URP 6931. Introduction to Urban Analytics

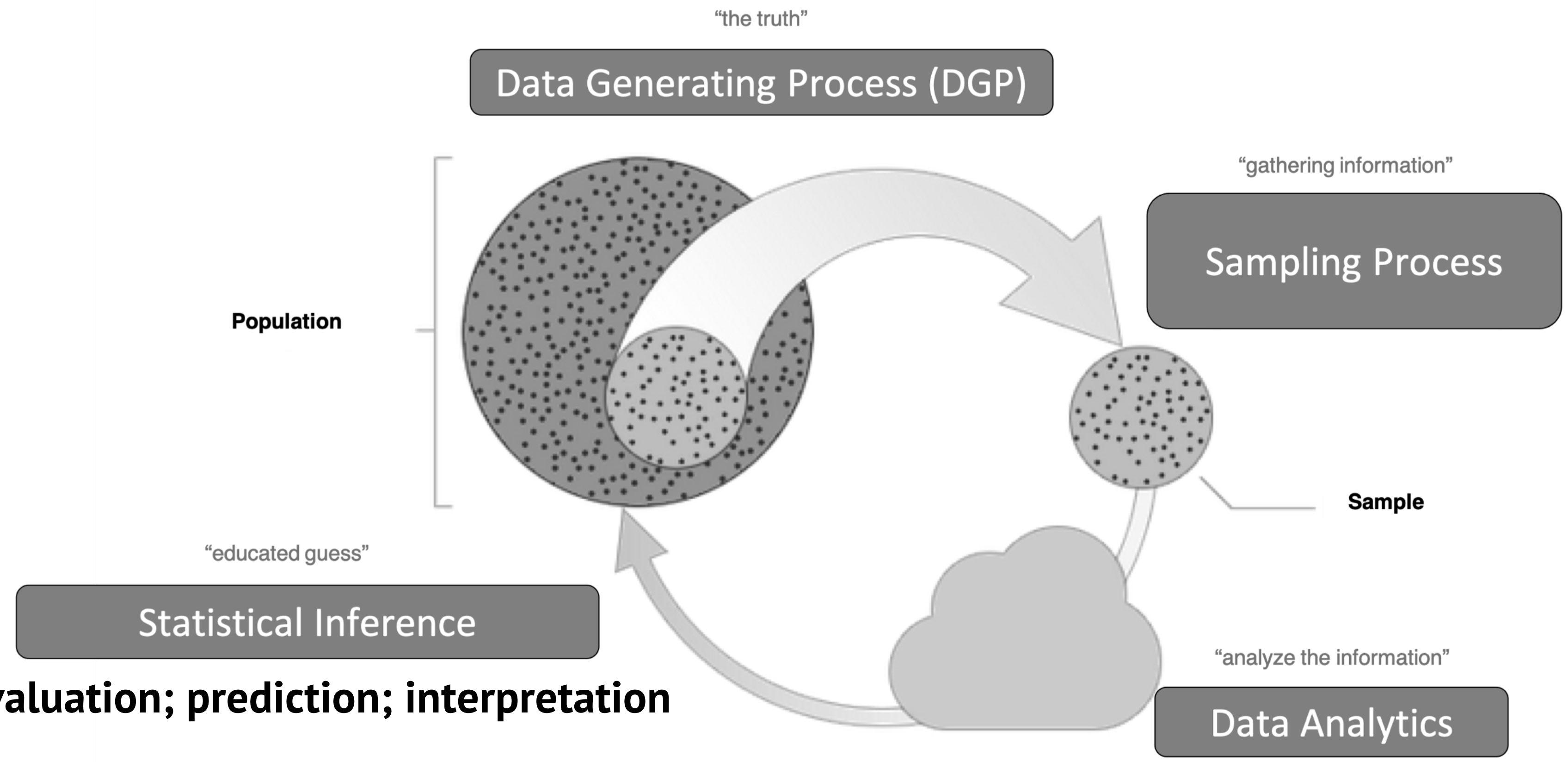
Lecture 07: Spatial autocorrelation and regression

Instructor: Shenhao Wang
Assistant Professor, Director of Urban AI Lab
Department of Urban and Regional Planning
University of Florida

Review lecture 06

Regression: recovering $E[Y|X]$

Random vs. non-random DGPs



Data: ~~i.i.d.~~
Graph: (V, E)
Graph Features: (X, A)

Evaluation; prediction; interpretation

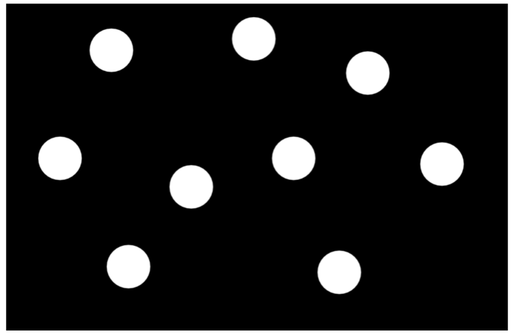
Power-law distribution; GIS visualization

Regression: $\hat{E}[Y|X] = \sigma(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki})$

Estimation/Training/Learning: OLS & MLE

Review: two questions in urban network analysis

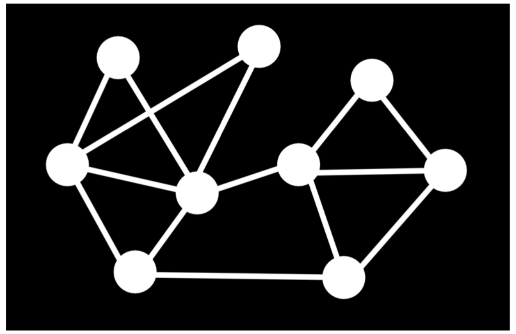
Nodes



Question 1:

What are the nodes and node features?

Graph (nodes + edges)



Question 2:

What are the edges and edge features?

Outline in lecture 07

[Focus on only part 1 & 2 today]

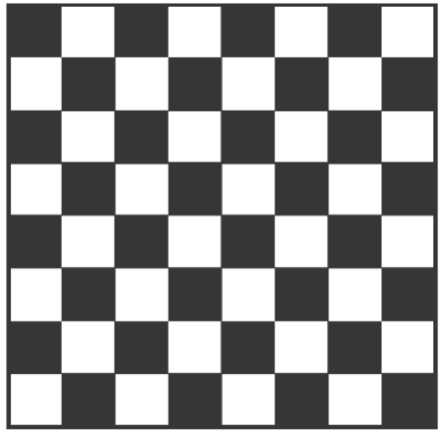
1	2	3	4	5
Spatial autocorrelation (X, A)	Spatial regressions for fitting node features $(X, A) \rightarrow y_i$	Gravity models for fitting edge features $(X, A) \rightarrow y_{ij}$	Community detection algorithms $(X, A) \rightarrow z_i$	Summarizing urban network analysis

Part 1. Spatial autocorrelation

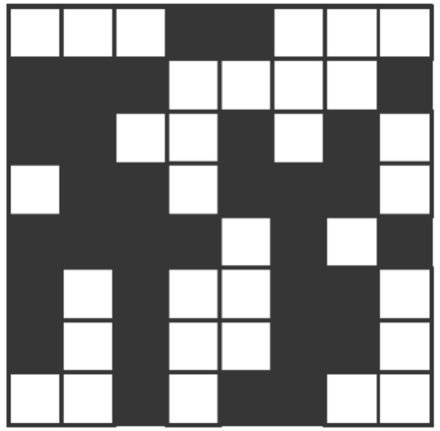
- It is the most common descriptive statistics in spatial analysis.
- It resembles the standard correlation coefficient.

