Massachusetts Institute of Technology - Department of Urban Studies and Planning

11.520: A Workshop on Geographic Information Systems

11.188: Urban Planning and Social Science Laboratory

Lecture 7: Spatial Analysis (Vector Models)

October 19, 2005, Joseph Ferreira, Jr.
(based, in part, on notes by Visiting Prof. Zhong-Ren Peng in Fall 2003)

Major Topic: Spatial Analysis (using Vector-based Data Models)

- Taking advantage of GIS to handle spatial operations
- Spatial 'Join' issues:
  - Combine data from different sources based on ***location*** & ***proximity***
  - Present results so that they can by visualized and/or saved for further use
- Examples of spatial 'joins'
  - Which Cambridge house sales fall within which census blockgroups
    - Point-in-polygon operation
    - Because we want to know about the neighborhood surrounding lower-price homes...
  - Which blockgroups are in which Mass Town?
    - Polygon-in-polygon operation
    - Tricky because polygon boundaries may not line up (sliver & gap issues)
  - Which roads are in the flood plain?
    - Line-in-polygon operation

        - Should you cut up the road segments or list line segments with part (>50%) inside?
    - Which shopping centers are close to major highways
        - How do we represent shopping centers and highways?
- There are many interesting ***Proximity*** questions

General spatial operations

- Equals – are the geometries the same?
- Disjoint – do the geometries share a common point?
- Intersect – do the geometries overlap?
- Touch – do the geometries intersect at their boundaries?

Buffering Vector Data

- Buffering creates new polygon features by computing which areas are within a certain specified distance of the selected object.
- Buffering may define a new spatial object whose boundary enclosed all the area within the buffer
    - the geometric computations can be complex
    - beware of islands, slivers, etc.
    - involves choices when buffers of selected features overlap (dissolve...)
- Buffer Input: Points, lines or polygons.
- Buffer Output: Polygon(s).

Buffering Raster Data

- Classify grid cells according to whether they lie inside or outside the buffer.
- In addition to distance, you can also use any function of grid cell values (e.g. travel speed) as a buffer criteria.
- The result is a new table of cell values for each grid cell - e.g., a raster map distinguishing all grid cells with, say, less than 30 minutes travel time from a particular point (e.g., a city center).
- More on raster modeling and analysis after next Monday's vector analysis lab.

Point in Polygon

- Whether a point lies inside or outside a polygon.
- In the case of many points and many polygons, the task is to assign points to polygons
- Common operation:
  - Many mailing lists and GPS readings are 'geocoded' into point files
  - You need point-in-polygon operation to **tag the attributes of the point map layer** with characteristics of the places they are 'in'
  - Sometimes, this is a way of handling difficulty polygon-on-polygon operations such as the blockgroup-in-town issue in HW#2
    - Compute centroids of blockgroups
    - Do point-in-polygon operations to see which blockgroup centroids fall inside which towns
  - ArcGIS operation

        - 'intersect' the two coverages in order to add polygon attribute information to point attribute table
        - 'Desktop' ArcGIS tools have less sophisticated capabilities than 'workstation' and 'arcedit' tools (that cost additional money for the extra tools in ArcToolbox)

