Lecture 9: Vector Data Operations/Analysis

Geography 373
Fall, 2006

Contents of Lecture
- Topology
- GIS Database Errors
- Edge Matching
- Centroid
- Spatial Analysis
- Network Analysis
- Basic Operations
- Database & DBMS
- SQL

Topology
- Topology explicitly defines spatial relationships
- Creating and storing topological relationships has a number of advantages
  - Data is stored efficiently, so large data sets can be processed quickly
  - Topology facilitates analytical functions, such as
    - modeling flow through the connecting lines in a network
    - combining adjacent polygons with similar characteristics
    - identifying adjacent features
    - overlaying geographic features

GIS Database Errors
- Most GIS database errors result from improper input
  - Pushing wrong button on digitizing puck
  - Typing errors during attribute input
  - Problems with input documents
- Three types of error
  1. Entity error
  2. Attribute error
  3. Entity-attribute agreement error

GIS Database Errors
- Primary with vector data
  - e.g. missing entities, incorrectly place entities, disorder entities
- Positional errors
- Topological errors
  - Topology building detects obvious entity errors and allows to check if
    - All entities entered
    - No extra entities entered
    - Entities are in the right place and are correct shape and size
    - All polygons have a single label
    - All entities are within the boundary set

Topological Errors Editing
- Effect of tolerance on topological cleaning
- Topological ambiguities in raster to vector conversion
GIS Database Error

2. Attribute error
   - Difficult to detect
     - Why?
   - Raster
     - Easily detected by blank or very different grid cell value
   - Result from wrong input, typing errors
   - Vector
     - Can be detected by looking at attribute table for missing row
     - Result from forgetting to fill in attribute table

Edge Matching

- Working on more than one tile at a time to make sure there is a correct match between features that extend beyond the boundaries
- Occurs when two adjacent coverages are physically linked
- Projection is one of the main sources of difficulty to match

Centroid

- Point that occurs at exact geographic center of an area
- Involve complex calculations for complex polygons
- Centroids can also be used to signify different distributions
  - Geographic center (based on geometry)
    - Center of large county may not adequately reflect skewed population distribution
  - Mean center (center of gravity)
    - Reflects better where the population really is

Spatial Analysis

- Studying the locations and shapes of geographic features and the relationships between them
- Basics
  - What: nature of an entity by its attributes
  - Where: by geographical location or coordinates
  - Spatial relation: between different entities
    - Proximity, connectivity, adjacency, containment

Spatial Analysis

- Select based on attribute queries (i.e., zoning = "Multi-use")
- Select based on location (within a distance of, intersects, completely within, touching, not within, and many more)
- Create Buffers based on values or attributes
- Clip one data set using another data set
- Merge multiple data sets into a single data set
- Union multiple data sets into a single data set
- Join attributes of data sets together based on spatial relationships (Spatial Join)
- Utilize more than 90 tools in the geoprocessing framework in dialogs, models, and scripts
Network Analysis

- **Network:**
  - An interconnected set of lines representing geographic features such as roads, wires, pipes, or cables through which resources can be moved
  - Can be directed or undirected (one way vs. two way street)
  - Straight line: interstate highway
  - Braiding: stream network
  - Circuits: street patterns
- Any method of calculating locations and relationships in a network, usually in order to study or model connectivity, rate of flow, or capacity
  - Shortest route (path)
  - Optimal delivery route
  - Service area
  - Accessibility

Basic Operations

- **Attributes operations**
  - Query based
    - Logical conditions
      - Boolean logic: OR, AND, NOT
    - Mathematical: arithmetical
  - Location operations
    - Buffer based: “Create Buffer”
      - Find all the students within 0.1 mile of library?
    - Feature based: “Select by Feature”
    - Distance based
      - Find the closest hospital from MSU campus?
- **Topological operations**
  - Overlay based
    - Spatial join: merge, intersect, ...
  - Spatial relationship based
    - Direction: Find the road of right side of library?

Attribute Operations: Logical & Boolean

- **Logical conditions**
  - =, >, <, >=, <=, <>
- **Boolean logic**
  - AND
    - Intersection
  - OR
  - Union
  - NOT
    - Complement
    - Difference

Attribute Operations: Arithmetical

- **Arithmetical operation**
  - Population change in Blue Earth County
    - [population in year 2000] - [population in year 1990]
  - Wet land space in Blue Earth County
    - [lake] + [river] + [pond]
  - Population density in year 2000 in Blue Earth County
    - [population in year 2000] / [area]

Topological Operations: Overlay, Join

- **Layer 1:** land use
- **Layer 2:** land cover
- **Layer 3:** combined layer of Layer 1 & 2
  - Land use / cover information
- **Backward possible:** disjoin

Topological Operations: Merge

- **Important to identify possible spatial extend** for each function with respect to input and output
  - Union
    - Output extend may be bigger than input extends
  - Identity
    - Output extend is never bigger than input extends
    - Keep the input 1 extend
  - Intersect
    - Output extend is never bigger than input extends
    - May not keep the input extends
Spatial Join by Merge

Input 1, Input 2, Output

Database Management Systems

- DBMS
- Database: A collection of related data
- Computer program to create and maintain a database
- Advantages
  - Allows quick access to data
  - Allows edit data (add, delete, update)
  - Protects data from corruption
  - Storage and retrieval
  - Provide interface to data for users
  - Multiple users access
- GIS links attributes and spatial data

Database Model

- Relational model (RDBMS)
- Hierarchical model
- Network model
- Object-oriented model

RDBMS

- Relational Database Management System
- Collection of tables
  - Based on multiple flat files for records
  - Each table stored as a separate file
  - Data stored in simple records (tuples)
    - Tuples grouped together in two-dimensional tables: relations
- SQL
  - Structured Query Language
  - Rules define relationships between files
- Normalization
  - Reduces redundancy

RDBMS Components

- Tuples
  - Records
- Attributes
  - Fields
- Relations
  - Tables
- Primary Key: unique identifier
  - ID

Flat File Database

<table>
<thead>
<tr>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>Attribute 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record 1</td>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>Record 2</td>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>Record 3</td>
<td>Value</td>
<td>Value</td>
</tr>
</tbody>
</table>
RDBMS

- Relation Schema
  - Student
    - Name, SSN, HomePhone, Address, OfficePhone, Age, GPA

---

RDBMS

- School database
  - Student
  - Course
  - Section
  - Grade_Report
  - Prerequisite

---

RDBMS

- In School database
  - Two entities

---

Structured Query Language (SQL)

- A syntax for defining and manipulating data in a relational database
- Developed by IBM in 1970s
- Became an industry standard for query language in most relational database management systems
- Query: a statement or logical expression used to select features or records from a database
  - Spatial query: selecting geographic features by where they are in relation to each other

---

SQL Query

- Standard interface to relational database
- Access data behind the map by "sort", "renumber", "subset", "search", etc.
- Select
  - Attribute list: a list of attribute names whose values are to be retrieved by the query
- From
  - Table list: a list of relation names required to process the query
- Where
  - Condition: a conditional boolean search expression that identifies the tuples to be retrieved by the query

---

SQL Query

- Retrieve the name and home phone number of the student whose GPA >= 3.5
  - Select: Name, HomePhone
  - From: STUDENT
  - Where: GPA >= 3.5
- Query Result
  - Dick Davidson, num
  - Charles Cooper, 376-9821
Hierarchical Database

USA
  - California
  - Minnesota
    - Mankato
    - Minneapolis
  - Ohio
    - St. Cloud

Homework

- Read Chapters 3, 4, 9, 19