Check your intuition: What is the sign of the spatial autocorrelation in the three examples below?

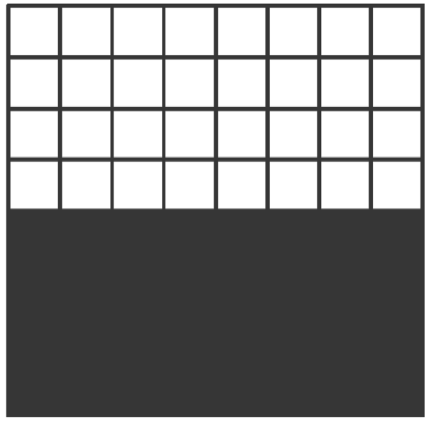
- But first of all, what are the nodes and node features? And what are the edges and edge features? [Whiteboard 3*3 example] We use +1 represent the black blocks, and -1 represent the white blocks. Edges are typically defined with Rook's rule.
- In fact, this network perspective (grid nodes and edges) applies to images with each node representing a pixel.



Negative autocorrelation



No autocorrelation

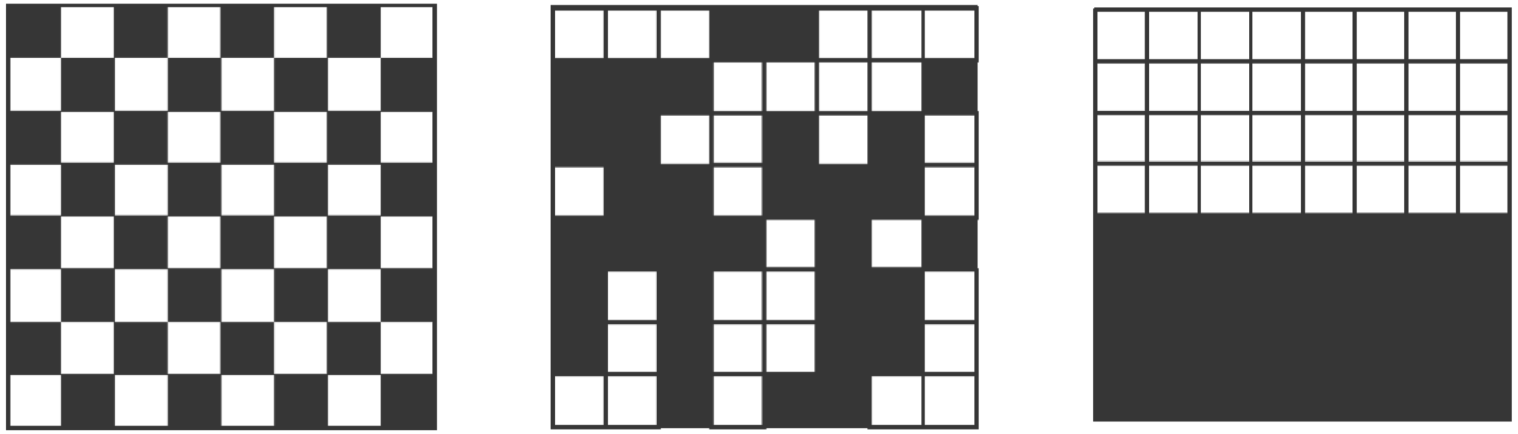


Positive autocorrelation

What is the consequence of spatial autocorrelation in a linear regression?

It leads to the violation of the assumptions. e.g. i.i.d., A1 and A2.

The violation of i.i.d. is more obvious but A1 and A2 are less so



How to measure the spatial autocorrelation?

Moran's I

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i}^{N}(x_i - \bar{x})^2}$$

Where

- N is the number of spatial units
- x is the variable of interest. e.g. property values.
- $\bar{x}$ is the mean of x
- w_{ij} is the matrix of spatial weights with zeros on the diagonal (i.e., $w_{ii} = 0$).

For simplicity, let's replace the weights w_{ij} by adjacency a_{ij}, because a_{ij} is a specific case of continuous weighting. Then the formula becomes:

$$I = \frac{N}{|A|} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i}^{N}(x_i - \bar{x})^2}$$

How to measure spatial autocorrelation?

Let's further reshuffle the terms

$$I = \frac{N}{|A|} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i}^{N} (x_i - \bar{x})^2}$$

as

$$I = \frac{N}{\sum_{i}^{N} (x_i - \bar{x})^2} * \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x})}{|A|}$$

Notes

- It is named as spatial autocorrelation because of its similarity to the correlation coefficient: $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$. The key difference lies in the autocorrelation.
- The first term $\frac{N}{\sum_{i}^{N}(x_i - \bar{x})^2}$ normalizes the metric by the variance, which is quite similar to the correlation coefficient.
- In the second term, $1/|A|$ is essentially another normalization term to deal with the a_{ij} in the numerator.
- The only critical term is $\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x})$, but what is it?

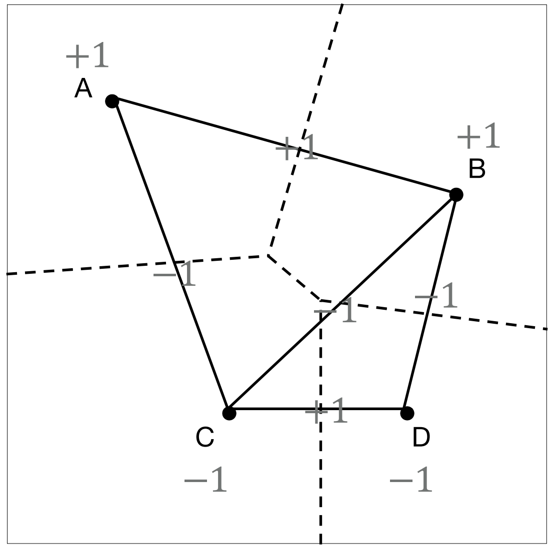
How to measure spatial autocorrelation?

