SIG/GIS: Systèmes d’Information Géographiques

Geographical Information Systems

Principles

THE NATURE OF GEOGRAPHIC DATA

Outlines

- 1- Representing Geography
- 2- **The nature of Geographic Data**
- 3- Geo-referencing
- 4-Uncertainty

The nature of Geographic Data

Overview

- Elaborates on the spatial is special
- Focuses on how phenomena vary across space and the general nature of geographic variation
- Describes the main principles that govern scientific sampling, how spatial variations is formalized and measured as spatial autocorrelation

The nature of Geographic Data

- Learning Objectives
  - How Tobler's First Law of Geography is formalized through the concept of spatial autocorrelation;
    - Tobler's First Law of Geography: Everything is related to everything else, but near things are more related than distant things.
  - The relationship between scale and the level of geographic detail in a representation;
  - The principles of building representations around geographic samples;

The nature of Geographic Data

- Learning Objectives
  - How the properties of smoothness and continuous variation can be used to characterize geographic variation;
  - How fractals can be used to measure and simulate surface roughness.

The nature of Geographic Data

- Introduction
  - 1- Spatial is special
    - Many geographic data are correctly thought of as sample observations selected from the larger universe of possible observations that could be made
    - This chapter describes the main principles that govern the scientific sampling and the principles invoked in order to infer the gaps between samples
    - Be aware of the nature of spatial variation
  - **Here we learn how this is formalized, and measured as spatial autocorrelation**

The nature of Geographic Data

- Introduction
  - 2- The level of detail that is apparent at particular scales of analysis

- Spatial Autocorrelation and scale
  - Measures of spatial and temporal autocorrelation are scale dependent
  - The issue of sampling interval is of direct importance in the measurement of spatial autocorrelation
  - When the pattern of spatial autocorrelation at the coarser scale is replicated at the finer scale, the overall pattern exhibits the property of self-similarity

The nature of Geographic Data

- Previous principles
  - That the representations we build in GIS are of unique places
  - That our representations of them are necessarily selective of reality, and hence incomplete
  - That in building representations, it is useful to think of the world as either
    - comprising continuously varying fields or
    - as an empty space littered with crisp and well-defined objects

The nature of Geographic Data

- Assert three further principles
  - That proximity effects are key to understanding spatial variation, and to joining up incomplete representations of unique places;
  - That issues of geographic scale and level of detail are key to building appropriate representations of the world;
  - That different measures of the world co-vary, and understanding the nature of covariation can help us to predict

The nature of Geographic Data

- The fundamental problem revisited
  - Distinguishes between controlled variation, which oscillates around a steady state, and uncontrolled variation.
  - Some applications address controlled variation,
    - such as utility management
  - Others address uncontrolled,
    - such as those studying longer term processes

The nature of Geographic Data

- The fundamental problem revisited
  - Introduces the concept of time series and acknowledges the concept of temporal autocorrelation.
  - Our behavior in space often reflects past patterns of behavior, thus it is one dimensional, need only look in the past
  - Spatial heterogeneity is the tendency of geographic places and regions to be different from each other
  - Occurs in both form and process

