Spatial and Regional Economic Analysis Mini-Course

Author: Vicente Rios

- **Introduction to Spatial Economics**
- **Spatial Weight Matrices**
- **Other Weight Matrices**

Why do we need spatial econometrics?

- Motivations for going spatial

    - **Independence assumption not valid**

    The attributes of observation *i* may influence the attributes of *j*.

    eg: growth in *i* affects growth in *j*

    - **Spatial heterogeneity**

    The magnitude and direction of an effect may vary across space.

    eg: government spending in *i* does not have the same stimulus than in *j*

    - **Accuracy!**

Spatial heterogeneity

- **Spatial econometrics deals with spatial effects**

  (I) Spatial heterogeneity

Definition:

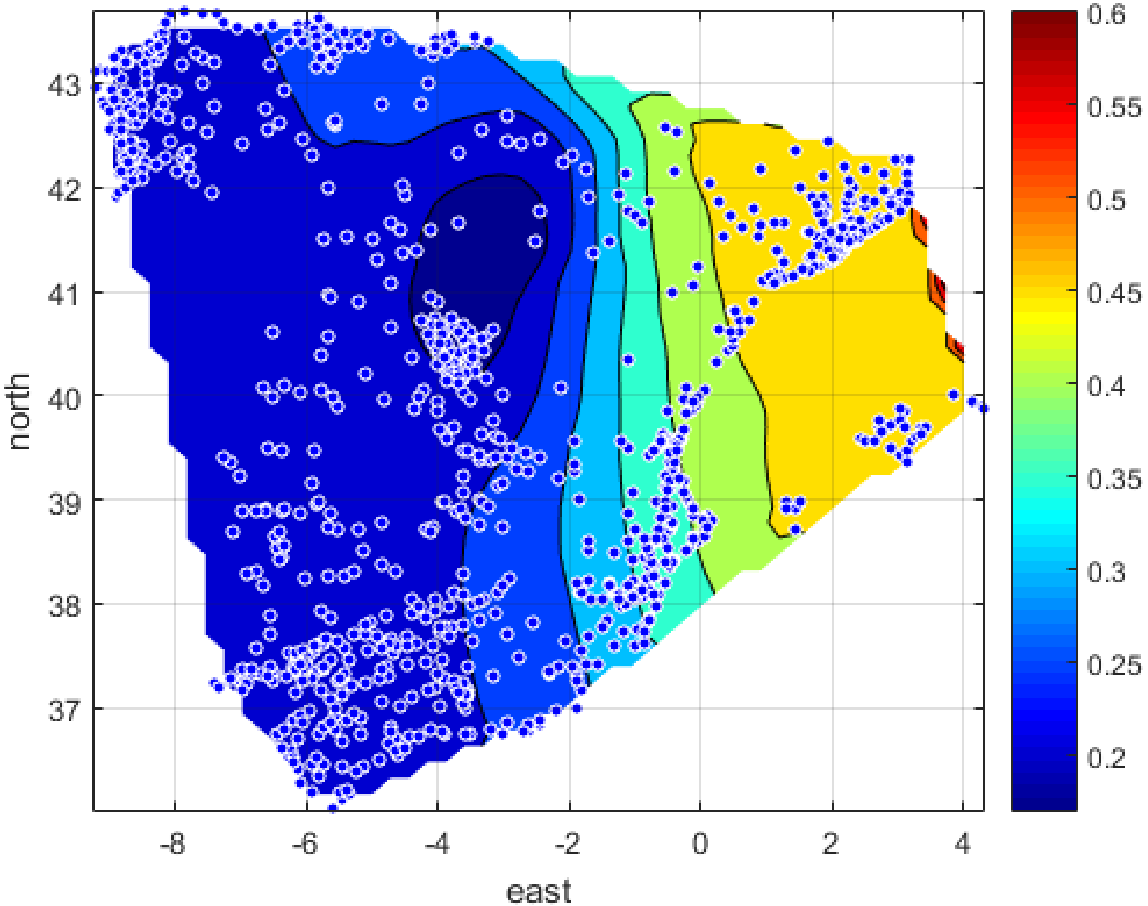
Spatial heterogeneity relates to a differentiation of the effects of space over the sample units. Formally, for spatial unit i:

$$y_i = f(x_i)_i + u_i \rightarrow y_i = \beta_i x_i + u_i$$

Lack of stability over the geographical space

Spatial heterogeneity

$$Inequality_i = f(Creative\_Class_i)_i + u_i$$



Spatial dependence

- **Spatial econometrics deals with spatial effects**

  (II) Spatial dependence

Definition:

What happens in i depends on what happens on j. Formally, for spatial unit i:

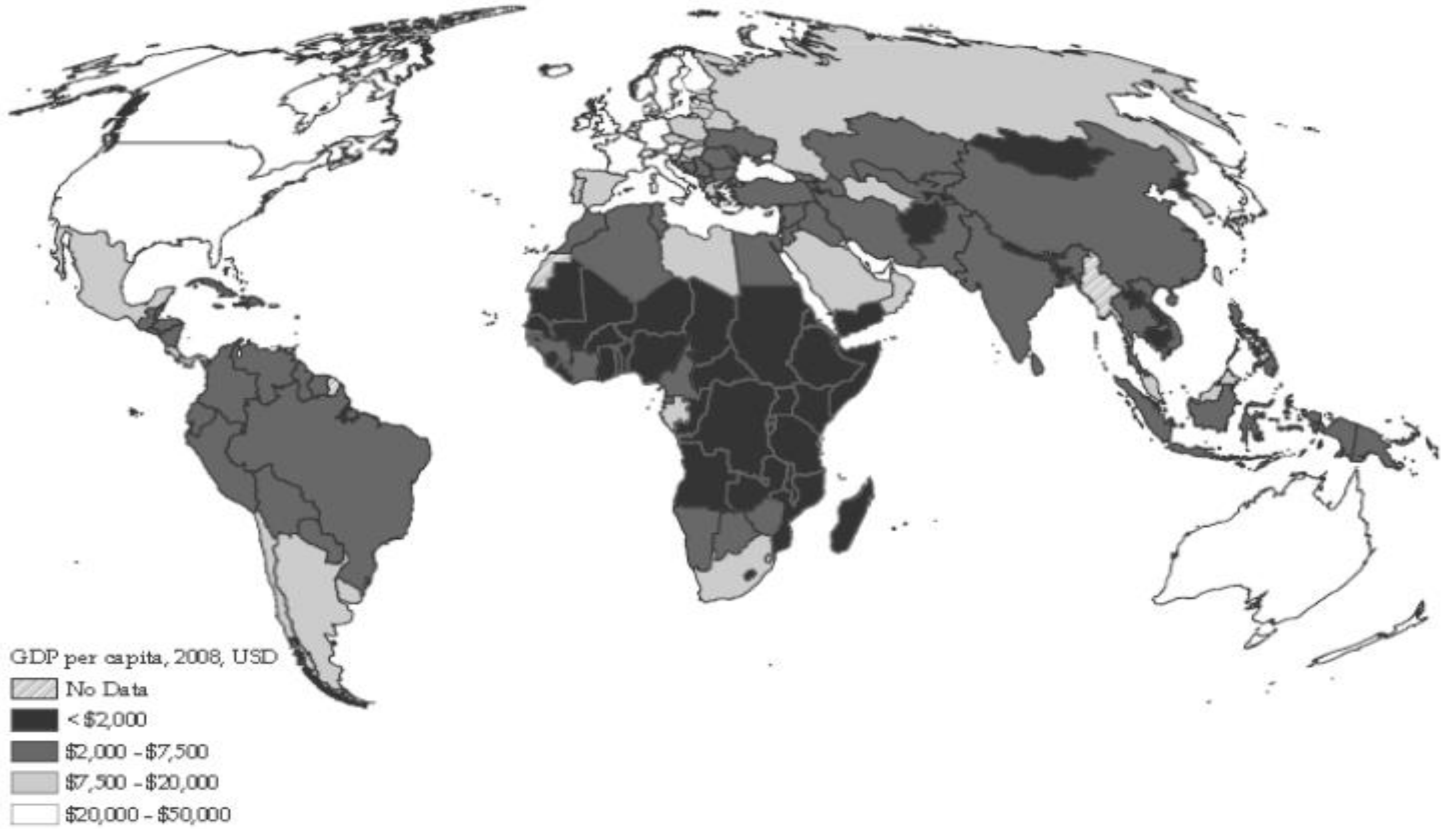
$$y_i = f\left(y_i, y_j\right) + u_i \,\forall\, i \neq j$$

Spatial dependence is a special case of cross-sectional dependence

Correlation structure is derived from a specific ordering, determined by the spatial arrangement in geographic space or in network spacee

Spatial dependence

- GDP per capita

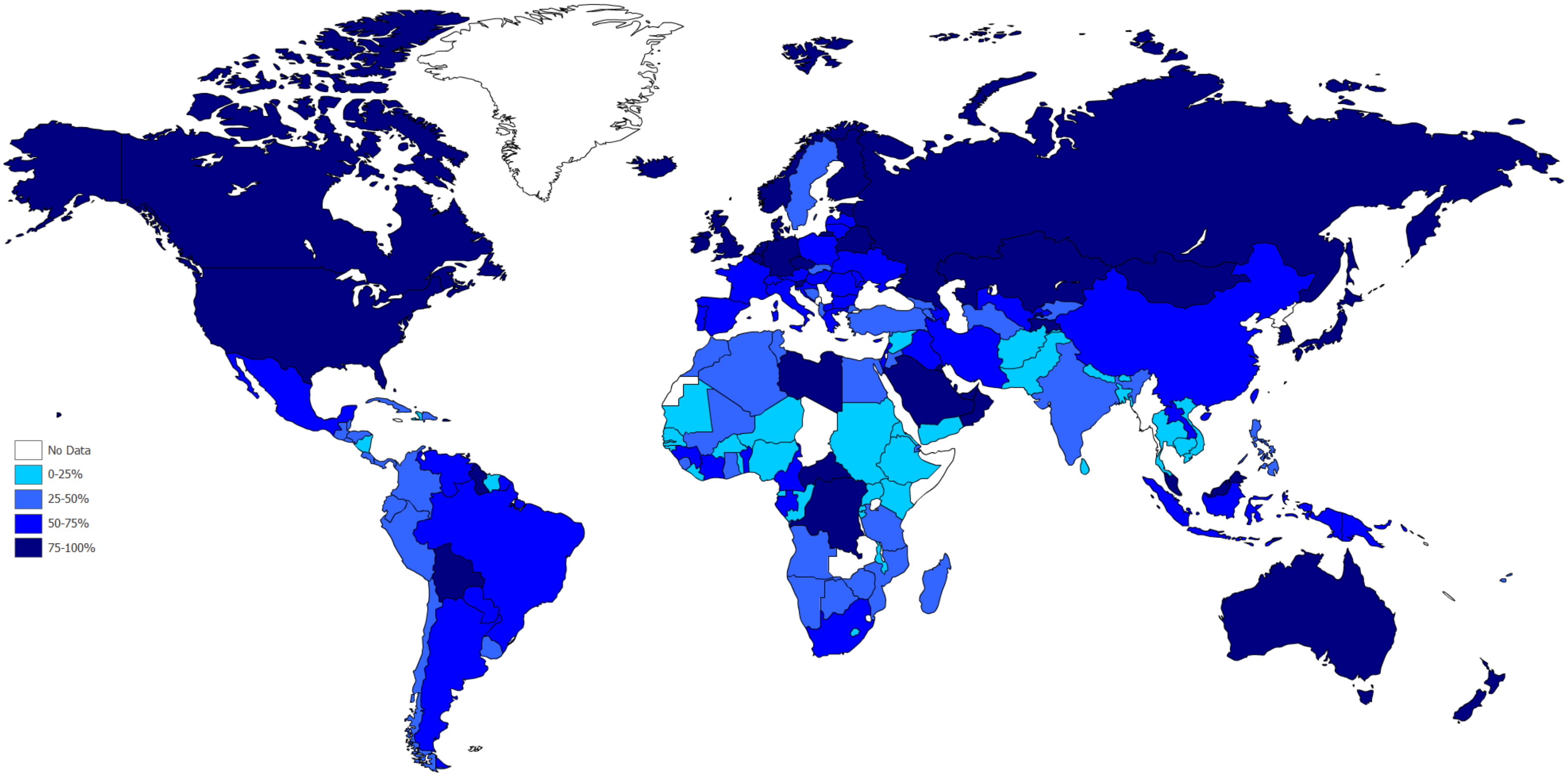


Spatial dependence

- Night Lights

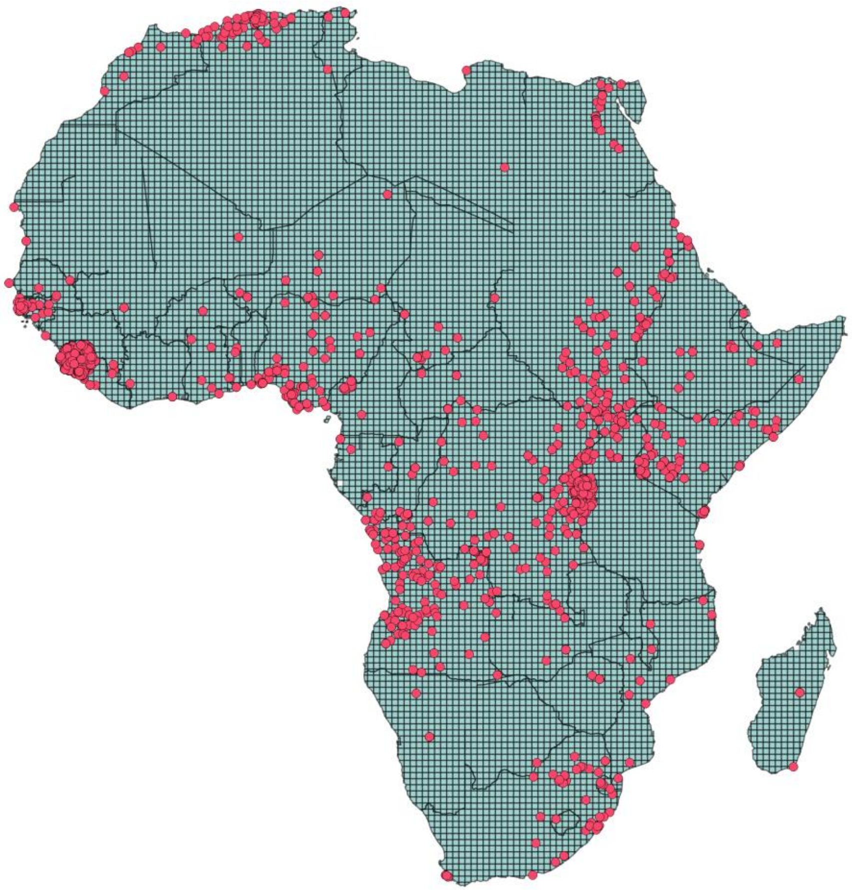
Spatial dependence

- CO2 emissions around the globe



Spatial dependence

- **Conflict events in Africa (with > 25 deaths, revolutions, riots, etc)**



Why do we need spatial econometrics?