Moran's I

$$I = \frac{N}{|A|} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i^N (x_i - \bar{x})^2}$$

Interpretation

- For $a_{ij} = 1$, we compute the pairwise correlation between nodes i and j as $a_{ij}(x_i - \bar{x})(x_j - \bar{x})$, which is the standard formula for correlation.
- For $a_{ij} = 0$, then nodes i and j are not correlated since $a_{ij}(x_i - \bar{x})(x_j - \bar{x}) = 0$.
- In other words, Moran's I first computes the pairwise correlation as an edge feature, and then sum over all the edges.
- $I \in [-1, +1]$. When $I > 0$, it indicates positive spatial autocorrelation. When $I < 0$, it indicates negative spatial autocorrelation.



Example

- x represents the property value.
- $x_A = +1$, $x_B = +1$, $x_c = -1$, and $x_d = -1$. e.g. high and low property value
- Then we can compute $a_{ij}(x_i - \bar{x})(x_j - \bar{x})$. And sum them over.
- $I = \frac{4}{10} * \frac{-1}{4} = -0.1$
- Therefore, the network has a weakly negative spatial autocorrelation.

$A =$

	A	B	C	D
A	*	1	1	0
B	1	*	1	1
C	1	1	*	1
D	0	1	1	*

Intuitively, what are the spatial autocorrelation of the three examples?

Moran's I

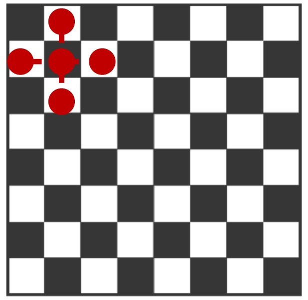
$$I = \frac{N}{|A|} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i}^{N} (x_i - \bar{x})^2}$$

How to get your intuition?

Random Walk!

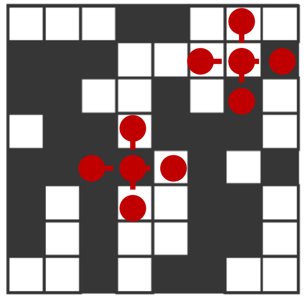
$$I = -1$$

- Every "edge feature" equals to -1.
- In other words, when you randomly walk to another node, you always arrives to the opposite of the node.
- Strongly negative spatial autocorrelation



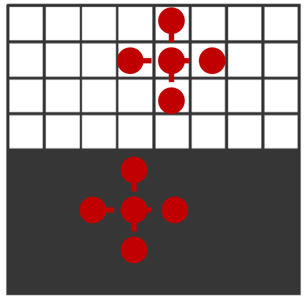
$$I \approx 0$$

- The average of "edge feature" approximates 0.
- In other words, when you randomly walk to another node, it is quite unclear whether you will arrive at the same or the opposite of your current node.
- No spatial autocorrelation



$$I \approx 1$$

- The majority of the "edge feature" equals to +1.
- In other words, when you randomly walk to another node, you always arrive at another node with the same value as your current node.
- Strongly positive spatial autocorrelation



Local spatial autocorrelation

Local Moran's I:

$$I_i = \frac{N}{\sum_i^N (x_i - \bar{x})^2} * (x_i - \bar{x}) \sum_j^N w_{ij}(x_j - \bar{x})$$

Notes

- I_i is specific to each node i.
- Again, $\frac{N}{\sum_i^N (x_i - \bar{x})^2}$ is a normalization term.
- The global Moran's I is related to the local Moran's I by $I = \sum_i^N \frac{I_i}{N}$.
- With the intuition in (1) pairwise correlation as an edge feature, and (2) random walk for intuitive check, it should be relatively easier to understand this formula. Let's replace w_{ij} by a_{ij}, then

$$I_i = \frac{N}{\sum_i^N (x_i - \bar{x})^2} * (x_i - \bar{x}) \sum_j^N a_{ij}(x_j - \bar{x})$$

How to measure local spatial autocorrelation?

Local Moran's I for node i

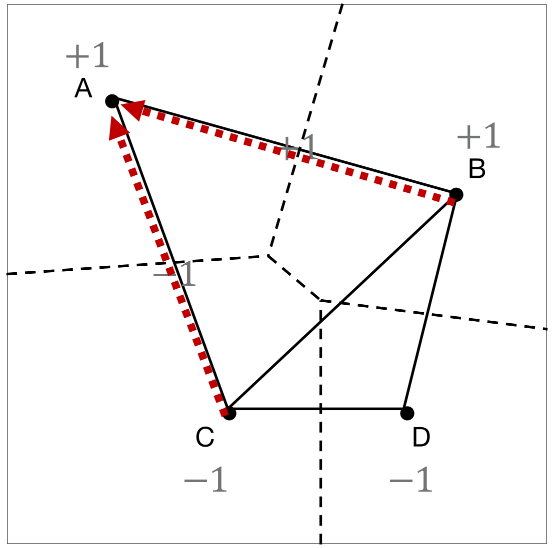
$$I_i = \frac{N}{\sum_i^N (x_i - \bar{x})^2} * (x_i - \bar{x}) \sum_j^N a_{ij}(x_j - \bar{x})$$

Computing I_A

- For $a_{ij} = 1$, we compute the pairwise correlation between node A and others as $a_{ij}(x_i - \bar{x})(x_j - \bar{x})$, which is the standard formula for correlation.
- Then we sum over all the pairwise correlations, which pass information towards node A.