The nature of Geographic Data

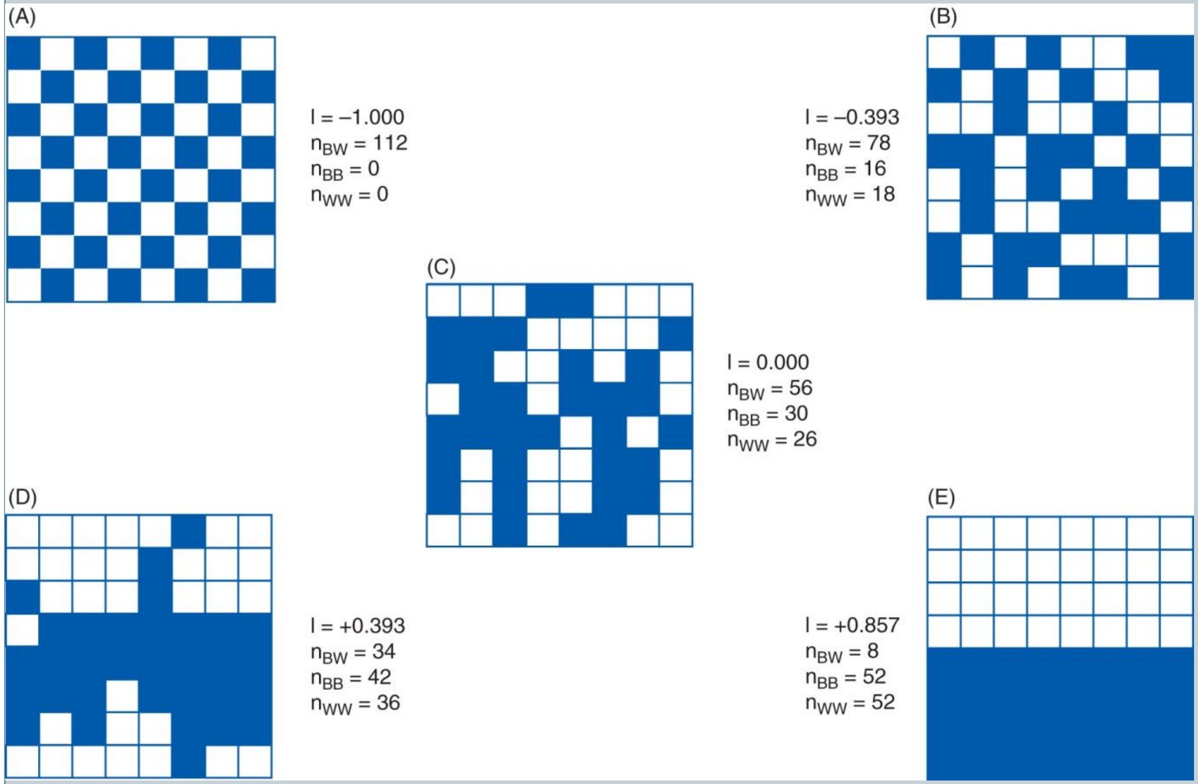
- The principles addressed in this chapter help answer questions about what to leave in and what to take out of digital representations

- Scale and spatial structure help determine how to sample reality and weight sample observations to build representations

The nature of Geographic Data

- Spatial Autocorrelation and scale
  - Spatial autocorrelation measures attempt to deal simultaneously with similarities in the location of spatial objects and their attributes
    - If the features that are similar in location are also similar in attributes, then the pattern as a whole is said to exhibit
      - **positive spatial autocorrelation**
    - If features which are close together in space tend to be more dissimilar in attributes than features which are further apart (in opposition to Tobler's Law), then there is
      - **negative spatial autocorrelation**
    - If attributes are independent of location
      - **Zero autocorrelation occurs**

Field arrangement of blue and white cells



(A) extreme negative spatial autocorrelation

(B) a dispersed arrangement

(C) spatial independence

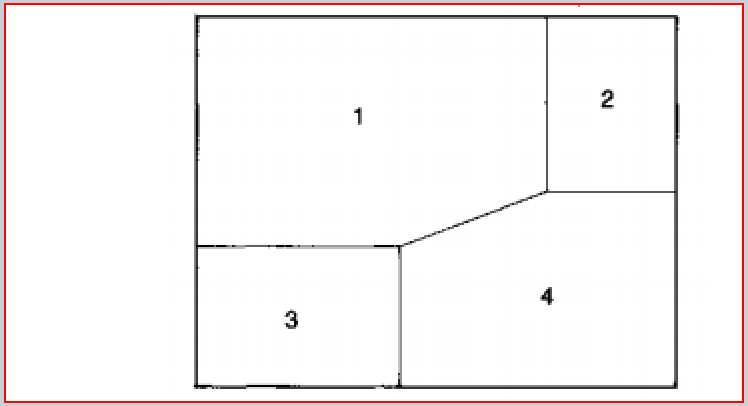
(D) spatial clustering

(E) extreme positive spatial autocorrelation

(Source: Goodchild 1986 CATMOG, GeoBooks, Norwich)

The nature of Geographic Data

- Spatial autocorrelation is determined both by similarities in position, and by similarities in attributes.
- Example



Autocorrelation Moran Index

- Different values

$n = 4$

$z_1 = 3$

$z_2 = 2$

$z_3 = 2$

$z_4 = 1$

$$W = \begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

$$\bar{z} = \sum_i z_i / n = 8/4 = 2$$

$$c_{ij} = (z_i - \bar{z})(z_j - \bar{z})$$

e $\bar{z}$ denotes the mean of the attribute variable, as before.

