Routes – Modeling, Storage and Analysis

Technical Report

April Jensen and Marina Gavrilova

SPARCS Lab
CPSC Department
University of Calgary

Routes – Modeling, Storage and Analysis

The modeling and analysis of routes and networks has been studied widely in computer science and other disciplines due to its large number of applications. Some of these include the planning of transportation routes, Internet routing and the location of emergency facilities. This paper will include a look at some topology-based methods for storing and analyzing routes, including the use of ArcGIS, as well as a look at Voronoi diagrams. Some potentially relevant network analysis methods will be discussed as well.

The vector model works well for storing and visualizing routes. A vector has magnitude and direction and can be represented graphically by a line segment [5]. A point is the fundamental primitive in the vector model and objects are created by joining points by straight lines [12]. Measures in the vector system are done by performing calculations on coordinates, rather than counting cells as in the raster system [12], and this may make them more accurate.

In the area of GIS, a number of vector data formats exist and are used, for example the TIGER format and the TIN, which makes use of the Delaunay triangulation and is used to model elevation [16]. The arc-node model is very widely used (according to [5]) and it is well-suited to representing linear features such as routes as well as being efficient for storage [1].

Lines or polygons can be represented as a sequence of coordinates [12]. In the arc-node model, arcs (lines) are used as a basic unit and are defined by a start point and end point (called “nodes”). Arcs then join only at these nodes, and areas (“polygons”) can then be built up from the arcs [16]. In this model, arcs have attributes stored with them, for example “left” and “right” which keep track of which polygon is on each side of the arc [12]. A “node topology table” defines nodes by which arcs they belong to; an “arc topology table” stores points designated as the start or end points of arcs, and the “polygon topology table” stores the arcs that make up each polygon’s boundaries [5].

ArcGIS has 2 storage formats for vector data – the “shapefile” and the “coverage”, both of which store vector information in binary files [15]. Shapefiles store geometric information (shapes as a set of coordinates, [16]) and support line, point and area features but only one type per file [16]. They make use of optional files as well as 3 mandatory files which include a “.shp” file (which stores the geometry), an index file (“.shx”) and a database or “.dbf” file (which stores the shape attributes). Because they lack topology information there is less overhead so they are faster to draw and edit, however this limits spatial analysis [16]. To perform spatial or network analysis, the shapefile can be converted to a coverage and a topology can be built.

In the study of routes and networks, topology is often very important. It refers to spatial relationships between features based on a set of rules [8]. “Strict” topology properties are not changed by distortions, for example the “adjacency” relationship [5]. Some properties that are often important and may need to be preserved include proximity, connectivity (path relationships) and adjacency. Other interests may include distance and direction, distribution patterns and neighborhoods [5]. There are various types of topology, including route, arc-node and region [8], and it can be useful not only for analysis but also for supporting queries, defining and enforcing data rules and other applications [8]. In routing, path connectivity and path information are of particular interest [20].

In ArcGIS (specifically ArcInfo), vector topology data is stored in a coverage, which is the standard format for spatial analysis [15]. These contain feature classes, which can be primary (arcs, nodes, polygons, label points) or composite (routes, regions etc., built from primary features) [15]. Multiple classes can be stored in one coverage, and they can be exported to an ASCII file [15]. A feature in a coverage is represented by a unique “feature number” used to link spatial attribute data [1], as well as a sequence number that is automatically generated.

Points are stored as (x, y) coordinates in a “LAB” file, and a PAT (Point Attribute Table) holds data about them [1]. Similarly, arcs are stored in an ARC file with one record per arc and information such as “from-node”, “to-node”, and left and right polygon numbers. An AAT (Arc Attribute Table) holds their descriptive information [1]. Node coordinates are not stored explicitly but are part of arcs [1].

Coverages in ArcInfo make it possible to determine relationships such as connectivity, area and adjacency [16]. For example, points are contiguous (adjacent) if they share a common arc, or connected if they share a common node [1]. It is also possible to calculate distances between features or to find the nearest feature [1]. The coverage stores line and polygon features topographically in order to optimize storage by decreasing redundancy. It also allows for operations such as overlay [1].

Routes are defined as paths along existing linear features. Based on arcs, they can go from one point to another as well as be disconnected or form circuits [1]. Routes can be created using ArcToolbox and written to a shapefile [4], or if a shapefile containing them already exists it can be converted to a coverage. It is straightforward to convert from a coverage to a shapefile, however conversion from a shapefile to a coverage can require additional processing and involves more limitations [15]. Instructions on how to perform this conversion are found in [18].