Why do we need spatial econometrics?

- Important aspect when studying spatial units (cities, regions, countries)

  [i] Potential relationships and interactions between them

  [ii]Assumption of economies evolving as isolated entities is not reliable (co-evolution)

- Example of Interdependence: local government spending

  [i] Analyze regions/cities as independent units?

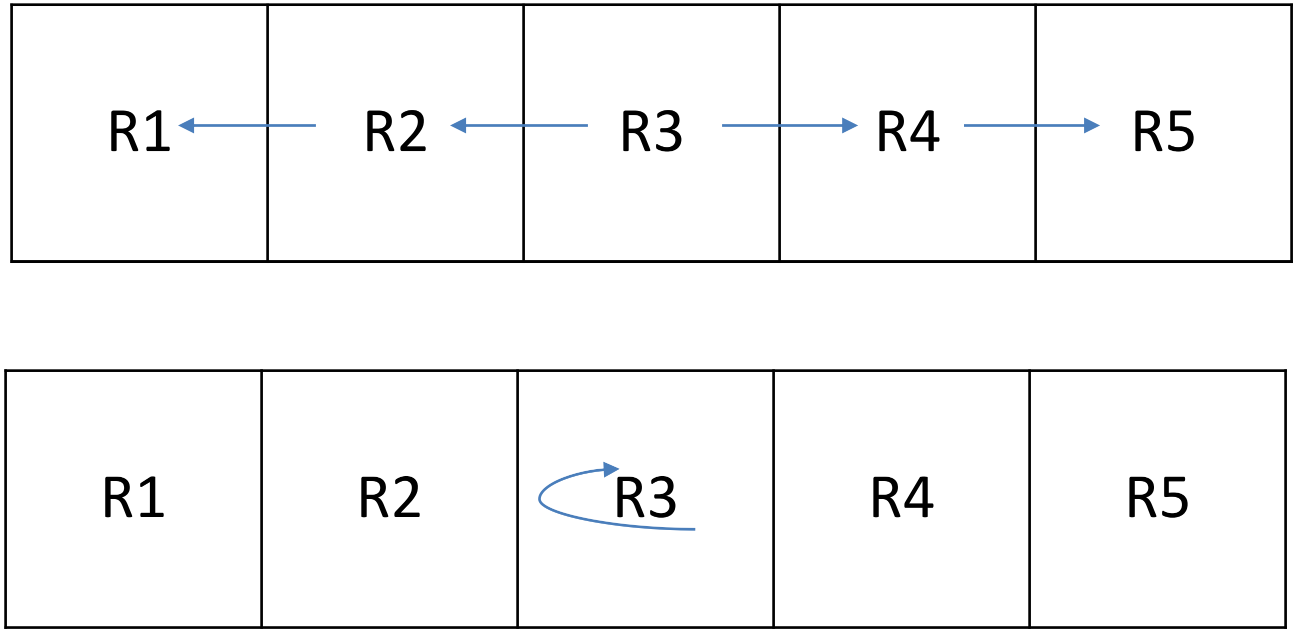
  [ii] No, Regions/cities are spatially/game-theoretically interrelated

  [iii] Existence of fiscal policy spillovers/externalities

  An increase in government spending in Pisa will affect government spending in neighboring cities but the impact will be lower for more distance cities

Why do we need spatial econometrics?

- Fiscal policy spillovers



Why do we need spatial econometrics?

- Assume we live in a world with 5 cities:

Pisa

Livorno

Lucca

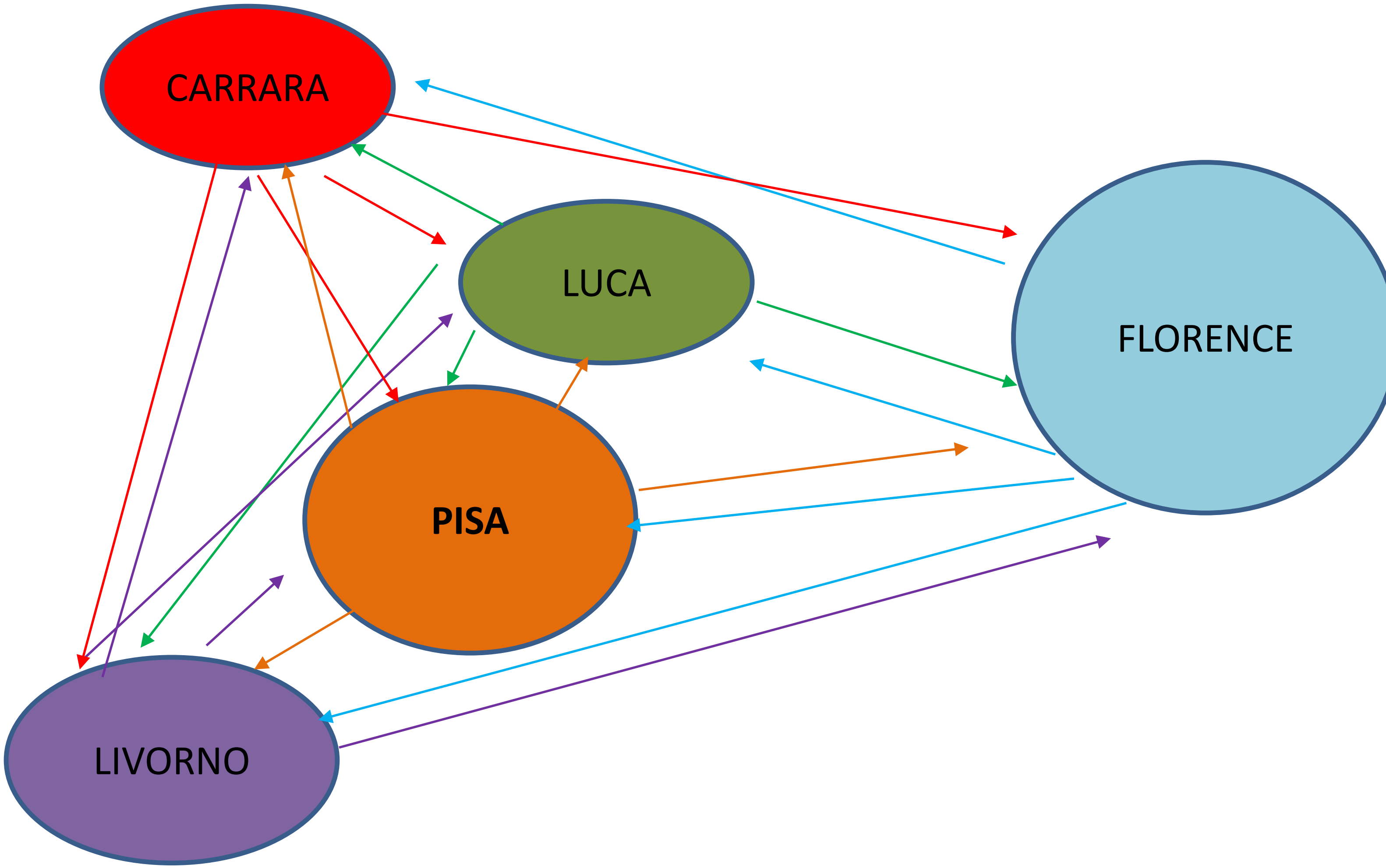
Carrara

Florence

- Each city has a different budget to spend the money of tax-payers in some service (i.e, sport centers, roads, etc)

- Decision on how much to spend in Pisa depends on what the other cities are going to spend, the decision on how much to spend in Livorno depends on what the others are going to spend….

Why do we need spatial econometrics?



Example of Fiscal Interaction Model

- Assume agents in each city i have the following utility function

$$V = v(G_i, C_i, \widetilde{G}_j)$$

where:

G_i denotes the consumption/expenditure in the public good in *i*

C_i denotes the consumption/expenditure in the private good in *i*

$\widetilde{G}_j$ denotes the consumption/expenditure of the public good in neighboring cities $j \neq i$. Neighbor's expenditure is a weighted average of the expenditure in other cities:

$$\widetilde{G}_j = 1/n \sum_{j=1}^{n} w_{ij} G_j$$

where the weights $w_{ij} \in (0,1)$ are a negative unknown function of geographical distance between *i* and *j*

Example of Fiscal Interaction Model

- Each municipality faces the following budget constraint:

$$pG_i + C_i = Y_i^d + T_i$$

where:

p is the price of the public good and the price of the private good has been normalized to 1 (i.e, numerarie)

Y_i^d is the disposable income after taxes

T_i are the transfers received from upper tier levels of government (regions, central gov)

We are going to model the optimal solution to the problem of an agent that wants to get maximum felicity levels from G and C

Example of Fiscal Interaction Model

We assumed a general concave/monotonic utility function V in G and C

$$V = v\left(C_i, G_i, \widetilde{G}_j\right)$$

To get a numerical closed solution we will assume V has the following form:

$$V = {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta} + \theta \widetilde{G}_j$$

yielding the following Lagrangian:

$$L = {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta} + \theta \widetilde{G}_j + \lambda[Y_i^d + T_i - pG_i + C_i]$$

F.O.C

$[G_i]$:$\lambda p = \beta {C_i}^{\alpha} {G_i}^{\beta-1} {\widetilde{G}_j}^{\delta}$

$[C_i]$: $p = \alpha {C_i}^{\alpha-1} {C_i}^{\beta} {\widetilde{G}_j}^{\delta}$

$[G_j]$: $\theta \; = \delta {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta-1}$

Example of Fiscal Interaction Model

Combining

$[C_i]$: $\lambda = \beta {G_i}^{\alpha} {C_i}^{\beta-1} {\widetilde{G}_j}^{\delta}$

$[G_i]$: $\lambda p = \alpha {G_i}^{\alpha-1} {C_i}^{\beta} {\widetilde{G}_j}^{\delta}$

We get that in the optimum:

$$G_i = \frac{1}{p}\frac{\beta}{\alpha} C_i$$

Now we are going to solve for C in the FOC of G_j to get an expression on how spending elsewhere affects spending in i

We can re-arrange

$$\theta = \delta {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta-1}$$

$$C_i = \left[\frac{\theta}{\delta} {G_i}^{\beta} {\widetilde{G}_j}^{\delta-1}\right]^{\frac{1}{\alpha}}$$

Example of Fiscal Interaction Model

If we plug $C_i = \left[\frac{\theta}{\delta} G_i{}^{\beta} \widetilde{G}_j{}^{\delta-1}\right]^{\frac{1}{\alpha}}$ into $G_i = \frac{1}{p}\frac{\beta}{\alpha} C_i$

we obtain:

$$G_i = \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\beta}} \left(\frac{1}{p}\right)^{\frac{\alpha}{\beta}} \left(\frac{\theta}{\delta}\right)^{\frac{1}{\beta}} \widetilde{G}_j{}^{\frac{\beta}{1-\delta}}$$

Therefore, the optimal reaction in the spending decision of a municipality i depends on the spending in j as long as:

$$\frac{\partial G_i}{\partial \tilde{G}_j} = \frac{\beta}{1-\delta} \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\beta}} \left(\frac{1}{p}\right)^{\frac{\alpha}{\beta}} \left(\frac{\theta}{\delta}\right)^{\frac{1}{\beta}} \widetilde{G}_j{}^{\frac{\beta}{1-\delta}-1} \neq 0$$

From the concavity assumption on C_i and G_i (which is a classic in utility modeling) $\beta, \alpha \in (0,1)$. The sign will depend on θ and δ

Example of Fiscal Interaction Model

- In general, we have three type of spatial interactions among governments (in spending):

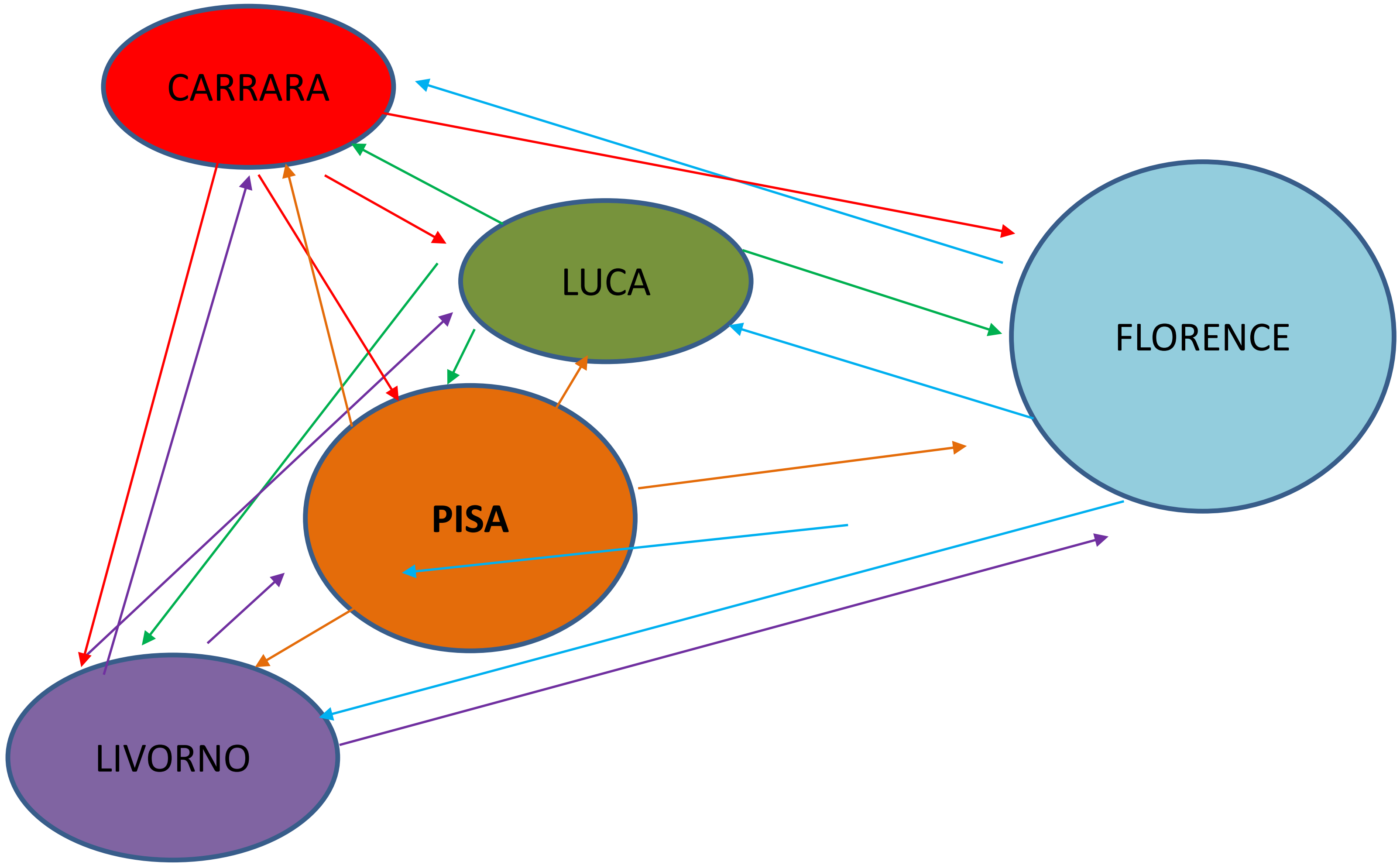
$$\frac{\partial G_i}{\partial \tilde{G}_j} = \rho < 0 \rightarrow \text{Strategic Substitutes (free-riding?)}$$

$$\frac{\partial G_i}{\partial \tilde{G}_j} = \rho > 0 \rightarrow \text{Complementarity (self-reinforcing)}$$

$$\frac{\partial G_i}{\partial \tilde{G}_j} = \rho = 0 \rightarrow \text{Independent (classical analysis)}$$

Spatial dependence

How would you model this econometrically?