The remaining terms in the Moran index are again designed to constrain it to a fixed range:

$$I = \sum_i \sum_j w_{ij} c_{ij} / s^2 \sum_i \sum_j w_{ij} \qquad (8)$$

where s^2 denotes the sample variance

$$\sum_i (z_i - \bar{z})^2 / n$$

Autocorrelation Moran Index

Example calculation of the Moran index:

$$c_{ij} = (z_i - \bar{z})(z_j - \bar{z}) \qquad c = \begin{matrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{matrix}$$

$$\sum_i \sum_j w_{ij} c_{ij} = -2$$

$$s^2 = \sum_i (z_i - \bar{z})^2 / n = 2 / 4 = 0.5$$

$$I = \sum_i \sum_j w_{ij} c_{ij} / s^2 \sum_i \sum_j w_{ij} = -2 / 0.5 \times 10 = -0.400$$

Autocorrelation Moran Index

- Do same calculation for the example of 64 cells (you can program !!!!!!!)

Scale effects

- if the measures of the blue/white property were made at intervals that did not coincide with the dimensions of the squares of the chess boards in Figure 4.1,
- then the spatial autocorrelation measures would be different.
- Thus **the issue of *sampling interval* is of direct importance in the measurement of spatial autocorrelation**

The meaning of Scale

- 1- *Scale is in the details.*
  - Many scientists use scale in the sense of spatial resolution, or the level of spatial detail in data. Data are fine-scaled if they include records of small objects, and coarse-scaled if they do not
- 2- *Scale is about extent.*
  - Scale is also used by scientists to talk about the geographic extent or scope of a project:
    - a large-scale project covers a large area,
    - and a small-scale project covers a small area.

The meaning of scale

- *3- The scale of a map.*
  - Geographic data are often obtained from maps, and often displayed in map form.
  - Cartographers use the term scale to refer to a map's *representative fraction* (the ratio of distance on the map to distance on the ground
    - large scale corresponds to a large representative fraction, in other words to plenty of geographic detail.

The meaning of scale

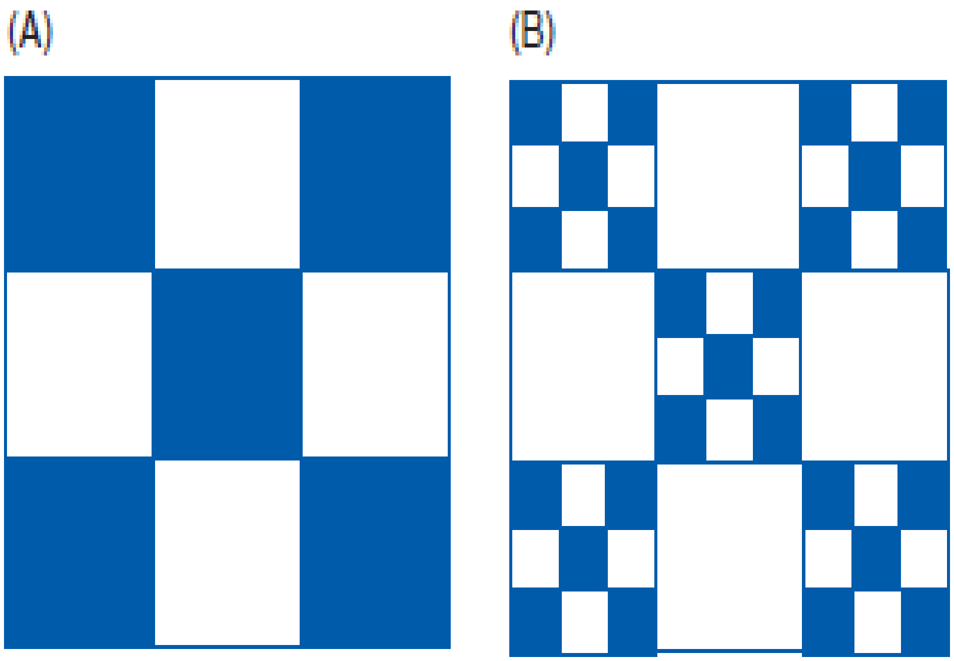


Figure 4.2 A Sierpinski carpet at two levels of resolution: (A) coarse scale and (B) finer scale

Figure 4.2A presents a coarse-scale representation of attributes in nine squares, and a pattern of negative spatial autocorrelation. However, the pattern is self-replicating at finer scales, and in Figure 4.4B, a finer-scale representation reveals that the smallest blue cells replicate the pattern of the whole area in a recursive manner. The pattern of spatial autocorrelation at the coarser scale is replicated at the finer scale, and the overall pattern is said to exhibit the property of *self-similarity*. Self-

Spatial Sampling

- Sample frame : the universe of eligible elements of interest
- Sampling : the process of selecting points from continuous field or selecting some discrete objects while discarding others
- Sampling can determine the quality of the representation
  - How do we ensure a good sample?