Example

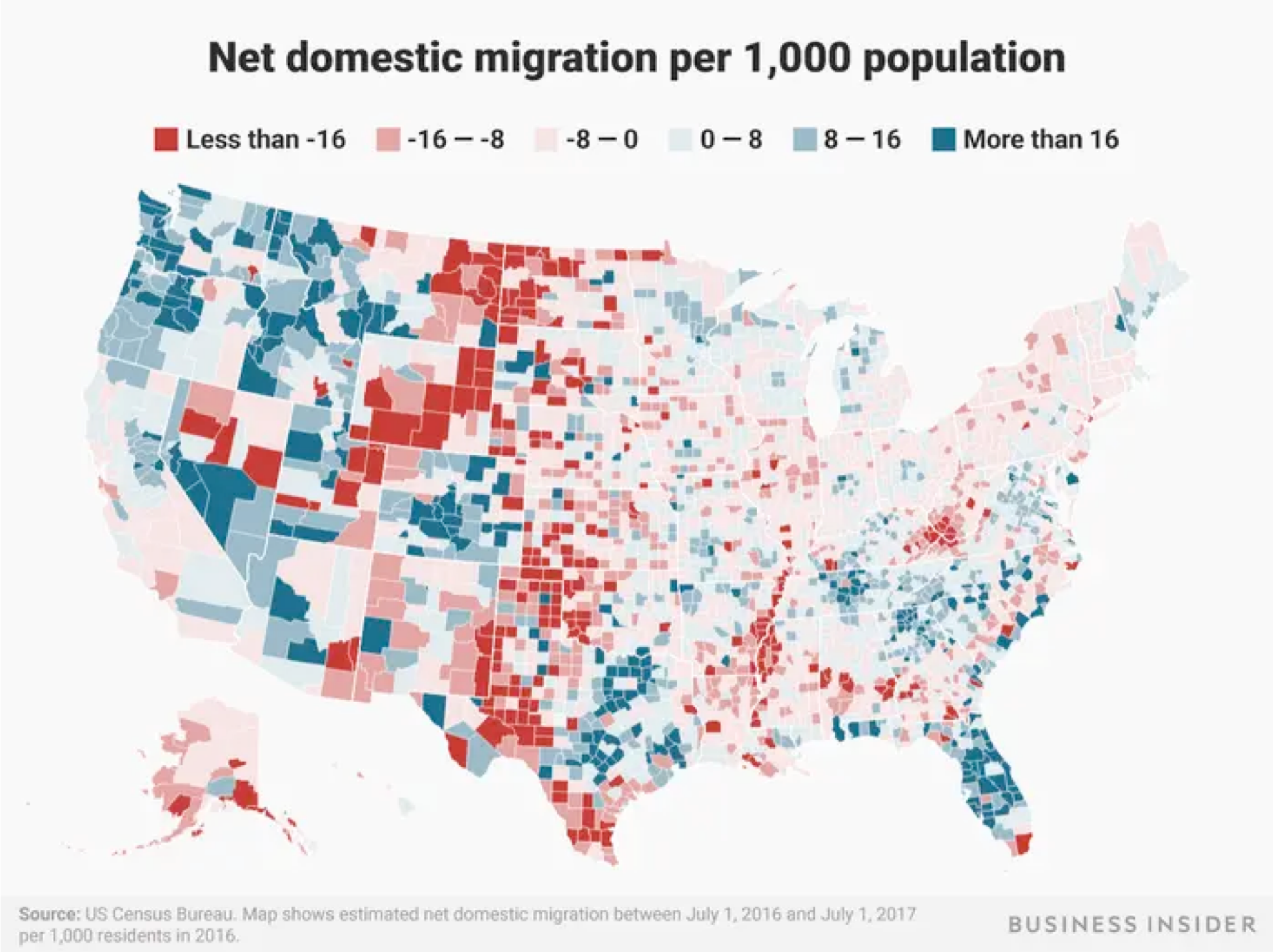
- x represents the property value.
- $x_A = +1$, $x_B = +1$, $x_c = -1$, and $x_d = -1$. e.g. high and low property value
- Then we can compute $a_{Aj}(x_A - \bar{x})(x_j - \bar{x})$. And sum them over.
- $I = \frac{4}{10} * 0 = 0$
- Therefore, locally node A does not have any spatial autocorrelation.
- Check it with the random walk intuition.



$A =$

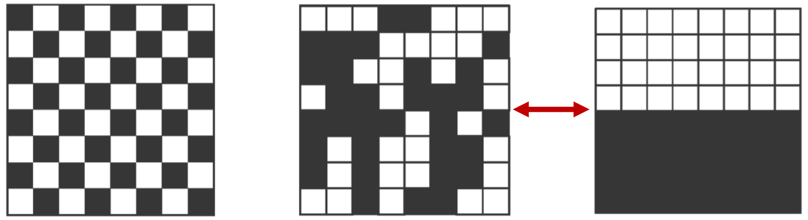
	A	B	C	D
A	*	1	1	0
B	1	*	1	1
C	1	1	*	1
D	0	1	1	*

Intuitively check the spatial autocorrelation

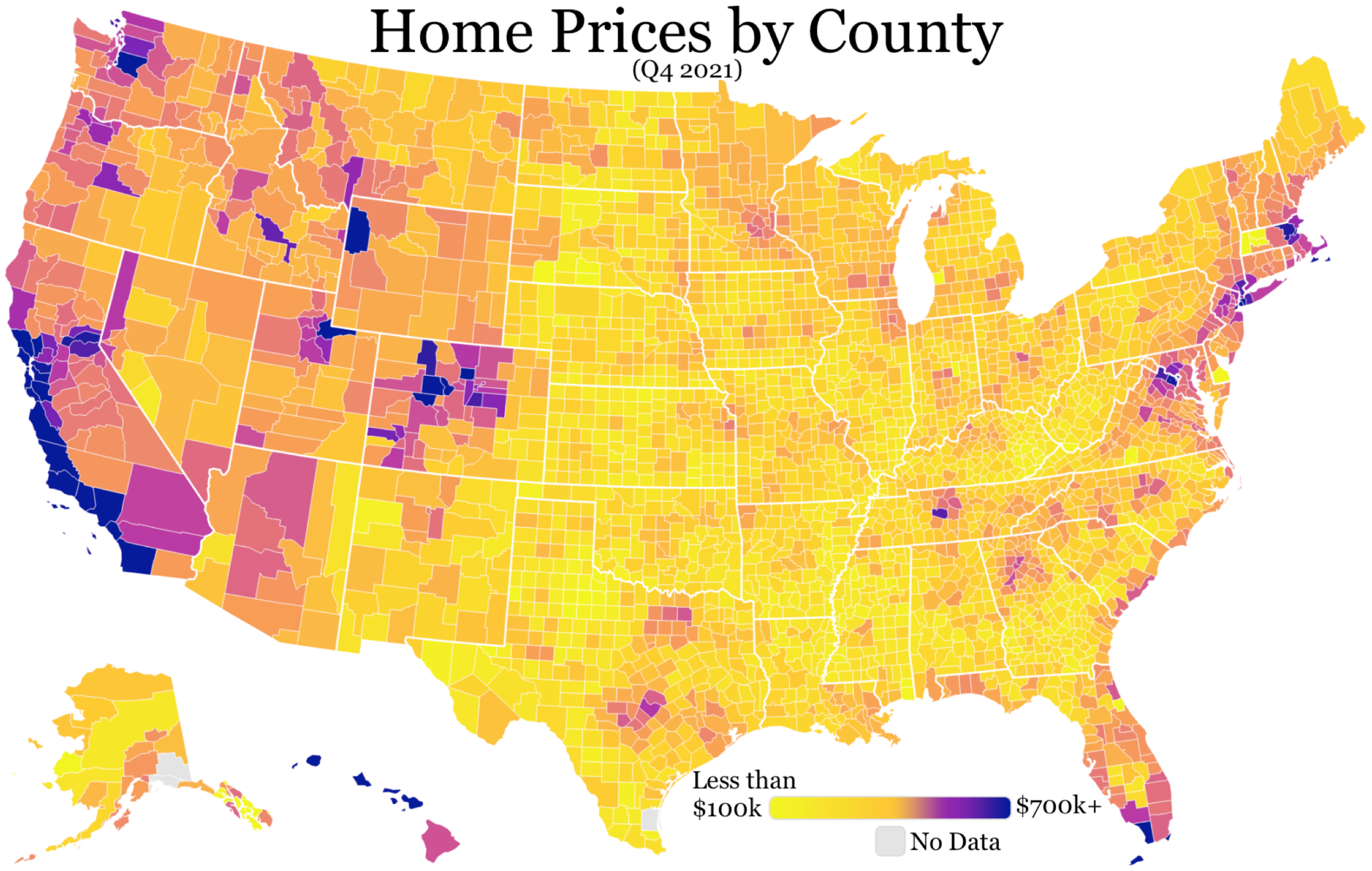


Define the nodes and edges
Formal approach: Moran's I
Intuitive approach: random walk
My guess: weakly positive spatial autocorrelation.