Spatial dependence

- Using our previous example, we would like to estimate:

$$y_1 = \beta_{21}y_2 + \beta_{31}y_3 + \beta_{41}y_4 + \beta_{51}y_5 + u_1$$

$$y_2 = \beta_{12}y_1 + \beta_{32}y_3 + \beta_{42}y_4 + \beta_{52}y_5 + u_2$$

$$y_3 = \beta_{13}y_1 + \beta_{23}y_2 + \beta_{43}y_4 + \beta_{53}y_5 + u_3$$

$$y_4 = \beta_{14}y_1 + \beta_{24}y_2 + \beta_{34}y_3 + \beta_{54}y_4 + u_4$$

$$y_5 = \beta_{15}y_1 + \beta_{25}y_2 + \beta_{35}y_3 + \beta_{45}y_4 + u_5$$

where β_{ij} is the effect on government spending y of city i on city j
What is the problem with this modelling strategy?

K > N!!

Under standard econometric modeling, it is imposible to model such interdependence

Spatial dependence

- Spatial econometrics approach to model optimal spending reaction functions

$$y_i = \rho \sum_{j=1}^{n} w_{ij}\, y_j$$

This problem is solvable: 1 parameter and 5 data points.

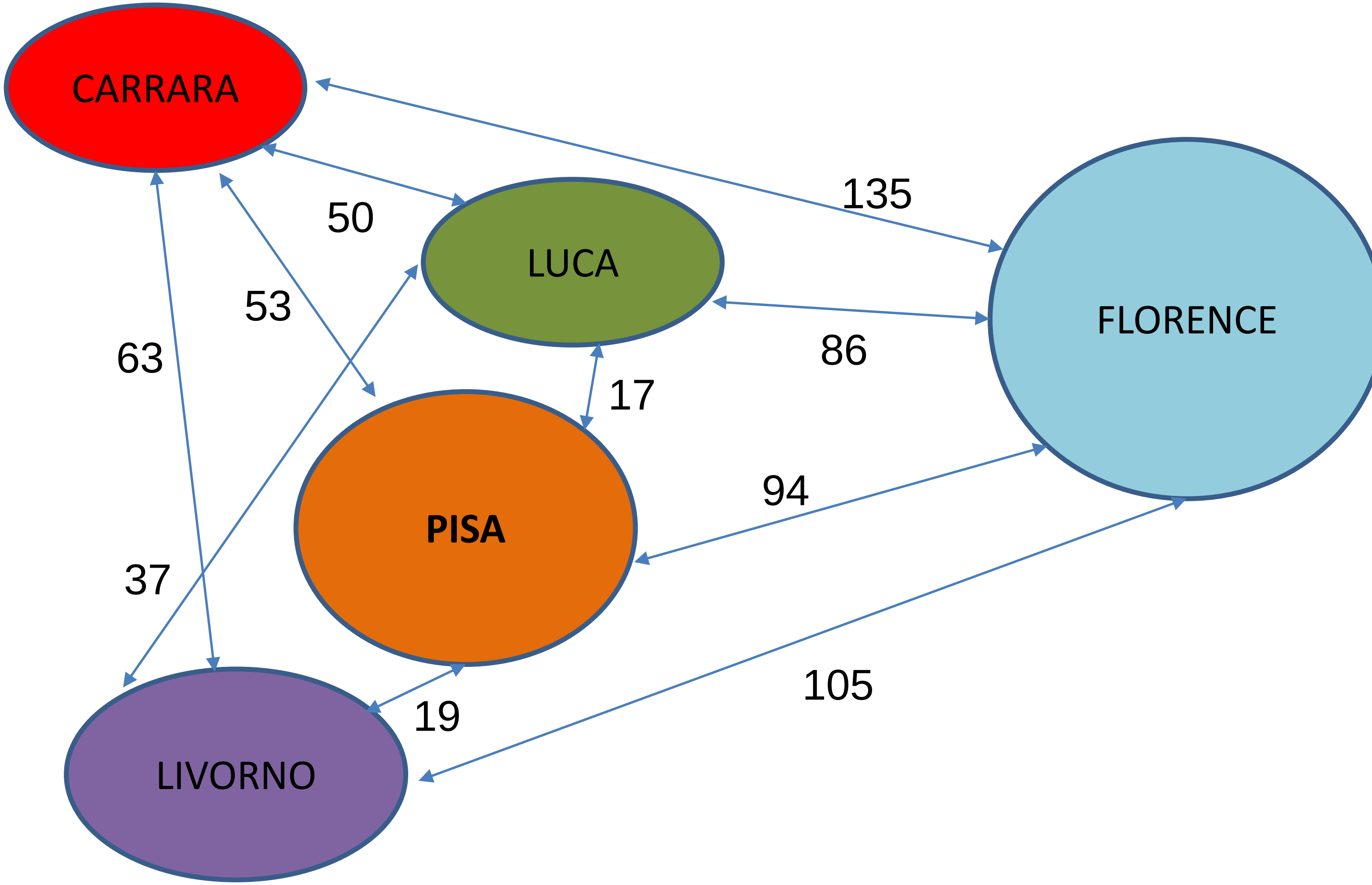
In the first two scenarios where $\rho \neq 0$, complementary/substitute relationships for each *i* depend on his neighbors and the influence they exert, measured by w_{ij}.

$w_{ij} = f(-d_{ij})$ negative function of distance and informative on how *i* is connected to the rest of the system

However, w_{ij} has to be pre-specified in advance (it is fixed) and usually is not estimated!

Spatial dependence

Distances between our 5 cities



Spatial autocorrelation

- **Distance matters**

Key Point:

First law of geography of Tobler*: "everything is related to everything else"*, but near things are more related than distant things

This first law is the foundation of the fundamental concept of spatial dependence and spatial autocorrelation

Spatial autocorrelation

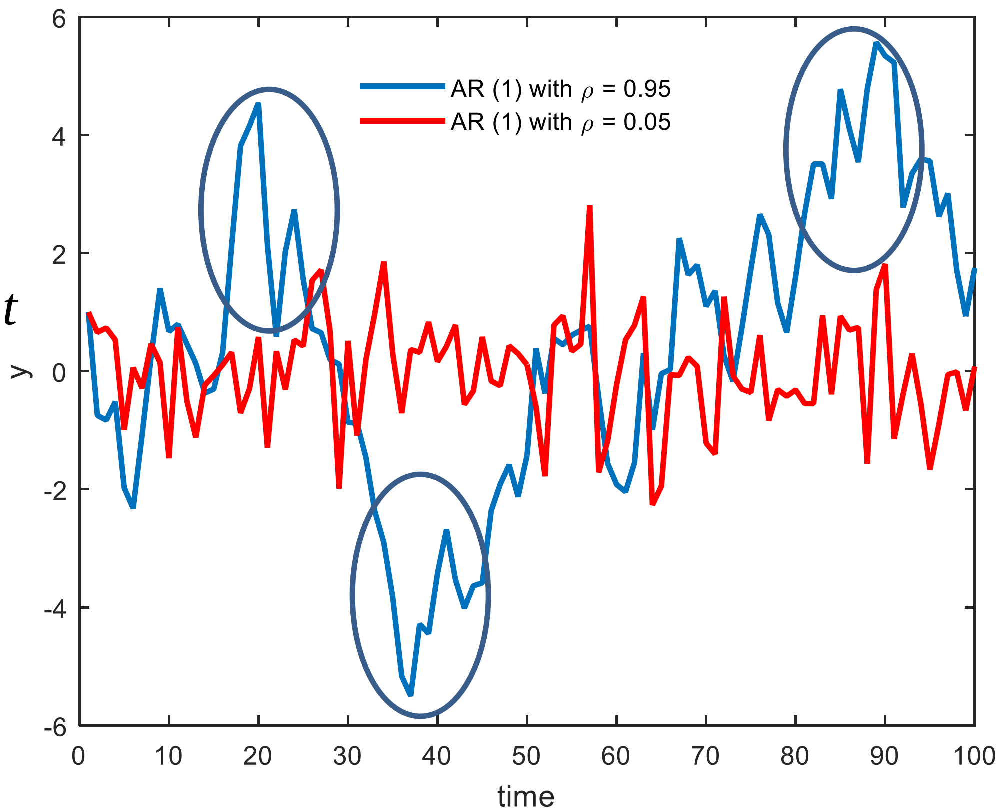
- **Spatial Autocorrelation**
  - Autocorrelation: the correlation of a variables with itself
  - Time series: the value of a variable at time t depends on the value of the same variable at time *t-1*
  - AR(1) model $y_t = \rho y_{t-1} + u_t$

High temporal correlation?

When y is high in period *t-1*, it is very likely that it will be high in *t*

Low temporal correlation?

When y is low in *t-1*, it is difficult to know what would be the value in *t*



Spatial autocorrelation

- **Spatial Autocorrelation**
  - Autocorrelation: the correlation of a variables with itself
  - Time series: the value of a variable at time *t* depends on the value of the same variable at time *t-1*
  - Space: **the correlation between the value of the variable (or attribute) at two different locations i and j**

Definition:

Correlation between the same attribute at two (or more) different locations

Value similarity coincides with location similarity

It can be positive and negative

- **Definition of positive spatial autocorrelation**

Observations with high (or low) values of a variable tend to be clustered in space

Figure: Positive spatial autocorrelation

1	1	
1	1	

Spatial autocorrelation

- **Definition of negative spatial autocorrelation**

Locations tend to be surrounded by neighbors having very disimilar values

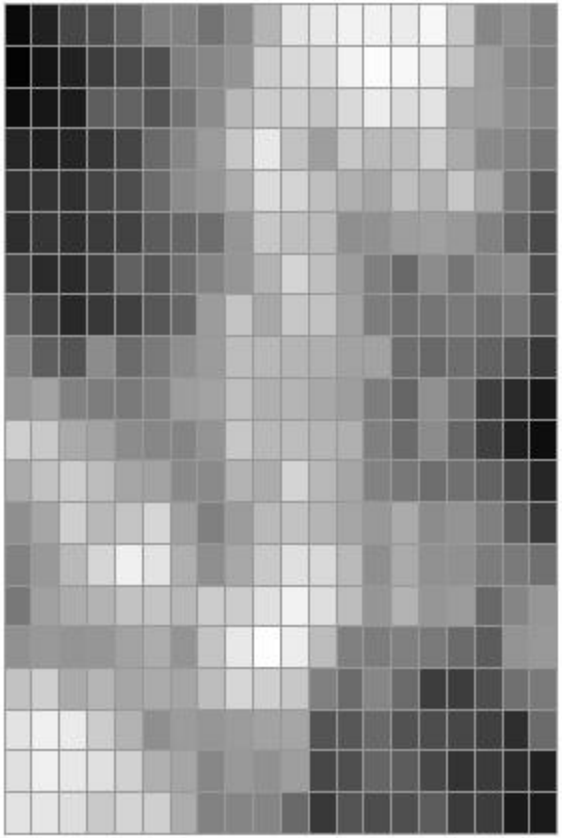
Figure: Negative autocorrelation

1		1
	1	
1		1

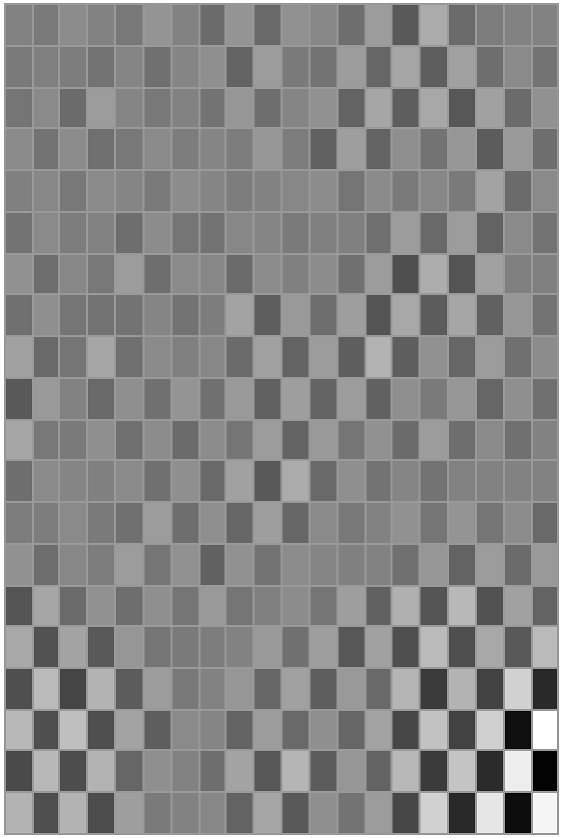
Spatial autocorrelation

- Other examples

Positive spatial autocorrelation



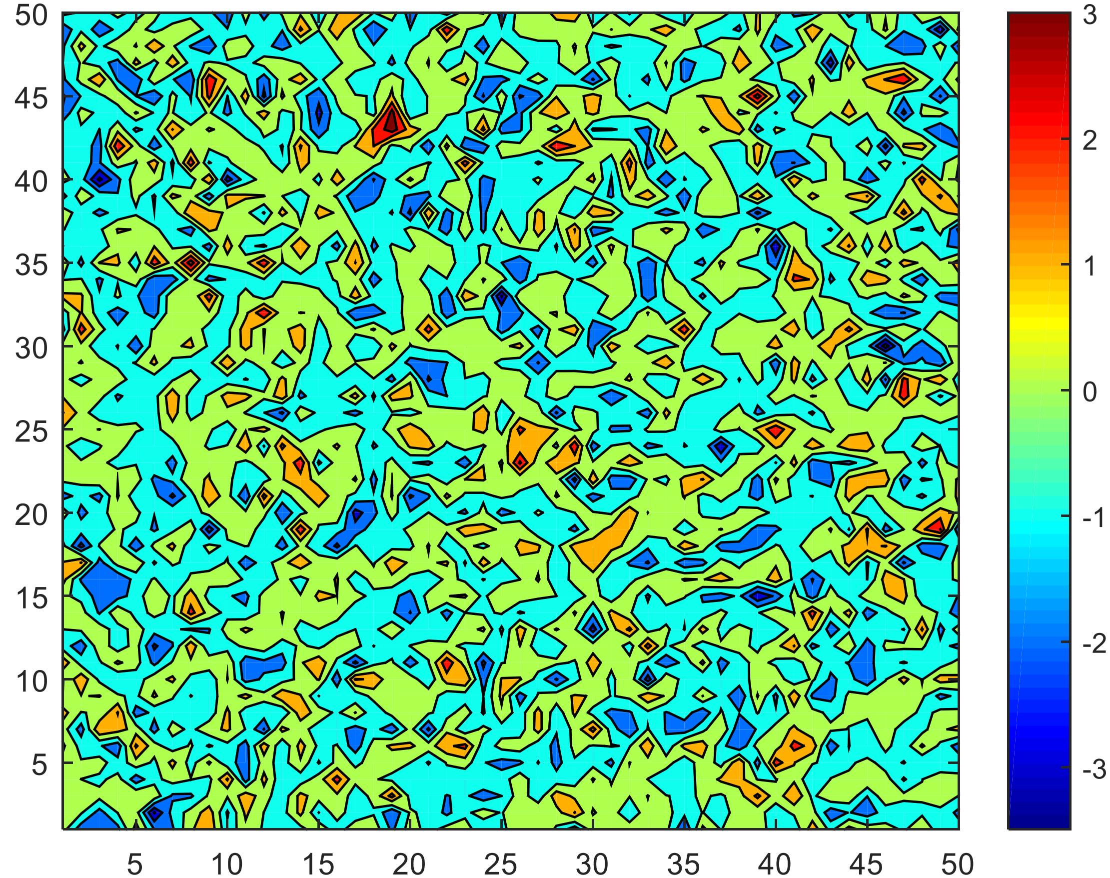
Negative spatial autocorrelation



Spatial autocorrelation

- **Spatial randomness**

When none of the two situations occur



Spatial autocorrelation

- **There are two main sources of spatial autocorrelation** (Anselin, 1988)
  - Measurement errors/Nuisance/Spatially correlated shocks
  - Substantive spatial interactions

