POLYGONIZATION AND TOPOLOGICAL EDITING AT THE BUREAU OF THE CENSUS

DAVID MEIXLER
ALAN J. SAALFELD
BUREAU OF THE CENSUS
WASHINGTON, D.C. 20233

ABSTRACT

In 1983 and 1984, the Bureau of the Census developed a computer program to polygonize digital map data (organize linear feature information into polygons) and to validate the topological and geometric correctness of the nodes, chains and polygons. The program evolved from an earlier planar sweep program, and eliminated many of its geometric dependencies. The program runs in both an insertion and edit mode. In the insertion mode, the program is used to generate the elementary topological polygons for over 50,000 maps covering nearly 3,000,000 square miles of the contiguous 48 states. This insertion process is a necessary component in the joint United States Geological Survey-Bureau of the Census project to produce the National Digital Cartographic Data Base. At critical points in the work flow, the edit mode is used to verify the underlying topological soundness of the file structure.

INTRODUCTION

The TOPOLY program evolved from an earlier planar sweep program and still retains some of its characteristics. A planar sweep program labels each unique polygon in a graph by visiting each node in the order it would be visited by a line sweeping across the graph [Nievergelt and Preparata, 1982].

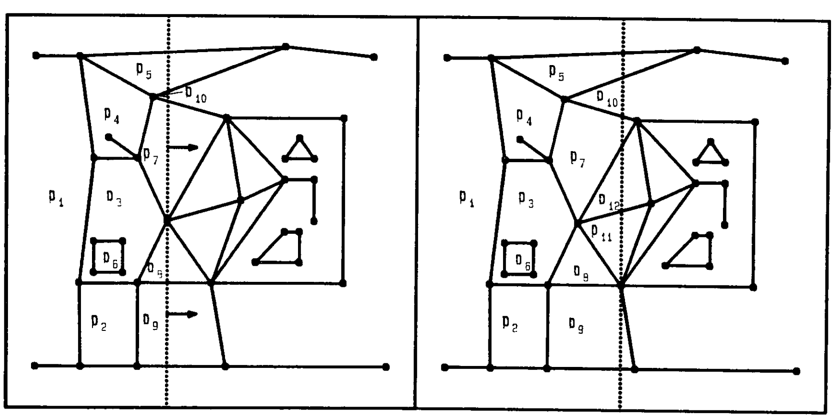

FIGURE 1 PLANAR SWEEP ALGORITHM ILLUSTRATED

When at a node, the algorithm will ensure that any new polygon labels placed upon the graph are consistent with existing labels previously applied. The planar sweep algorithm keeps track of the regions currently under this line and can also check for undiscovered intersections of the lines or resolve discontiguous parts of the graph (islands) that are uncovered as the line sweeps over the graph.

The TOPOLY program is more efficient with the Census Bureau file structure and needs no overhead for keeping track of the regions under the sweep line. Since intersection checks are done in a separate process at the Census Bureau, an intersection testing capability was not needed in the polygonization process. However, when the Census Bureau staff adopted the unordered choice of nodes on which to apply consistent polygon labels, they still needed processes to detect and resolve the disconnected complexes in the graph. The detection process fit with other polygon analyses being done. The resolution of the disconnected pieces became a new process. The program thus has four parts: the initial label placement on the sides of the chains; the analysis of the resulting polygons; the resolution of the discovered islands; and the modified Euler edit at the end of processing.

POLYGON LABELING

The insertion of polygon labels in the graph is a simple and fast process. All the nodes in the file are visited in whatever order they are stored in the file. The chains (1-cells) attached to them are collected and ordered by angle of emanation from the node; this ensures that the facing sides of adjacent chains are labeled consistently.

