure of dependence.

Sparse vs. Dense Matrices

Sparsity carries a number of substantive and computational advantages:

- Dense matrices are noisy and contain a potentially large number of irrelevant connections.
- Dense matrices will bias downward indirect effects of a change in observation j (the individual weights of non-zero entries in row-standardized weight matrices will be smaller).
- Dense matrices can be computationally intensive to the point that even simple matrix operations are infeasible.

Sparse vs. Dense Matrices

Consider the following example with 2000 U.S. Census data:

Tracts

$n = 65,443$

31.90 GB of storage required for dense matrix, .01 GB for sparse matrix.

Block Groups

$n = 208,790$

324.80 GB of storage required for dense matrix, .03 GB for sparse matrix.

Blocks

$n = 8,205,582$

501,659.33 GB of storage required for dense matrix, 1.10 GB for sparse.

Here, dense and sparse matrices have n^2 and $6/n$ nonzero elements, respectively. For spatially random data on a plane, each unit will have an average of 6 contiguity neighbors (LeSage and Pace 2009).

Ordering of Weights Matrix

Ordering of rows and columns matters greatly for computation times.

- Consider an $n \times n$ permutation matrix $\mathbf{P}$, which has exactly one entry 1 in each row and each column and 0's elsewhere. Each permutation matrix can produce a reordered weights matrix $\mathbf{W}_P$, by the operation $\mathbf{W}_P = \mathbf{PWP}'$.
- Note that $\mathbf{P}^{-1} = \mathbf{P}'$, $|\mathbf{P}| = 1$ and $|\mathbf{P}(\mathbf{I}_n - \rho\mathbf{W})\mathbf{P}'| = |\mathbf{P}||\mathbf{I}_n - \rho\mathbf{W}||\mathbf{P}'| = |\mathbf{I}_n - \rho\mathbf{W}| = |\mathbf{I}_n - \rho\mathbf{PWP}'|$
- Thanks to these properties, log-determinant calculation and other matrix operations will not be affected by the reordering of $\mathbf{W}$.
- But computation times for these operations are affected.

Ordering of Weights Matrix

Efficiency is increased if ordering is **geographic** (north-south or east-west)

- This ordering concentrates nonzero elements around the diagonal, which reduces the bandwidth of a matrix ($max|i-j|$ for nonzero elements).
- For a sample of 62,226 U.S. Census Tracts, calculation of a single log-determinant requires over 12 GB of memory for a randomly ordered weights matrix, making calculation infeasible on most machines.
- The same operation takes less than a minute for a geographically-ordered matrix (LeSage and Pace 2009).

Examples in R

Switch to R tutorial script. Section 3.