Some researchers treat space as nuisance → Spatial filters

The **second one** is of **much more interest**

The essence of regional sciences and new economic geography is that location and distance matters

What is observed at one point is determined by what happen elsewhere in the system

Spatial Weight Matrix

- **Tobler's first law of geography**

"everything depends on everything else, but closer things more so"

- Important ideas
  - Existence of spatial dependence
  - Structure of spatial dependence
    - Distance decay
    - Closeness = Similarities

Spatial Weight Matrix

One **crucial issue** in spatial econometric is the problem of formally **incorporating spatial dependence into the model**.

- **Question**

What would be a good criterion to define closeness in space?

Or, in other words, how to determine which other units in the system influence the one under consideration?

Spatial Weight Matrix

- The device typically used in spatial analysis is the so-called spatial weight matrix, or simply **W** matrix

- Impose a structure in terms of what are the neighbors for each location

- Assigns weights that measure the intensity of the relationship among pair of spatial units

Spatial Weight Matrix

- **Definition of W matrix**

Let n be the number of spatial units. The spatial weight matrix, **W**, a n x n positive *symmetric* matrix with element $w_{i,j}$ at location *i,j*. The values of $w_{i,j}$ or the weights for each pair of locations are assigned by some preset rules which defines the spatial relations among locations. By convention $w_{i,j}$ = 0 for the diagonal elements.

$$\boldsymbol{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n-1} & w_{1,n} \\ w_{2,1} & w_{2,2} & & w_{2,n-1} & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{n-1,1} & w_{n-1,2} & & w_{n-1,n-1} & w_{n-1,n} \\ w_{n,1} & w_{n,2} & \cdots & w_{n,n-1} & w_{n,n} \end{bmatrix}$$

Entry $w_{i,j} \neq 0$ if i and j are "connected"

The tricky part is how the word connected is defined.

Spatial Weight Matrix

- **Two main approaches**
  - Contiguity
  - Based on distance

Spatial Weight Matrix

- **Weights based on boundaries**

  - The availability of poligon or lattice data permits the construction of contiguity-based spatial weight matrices. A typical specification of the contiguity relationship in the spatial weight matrix is:

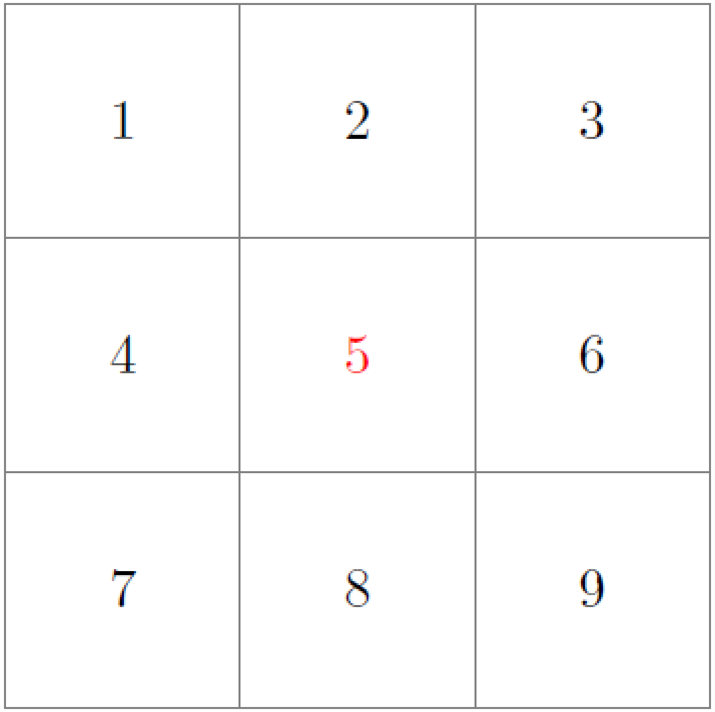
$$w_{i,j} \begin{cases} 1 \text{ if i and j are contiguous} \\ 0 \text{ if i and j not contiguous} \end{cases}$$

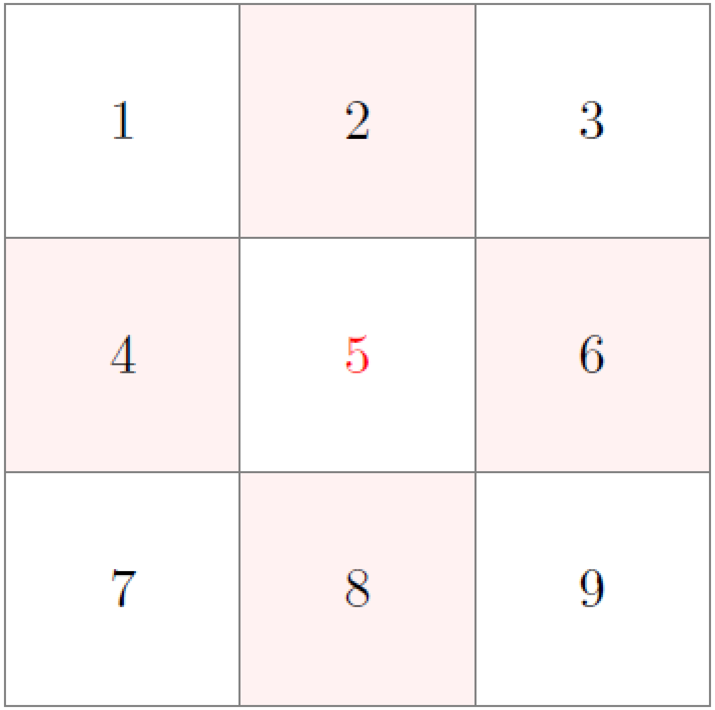
  - Binary contiguity:
    - Rook criterion (Common border)
    - Bishop criterion (Common vertex)
    - Queen criterion (Either common border or vertex)

Spatial Weight Matrix

Who are the neighbors of region 5?

Criteria: Rook contiguity (border)





Common border: 2, 4, 5, 6

Spatial Weight Matrix

Who are the neighbors of region 5?

Criteria: Bishop contiguity (vertex)

1	2	3
4	5	6
7	8	9

1	2	3
4	5	6
7	8	9

Common vertex: 1, 3, 7, 9

Spatial Weight Matrix

Who are the neighbors of region 5?

Criteria: Queen contiguity

1	2	3
4	5	6
7	8	9

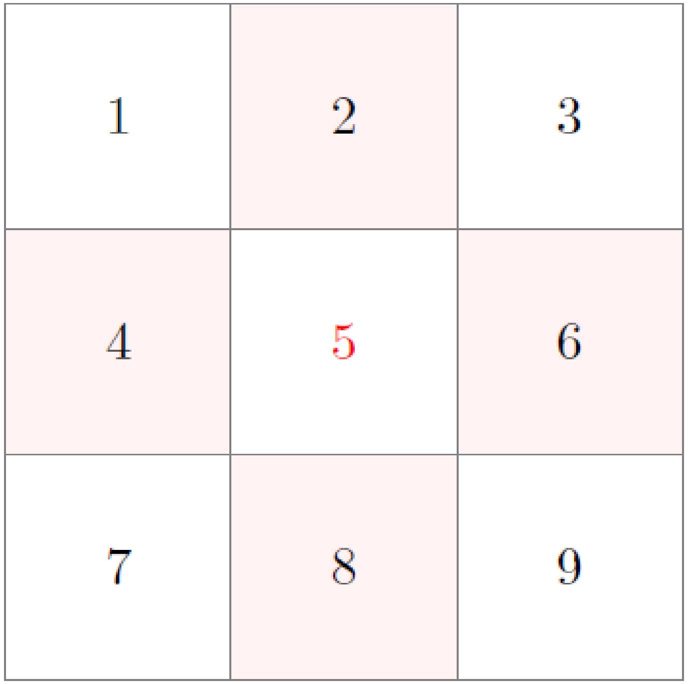
1	2	3
4	5	6
7	8	9

Common border and common vertex: 1, 2, 3, 4, 5, 6, 7, 8 and 9

Spatial Weight Matrix

- Rook contiguity (common border)

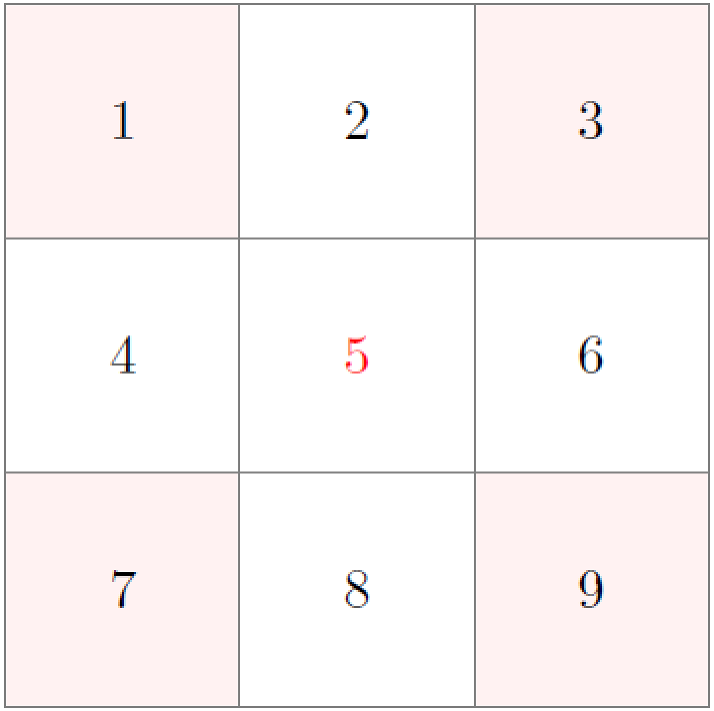
- $$\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$



Spatial Weight Matrix

- Bishop contiguity (common vertex)

- $$\boldsymbol{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Higher-order neighbors

- How to define higher order neighbors?
  - We might be interested in the neighbors of the neighbors of spatial unit i
- We define the higher-order spatial weight matrix l as $\mathbf{W}^{l}$

Spatial weight of order l=2 is given by: $\mathbf{W}^{2} = \mathbf{W}\,\mathbf{W}$

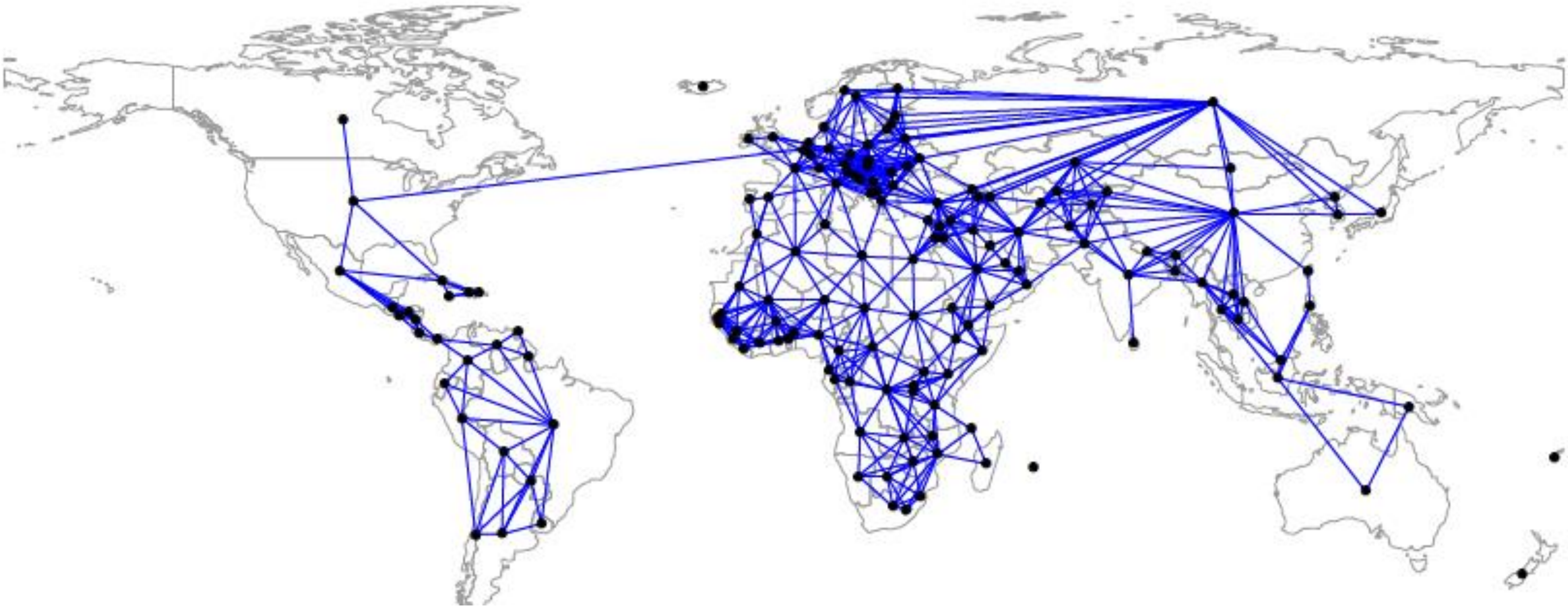
Spatial weight of order l=3 is given by $\mathbf{W}^{3} = \mathbf{W}\,\mathbf{W}\,\mathbf{W}$