Spatial Sampling

- Any geographic representation is a kind of sample

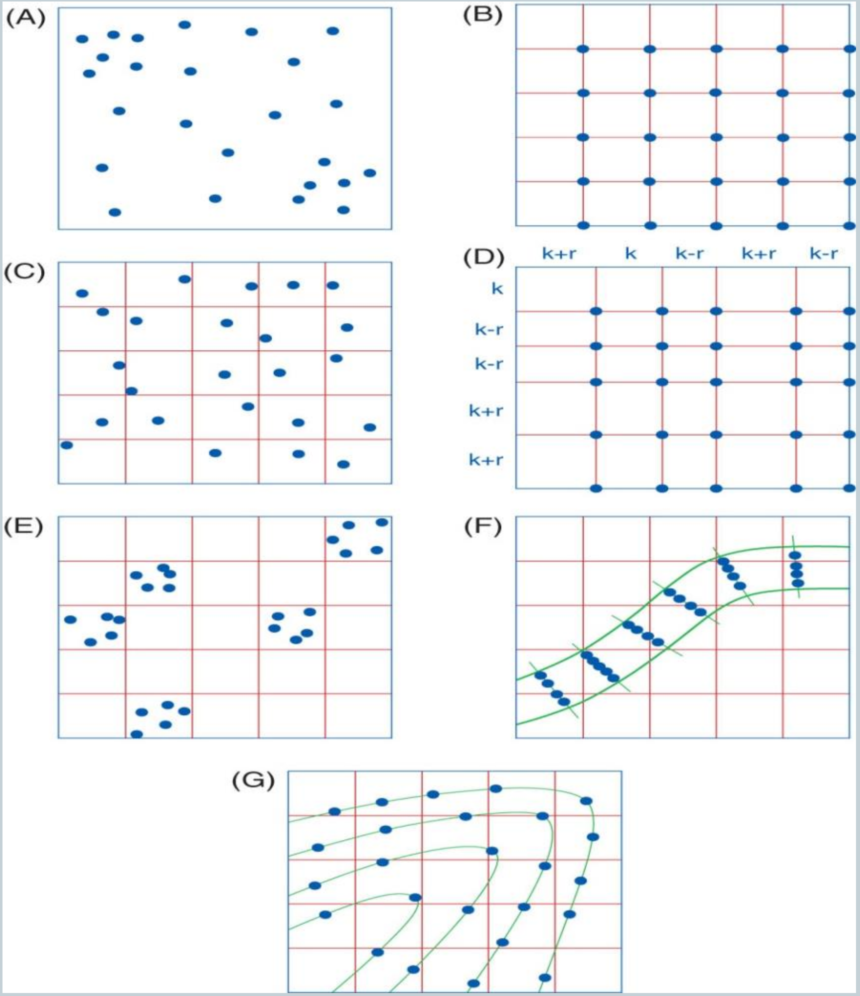
- Procedures of *statistical inference* allow us to infer from samples to the population from which they were drawn

Spatial Sampling

- Randomness is important in sound sample design
  - Simple random sampling
    - Each element in the sample frame is assigned a unique number
    - Selection is done using a random number generator

- Probability theory
  - Distribution of the values

Sampling design



Spatial sample designs

(A)simple random sampling

(B) stratified sampling

(C) stratified random sampling;

(D) stratified sampling with random variation in grid spacing

(E) clustered sampling

(F) transect sampling

(G) contour sampling

Sampling Design

- Types of sampling designs include:
  - simple random sampling,
  - spatially systematic sampling (problems if the sampling interval and spatial structure coincide so that the sample frame exhibits *periodicity*),
  - stratified random sampling, periodic random changes in the sampling grid,
  - clustered sampling,
  - sampling along transects

Sampling Design

- In any application, where the events or phenomena that we are studying are spatially heterogeneous, we will require a large sample to capture the full variability of attribute values at all possible locations
- Thus it may be sensible to partition the sample frame into sub-areas, based on our knowledge of spatial structure
- **Stratified sampling designs attempt to allow for the unequal abundance of different phenomena on the Earth's surface**.