Which following diagram is most similar to the map?

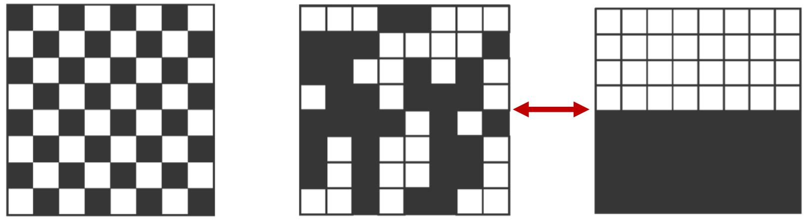


Intuitively check the spatial autocorrelation



Define the nodes and edges
Formal approach: Moran's I
Intuitive approach: random walk.
My guess: positive spatial autocorrelation.

Which following diagram is most similar to the map?



Pay attention to a specific math formula here:

$$z_i = \sum_{j=1}^{N} a_{ij} z_j$$

Notes

- Let $z_j = x_j - \bar{x}$, we get the key component in the formula to compute the local Moran's I:

$$I_i = \frac{N}{\sum_i^N (x_i - \bar{x})^2} * (x_i - \bar{x}) \sum_j^N a_{ij}(x_j - \bar{x})$$

- Intuitively, it represents the message passing from the neighboring nodes of node i towards node i. In other words, it collects the local information in a network.

Take-aways from

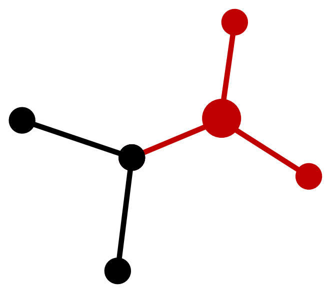
Spatial autocorrelation

Notes

1. Formula for Moran's I as an analogy of correlation coefficient in space.
2. Intuitively examine the spatial autocorrelation with random walk – quite useful in practice!
3. Formally examine the spatial autocorrelation with local or global Moran's I.
4. Intuition into the formula: $\sum_j^N a_{ij}(x_i - \bar{x})(x_j - \bar{x})$ - it passes messages from neighbors of node i to node i.

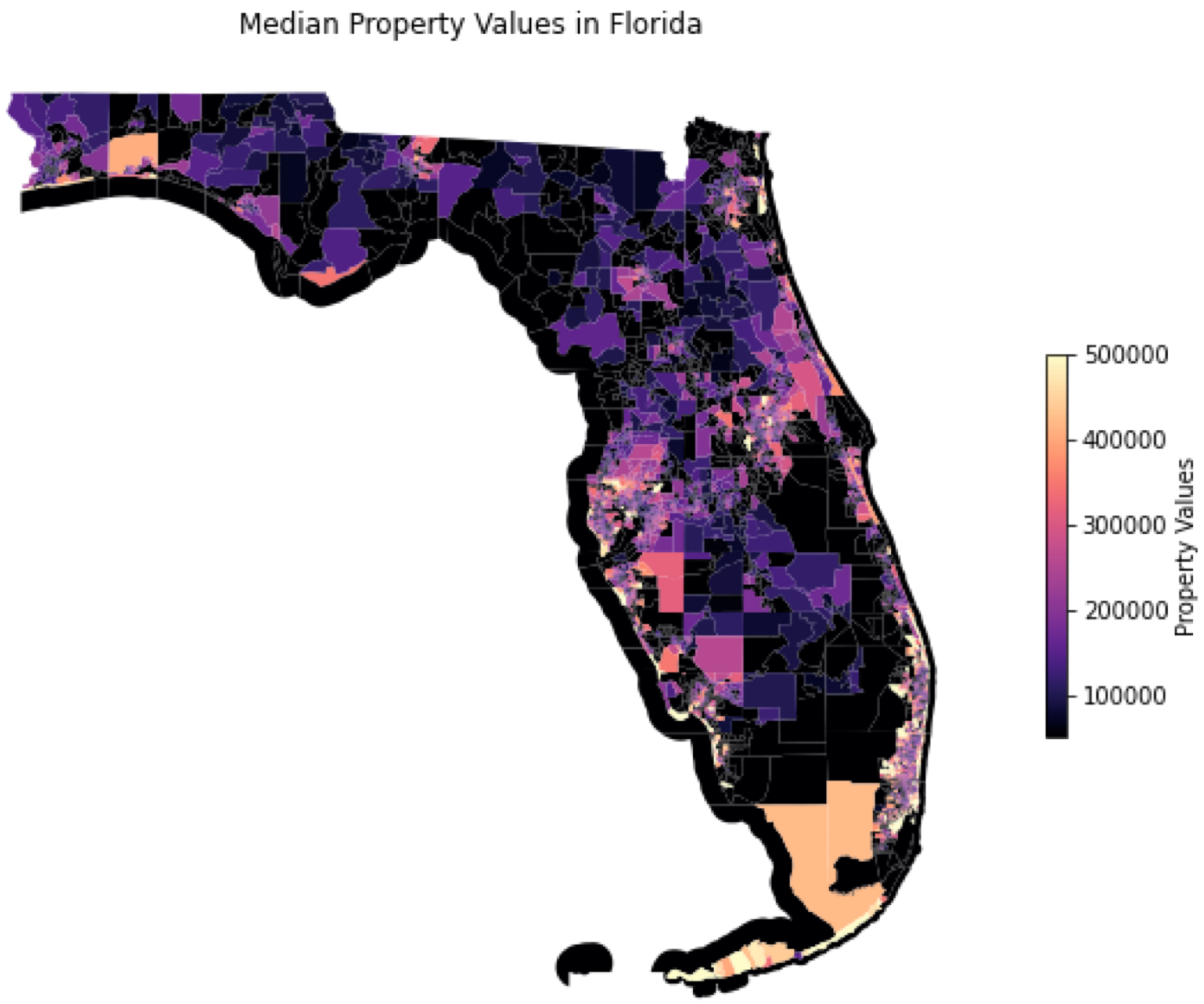
Part 2. Spatial regressions for fitting node features