Rook's case(border)

		k		
	k	*j*	*k*	
k	*j*	*k*	*j*	*k*
	k	*j*	*k*	
		k		

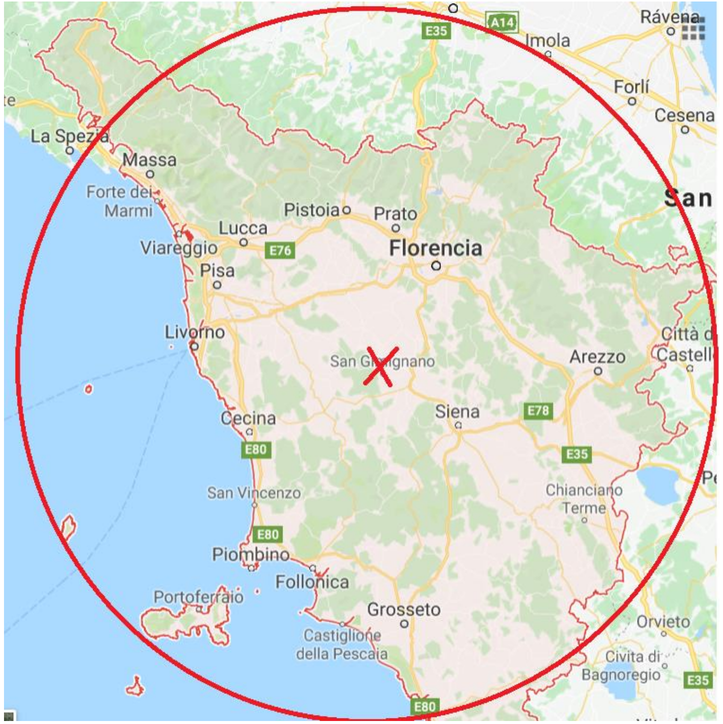
Spatial Weight Matrix

- Contiguity neighbors World **W** (with 500km of snap distance)



Spatial Weight Matrix

- Weights may be also defined as a function of the distance between region i and j, d_{ij}
- d_{ij} is usually computed as the distance between their centroids or other important unit
- Example of centroid:



Spatial Weight Matrix

Let x_i and x_j be the longitude and y_i and y_j the latitude coordinates for region i and j respectively:

- Minkowski metric of distance

$$d_{ij}^{p} = \left(\left|x_i - x_j\right|^{p} + \left|y_i - y_j\right|^{p}\right)^{1/p}$$

- Euclidean metric of distance (p=2)

$$d_{ij}^{e} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}$$

- Manhattan metric of distance (p=1)

$$d_{ij}^{m} = \left|x_i - x_j\right| \quad + \left|y_i - y_j\right|$$

Spatial Weight Matrix

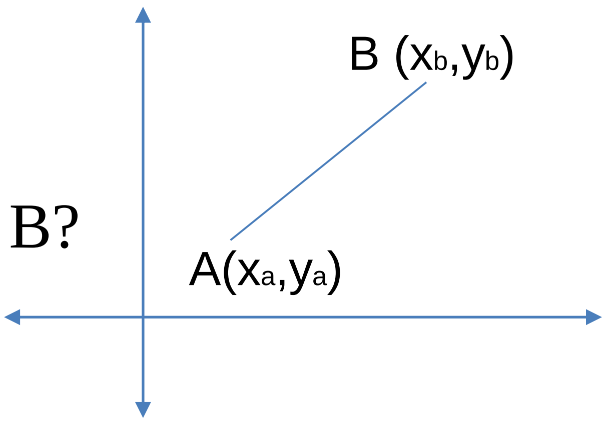
- **Euclidean distance** is the shortest distance in a plane

Numerical example:

A(1,1) and B(3,4)

What is the Euclidean distance between A and B?

Plug the numbers in:



$$d_{AB}^{e} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

$$d_{ij}^{e} = \sqrt{(1-3)^2 + (1-4)^2} = \sqrt{4+9} = 3.605$$

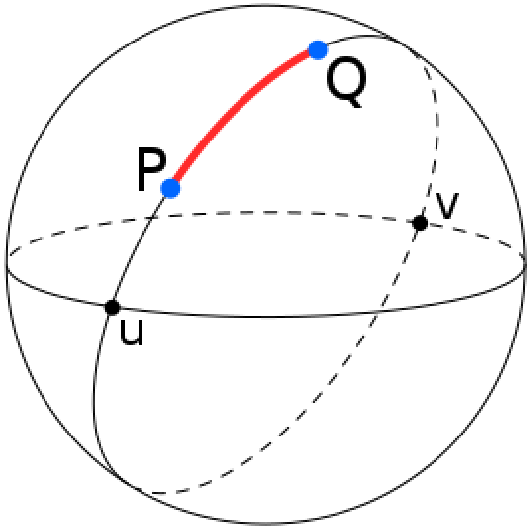
but not necessarily the shortest if you take into account the curvature of the earth

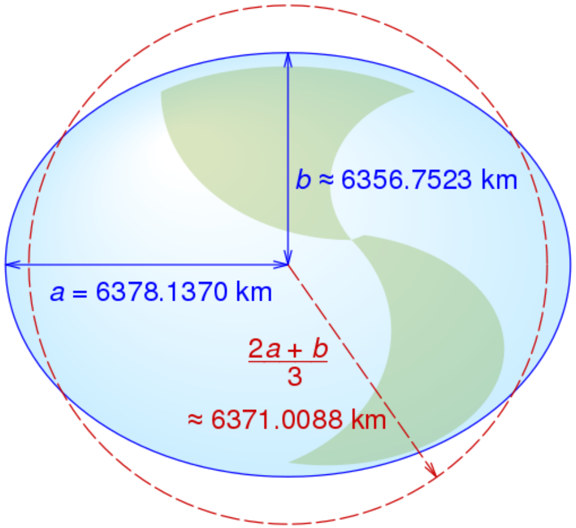
Spatial Weight Matrix

- **Great circle distance**
  - Is the shortest distance between two points in a sphere
  - Let x_i and x_j be the longitude and y_i and y_j the latitude coordinates for region i and j respectively:

$$d_{AB}^{e} = r\ x\ arccos^{-1}\left[cos\left|x_i - x_j\right|cosy_i cosy_j + siny_i siny_j\right]$$

Where r is the Earth radius (r = 6371 km)





W based on distance

- **K-nearest neighbors**

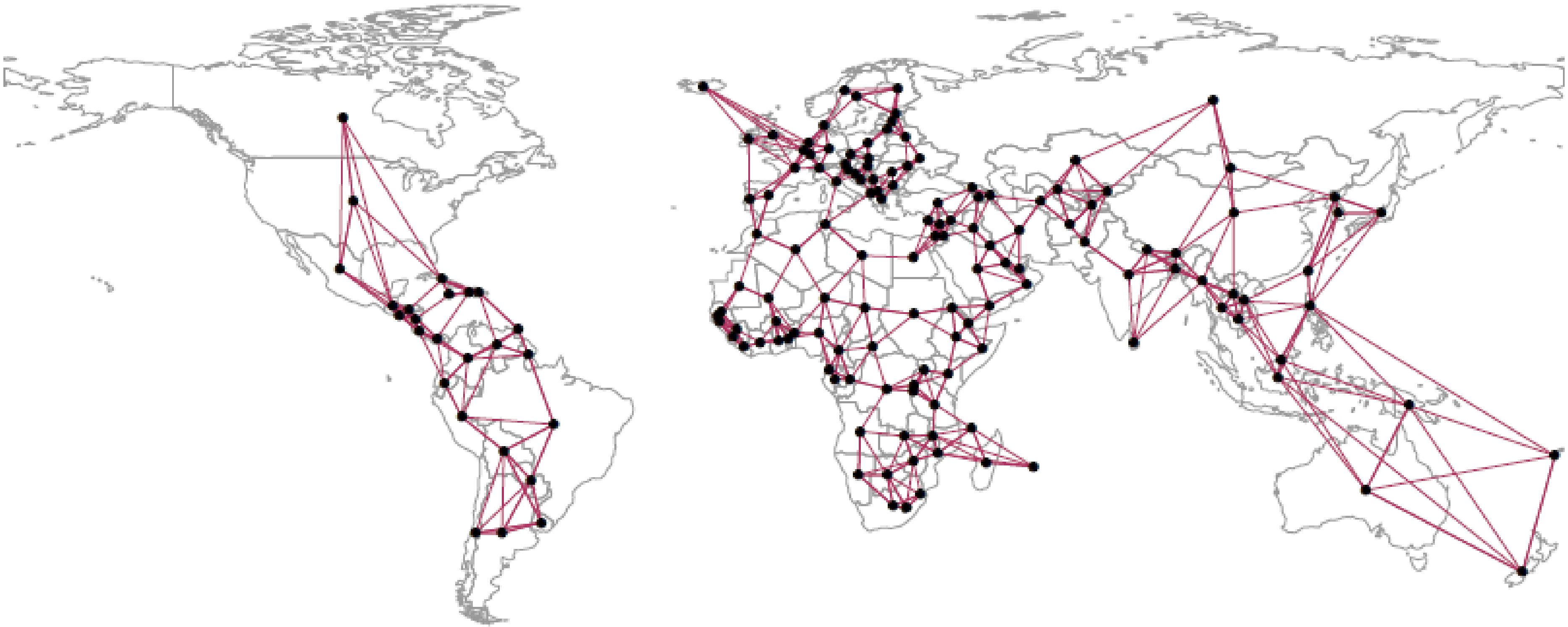
Definition:

Let centroid distances from each spatial unit *i* to all units j $\neq i$ be ranked as follows: $d_{ij(1)} < d_{ij(2)} < \cdots < d_{ij(n-1)}$ Then for each $k = 1, \ldots, n-1$, the set $N_k(i) \in \{j(1), j(2), \ldots, j(k)\}$ contains the *k* closest units to *i* (where for simplicity we ignore ties). For each given *k* , the *k*-nearest neighbor weight matrix, **W**, then has spatial weights of the form:

$$\boxed{w_{ij} = \begin{cases} 1 & , \quad j \in N_k(i) \\ 0 & , \quad otherwise \end{cases}}$$

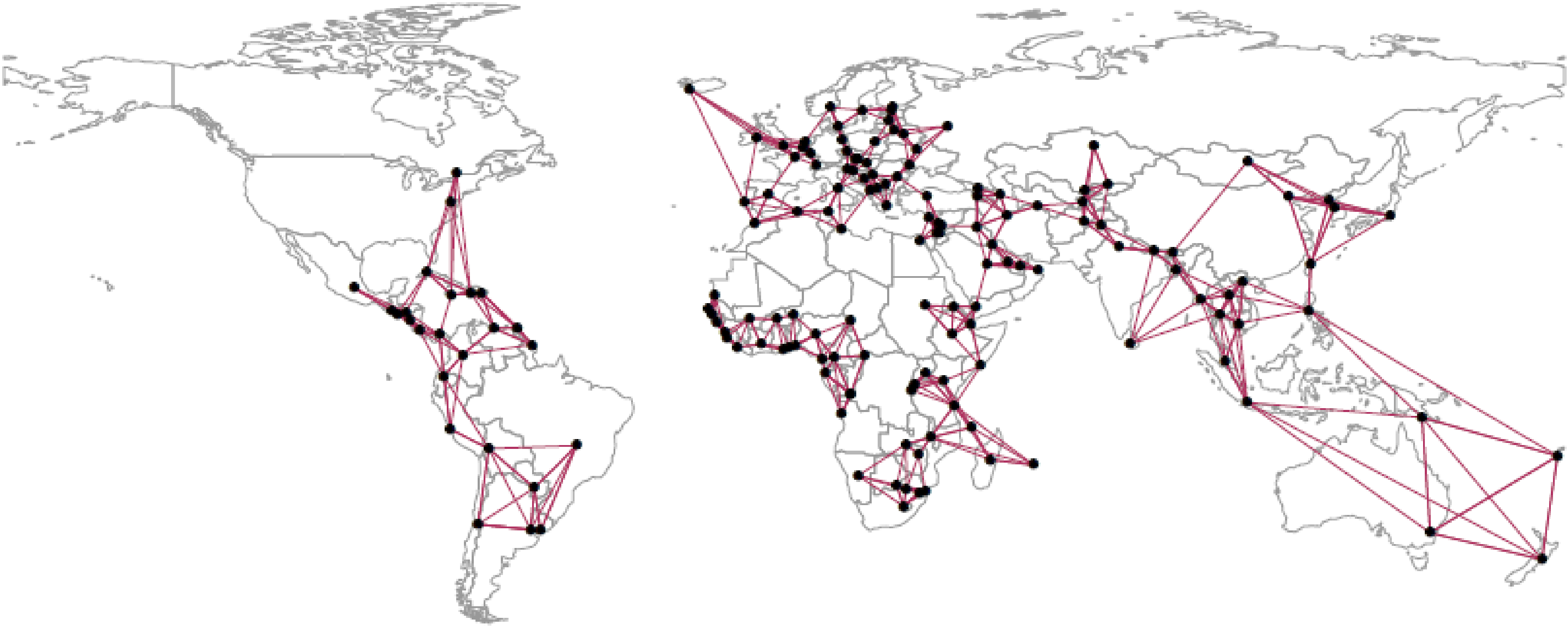
W based on distance

- k = 4 Nearest Neighbors World **W** (polygon centroids)



W based on distance

- k = 4 Nearest Neighbors World **W** (capital cities)



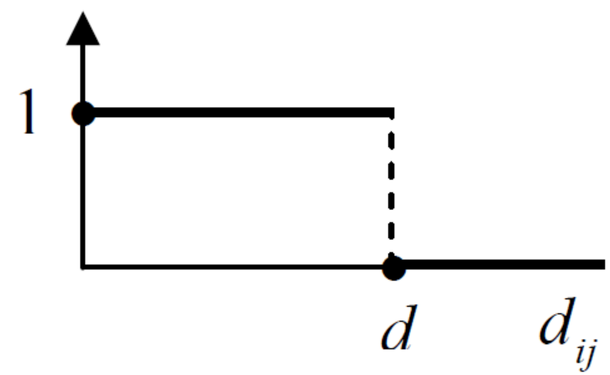
W based on distance

- **Radial distance/Distance band matrices**

Definition: If distance itself is an important criterion of spatial influence, and if *d* denotes a threshold distance (*or bandwidth*) beyond which there is no direct spatial influence between spatial units, then the corresponding radial distance weight matrix, **W** , has spatial weights of the form:

$$\boxed{w_{ij} = \begin{cases} 1 & , \; 0 \leq d_{ij} \leq d \\ 0 & , \; d_{ij} > d \end{cases}}$$

for $i \neq j$.



