# Algorithms for GIS csci3225 

Laura Toma

Bowdoin College

## Spatial data types and models

## Spatial data in GIS

satellite imagery

networks

planar maps

surfaces

point cloud (LiDAR)



Spatial data in GIS

How to represent it?

## Vector model

- Data is represented using points, lines and polygons
- Useful for data that has discrete boundaries, such as streets, maps, rivers


## Raster model

- Data is represented as a surface modeled by a matrix of values (pixels)
- Useful for "continuous" data (ie data that varies continuously) such as satellite imagery, aerial photographs, surface functions such as elevation, pollution, population


Real World



## Both models can be used for all data


http://gsp.humboldt.edu/olm/Lessons/GIS/08\ Rasters/Images/convertingdatamodels2.png

## Both models can be used for all data


http://www.newdesignfile.com/postpic/2013/04/vector-vs-raster-data_132177.jpg

World is often modeled as a collection of vector and raster layers
streets, parcels, boundaries usually as vector
elevation, land usage usually as raster


## Spatial data in GIS

satellite imagery

planar maps

VECTOR
point cloud (LiDAR)


## VECTOR

## Data structures for networks

## Data structures for networks?

- How about this:
- list of points, each point stores its coordinates
- list of segments, each segment stores pointers to its vertices

- What if you wanted to traverse a path starting at point 0?
- search through the segment list looking for a segment that starts from a; you find $(0,1)$
- search through the segment list looking for another segment from 1; you find (1,2)


## Data structures for networks?

- How about this:
- list of points, each point stores its coordinates
- list of segments, each segment stores pointers to its vertices


Spaghetti data structure (like spaghetti, no structure, messy, inefficient eating)

## Data structures for networks

- Need a topological data structure that allows to traverse paths efficiently
- Wait, a network is a graph!
- Use adjacency list

- In practice, this adjacency list needs to be built
- From raw data


## Exercise

```
Assume you download US road data. It comes as a file that has the
following format
    - first the number of vertices and the number of edges
    - then all the vertices and their geometric coordinates
    - then all edges, where an edge is given through the indices of its
        vertices.
Sketch how you would build an adjacency list from it.
Analyze function of |V| vertices and |E| edges.
```

```
4
3
(1.1, 2.3)
(3.4, 2.1)
(2.6, 1.8)
(1.4,8.2)
(0,1)
(1,2)
(2,3)
```


## Spatial data in GIS


point cloud (LiDAR)


Data structures for surfaces

## Surfaces can have different topologies



## Surfaces in GIS

- GIS deals with the surface of the Earth (or Mars, or...)
- The Earth is round, and its surface has the topology of a 2D sphere



## Terrains

- A terrain is a function of two variables, $z(x, y)$. Meaning that, for a given $(x, y)$, there is a unique $z(x, y)$
- Put differently, a terrain is a surface in 3D such that any vertical line intersects it in at most one point (xy-monotone)



## Terrains



## Not terrains



## Terrains

- Most often terrains represent elevation, but they can also represent other Earth surface functions like rainfall, population, solar radiation, ...


Global Annual Average $\mathrm{PM}_{25}$ Grids from MODIS and MISR Aerosol Optical Depth (AOD), 2010: North America
Satellite-Derived Ervironmental Indicators

 (AOD) data sc:s provide annua 'snef snots' sf particulste ma'tor 2.5 in latitalers ir smaler in diameter Fom 200-2010. Exiusu et to fi"e partides is as $50 c$ ated with premature dath as well as increased morbilly fom respiratory ant cartiovascular disease, espec ally in the e derly, ycung
 de rived from Mcderate Resolution Imaza no Spectrorad ometer (MODIS) anc PA lioangle Imagiog GpertroRsdiometer (VISR) \&e-osol Opt cal Desth (AOD) da:a. The rastecerid call size is approximstels 50 sq . km at the equator, and the wxilut is from $70^{\circ} \mathrm{N}$ tu $60^{\circ} S$ latitude
 Canct Lex Jusenz.en a Evint




旬 2013 The Thusters of C-:kith a IJriversity in the Cily of New York

## Modeling terrains: Digital terrain models

- In practice, terrain data comes as a set of sample points $\{(x, y)\}$ and their sampled $z$-values
- A digital terrain model = sample points + interpolation method + data structure



## With DTM we can do terrain modeling



## Digital terrain models

|  |  |  |
| :---: | :---: | :---: |
|  |  |  |
|  | Rasters |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |


images from Herman Haverkort

## Terrain as a raster (grid)

A raster terrain is a matrix of (elevation) values


- Samples
- uniform grid
- Data structure
- matrix
- Interpolation method
- nearest neighbor, linear, bilinear, splines, krigging, IDW, etc