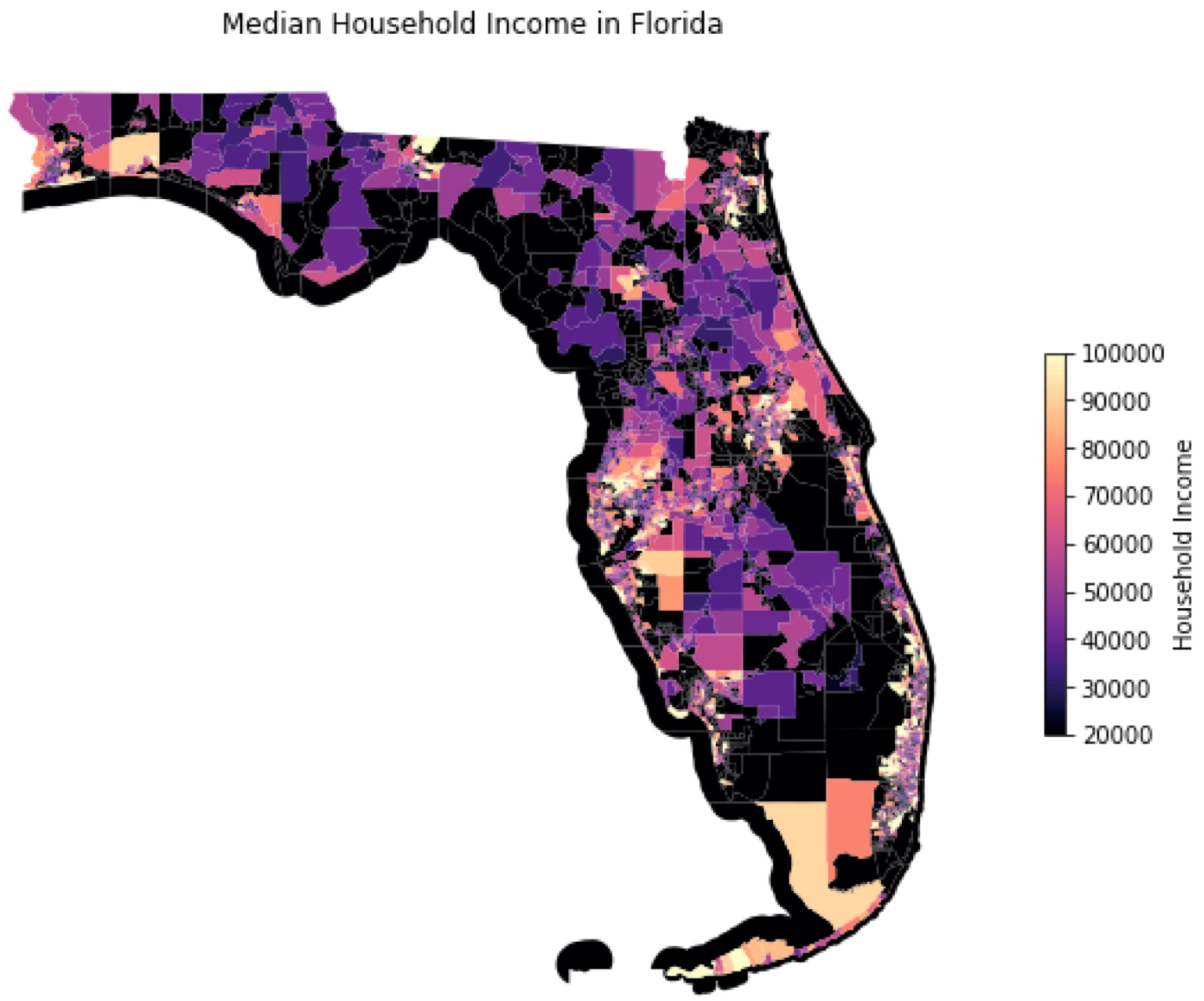
- Moran's I describes the spatial autocorrelation, and spatial regressions can leverages the spatial autocorrelation to enhance a baseline linear regression.
- Spatial regressions: combining edge & node features to analyze node features.



What is the problem with the previous linear regression?

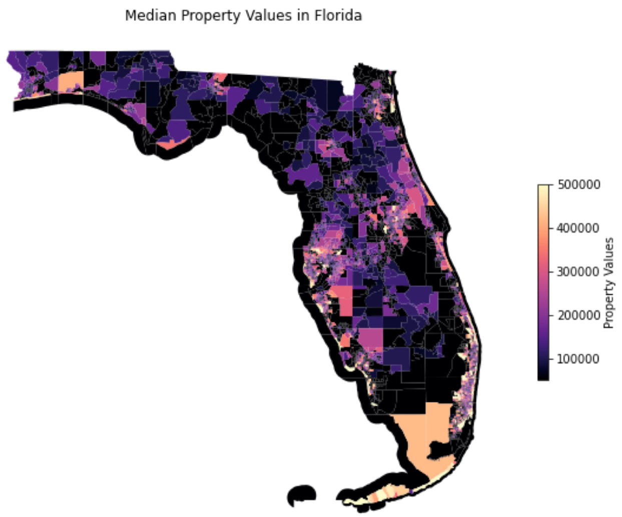
Both property value and income have spatial autocorrelation.

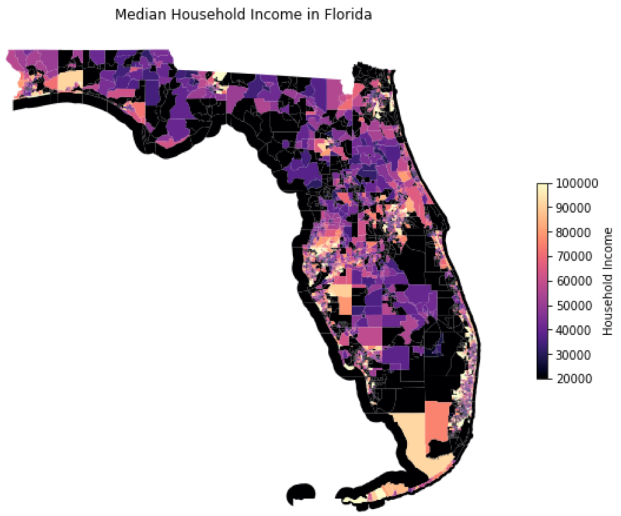




What is the problem with the previous linear regression?

Both property value and income have spatial autocorrelation.





Previously in univariate linear regressions,

- Each node has node features (property values and income).
- We use the income (x) of node i to predict the property value (y) of node i. Regression: $y_i \sim x_i$

However, with the spatial autocorrelation, we have the following hypotheses:

1. Is it possible that the property value (y) of node i can be predicted by not only its own income x_i but also the income from node i's neighboring nodes ($x_{N(i)}$)?
2. Is it possible that the property value (y) of node i can be predicted by not only its own income x_i but also the property values from node i's neighboring nodes ($y_{N(i)}$)?
3. How to represent the information passing from the neighboring nodes?

To simplify the discussion, let's step back to the univariate regression and ask:

How to represent the information passed from the neighboring nodes in a regression?