Typically the threshold *d* is given by some statistic of the distribution of distances in the sample (i.e, median, first quartile, etc)

W based on distance

- **Inverse power distance matrices**

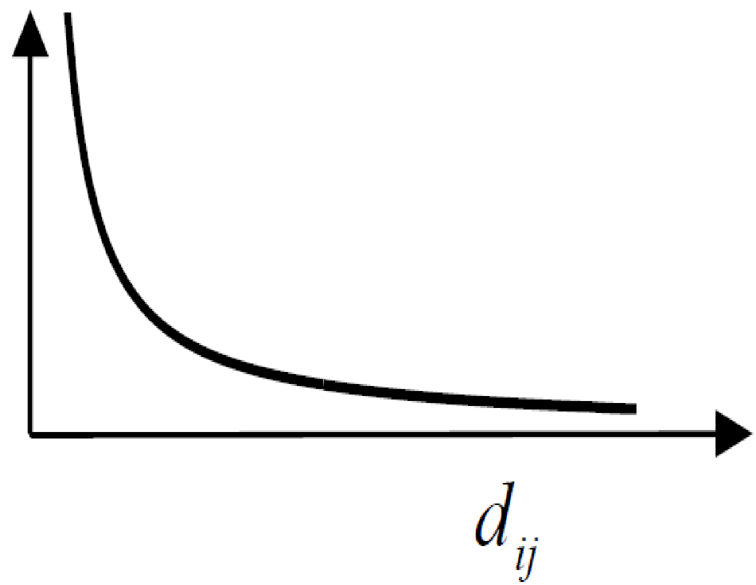
Definition: The **strength of interactions decreases as distance increases**. If there are believed to be diminishing effects, then one standard approach is to assume that **weights are a negative power function of distance** of the form:

$$\boxed{w_{ij} = d_{ij}^{-\alpha}}$$

for $i \neq j$.

Typically, $\alpha = 1$ or $\alpha = 2$

The higher is α the faster the decrease



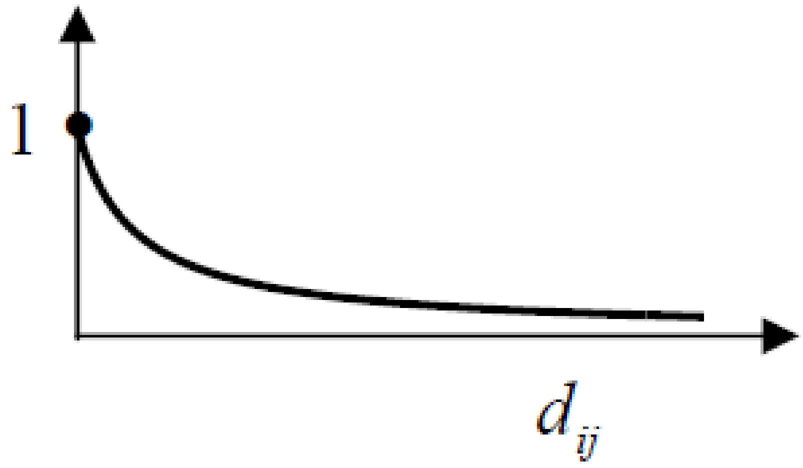
W based on distance

- **Negative exponential**
- **Definition:** An alternative to negative power functions are *negative exponential functions* of distance of the form:

$$\boxed{w_{ij} = \exp(-\alpha \, d_{ij})}$$

for $i \neq j$.

Typically, $\alpha = 1\%$ up to $\alpha = 10\%$ → faster decrease than power inverse



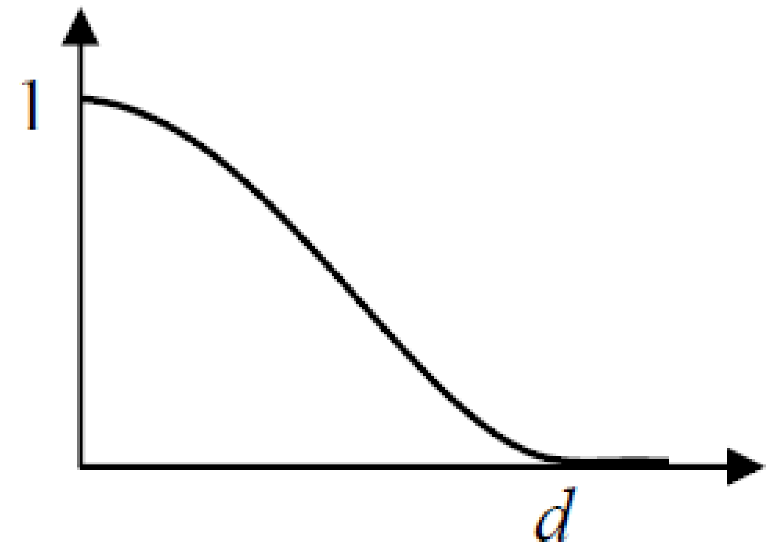
W based on distance

- **Double-power distance weights**

Definition:

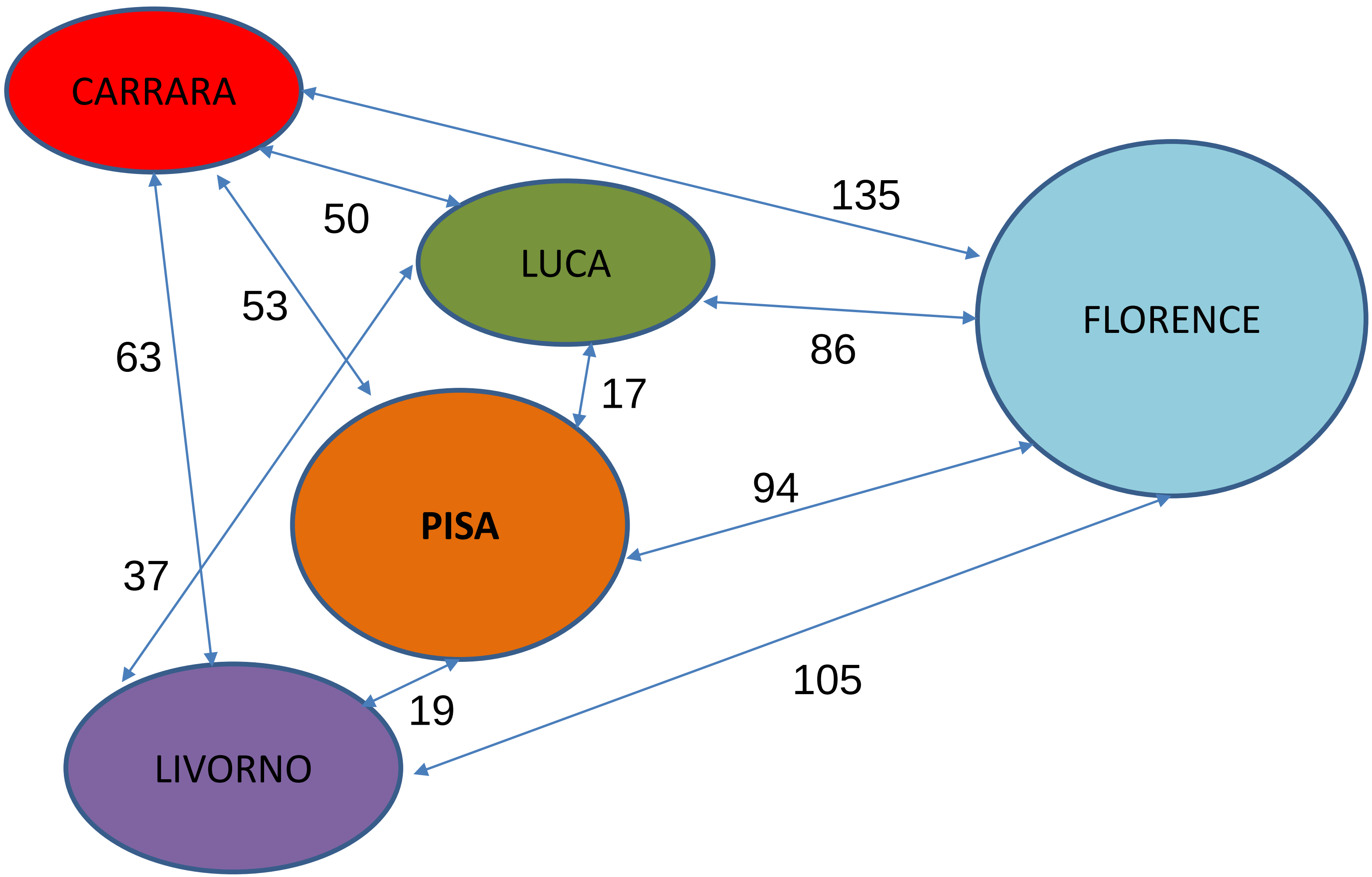
A somewhat more flexible family incorporates finite bandwidths with *"bell shaped"* functions. If *d* again denotes the maximum radius of influence (*bandwidth*) then the class of *double-power* distance weights is defined for each positive integer *k* by

$$w_{ij} = \begin{cases} \left[1-\left(d_{ij}/d\right)^k\right]^k & , \; 0 \le d_{ij} \le d \\ 0 & , \; d_{ij} > d \end{cases}$$



W based on distance

Distances between our 5 cities



W based on distance

This is our initial matrix of geographical distances

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0,0	94,022	135,231	105,481	86,408
Pisa	94,022	0,0	53,719	19,916	17,227
Carrara	135,231	53,719	0,0	63,132	50,960
Livorno	105,481	19,916	63,132	0,0	37,142
Lucca	86,408	17,227	50,960	37,142	0,0

We are going to see how this matrix of distances becomes a W.

Inspirtation: Tobler's law.

Use of some of the previous functional forms to define the weights between i and j

Examples:

k-nearest neighbors = 3

inverse power distance with α=2 to reflect a gravity function

inverse power distance with α=2 with cut-off d at 90km

W based on distance

- Original matrix of geographical distances

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0,0	94,022	135,231	105,481	86,408
Pisa	94,022	0,0	53,719	19,916	17,227
Carrara	135,231	53,719	0,0	63,132	50,960
Livorno	105,481	19,916	63,132	0,0	37,142
Lucca	86,408	17,227	50,960	37,142	0,0

- W based on the three nearest neighbors (we do it by rows)

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	1	0	1	1
Pisa	0	0	1	1	1
Carrara	0	1	0	1	1
Livorno	0	1	1	0	1
Lucca	0	1	1	1	0

This matrix is not symmetric. Florence is connected to Pisa, Livorno and Lucca and it is not a neighbor of Carrara, but none of these cities has Florence as a 3-nearest neighbor

W based on distance

- Original matrix of geographical distances

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	94,022	135,231	105,481	86,408
Pisa	94,022	0	53,719	19,916	17,227
Carrara	135,231	53,719	0	63,132	50,960
Livorno	105,481	19,916	63,132	0	37,142
Lucca	86,408	17,227	50,960	37,142	0

W based on the gravity function with α=2

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,000113	0,000055	0,000090	0,000134
Pisa	0,000113	0	0,000347	0,002521	0,003370
Carrara	0,000055	0,000347	0	0,000251	0,000385
Livorno	0,000090	0,002521	0,000251	0	0,000725
Lucca	0,000134	0,003370	0,000385	0,000725	0

This W is symmetric

W based on distance

W based on the gravity function with α=2

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,000113	0,000055	0,000090	0,000134
Pisa	0,000113	0	0,000347	0,002521	0,003370
Carrara	0,000055	0,000347	0	0,000251	0,000385
Livorno	0,000090	0,002521	0,000251	0	0,000725
Lucca	0,000134	0,003370	0,000385	0,000725	0

W based on the gravity function with α=2 with cut-off at 90km

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,00000	0,00000	0,00000	0,00013
Pisa	0,00000	0	0,00035	0,00252	0,00337
Carrara	0,00000	0,00035	0	0,00025	0,00039
Livorno	0,00000	0,00252	0,00025	0	0,00072
Lucca	0,00013	0,00337	0,00039	0,00072	0

This W is symmetric

W normalization

- The row-standardized matrix is also known in the literature as the **row-stochastic matrix**

- Definition: A real n x n matrix **W** is called **Markov** matrix or row-stochastic matrix if:

$$w_{ij} \geq 0 \text{ for } 1 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{n}$$

$$\sum_j^N w_{ij} = 1 \text{ for } 1 \leq \mathrm{i} \leq \mathrm{n}$$

W normalization

- **W’s are used to compute weighted averages** in which more weight is placed on nearby observations than on distant observations

- The elements of a row-standardized weight matrix equal:

$$w_{ij}^{s} = \frac{w_{ij}}{\sum_{j}^{N} w_{ij}}$$

- This ensures that **all weights are between 0 and 1** and facilitates the interpretation of operation with weight matrix as an averaging of neighboring values

- Under row-normalization, **the sum of elements of each row add up to 1**

- The row-standardized weight matrix also **ensures that the spatial parameter in many spatial stochastic processes are comparable** between each other

- **It is relative and not absolute distance what matters**

They are **not longer symmetric**

W normalization

- Original distance matrix

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0,0	94,022	135,231	105,481	86,408
Pisa	94,022	0,0	53,719	19,916	17,227
Carrara	135,231	53,719	0,0	63,132	50,960
Livorno	105,481	19,916	63,132	0,0	37,142
Lucca	86,408	17,227	50,960	37,142	0,0

- Row-standardized 3-nearest neighbors

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,33333333	0	0,33333333	0,33333333
Pisa	0	0	0,33333333	0,33333333	0,33333333
Carrara	0	0,33333333	0	0,33333333	0,33333333
Livorno	0	0,33333333	0,33333333	0	0,33333333
Lucca	0	0,33333333	0,33333333	0,33333333	0

- Row-standardized gravity matrix

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,289	0,140	0,230	0,342
Pisa	0,018	0	0,055	0,397	0,531
Carrara	0,053	0,334	0	0,242	0,371
Livorno	0,025	0,703	0,070	0	0,202
Lucca	0,029	0,730	0,083	0,157	0

W normalization

Row-standardized gravity matrix

	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,289	0,140	0,230	0,342
Pisa	0,018	0	0,055	0,397	0,531
Carrara	0,053	0,334	0	0,242	0,371
Livorno	0,025	0,703	0,070	0	0,202
Lucca	0,029	0,730	0,083	0,157	0

