TOPOLOGY IN THE TIGER FILE

Gerard Boudriault
Geography Division
U.S. Bureau of the Census
Washington, D.C. 20233

ABSTRACT

The topological structure in the Topologically Integrated Geographic Encoding and Referencing (TIGER) File is a realization of James Corbett's topological model of two-dimensional feature networks. The TIGER File's topology conforms to the model and adopts specific conventions. Software that maintains conformity to the model is essential.

INTRODUCTION

The Bureau of the Census is developing a geographic information system known as the TIGER System. The first two letters of the acronym stand for the words, Topologically Integrated. It is not for reasons of creating a clever name that Topologically Integrated is included. Topologically integrated is, in fact, the key to the power and beauty of the TIGER File. The topological structure contained in the TIGER File integrates geographic features, ZIP codes, address ranges and the myriad of political and statistical areas into a self-consistent Geographic Encoding and Referencing System. Without the integrating function of the TIGER File's topological structure, the geographic information needed to support the 1990 decennial census would likely be reduced to a disparate set of computer files and traditional maps.

The topological structure in the TIGER File is an application of the theoretical model of two-dimensional networks developed by James Corbett. Corbett's model uses the topological entities called 0, 1 and 2-cells to represent visible and non-visible features on the earth's surface. The model contains a set of topological rules in addition to the entities. This model is generally known in the automated cartography field. The TIGER File incorporates the basic theory of 0, 1 and 2-cells, but as a specific implementation of the theory, the TIGER File design adds unique rules and conventions to the theory.

Software designed for the TIGER File must be capable of maintaining the topological structure and rules or the TIGER File would become invalid. This essential requirement implies the construction of a software library whose algorithms and functions are determined by the model. Strict adherence to the rules is a prerequisite to the successful functioning of the TIGER System as a nationwide geographic information system.

TOPOLOGICAL ENTITIES

Corbett's topological model of features on the earth's surface is based on the representation of linear features, their intersections, and the areas they bound by the mathematical objects-- lines, points and areas respectively. The model specifically originates from the area of mathematics called combinatorial topology which uses the terms Ø-cell, 1-cell and 2-cell in place of point, line and area. Combinatorial topology delineates the sets of both global and local relationships that exist between these cells. The ability to derive the relationships between the cells means a corresponding ability to derive the relationships among the features that the cells represent. Hence the power and dynamic of the model.

1-cell

A 1-cell is defined as a sequence of at least two points where each point has a coordinate in the chosen coordinate space. In addition to the point sequence, the 1-cell implicitly contains the points between (in the sense of analytic geometry) adjacent points in the sequence. In parametric form a 1-cell consists of the set of points,

$$\{X \mid X=P_1, P_2, \ldots P_n \text{ where } P_i <> P_{i-1} \text{ and}$$
$$X=P_1+(P_2-P_1)t,$$
$$X=P_2+(P_3-P_2)t,$$
$$\ldots$$
$$X=P_{n-1}+(P_n-P_{n-1})t \text{ for } 0<t<1\}.$$

The first and last points of the sequence are by definition Ø-cells, the first point is the "from" Ø-cell of the 1-cell and the last point is the "to" Ø-cell of the 1-cell. Every 1-cell is thereby bounded by its from and to Ø-cells. The points of a 1-cell other than its from and to Ø-cells are the curvature points and they serve to trace the shape of the feature identified with the 1-cell. Other terms commonly used for curvature points are shape points, curve points, and intermediate points. The TIGER File design imposes no limit on the number of curvature points that a 1-cell can possess.

A pair of adjacent points and the implied points between them comprise a line segment or vector. Using the concept of vector, a 1-cell may be conceived as a string or chain of vectors. The 1-cell definition's requirement that adjacent points are not equal excludes the legitimacy of a vector having a length of zero. By extension a 1-cell also may not have a length of zero. Zero length vectors and 1-cells are called degenerate and are forbidden in the TIGER File.

A 1-cell for which the from and to Ø-cells are the same is called a loop 1-cell. A loop 1-cell must possess at least two curvature points to form a simple closed curve. Loop 1-cells are legal in a TIGER File and are the most efficient means of representing unconnected lakes and so forth.