Distance Decay

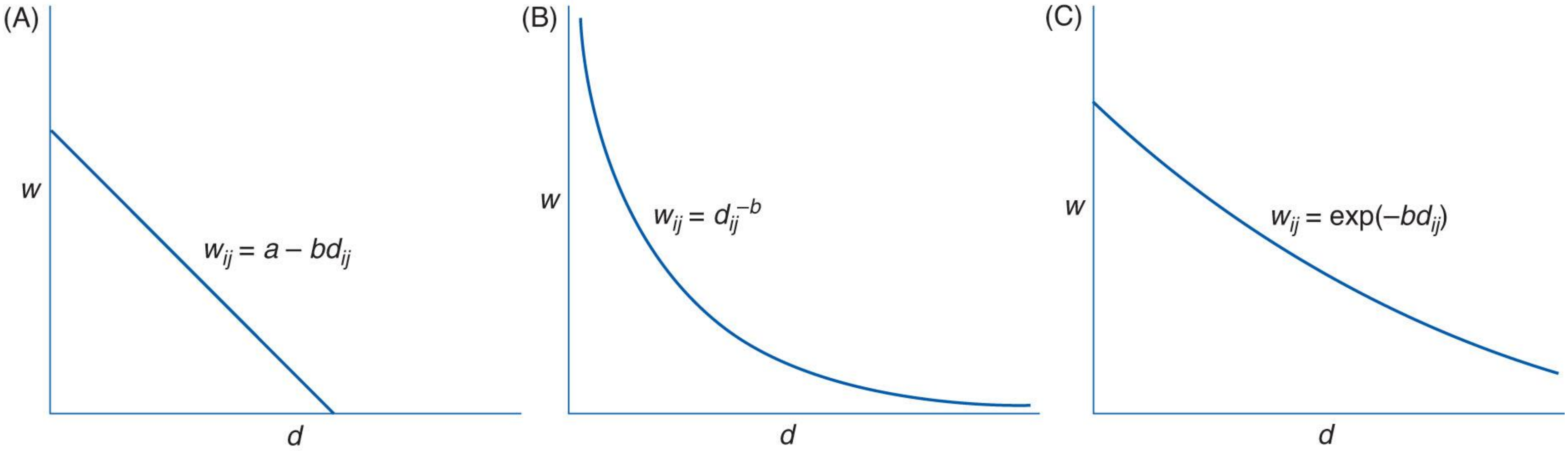
- Discusses the attenuating effect of distance and the need to make an informed judgment about an appropriate *interpolation* function and how to *weight* adjacent observations.
- In mathematical terms, we take b as a parameter that affects the rate at which the weight w_{ij} declines with distance:
  - A small b produces a slow decrease, and a large b produces a more rapid one.

The attenuating effect of distance

(A) linear distance decay,

(B) negative power distance decay,

(C) negative exponential distance decay,



Distance Decay

- linear, negative power, and negative exponential distance decay equations and graphs

- Notes that with these equations, the effects of distance are presumed to be regular, continuous, and *isotropic* (uniform in every direction)