Grids with nearest neighbor interpolation


Grids with nearest neighbor interpolation


Grids with nearest neighbor interpolation


## Linear interpolation



## Linear interpolation



Terrain: mesh of triangles on grid points

Raster with nearest neighbor vs linear interpolation

http://c1.zdb.io/files/2009/03/10/9/9700e9183d96ccb416b81b053887fef0.gif

## Grids in practice

- Grid elevation data can be obtained from aerial imagery
- image = raster
- SAR interferometry: by combining Synthetic Aperture Radar (SAR) images of the same area it is possible to go from color to elevation maps
- Massive amounts of aerial imagery available
- Grid elevation data from LiDAR point clouds

point cloud (LiDAR)



## Grids in practice

- Elevation data sources
- GTOPO dataset
- whole Earth at 1 km resolution
- http: ???
- SRTM 90m elevation data for entire world
- http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp
- can download tiles anywhere in the world
- SRTM 30m data available for the entire USA (50+GB)
- Recently, elevation from LiDAR point clouds
- below 2 m resolution
- Huge!
- Grids available in a variety of formats


## Grid arc-ascii format



## Exercise

Consider an area of 300 km -by-300km to be represented as a raster (grid) at:
A. 100m resolution
B. 10 m resolution
C. 1 m resolution

How big (how many points) is the grid in each case?

Answer:

## Terrain as a TIN (triangulated irregular network)



- Samples
- points arbitrarily distributed, variable resolution
- Interpolation method
- linear
- Data structure
- need a topological structure for triangular meshes


## Why TINs?


uniform resolution $=$ waste on flat areas

variable resolution ==> fewer points

The differences bewtween a DEM and a TIN data set



http://www.staff.city.ac.uk/~jwo/landserf/landserf230/doc/userguide/images/figure3.12.jpg

## Topological data structures for TINs



The 2D projection of a triangulated terrain is a triangulation.
$\downarrow$
A TIN is equivalent to a planar triangulation, except points have heights

## Topological data structures for TINs



The 2D projection of a triangulated terrain is a triangulation.
What do we expect to do on a TIN?

- walk along an edge/triangle path
- given an edge, find the two faces that are adjacent to this edge

A good data structure for TINs $\longleftarrow$ should do all these fast

- walk along the boundary of a face (triangle)
- find all edges and all triangles incident to a point


## Topological data structures for TINs

## Edge-based

- arrays of vertices, edges and triangles


## Triangle-based

- arrays of vertices and triangles (edges are not stored explicitly)
- every vertex stores:
- its coordinates
- every edge stores:
- 2 references to its adjacent


## geometry every vertex stores: <br> - its coordinates

- every triangles stores:
- 3 references to its incident vertices
vertices


## topology

- 2 references to its adjacent triangles
- every triangle stores:
- 3 references to its 3 edge
- 3 references to its adjacent triangles
- Note: CGAL uses triangle-based
- These are simplified versions of more general structures for arbitrary meshes which we might study later (the half-edge and quad-edge data structures)


## Data structures for TINs


$\begin{array}{ll} & \text { vertex } \\ \text { edge } \\ \text { triangle }\end{array}$

edge-based

triangle-based

## Data structures for TINs


$\begin{array}{ll} & \text { vertex } \\ \text { edge } \\ \text { triangle }\end{array}$

edge-based

triangle-based

Analysis? Is one better than other?

- Storing topology:
equivalent
- Memory:??


## Data structures for TINs

```
How much memory does a topological structure for a TIN need?
    A. edge-based
    B. triangle-based
Denote
    n = number of points in the TIN
    e = number of edges
    f = number of triangles (faces)
```

There is a formula that connects e, $f$ and $n$.

Detour through planar graphs and Euler characteristic

Let P be a set of points in the plane.
A triangulation is a partition of the plane into regions such that all regions are triangles.

one possible triangulation of $P$

## planar graph

## $\downarrow$

A triangulation is a graph: $V=$ the points, $E=$ the edges


## Planar graphs

A graph is called planar if it can be drawn in the plane such that no two edges intersect except at their endpoints.

Such a drawing is called a planar embedding of the graph.

Example: $\mathrm{V}=\{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}\}, \mathrm{E}=\{(\mathrm{a}, \mathrm{b}),(\mathrm{a}, \mathrm{c}),(\mathrm{a}, \mathrm{d}),(\mathrm{b}, \mathrm{c}),(\mathrm{c}, \mathrm{d}),(\mathrm{d}, \mathrm{a})\}$

Drawing the graph in the plane is called embedding.
Let's come up with different embeddings of the graph.

## Planar graphs

A graph is called planar if it can be drawn in the plane such that no two edges intersect except at their endpoints.

Such a drawing is called a planar embedding of the graph.


Two drawings of the same graph.
Since there exists a planar embedding, the graph is planar.

## Planar graphs

A graph is called planar if it can be drawn in the plane such that no two edges intersect except at their endpoints.

Such a drawing is called a planar embedding of the graph.

Note: Edges can be represented as simple curves in the drawing

http://people.hofstra.edu/geotrans/eng/methods/img/planarnonplanar.png

## Planar graphs

A planar graph introduces a subdivision of the plane into regions called faces, which are polygons bounded by the graph's edges.

planar graph

triangulation

All faces (except the "outside" face) are triangles

## Euler formula

[Euler] The following relation exists between the number of edges, vertices and faces in a connected planar graph: $v-e+f=2$.