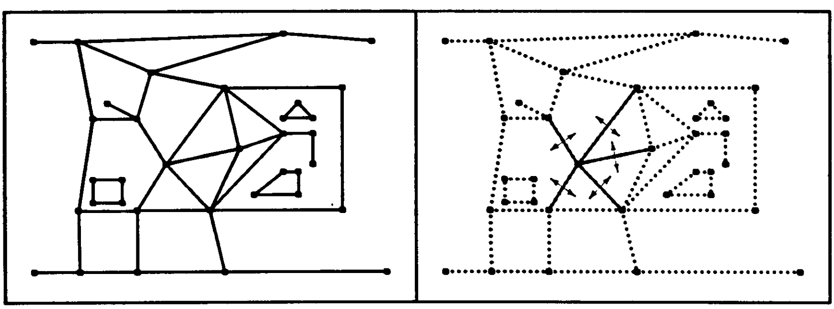

FIGURE 2 UMBRELLA EDIT
COLLECTING CHAINS ABOUT A NODE

If no label exists between the chains, a new label is created and inserted on both. If the labels are different, one of the labels is preserved and the other deleted. All chains having the deleted label are reassigned the preserved label. If the labels are already consistent, the next pair of chains is visited. These are the same rules applied in a planar sweep process. The important difference is that the nodes are visited independently of their position. Thus the program is called TOPOLY, for TOPOlogical POLYgonization.

The labeling process and the umbrella edit are equivalent. When running in edit mode, this section of the program reverts to an umbrella edit [White, 1984]. This edit examines the chains emanating from every node to ensure that the polygon labels on facing sides of the chains are consistent. For example, if angles of emanation of the chains increase when measured counter-clockwise, then the left polygon label on one chain should agree with the right polygon label on the next chain. Similarly, the left label on this next chain must agree with the right label on the subsequent chain. This edit treats each node as the hub of an umbrella and tries to ensure that the order of the emanating spokes (chains) is consistent with the labels of the area between them.

If this edit is done on every node of the graph, and each node is consistent, then the graph is guaranteed to be labeled consistently. However, this consistency of labeling does not guarantee a topologically sound file. The most obvious case of an unsound file would be the case where the entire file has the same polygon label. Assuming the graph has at least one cycle, then there must be more than one polygon. However, the umbrella edit will not detect two atomic polygons with the same label. It will detect the lack of consistency around individual nodes. Thus failure of the umbrella edits insures that the file is unsound, but passing it does not insure topological consistency.

POLYGON ANALYSIS

The major process in the polygon analysis phase is the Kirchoff routine. A Kirchoff analysis is done separately on each individual polygon in a file. The basic Kirchoff procedure counts the number of cycles and acycles in a set of chains [Prather, 1976]. In a complex graph, the number of independent cycles is a fixed number N. However, the N cycles themselves, i.e. the chains that constitute them, are not uniquely determined. In other words, there may be many sets of N independent cycles for a particular complex graph.

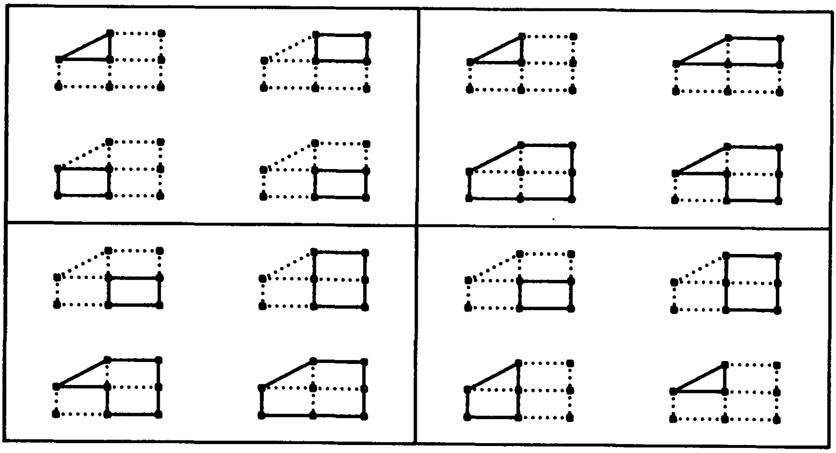

FIGURE 3 SETS OF INDEPENDENT CYCLES
OF A COMPLEX GRAPH

For a topologically sound polygon, the cycles themselves are determined uniquely. In fact, every boundary chain belongs to a cycle. Nonboundary (internal) chains do not form cycles. Additionally, all boundary cycles are independent. One and only one cycle may be recognized as the outer boundary of that polygon. Any other cycles will constitute "inner boundaries", i.e. the chains that surround a "hole" in the polygon.