When points are entered, the topology must be “built”, the relationships calculated and encoded [12]. An algorithm for building full topology is described in [13], designed specifically for intelligent transport systems and containing tables with data such as “from-node”, “to-node” and area information [13]. Similarly, in ArcInfo, tables contain the Arc ID, from-node, to-node and other information [12]. Traditionally, topology is stored in tabular format but can also be stored in Voronoi diagrams [13].

When dealing with routes, it is important to be able to create, display, edit and analyze them. These are all possible in ArcGIS. Creating routes can be done using ArcToolbox and can be written to a shapefile [4]. They can also be added to an existing map using ArcMap [4]. There are other ways to add routes as well, such as using ArcTools and ArcEdit [1]. It is possible to create simple or complex routes [4] from existing lines.

A route is an ordered collection of sections of arcs that represent a linear feature [1]. An RAT (Route Attribute Table) contains information about the routes, and an SEC (Section Attribute Table) contains information about the sections that make up the routes, for example how far along an arc the section starts and ends (routes do not have to begin and end at nodes) [1]. A “route location” can be a point on a route or a section along one (for example “mile 2 to mile 4”) [4].

ArcInfo enables a process called “dynamic segmentation” [1]. It builds upon the arc-node model to allow for modeling of “events” along a linear feature. The segmented event data is stored in “event tables” and the events can be point, line or continuous [1]. Dynamic segmentation makes use of event tables as well as section and route data. Some examples include modeling the location on a route where certain events occurred, or assigning certain linear attributes to part of an arc [1]. Only one measurement system may be used at a time [4].

Editing routes is another important operation and can be done using ArcEdit [1]. Either routes or sections can be edited, and routes can be added, deleted, realigned, copied etc. [1]. Another option for editing routes is ArcTools, which can also be used to add routes, delete them, split them, change route directionality etc. [1].

Displaying routes on a map is possible using ArcPlot. Routes can be drawn and have text placed along them [1]. Using a “simple display,” points and arcs can be used to display all objects or a subset of them [12]. As well, an overlay can be performed, for example buffering is a common application where this is done [12]. It is also possible to display where events occurred along a route using event data [4]. Using ArcPlot, a map can be made from ArcInfo files. A coverage file is all that is needed for a simple map, but to make a more sophisticated one graphics files, font files and others can be used [1].

Various types of network analyses are usually the goal in the study of routes. Spatial analysis looks at such measures as distance, direction and length, as well as patterns such as path density (the extent to which paths cover an area) or path frequency (the number of paths per unit area) [19]. Vector analysis can be used to measure flow [19].

Topological analysis deals only with connections rather than direction, length etc. [19]. Depending on the application, the goal may be to find a shortest path, an optimal location etc. and there are various algorithms to do this [19]. There are a number of possible measures for use in network analysis. “Degree” refers to the number of edges ending at a node [2]. “Centrality” refers to how important a vertex is for transport in the network [2]. “Distance statistics” measure distances between vertices. “Neighborhoods” refer to the set of all vertices within a certain distance of a particular vertex [2]. “Graph similarity” is a measure of the distance between 2 graphs. It requires a distance metric and can check for differences in the lengths of corresponding paths [2]. There a number of other measures and types of analysis.

An extension to ArcView called “Network Analyst” can be used to perform some network analysis [7]. It can help solve common problems such as finding the best route or closest facility [7]. It can also determine if one location is linked to another (connected), according to [6]. The Arc NETWORK extension to ArcInfo can be used to model and analyze linear networks [1]. For example, it can solve allocation and shortest path problems, trace paths and perform spatial interaction and gravity modeling [1], as well as formulate directions. In addition to the basic ArcGIS functions, scripts can be downloaded [9] or written to enhance functionality. Some examples of these are scripts that calculate the nearest node or Euclidean distances [9]. Routes can be classified in ArcGIS using a CLASS command which allows a user to group them based on a specific item in their attribute table [1].

The arc-node model has limitations for example in the handling of proximity analyses such as neighborhood analysis [1]. As a result, Voronoi diagrams have been viewed as a potential alternative for storage and modeling. According to [1], there is a command in ArcGIS to generate a coverage of Voronoi polygons from a point coverage. Advantages of these diagrams are that they can preserve topological information, represent networks, as well as carry out proximity operations such as nearest-object queries [11]. They can also model data of any density according to [10], and have been used in modeling fluid flow, collision detection, buffer zone modeling and other applications [10].