Row-standardized gravity matrix + cut-off

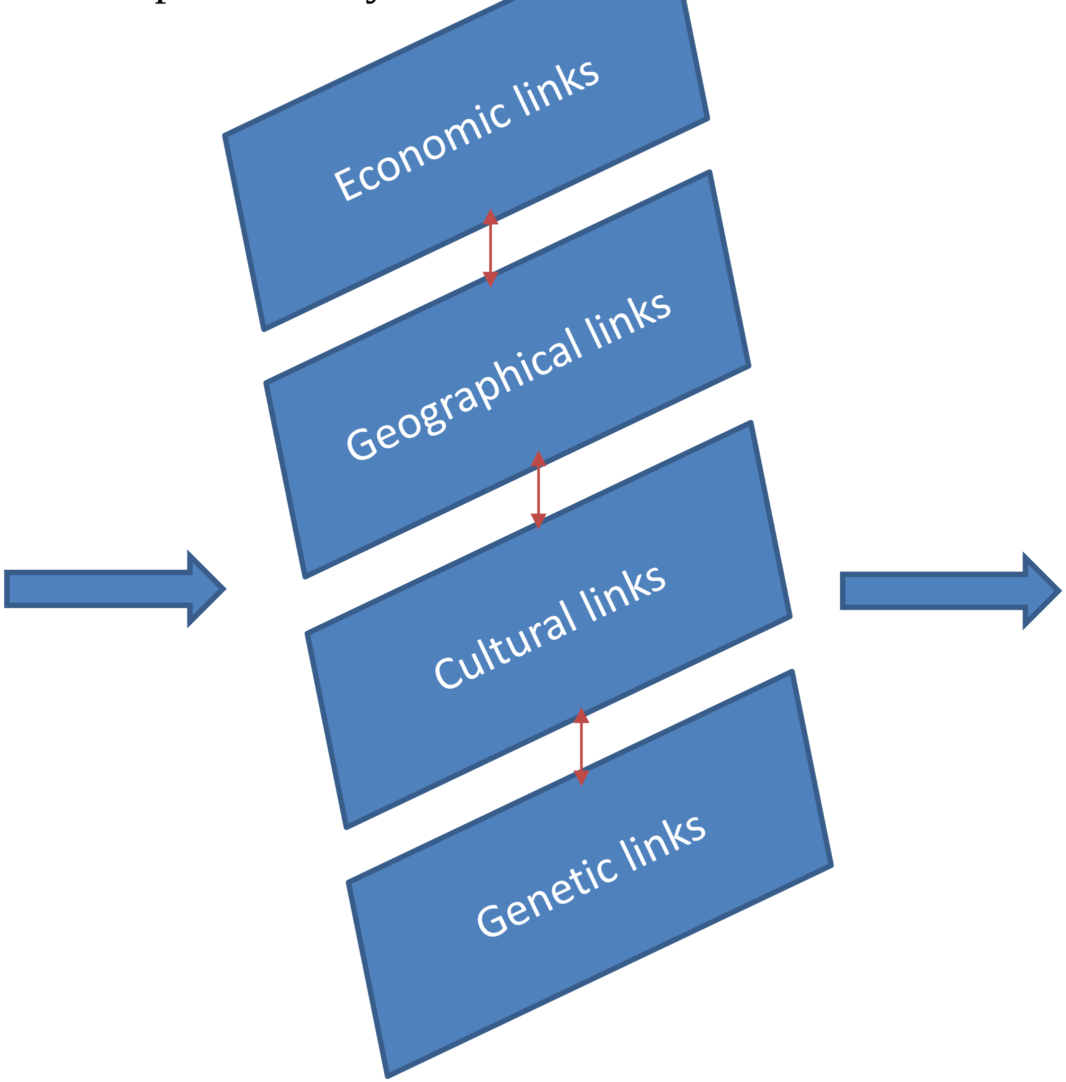
	Florence	Pisa	Carrara	Livorno	Lucca
Florence	0	0,000	0,000	0,000	1,000
Pisa	0,000	0	0,055	0,404	0,540
Carrara	0,000	0,352	0	0,255	0,391
Livorno	0,000	0,720	0,071	0	0,207
Lucca	0,029	0,730	0,083	0,157	0

Space is more than geography

Set of interdependent layers

How does a change of x in unit i at period t

Δx_{it}^{k}



Affects other variable y in another unit j at period t+h

Δy_{jt+h} ?

Other W matrices

- Geographical distance is key explaining the strength of interactions among agents, countries, regions, etc. But not the only force.

- Other forms of distance-based interactions/linkages (Spolaore and Wacziarg, 2016) :
  - **Genetics**
  - **Culture: languages**
  - **Culture: religions**
  - **Culture: social values**
  - Economy: productive specialization/technology
  - Economy: trade linkages

Memetic distances

Other W matrices

- Until recently, most of spatial analysis just considered models with one or two W

- Increasing focus in modeling interdependence of events at different spatial units as a function of a complex set of concepts of distance:

$$y_i = \rho \sum_{j=1}^{n} w_{ij}^{H} y_j$$

$$w_{ij}^{H} = f(w_{ij}^{GEO}, w_{ij}^{TRADE}, w_{ij}^{CULTURE}, w_{ij}^{GENES})$$

- Examples of Hybrid W:
  - LeSage and Deberasy (2017) → estimation of convex combinations
  - Fiaschi et al (2017) → frequentist model averaging
  - Rios and Gianmoena (2017) → bayesian model averaging

Other W matrices

- **Genetic distance-based interactions**

- In order to capture global differences in gene frequencies between populations, geneticists have devised summary measures, called genetic distances.
- One of the most widely used measures of genetic distance, suggested by Sewall Wright (1951), is called FST:

$$FST = \frac{V_p}{p(1-p)}$$

where:

Vp is the variance between gene frequencies across populations

p their average gene frequencies.

In general, 0 <FST< 1

In particular, FST = 0 when the frequencies of the genes are identical across populations ($\sigma = 0$) and FST =1 when one population has only genes the other one has not ($\sigma = p$)

Other W matrices

- **Due to random drift, FST genetic distance** has a very useful interpretation in terms of separation time, **defined as the time since two populations shared their last common ancestors** (since they were the same population)

FST:

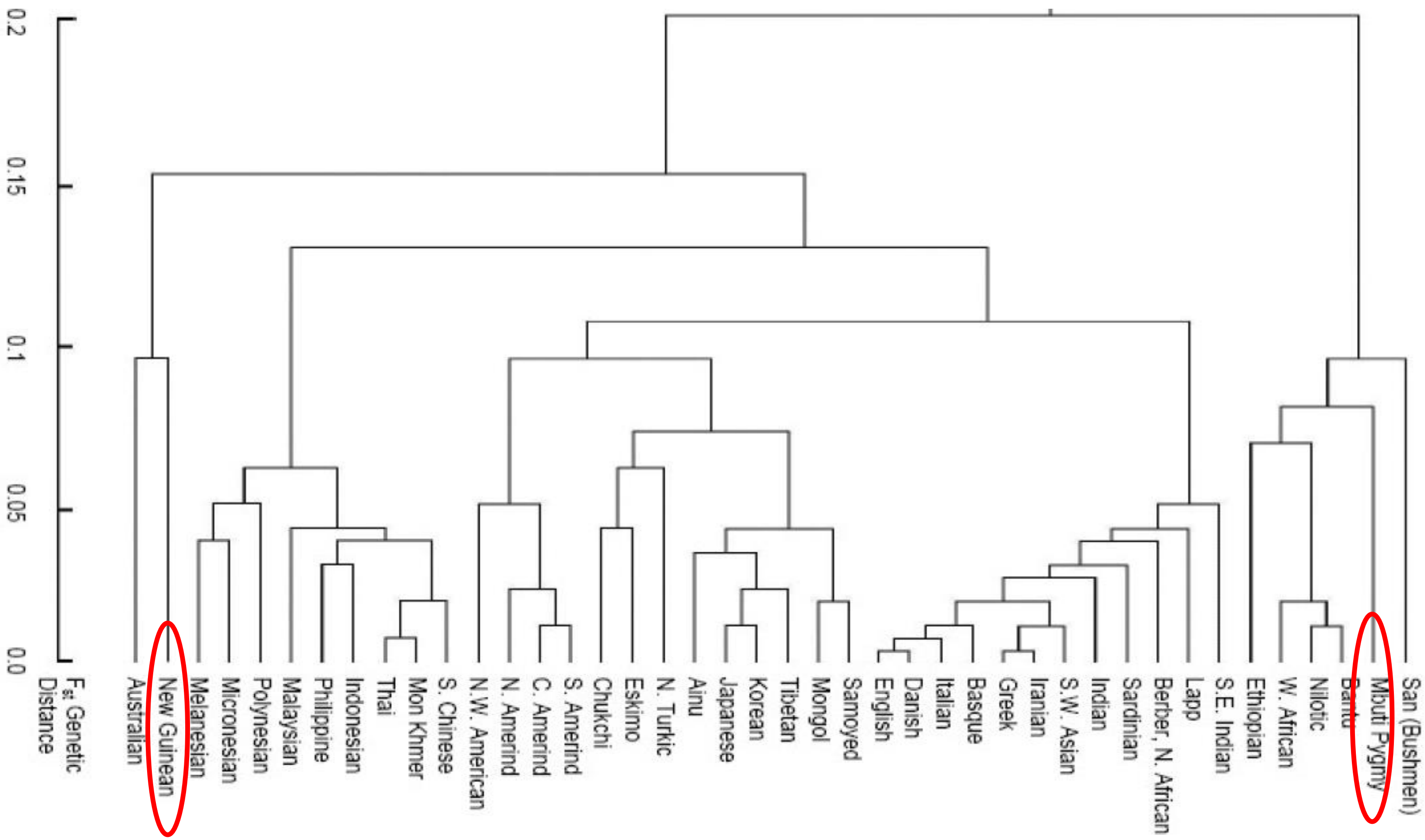
$$FST = 1 - e^{-\frac{t}{2N}}$$

t: is the separation time between two populations whose ancestors were part of the same population

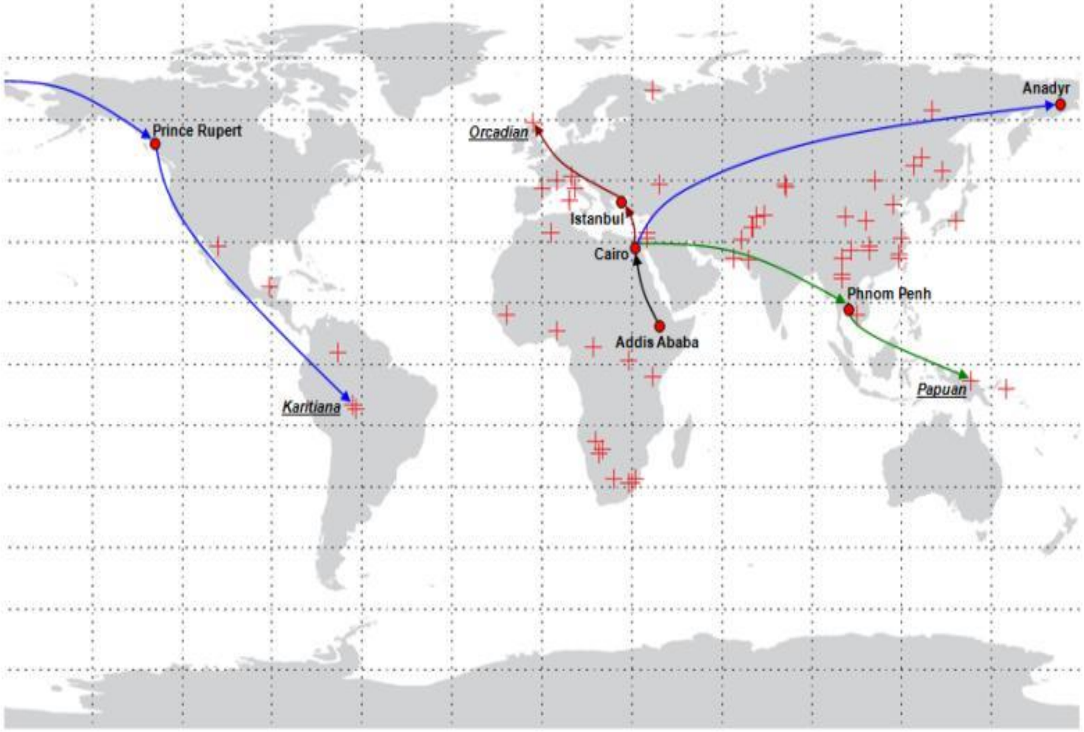
- This means that **the genetic distance between two cousin populations** is roughly **proportional to the time since the ancestors of the two populations split** and formed separate populations.

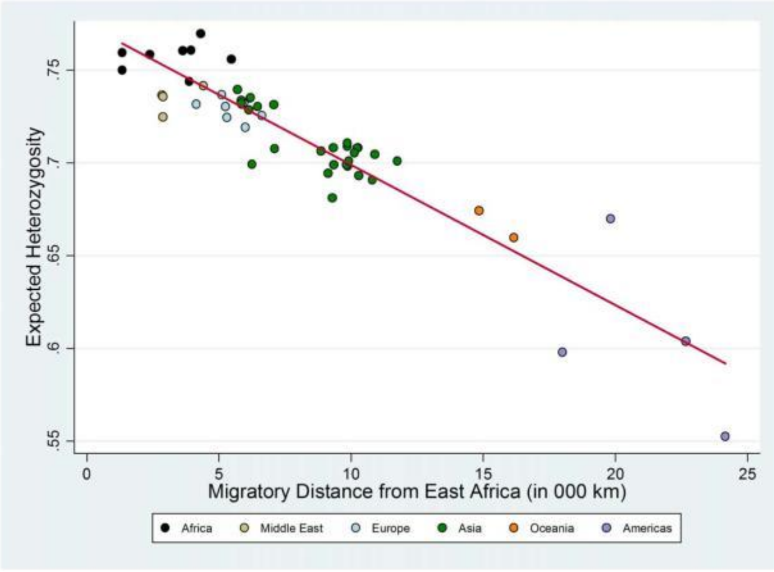
- Recently, FST (weighted) by populations within and a country and their sized developed by Spolaore and Wacziarg (2016)

- **Phylogenetic tree**



Out of Africa hypothesis





Other W matrices

- **Language distance-based interactions**

- To capture cultural distance (language) literature usually employs *language trees* or *lexicostatistics*

- **Language distance is strongly related with genetic distance as it is transmitted from parents to children within populations**, and because linguistic differentiation, like genetic differentiation, results over time from horizontal separation between populations.

- The *classification of languages into trees* is based on a methodology *borrowed from cladistics*. Linguists group languages into families based on perceived similarities between them.

- Sources: Ethnologue / Fearon (2003)

Other W matrices

- **French** is classified as:

"*Indo-European - Italic - Romance - Italo-Western* - Western - Gallo-Iberian - Gallo-Romance - Gallo-Rhaetian - Oil - Français."