The initial linear regression form is: $y_i = \beta_0 + \beta_1 x_i + u_i$

Three options regarding x_i, y_i, and u_i.

1. Spatially lagged regression using x: $y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} x_j + u_i$. Matrix form: $y = X\beta + \rho AX + u$
2. Spatially lagged regression using y: $y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} y_j + u_i$. Matrix form: $y = X\beta + \rho Ay + u$
3. Spatial error model (We won't cover it)
4. Intuitively, it is already discussed in Moran's I – because this math form $\sum_j a_{ij} x_j$ or $\sum_j a_{ij} y_j$ collects local information.

Spatially lagged regression using x

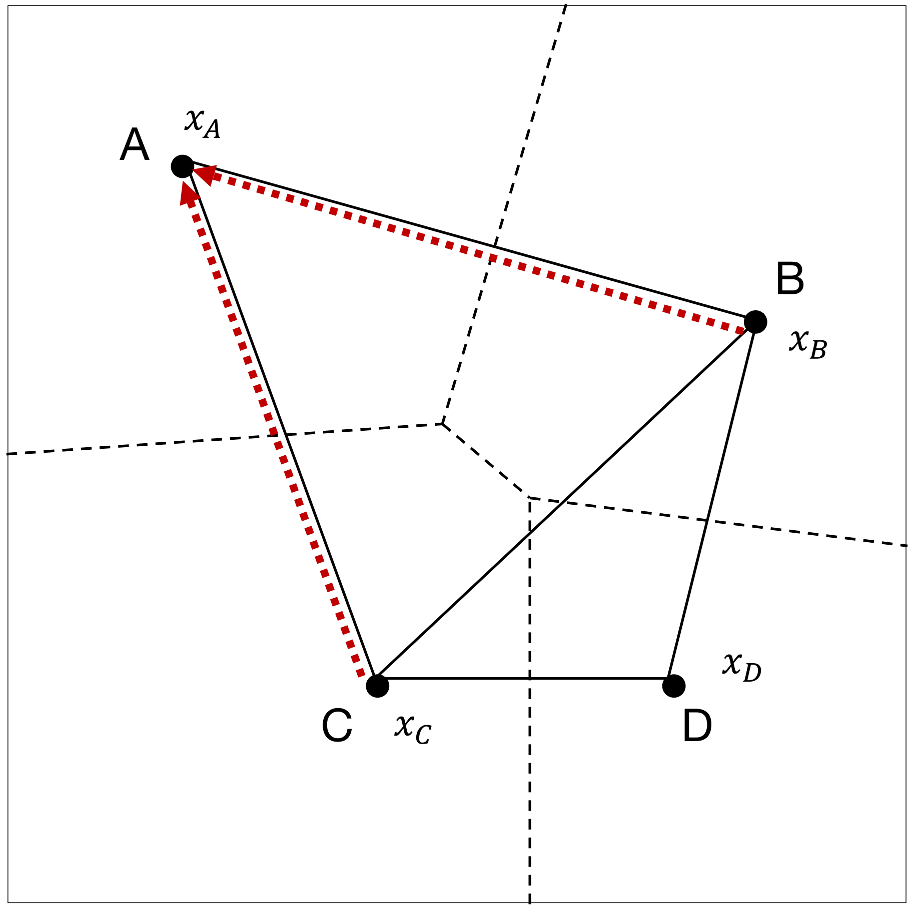
About

$$y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} x_j + u_i$$

Again, x is the independent variable (e.g., income), and y is the dependent variable (e.g., property value).

Interpretation

- Examine x_A. When $i = A$, what is the meaning of $\sum_j a_{ij} x_j$?
- $\sum_j a_{ij} x_j$ sums over the income over the neighboring nodes of node A.
- In other words, we pass the messages of income from node B and C to node A.
- $\sum_j a_{ij} x_j$ is called the spatial lag term.
- In practice, a_{ij} is often row-normalized to create the same magnitude between local and neighborhood messages.
- The regression means that the prediction of property value at node A depends on not only its own income, but also the income of its neighbors.
- Meaning of $\hat{\rho}$: whether the spillover effect of income is significant.



Consider this spatially lagged regression using x…

$$y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} x_j + u_i$$

Questions

1. How is this model related to linear regression?
2. How to estimate this regression?
3. Comparing a simple linear regression, how do you think about A1 and A2 assumptions?

Note: since this model is another version of linear regression, people do not seriously discuss it in spatial econometrics.

Spatially lagged regression using y

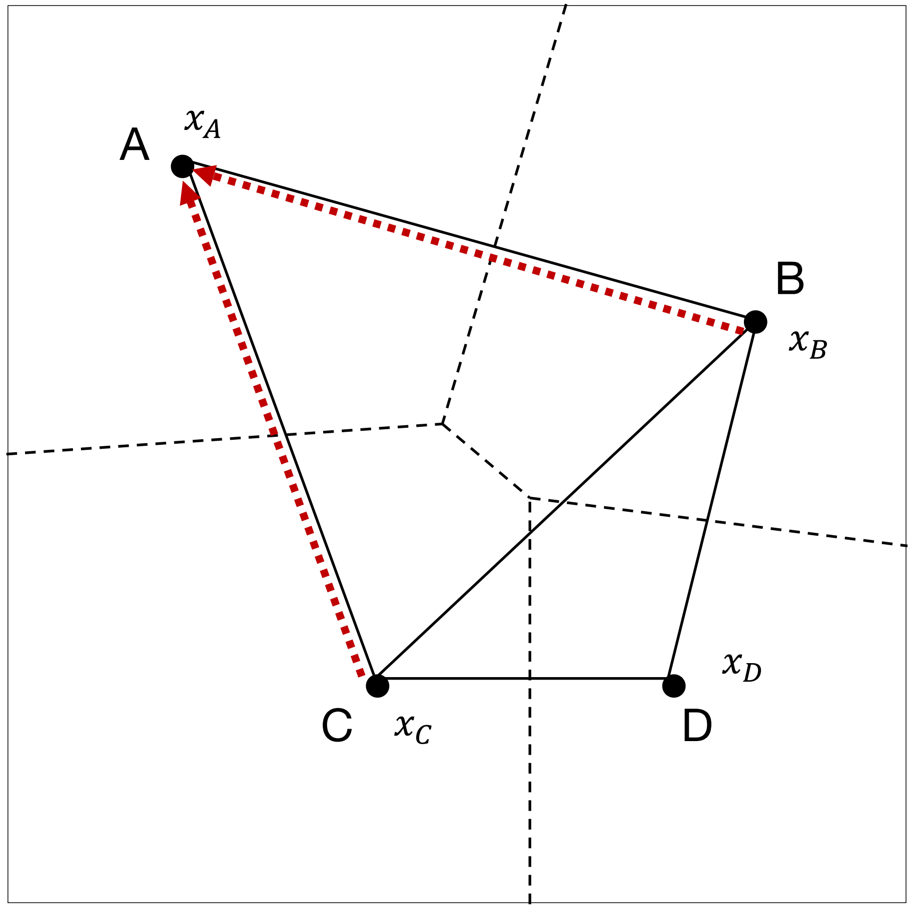
About

$$y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} y_j + u_i$$

Again, x is the independent variable (e.g., income), and y is the dependent variable (e.g., property value).