The 1-cell definition, it may be noted, does not mention 2-cells. At various points in TIGER File building it is valid to use a topological file whose sole topological entities are the zero-dimensional points and the one-dimensional lines with there being no set of areas coded to the sides of 1-cells. For these unpolygonized files, the Ø-cell and 1-cell entities are completely valid. (For example, the U.S. Geological Survey's 7.5-minute road overlay files were tagged without the benefit of 2-cells.)

Ø-cell

Ø-cells are zero-dimensional objects or points. In TIGER File topology, a Ø-cell is a point in the coordinate space and as such a Ø-cell possesses exactly one coordinate. The TIGER File's coordinate space is the spherical system of longitude and latitude. The coordinates are stored as fixed-point decimal values and have six digits of precision within a degree providing an average precision in absolute distance of 4 inches.

From the 1-cell definition it is known that the endpoints of 1-cells are Ø-cells. The TIGER File design also states the converse, namely that every Ø-cell is the endpoint of at least one 1-cell. This rule, in conjunction with the prohibition on degenerate 1-cells, disallows the use of Ø-cells to represent point features such as mountain peaks and radio towers unless the feature is coincident with a linear feature represented by a 1-cell. Note: The TIGER File design represents such free standing point features in a related landmark structure.

2-cell

2-cells are two-dimensional objects or finite regions of the coordinate space. A point is said to belong to a 2-cell if there is a two-dimensional sphere such that the point is its center and all of its interior points belong to the 2-cell. This ensures that 2-cells are continuous. The extent of a 2-cell is determined by the left and right sides of a set of 1-cells that is analogous to Ø-cells bounding 1-cells.

The complex of 1-cells in a TIGER File covers a finite region of the infinite projective plane. The unbounded region that surrounds the 1-cell complex is always represented by the 2-cell labelled with the number 1.

To depict correctly the many features that are unconnected from the rest of the feature network, the TIGER File design allows for the existence of 1-cell complexes unconnected from the remaining 1-cell complex. Unconnected complexes are called islands. No reason was found for linking an island to the surrounding complex through the addition of a false 1-cell. The TIGER File design also allows a single file to contain more than one primary complex, each surrounded by the unbounded 2-cell 1.

Typically a 1-cell separates two different 2-cells and such a 1-cell is called a boundary 1-cell. A 1-cell that is inside of a 2-cell is called an internal 1-cell and is bounded on both its left and right sides by the same 2-cell.

TOPOLOGICAL RULES

The preceding definitions are not sufficient to characterize the topological and geometric structure contained in the TIGER File. The TIGER File topology is more than simple sets of topological cells; it has rules that organize the sets into a unified consistent structure.

The files are subject to two basic rules. The rule of topological completeness requires that the topological relationships between cells are complete. The rule of topological-geometric consistency requires a consistent relationship between the geometric or coordinate placement of cells and the pure topological relationships of cells.

Topological Completeness

From the cell definitions it is known that 1-cells are related to Ø-cells since every 1-cell is bounded by its from and to Ø-cells and that 1-cells are related to 2-cells since every 1-cell is bounded by its left and right 2-cells. The boundary relationships in a TIGER File are its set of topological incidences. To ensure that the set of topological incidences is complete, a TIGER File must obey the following two rules:

1) The sides of 1-cells incident to a Ø-cell form a cycle.
2) The endpoints of 1-cells bounding a 2-cell form one or more disjoint cycles.

Topological-geometric Consistency

The rule of topological and geometric consistency states that the collection of topological cells must have coordinates that make the collection a disjoint partitioning of the coordinate space. The consistency rule may be decomposed into four conditions that must each be true of an entire TIGER File. The conditions are:

1) No two members of the combined set of Ø-cells and curvature points share the same coordinate.
2) No two vector interiors share a common coordinate.
3) No two 2-cell interiors share a common coordinate.
4) One cycle of a 2-cell's bounding 1-cells has an signed area greater than zero and surrounds the other cycles that have signed areas less than zero. The exception to this condition is 2-cell 1 that has one or more cycles each with an signed area less than zero.