Italian is classified as:

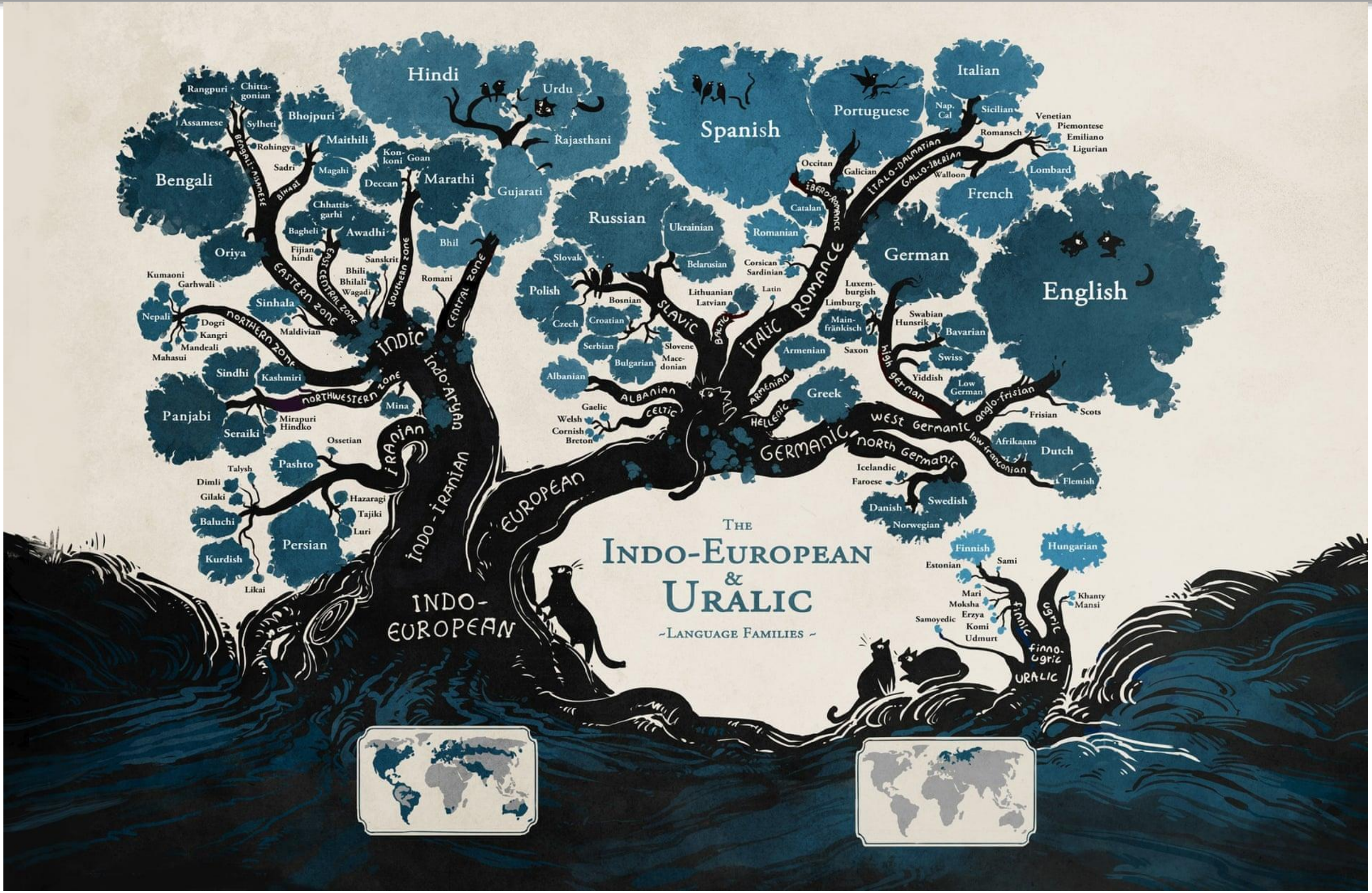
"*Indo-European - Italic - Romance- Italo-Western* - Italo-Dalmatian".

This can serve as the basis for characterizing the linguistic distance between French and Italian, because **Italian shares 4 nodes with French**.

Variation in the number of common nodes corresponds to variation in linguistic distance.

French and Italian, for instance, share no common nodes with non Indo-European languages, and are therefore at a higher linguistic distance from them than they are with each other.

Other W matrices



Other W matrices

- The most commonly employed language distance metric for two countries 1 and 2 is given by:

$$L^w = \sum_{i=1}^{I} \sum_{j=1}^{J} s_{1i} \, s_{2j} c_{ij}$$

where:

s_{1i} is the share of population in country 1 that speaks language i

s_{2j} is the share of population in country 2 that speaks language j

c_{ij} is the number of common nodes of j and i

- The drawback of tree-based measures is that linguistic distance is calculated on a discrete number of common nodes, which could be an imperfect measure of separation between language

Other W matrices

- **Language distance based on “cognate distance”**

Based on *lexicostatistics*, the branch of **quantitative linguistics classifying language groups based on whether words used to convey some common meanings** such as “ mother” or “table” are cognate, i.e. stem from the same ancestor word.

Two **languages with many cognate words are linguistically closer** than those with non-cognate words.

For instance, the words *“tavola”* in Italian and *“table”* in French both stem from the common Latin term *“tabula”*. They are therefore cognate.

Replicating this over a large number of meanings, the percentage of cognate words is a measure of linguistic proximity.

Other W matrices

- **Religious distance-based interactions**
- Let's say you are a roman-catholic..

Who is religiously closer? A hinduist or a jew?

Again using the number of nodes c_{ij}

and the shares of population s following religion i and religion j in a pair of countries

$$R^w = \sum_{i=1}^{I} \sum_{j=1}^{J} s_{1i}\, s_{2j} c_{ij}$$

1.0 Asia-born Religion
- 1.1 South Asian Religions
  - 1.11 Hinduism
- 1.2 Far Eastern Religions
  - 1.21 Taoism
  - 1.22 Buddhism
    - 1.221 Therevada
    - 1.222 Cao Dai
    - 1.223 Hoa Hao

2.0 Near Eastern Monotheistic Religion
- 2.1 Christianity
  - 2.11 Western Catholicism
    - 2.111 Roman Catholic
    - 2.112 Protestant
      - 2.1121 Anglican
      - 2.1122 Lutheran
      - 2.1123 Presbyterian
      - 2.1124 Methodist
      - 2.1125 Baptist
      - 2.1126 Calvinist
      - 2.1127 Kimbanguist
      - 2.1128 Church of Ireland
  - 2.12 Eastern Orthodox
    - 2.121 Greek Orthodox
    - 2.122 Russian Orthodox
      - 2.1221 Old Believers
    - 2.123 Ukranian Orthodox
      - 2.1231 Russian Patriarchy
      - 2.1232 Kiev Patriarchy
    - 2.124 Albanian Orthodox
    - 2.125 Armenian Orthodox
    - 2.126 Bulgarian Orthodox
    - 2.127 Georgian Orthodox
    - 2.128 Macedonian Orthodox
    - 2.129 Romanian Orthodox
- 2.2 Islam
  - 2.21 Sunni Islam
    - 2.211 Shaf'i Sunni
  - 2.22 Shi'I Islam
    - 2.221 Ibadi Shi'i
    - 2.222 Alevi Shi'i
    - 2.223 Zaydi Shi'i
  - 2.23 Druze
- 2.3 Judaism

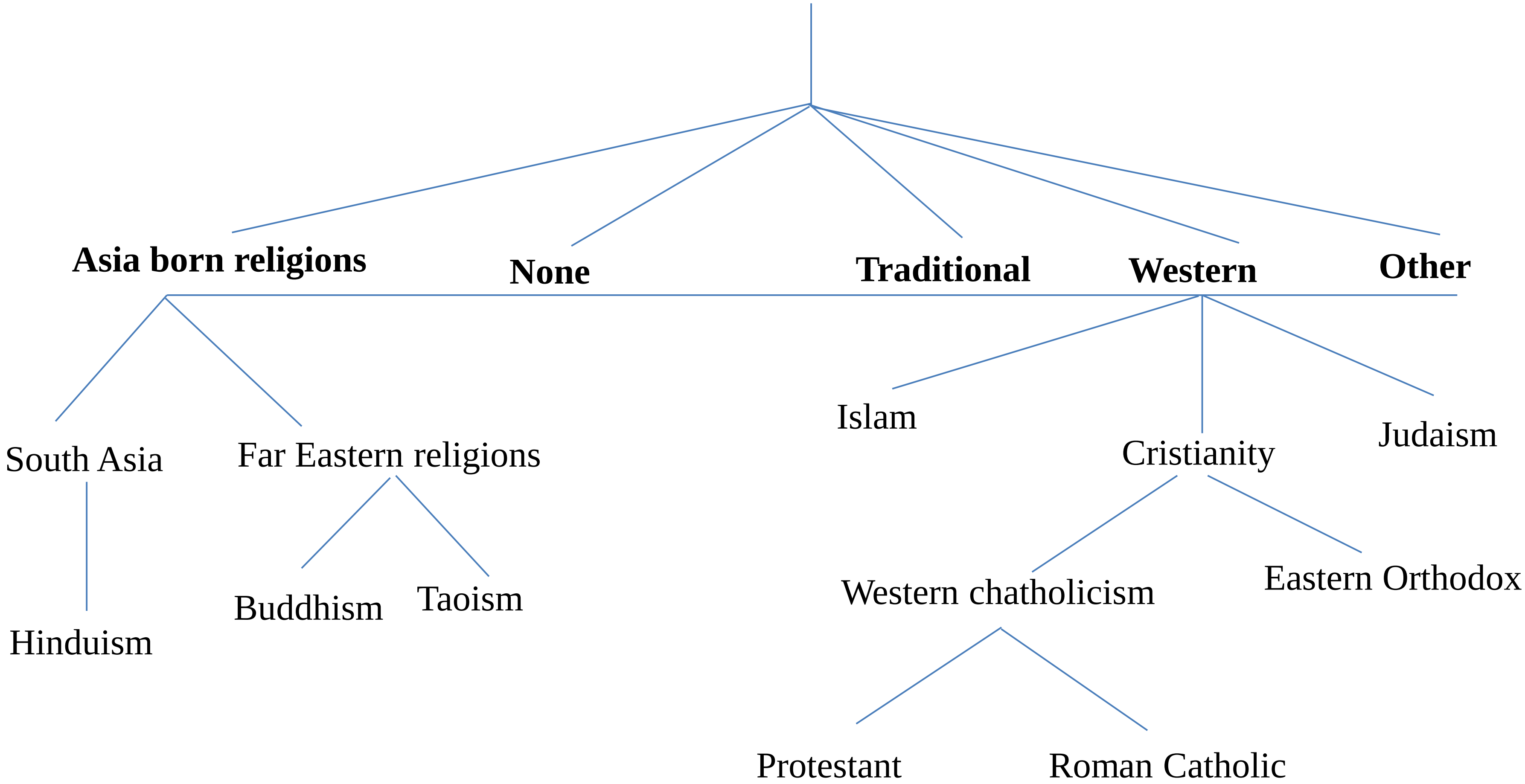
3.0 Traditional

4.0 Other

5.0 Assorted

6.0 None

Other W matrices



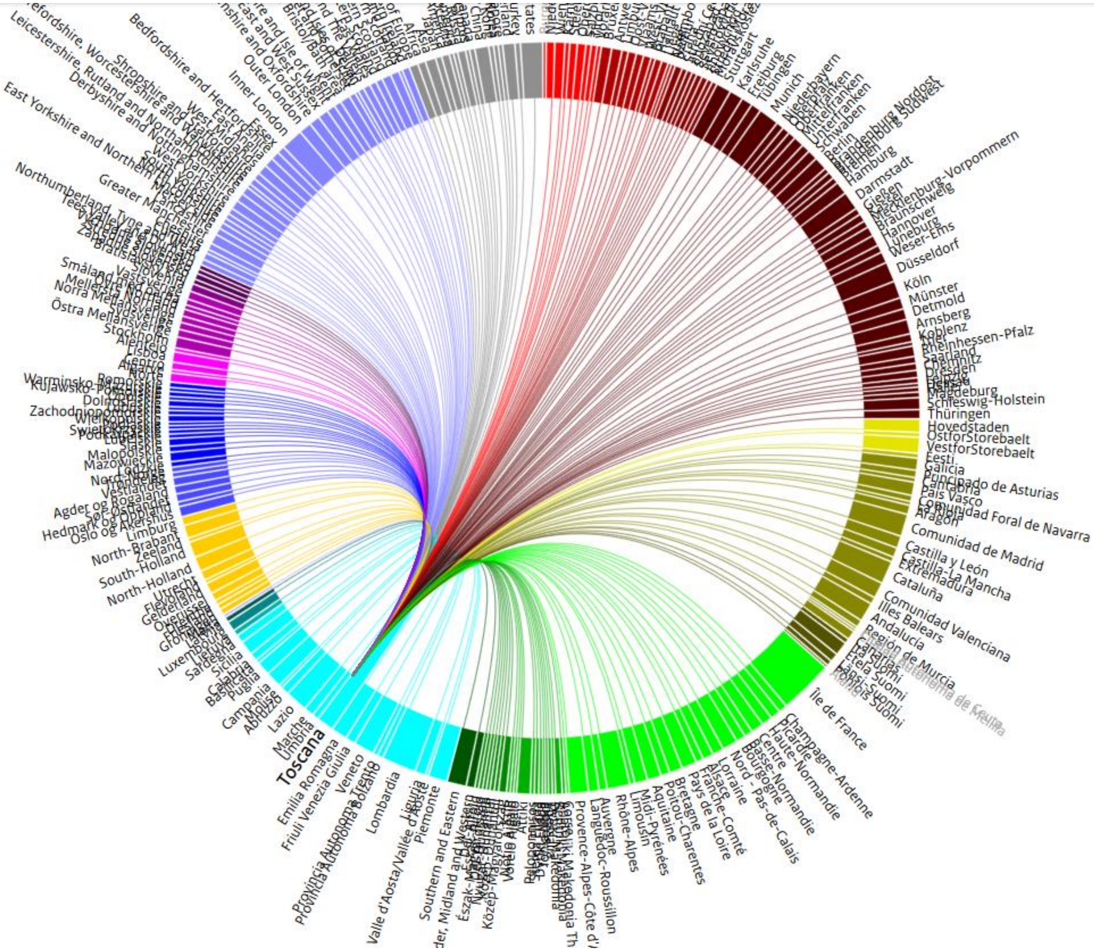
Other W matrices

- **Social values distance-based interactions:**

Indicator	Question
Community organizational life	
• Attention to environmental issues	Would you be willing to give up part of your income in order to prevent environmental damage?
• Perception of tax cheating	How many compatriots cheat on taxes?
Engagement in public affairs	
• Participation to political life	How important in your life is politics?
• Concern in regional affairs	Are you concerned with people from your own region?
• Commitment to social affairs	Participation in any social activity
Community volunteerism	
• Volontary work	Do you work unpaid for no organization? (This share is subtracted to 1 to obtain the share of people who works unpaid for any voluntary activity).
• Social volunteering	Do you belong to any welfare organization?
Informal sociability	
trust	Do you agree that 'Most people can be trusted'?

Other W matrices

- Trade links across European regions: http://themasites.pbl.nl/eu-trade/index2.html



TOSCANA, ITALY

Capital: Firenze
Population (2012): 3761616
GDP (2010): 103713 million Euro

Production: 212558 million Euro
Intranational exports: 188663 million Euro
International exports: 23895 million Euro

Use: 211915 million Euro
Intranational imports: 188988 million Euro
International imports: 22927 million Euro
see the competitors of this region

Top 10 Export destinations (in million Euro)

United States, United States	2566
Switzerland, Switzerland	2209
Lombardia, Italy	2126
Asia, Asia	1388
Emilia Romagna, Italy	1387
Rest of Europe, Rest of Europe	1278
Africa, Africa	1090
Piemonte, Italy	1062
Veneto, Italy	930
Russia, Russia	871