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)



A *showed involvence* possible let of options for a future "frequent services" bus network in inner justing fours of the Harbour







satellite imagery



networks

point cloud (LiDAR)



A along indicative pointere let of options for a future "pregnent services" bus nutwork in inder iducions foutde of the Raebour

Tager inductions correcters to and room the CBO, with services every 3 to 5 minutes, or better, in an

Other insures radial processional framily CRD with services every Similaries, whether all he



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



http://campusarch.msu.edu/wp-content/uploads/2011/10/raster-and-vector-model1.jpg



https://sqlserverrider.files.wordpress.com/2013/10/raster-vector-gis-i4.jpg

Both models can be used for all data



http://gsp.humboldt.edu/olm/Lessons/GIS/08%20Rasters/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



http://3.bp.blogspot.com/-A140pKBwSXU/VHvp73TcocI/AAAAAAAAAAAAK8/xoP1KQI5L-Y/s1600/Raster%2Band%2BVector%2BData.jpg

Spatial data in GIS

satellite imagery



planar maps VECTOR

surfaces



point cloud (LiDAR)





VECTOR

Data structures for networks



A *highly indicative* possible set of options for a future "frequent services" bus network in inner suburbs south of the Harbour

Major radial trunk corridors to and from the CBD, with services every 3 to 5 minutes, or better, all day





on-CBD frequent services, generally with services every 10 minutes, or better, all day, resigned to connect with the radial srervices, and each other, to provide a complete network w

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)



satellite imagery



planar maps



surfaces



point cloud (LiDAR)



A *highly indicative* possible set of options for a future "frequent services" bus network in inner suburbs south of the Harbour

- Major radial trunk corridors to and from the CBD, with services every 3 to 5 minutes, or better, all day
- Other frequent radial services to and from the CBD, with services every 5 minutes, or better, all day



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)

surface of Mars



Terrains



https://farfarer.com/blog/wp-content/uploads/2011/12/terrain_triplanar_iso.jpg

http://www.orefind.com/images/blog-figures/topo2_fig3.jpg?sfvrsn=0

Not terrains

http://paulbourke.net/







Terrains

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

Portland Helena **Bismarck** Minneapolis Boston Pierre **dilwauke** Buttalo Ontroit Lake City New York Pittsburgh Cheyenne Cleveland Chicago Sacramento ladelphia San Francisco San Jose - Loue Vorfolk Las Vegas Nashville Santa Fe Memphis Los Angeles Phoen Atlanta San Diego Houston New Orleans Wind Speeds and Population Density Tampa Miami Low Population High Population Low Wind High Wind ://www.thecartofish.com/blog/wp-content/uploads/2014/12/WindPopRaster.png

Population surface: $z(x,y) = population_at(x,y)$

Global Annual Average PM2.5 Grids from MODIS and MISR Aerosol Optical Depth (AOD), 2010: North America

Satellite-Derived Environmental Indicators



(AOD) data sets provide annual "snap shots" of particulate matter 2.5 micrometers or smaller in diameter from 2001-2010. Exposure to fine particles is associated with premature death as well as increased morbidity from respiratory and cardiovascular disease, especially in the elderly, young children, and those already suffering from these i lnesses. The grids were derived from Moderate Resolution Imaging Spectrorad ometer (MODIS) and Multi-angle Imaging SpectroRadiometer (MISR) Aerosol Optical Depth (AOD) data. The raster grid cell size is approximately 50 sq. km at the equator, and the extent is from 70°N to 60°S latitude.



Data Sou ce: Battelle Meinor al Institute, and Center for International Earl'inScience Information Natwork Center for International Earth (C.E.EIN)/Oclumbia University 2013. Clobal Annual Average PMEL5 Grids from MODIS and MISR Science Information Network Astrosol Optical Depth (AOD), 2001–2010. Palisades, NY NASA Socioeconomic Data and Applications Sam Isomer Lovinae (samer) Center (SEDAC). http://accac.clesin.co.umbia.edu/date/set/sdei-global-annual-avg-pm2-5-2001-2010. 2013 The Trustees of Columbia University in the City 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



pictures from Herman Haverkort

With DTM we can do terrain modeling



risks of floods, landslides, eruptions, erosion

costs/benefits of roads, buildings, dikes, hydropower plants, solar power plants

possible locations of (pre-)historical roads and settlements

analysis of animal behaviour and evolution

Herman Haverkort, http://haverkort.net/

Digital terrain models



images from Herman Haverkort

Terrain as a raster (grid)





- Samples
 - uniform grid
- Data structure
 - matrix
- Interpolation method ٠
 - 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



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

Grid arc-ascii format

391

ncols

472 nrows 271845 xllcorner yllcorner 3875415 cellsize 30 NODATA value -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 781 773 766 761 756 751 745 738 730 723 716 -9999 775 767 759 754 750 746 741 735 728 721 714 709 705 700 696 694 693 692 691 690 691 692 696 701 707 714 722 728 732 733 731 726 721 718 718 722 729 737 746 755 761 764 762 760 760 759 754 748 741 733 725 718 -9999 770.

Exercise

```
Consider an area of 300km-by-300km to be represented as a
raster (grid) at:
A. 100m resolution
B. 10m resolution
C. 1m 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.geophysik.uni-kiel.de/~sabine/BsAs2000/DEM-TIN.gif





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.

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

A good data structure for TINs should do all these fast

Topological data structures for TINs

Edge-based

- arrays of vertices, edges and triangles
- every vertex stores:
 - its coordinates
- every edge stores:
 - 2 references to its adjacent vertices topology
 - 2 references to its adjacent triangles
- every triangle stores:
 - 3 references to its 3 edge

Triangle-based

- arrays of vertices and triangles (edges are not stored explicitly)
- geometry every vertex stores:
 - its coordinates
 - every triangles stores:
 - 3 references to its incident vertices
 - 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)





• Memory: ??

```
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

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



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

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

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.



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



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



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. $\chi = 2$), and applies identically to spherical polyhedra. An illustration of the formula on some polyhedra is given below.

Name	Image	Vertices V	Edges <i>E</i>	Faces <i>F</i>	Euler characteristic: V – E + F
Tetrahedron		4	6	4	2
Hexahedron or cube	T	8	12	6	2
Octahedron		6	12	8	2
Dodecahedron		20	30	12	2
lcosahedron		12	30	20	2

Euler formula

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

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

by (Greek lower-case letter chi).

Euler characteristic - Wikipedia https://en.wikipedia.org/wiki/Euler_characteristic

Notes:

- For c connected components: v e + f 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;

Name	Image	Vertices V	Edges <i>E</i>	Faces F	Euler characteristic: V - E + F
Tetrahemihexahedron		6	12	7	1
Octahemioctahedron		12	24	12	0
Cubohemioctahedron		12	24	10	-2
Great icosahedron		12	30	20	2

From Euler formula to size of triangulations

TIN: n = nb of points e = nb. edges f = 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 3f = 2e.
- We get:
 - the number of faces in a triangulation with n vertices is f = 2n-4
 - the number of edges in a triangulation with n vertices is e = 3n-6
- If the outside face is not triangulated it can be shown that
 - e < 3n-6, f < 2n-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 3n-6 edges and at most 2n-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(n³), but are difficult to implement.

End Detour

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 <= 3n-6 and f <= 2n-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 —> 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 edgebased 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 < 3n and f < 2n
- Grid or TIN?