The “dynamic” version of the model allows for local updates to be made without rebuilding the whole structure [10], allowing for rapid updates which is good for moving data or other situations where updates are frequent. When a point becomes or ceases to be the neighbor of another point, the topology changes and an update occurs [11]. The sequence of updates can be stored as a list of events and can be used to simulate motion [10].

The network can be stored using the “quad-edge” data structure [14] and using add and delete operations to insert or remove points [10]. An algorithm for these is provided by [14], who designed a program to perform inserts and delete operations locally.

Route and network modeling has many applications in a wide range of fields including computer science, geography, traffic planning and surveillance, as well as many others. Because of this, it will likely continue to be an area of intense study as researchers continue to seek faster and more efficient approaches.

References

[1] ArcInfo Help Documentation (ArcDoc), ArcInfo Workstation, Contents. Accessed: 06/28/05.

[2] Bae, Sang Won; Chwa, Kyung-Yong (2004). Voronoi diagrams with a transportation network on the Euclidean plane. Lecture Notes in Computer Science. Algorithms and Computation: 15th International Symposium, ISAAC 2004, Hong Kong, China. December 20-22, 2004. Proceedings.

[3] Brandes, Ulrik; Erlebach, Themas (Eds.) Network Analysis, Methodological Foundations. Berlin: Springer-Verlag, 2005.

[4] Brennan, Patrick (2002). Linear Referencing in ArcGIS. Environmental Systems Research Institute (ESRI).

[5] Cho, George (1995). A self teaching student's manual for Geographic Information Systems (Module 3: GIS: An Introduction). Canberra: University of Canberra and CAUT.

[6] Echols, Nancy (2003). Finding the fastest route – a GIS network analysis project. GISC 4438. April 25, 2003. URL:http://gis.esri.com/library/userconf/proc03/p0320.pdf.

[7] Environmental Systems Research Institute Inc. (ESRI, 1998). ArcView Network Analyst – Optimum Routing, Closest Facility and Service Area Analysis.

[8] Environmental Systems Research Institute Inc. (ESRI, 2004). Overview of ArcGIS Topology, An ESRI White Paper. October, 2004. URL:www.esri.com/software/arcgis/about/literature.html (as of 06/23/05).

[9] ESRI Support Center. URL: http://arcscripts.esri.com (as of 06/27/05).

[10] Gold, C.M. (2000). Chapter 4: An algorithmic approach to marine GIS. In Marine and Coastal Geographic Information Systems (Ed.: Wright, D.; Bartlett, D.). Pg. 37-52.

[11] Gold, Christopher M. (1994). Three approaches to automated topology, and how computational geometry helps. Proceedings, Sixth International Symposium on Spatial Data Handling, 1994. Pg. 145-158.

[12] Haithcoat, Tim. Vector GIS. University of Missouri, Columbia. URL:http://msdisweb.missouri.edu/presentations/intro-to-gis-pdf/Vector.pdf (as of 06/27/05).

[13] He, Chaoying; Jiang, Jie; Han, Gang; Chen, Jun (2004). An algorithm for building full topology. XXth ISPRS Congress, 12-23 July 2004. Istanbul, Turkey. Youth Forum.

[14] Mostavi, M.A. (2003). Dynamic Voronoi/Delaunay methods and applications. Journal of Computers and Geosciences, V. 29, pg. 523-530.

[15] National Ice Charting Working Group. ESRI vector format options. URL:http://nsidc.org/noaa/iicwg/pdf/esri_vector_options.pdf (as of 06/23/05).

[16] Sutton, Paul. Chapter 3: The vector data model. Department of Geography, University of Denver. URL:www.geography.du.edu/sutter/courses/IntroGIS/PowerPoint%20Lectures/Lecture2.ppt (as of 06/23/05).

[17] Turk, T.; Gumusay, M.U. (2004). GIS design and application for tourism. XXth ISPRS Congress, 12-23 July 2004. Istanbul, Turkey. Pg. 485ff.

[18] University of Alberta Sciences – GIS. Calculating a distance matrix of locations along river networks. June 23, 2004. URL:www.biology.ualberta.ca/facilities/gis/uploads/instructions/AVRiverDistanceMatrix .pdf (as of 06/27/05).

[19] Unwin, David (1981). Introductory Spatial Analysis. London: Metheun, 1981. Pg. 67-110.

[20] Watson, Paul (2002). Topology and ORDBMS technology. Technology White Paper, Laser-Scan Ltd. January 2002.