IMPLICATIONS FOR SOFTWARE

At any given time, a TIGER File's topology necessarily conforms to the topological definitions and rules. Therefore all topological changes applied to a TIGER File must result in a file that still is valid topologically. The need to maintain topological completeness and topological-geometric consistency dictates the modus operandi of the software that manages topological changes.

The topology management software can be constructed as a library of functions that operate on cells. These are the functions normally found in automated systems such as add, delete, find, and move. Since the objects of the functions are topological cells or even parts of cells, examples of specific functions performed on a TIGER File would be add 1-cell, delete Ø-cell, move curvature point. The functions would be combined to form complete topological actions. For instance, the complete action of deleting a 1-cell requires as sub-functions the merge 2-cell function to fuse the two 2-cells that a boundary 1-cell separates and the delete Ø-cell function to eliminate a Ø-cell that no longer has incident 1-cells.

The maintenance of TIGER File topology must be through software alone. There must be no human component or function in the system that maintains the topological structure. The only function of a TIGER File user is that of retrieving and modifying the feature network, not the labelling of 1-cell sides or some such operation.

Linear Feature Insertion

The maintenance of topological consistency by software alone may be illustrated through a high-level description of an algorithm that inserts a new linear feature into a TIGER File. The algorithm's domain is the transformation of a chain of coordinate points into one or more new 1-cells resident in a TIGER File. It is not concerned with the digitization of the feature or with the attachment of attribute data after insertion.

The chain of coordinates is first edited to remove or modify illegal and undesirable portions of the chain. The edit must at least detect and resolve self-intersections within the chain. Self-intersections are resolved by the addition of points at positions in the chain where intersections exist. The new points must be marked as intersection points. Degenerate vectors must be eliminated. Optional modules of the edit are a thinning algorithm to reduce the number of superfluous points and algorithms to eliminate small knots and 'z' shapes. After completion of the chain editing, the original chain has been decomposed into substrings each of which is atomic relative to the entire chain.

The next module in linear feature insertion is the detection of all intersections between the new atomic chains of coordinates and 1-cells in the TIGER File. Where the new feature intersects the interior of a 1-cell, the 1-cell must be split into two 1-cells both ending at the intersection point. Likewise, where a 1-cell or Ø-cell intersects the interior of a chain, the chain must be split by adding the intersection point to the chain and marking it as an intersection point. Once the detection and resolution of intersections is complete, both the TIGER

File and the new linear feature reflect the intersections through the splitting of 1-cells and the addition of intersection points.

A tolerance distance should be used in the determination of intersections in addition to strictly solving simultaneous linear equations. The tolerance enables the snapping of the new feature to nearby Ø-cells and 1-cells.

The remaining module is the insertion of each atomic chain into the appropriate 2-cell. The atomic chain insertion requires an analysis of the 2-cell's 1-cell cycles in regards to the incidence of the new chain. A 2-cell labelled A is split by an atomic chain into two 2-cells, A and B, if and only if the chain connects Ø-cells on the same cycle or if the chain is a loop. When the chain is a loop the outward facing side is linked to the cycle of 2-cell A it is connected to or is added as a new cycle of 2-cell A if it is an unconnected island. When the new chain connects two Ø-cells on the same cycle, the cycle of 1-cell sides is correspondingly split at the two Ø-cells. Half of the cycle is attached to 2-cell A along with the consistent side of the new chain. The other half of the split 2-cell's cycle is completed by the other side of the new chain and is assigned to 2-cell B. To complete the 2-cell split, islands must be assigned to one or the other 2-cells. Using a point-in-polygon routine each of the cycles not split by the new chain must be determined to be inside 2-cell A or B. The point-in-polygon routine should also be used to assign point landmarks to the appropriate 2-cell. Since the child 2-cells only occupy area occupied by the parent 2-cell, they are linked to the different geographic areas and area landmarks that were associated with the original 2-cell.

REFERENCES

Corbett, James P., 1979, Topological Principles in Cartography, Technical Paper 48. Washington D.C.: Bureau of the Census.

White, Marvin S., Jr., 1984, Technical Requirements and Standards for a Multipurpose Geographic Data System: The American Cartographer, vol. 11, no. 1, pp. 15-26.