Interpretation

- Examine node A. When $i = A$, what is the meaning of $\sum_j a_{ij} y_j$?
- Well, $\sum_j a_{ij} x_j$ sums over the property values over the neighboring nodes of node A.
- In other words, we pass the messages of property value from node B and C to node A.
- $\sum_j a_{ij} y_j$ is called the spatial lag term.
- In practice, a_{ij} is often row-normalized to create the same magnitude between local and neighborhood messages.
- The regression means that the prediction of property value at node A depends on not only its own income, but also the property values of its neighbors.
- Meaning of $\hat{\rho}$: whether the spillover effect of property value is significant.



Consider this spatially lagged regression using y…

$$y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} y_j + u_i$$

Questions

1. How is this model related to linear regression?
2. How to estimate this regression?
3. Comparing a simple linear regression, how do you think about A1 and A2 assumptions?

Note: this spatially lagged regression using y (commonly referred to as spatial lag regression) deviates from the standard linear regression.

MLE for the spatially lagged regression using y (not required)

Matrix form of the model:

$$y = \rho W y + X\beta + \epsilon$$

Then

$$y - \rho W y = X\beta + \epsilon$$
$$y = (I - \rho W)^{-1} X\beta + (I - Wy)^{-1}\epsilon$$

which is no longer a linear-in-parameter regression.

To use MLE, we always need a full probabilistic assumption on the DGP. Let's assume:

$$\epsilon \sim N(0, \sigma^2 I)$$

A small detour – how to use MLE for a multivariate Gaussian distribution. e.g. $\epsilon \sim N(0, \sigma^2 I)$?

$$\log L = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\sigma^2 I| - \frac{1}{2\sigma^2}\epsilon'\epsilon$$

Using MLE to estimate the spatially lagged endogenous regression (not required)

Log-likelihood of the spatial lag model is

$$\log L = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 + \log|I - \rho W| - \frac{(y - \rho W y - X\beta)'(y - \rho W y - X\beta)}{2\sigma^2}$$

Then the $\hat{\beta}$ and $\hat{\rho}$ can be estimated as:

$$\hat{\beta}, \hat{\rho} = \underset{\beta,\rho}{\operatorname{argmax}} \log L$$

The estimation process is the same as the logistic regression (MLE).

Notes

- **Question (Review)**: What is the fundamental logic in MLE?
- The MLE process uses numerical method, which is not required to know in this course.
- As usual, you should learn to find the script in Python to obtain the result.
- Meanwhile, you should understand the interpretation of $\hat{\beta}$ and particularly $\hat{\rho}$.

Spatiotemporal Regression Model 1

$$y_{it} = \beta_0 + \beta_1 x_{it} + \rho \sum_j a_{ij} x_{jt} + \gamma x_{i,t-1} + u_{it}$$

Notes

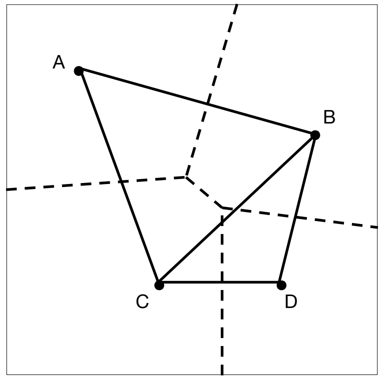
1. Temporal dimension is not the focus of this class. However, sometimes it is quite straightforward to incorporate the temporal dimension.
2. Here is a simple example about spatiotemporal regression with $\rho \sum_j a_{ij} x_{jt}$ representing the spatial effect and $\gamma x_{i,t-1}$ representing the temporal effect.
3. This is still a linear regression – so please don't get scared by any "spatiotemporal modeling"

Spatiotemporal Regression Model 2

$$y_{it} = \beta_0 + \beta_1 x_{it} + \rho \sum_j a_{ij} y_{jt} + \gamma y_{i,t-1} + u_{it}$$

Notes

1. Here is another example about spatiotemporal regression with $\rho \sum_j a_{ij} y_{jt}$ representing the spatial effect and $\gamma y_{i,t-1}$ representing the temporal effect.
2. This is not a linear regression.



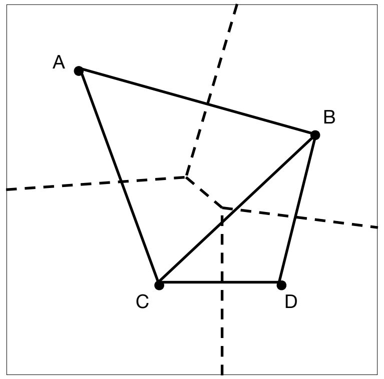
Discussion 1: meaning of linear algebra in spatial regressions

	A	B	C	D
A	*	1	1	0
B	1	*	1	1
C	1	1	*	1
D	0	1	1	*

$$Ax$$

Notes

- What is the meaning of Ax in the example? [Whiteboard example]
- This form (Ax and Ay) implicitly exists in $y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} x_j + u_i$ and $y_i = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} y_j + u_i$. In fact, the most common notations for spatial regressions use the matrix formula.
- It is the local message passing across the nodes. The matrix A becomes an operator to collect the local information.
- In fact, this intuition can be generalized to $A^k x$, which collects k-hop information around nodes.
- In practice, A is often row normalized to ensure the same magnitude between the spatial lag and the independent variable.



Discussion 2: meaning of linear algebra in spatial regressions

	A	B	C	D
A	*	1	1	0
B	1	*	1	1
C	1	1	*	1
D	0	1	1	*

$$x'Ax$$

Notes

- What is the meaning of $x'Ax$ in the example? [Whiteboard example]
- It collects all the edge information with a quadratic form. In fact, it can be used to represent Moran's I.
- The notation on the previous page represents node features, while this represents the sum of edge features.

Discussion 3: connecting to our discussions about A1 and A2

$$\text{DGP: } E[y_i|x_i] = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} x_j$$

$$\text{Model: } E[y_i|x_i] = \beta_0 + \beta_1 x_i$$

Notes

1. **Question**: What is the problem with using a linear regression for spatial data?
2. It violates the assumption of A1 and A2.
3. Actually we are following the same reasoning of enriching the model.
4. By the way, the notation for including $\rho \sum_j a_{ij} y_j$ is NOT $E[y_i|x_i] = \beta_0 + \beta_1 x_i + \rho \sum_j a_{ij} y_j$. But let's avoid further complicated notations in this class.