The 7 bridges of Konisberg problem
Euler, November 14, 1750

## WikipediA

The Free Encyclopedia

$$
V-E+F=2
$$

This equation is known as Euler's polyhedron formula. ${ }^{[1]}$ It corresponds to the Euler characteristic of the sphere (i.e. $\mathrm{X}=2$ ), and applies identically to spherical polyhedra. An illustration of the formula on some polyhedra is given below.

| Name | Image | Vertices <br> $\boldsymbol{V}$ | Edges <br> $\boldsymbol{E}$ | Faces <br> $\boldsymbol{F}$ | Euler characteristic: <br> $\boldsymbol{V}-\boldsymbol{E}+\boldsymbol{F}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Tetrahedron |  | 4 | 6 | 4 | $\mathbf{2}$ |
| Hexahedron or cube |  | 8 | 12 | 6 | $\mathbf{2}$ |
| Octahedron |  | 6 | 12 | 8 | $\mathbf{2}$ |
| Dodecahedron |  | 20 | 30 | 12 | $\mathbf{2}$ |
| Icosahedron |  | 12 | 30 | 20 | $\mathbf{2}$ |

## Euler formula

[Euler] The following relation exists and faces in a connected planar g

## Euler characteristic



In mathematics, and more specifically in algebraic topology and polyhedral combinatorics, the Euler characteristic (or Euler number, or Euler-Poincaré characteristic) is a topological invariant, a number that describes a topological space's shape or structure regardless of the way it is bent. It is commonly denoted by (Greek lower-case letter chi).

## Euler characteristic - Wikipedia

https://en.wikipedia.org/wiki/Euler_characteristic

Notes:

- For c connected components: v-e $+\mathrm{f}-\mathrm{c}=1$
- $v-e+f=2$ also true for any convex polyhedral surface in 3D
- $v-e+f$ is called the Euler characteristic, $X$
- $X$ is an invariant that describes the shape of space
- it is $X=2$ for planar graphs and convex polyhedra
- can be extended to other topological spaces

The surfaces of nonconvex polyhedra can have various Euler characteristics;


## WikipediA <br> The Free Encyclopedia

## From Euler formula to size of triangulations

$n=n b$ of points
e = nb. edges
$\mathrm{f}=\mathrm{nb}$ faces

- A triangulation is a planar graph, so $n-e+f=2$ [Euler]
- Furthermore, each triangle has 3 edges and each edge is in precisely 2 triangles (assuming the outside face is a triangle). This means $3 f=2 e$.
- We get:
- the number of faces in a triangulation with $n$ vertices is $f=2 n-4$
- the number of edges in a triangulation with $n$ vertices is $e=3 n-6$
- If the outside face is not triangulated it can be shown that
- $e<3 n-6, f<2 n-4$
- Intuition: Given n points, the planar graph with largest number of edges and faces is a complete triangulation.


## Theorem:

A triangulation with $n$ vertices has at most $3 n-6$ edges and at most $2 n-4$ faces.

## Known results

- Any planar graph has a planar straight-line drawing where edges do not intersect [Fary's theorem].
- A graph is planar iff it has no subgraphs isomorphic with K5 or K3,3 [Kuratowski's theorem].

- Any planar graph ==> has a dual graph.
- A graph is planar $<==$ it has a (well-defined) dual graph.
- Any planar graph has at least one vertex of degree $<=5$.
- Computationally: There are a number of efficient algorithms for planarity testing that run in $o\left(n^{3}\right)$, but are difficult to implement.


## End Detour

## Data structures for TINs

```
The problem: How much memory do we need to store a TIN into a
topological structure?
    - edge-based
    -triangle-based
Denote
    - n = number of points in the TIN
    - e = number of edges
    - f = number of triangles (faces)
```

Answer:
We'll use that: a triangulation with $n$ vertices has $e<=3 n-6$ and $f<=2 n-4$

## Grid or TIN?

## Grid

- Pros:
- implicit topology
- implicit geometry
- simple algorithms
- readily available in this form
- Cons:
- uniform resolution ==> space waste
- space becomes prohibitive as resolution increases


## TIN

- Pros:
- variable resolution
- space efficient (potentially)
- Cons:
- need to built the TIN (grid $\longrightarrow$ TIN)
- stored topology takes space
- more complex programming (pointers..);


## Grid or TIN?

```
We have an elevation grid for an area of 300km-by-300km at 1m
resolution. The elevation values are represented as floating point
numbers (4B).
A. How much space does the grid use (in GB)?
B. Assume the grid undergoes a process of simplification, so that
90% of the grid points are eliminated, leaving 10% of the points.
These points are represented as a TIN with a topological edge-
based structure. How much space does the TIN use (in GB)?
```

Answer:

## Grid-to-tin simplification



## Summary

- Data models: raster and vector
- Networks
- Terrains
- rasters
- TINs
- Topological structures for TINs
- Planarity
- Euler formula $V-E+F=2$
- A triangulation is a planar graph, and $e<3 n$ and $f<2 n$
- Grid or TIN?