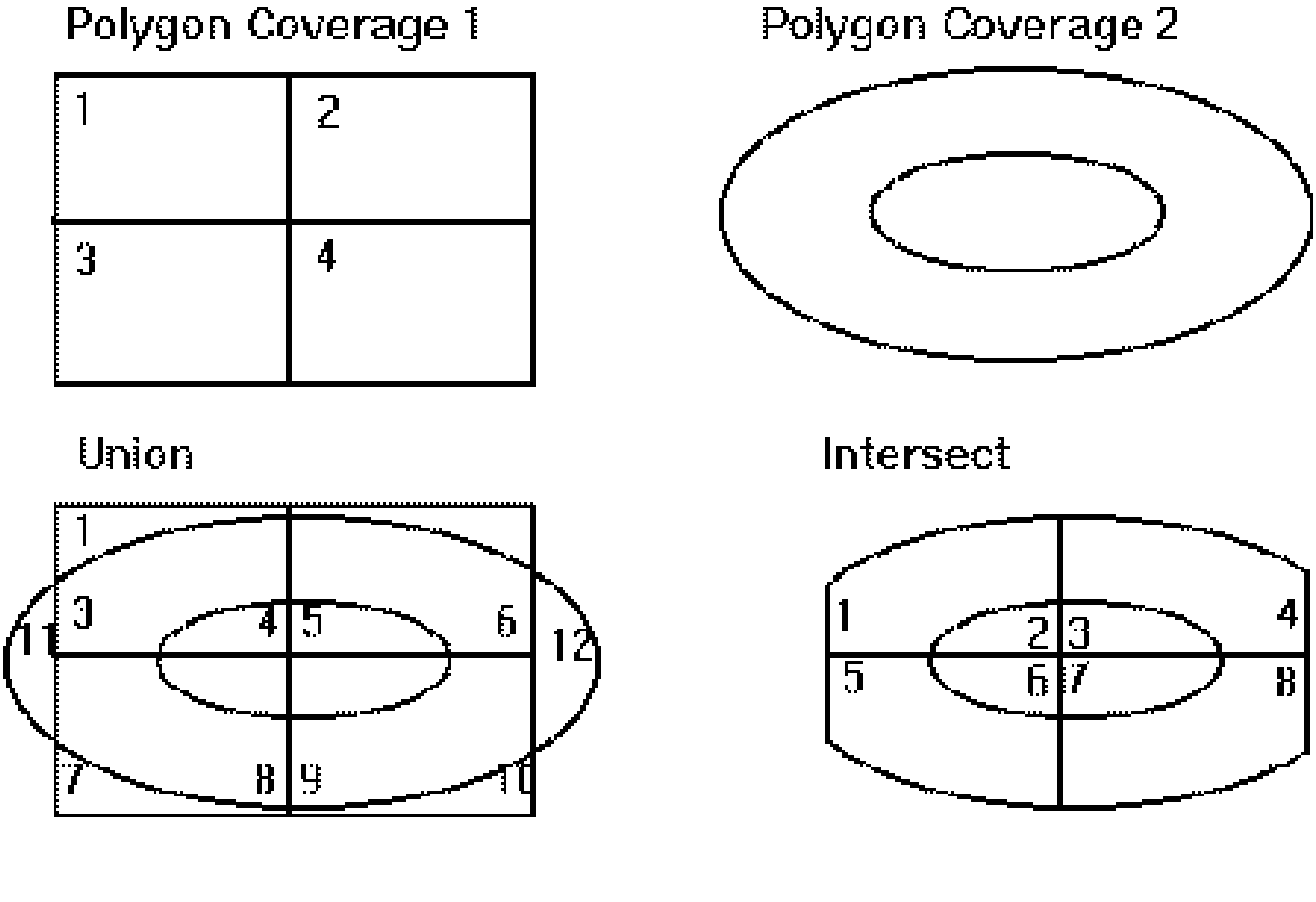
Line in Polygon

- Used to determine whether a line lies inside or outside a polygon.
- In the case of many lines and many polygons, the task is to assign line segments to polygons.
- ArcGIS operation:
- Intersect lines and polygons in order to tag attributes of line segments with polygon information
    - Since polygons aren't likely to cross the lines at the ends of the line segments, this operation can be complex if you wish to split lines at the polygon boundaries and assign the appropriate attributes to each new segment.

Polygon Overlay

- Many possible ways of intersecting polygons with one another
    - Intersection: Polygon 1 + polygon 2 = polygon 3 (with common areas)
    - Union: Polygon 1 + polygon 2 = polygon 3 (with both areas)
- Many issues about how to interpret the results
    - Handling **slivers** due to geometric errors or differences in degree of generalization of boundaries
    - Handling **attributes**

        - Some attributes apply to each fragment without change (e.g., in flood plain, in city...)
        - Some attributes need to be 'allocated' to each fragment (e.g., population in each part of blockgroup)
    - In next Monday's lab exercise we will have examples (e.g., estimate number of people living within a specified distance of MIT's biology building)



More spatial data processing

1. Dissolve features based on an attribute.
1. Merge layers together.
1. Clip one layer based on another.

Spatial Analysis Example - Locating a Shopping Center

1. Criteria:
    1. Within two mile of a major highway, but no less than one quarter mile away from a major highway.
    2. At least one mile away from a water body (e.g., rivers, lakes).
    3. Not in the flood zone.
    4. Close to a major residential area (at least 2500 households within 5 miles).
    5. Have a large piece of vacant land (at least 10 ac