Measuring distance effects as spatial autocorrelation

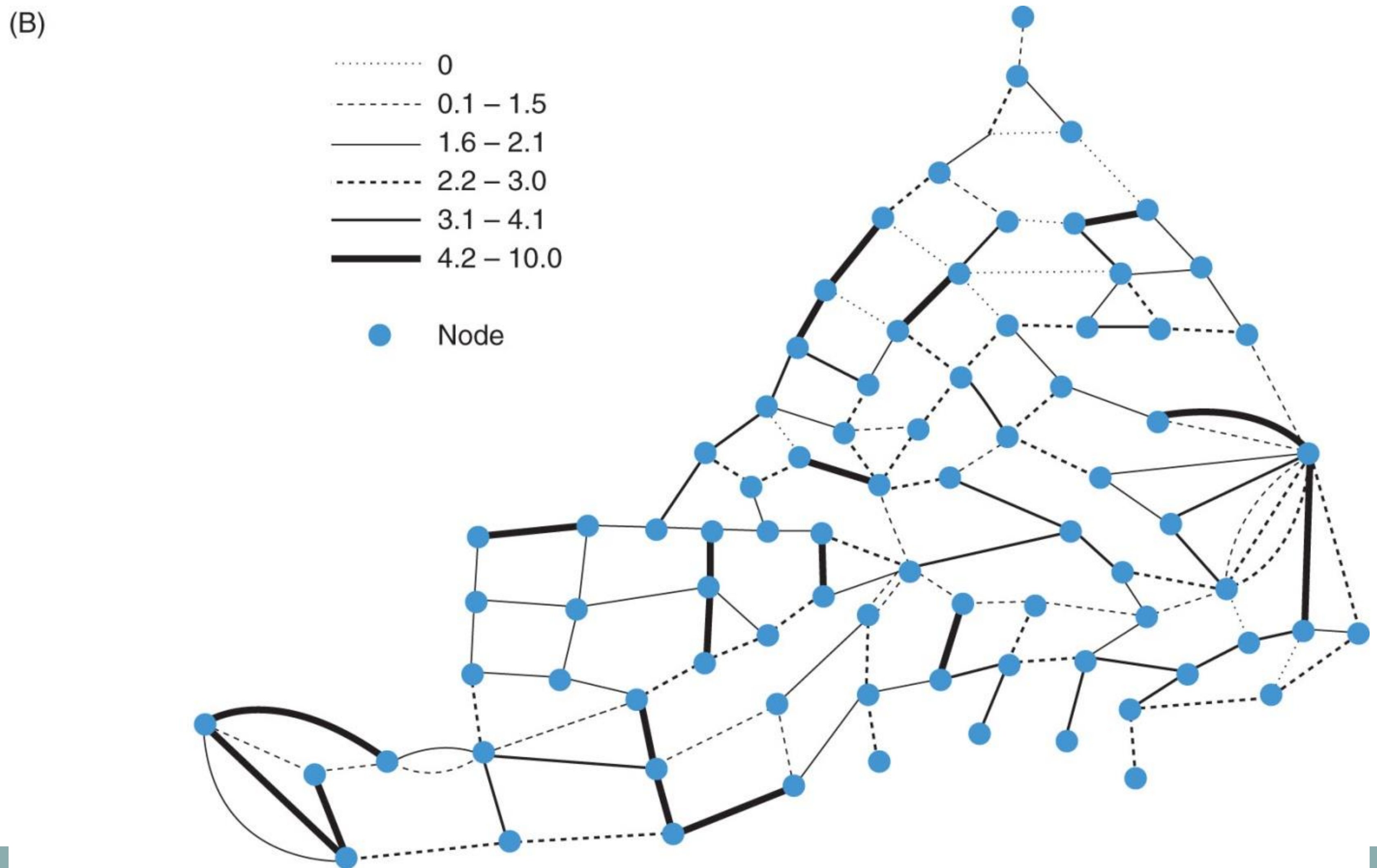
- *Induction* reasons from data to build up understanding, while *deduction* begins with theory and principle as a basis for looking at data

- Knowledge of the actual or likely nature of spatial autocorrelation can be used *deductively* in order to help build a spatial representation of the world.
- The *measurement* of spatial autocorrelation is a more i*nductive* approach to developing an understanding of the nature of a geographic dataset.

Measuring distance effects as spatial autocorrelation

- If the phenomenon is conceived as a field, then spatial autocorrelation measures the smoothness of the field using data from the sample points, lines, or areas that represent the field.
- If the phenomena of interest are conceived as discrete objects, then spatial autocorrelation measures how the attribute values are distributed among the objects, distinguishing between arrangements that are clustered, random, and locally contrasting.

(B) line data (accident rates in the southwestern Ontario provincial highway network)



Measuring distance effects as spatial autocorrelation

- present numbers of accidents for links of road over a lengthy survey period in the Southwestern Ontario, Canada, provincial highway network.
- Low spatial autocorrelation in these statistics implies that local causative factors (such as badly laid out junctions) account for most accidents,
- whereas strong spatial autocorrelation would imply a more regional scale of variation, implying a link between accident rates and lifestyles, climate, or population density.

Conclusion

An understanding of the nature of spatial data allows us to use induction (reasoning from observations) and deduction (reasoning from principles and theory) alongside each other to develop effective spatial representations that are safe to use.