<![endif]> Classic 'site suitability' analysis

<![endif]> ***Decompose task into many ArcGIS analysis steps***

Data Required

- Highway layer (linear feature),
- Vacant land layer (polygon feature),

1. Rivers layer (linear or polygon feature),
1. Lakes layer (polygon feature),

- Land use layer (polygon feature),

1. Flood zone layer (polygon feature),
1. Census layer (blockgroup level, polygon feature).

Spatial Analysis Process

- Select vacant land parcels larger than 10 acres, result: **vacantlg10**.
- Buffer highway layer with 2 mile and .25 mile, result: **hw2** and **hwquarter**.
- Buffer Rivers and Lakes layers with 1 mile, result: **river1** and **lake1**.
- Intersect vacantlg10 with hw2 = **vacanthw1**, useful land.
- Union hwquarter with river1, lake 1 and flood (need to do two layers at a time), result: **notusable**, (land not meeting the criteria)
- remove notusable from vacanthw1, result: **usable**.
- intersect usable with census blockgroup layer (touch the boundary of), result: **usableblock**.
- calculate and reselect usableblock with number of total household larger than 2500, result: **bigusableblock**.

Additional criterion

1. Consider an additional criterion: “The median household income of the proximate residential area should be more than $30,000”?
2. Refine 'major residential area' criterion
    3. Require enough people within a buffer area around site
    4. Require proximity to dense residential area

ArcGIS Example from Homework #2: find which blockgroups are in which town

- We can do a 'Spatial Join' involving polygon-to-polygon operations but that won't work well because of the many 'sliver' issues along the edges of the blockgroups and town boundaries

- Instead, compute centroid of blockgroups and then do a point-in-polygon operation to find surrounding town
- Computing blockgroup centroids (interior point that is close to 'middle')
  - Not part of 'shapefile' data model - which only stores boundaries
  - Use ArcGIS help to find a way: look up 'Centroid - calculating for polygons' in ArcGIS help
    - Select blockgroups in and around the 5 towns and save to a new shapefile (*blkgrp_start.shp*)
    - add X,Y fields to the attribute table of *blkgrp_start* (as double precision numbers)
    - Calculate new X values using the 'advanced' button and this VBA script:

      ```
      Dim dblX As Double
      Dim pArea As IArea
      Set pArea = [Shape]
      dblX = pArea.Centroid.X
      ```

    - Repeat for Y and export the edited table - call it blkgrp_c.dbf
    - Add blkgrp_c to the data frame and create a new point shapefile **'blkgrp_c Events'** of these X,Y centroid points using Tools/Add-X-Y-data option
  - Finally, Use the point-in-polygon **spatial join** operation to attach town attribute information to the point attribute table 'blkgrp_c Events'.
    - Right-click on the **'blkgrp_c Events'** layer and choose the **'join'** option from the **'joins & relates'** menu item,
    - Instead of joining attributes based on relating two tables, we want to 'join data from another layer based on spatial location'
    - The layer we want to join to the selected layer (**blkgrp_c Events)** is the town layer (which we called **cambridge5** when we saved the 5-town boundaries above),

- This operation will add the columns in the **cambridge5** attribute table to the columns in the **'blkgrp_c Events'** attribute table where the rows are matched based on the town that contains each block group centroid. The results are stored in a new shapefile with the same block group centroid points and the expanded attribute table.
- You won't be able to do the spatial join unless each of the layers you are joining have a known spatial reference system - because the spatial join won't be meaningfull unless both layers can have all their coordinate expressed (or transformed) into the same coordinate system.
- This expanded attribute table has the blk_key in one column and, now, the town name and code in another column. So, we turned off all the columns except for blk_key and Town and and selected only those blockgroups in the 5 towns. Exporting this table produced the block-group-to-town (**blkgrp2t.dbf)** lookup table that we have included in the 11.520 data locker for your use in HW#2.

Homework #2: This will take some time! Spread out the work!

- start early, focus on spatial operations for suitability analysis
- save often, take care in naming intermediate files, same a *.mxd project files with different version numbers as you move along.
- expect to redo overlays a few times before you like result (so you need to remember the saved pieces and combine them into new projects)
- some numbers in census blockgroup data table may be incorrect (with our new SDE

data layers and summary level '150' instead of '090'); staff will double check and email later this week

-

Last modified 19 October 2005. [jf]