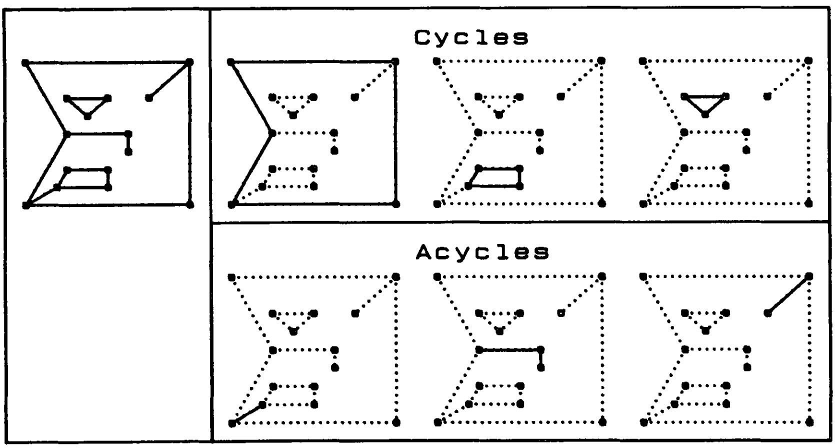

FIGURE 4 UNIQUE GENERATING CYCLES AND ACYCLES
OF A TOPOLOGICALLY SOUND POLYGON

All of the chains with the same polygon label are submitted to a Kirchoff routine, and its analysis is the basis for confirming the integrity of the polygon. For example, internal (nonboundary) chains on a polygon are known from the fact that the left and right polygon labels for each of these internal (non-boundary) chains on a polygon are known from the fact that the left and right polygon for each of these chains are the same. Similarly, all boundary chains are known by the fact that the cobounding polygons have different labels. A Kirchoff routine analyzing this set of chains determines which chains form cycles and which are acyclic. All chains in the acycles are confirmed to be internal and all chains forming cycles are guaranteed to be boundary chains. The cycles are ordered, coordinates are extracted, and the information is sent to a routine that computes the area, perimeter, centroid, shape, and direction of traversal in one computation.

Orientation is used to identify inner and outer boundary cycles. The direction (either clockwise or counter-clockwise) of the cycle is compared with the side on which the polygon lies. When walking around the outer boundary of a polygon in a clockwise fashion, then the polygon must be on the right side. Similarly, a counter-clockwise traversal around an inner boundary keeps the polygon on the right side. If the direction changes, then the sides reverse. Thus to walk counter-clockwise around an outer boundary, keep the polygon on the left. The cycles returned by a Kirchoff analysis are examined to ensure consistent labeling on a specific side. The direction of the cycles also is examined to insure that there is one and only one outer boundary and that the area it encompasses is greater than that of all inner boundaries.

Another modification of the Kirchoff routine counts the number of discontiguous components in the chains associated with each polygon. This normally will be one. However, some polygons may have more then one component. The number of separate components in the graph is summed to be used later for the modified Euler edit described below. Also, the disconnected parts created in the initial labeling are recognized for the following island resolution process.

ISLAND RESOLUTION

During the labeling process, disconnected components will acquire a separate set of polygon labels. Except for the outer regions of the separate components, these labels are valid polygons with an outer boundary and zero or more inner boundaries. These valid labels can be ignored for island resolution. However, the outer region of each component must be resolved to agree with the other components of the graph. The polygon labels that form the outer boundary of a component of the graph will form what Kirchoff recognizes as an inner boundary and will have no outer boundary. The region encompassed by this boundary is sometimes referred to as a "hole" or island in the main component. The largest of these islands is recognized as the outer region of the entire map. Recognizing and labeling this area as the "first polygon" in the file is a requirement of the joint agreement with the U.S. Geological Survey.

Once this island is detected it is resolved to agree with the main component of the graph. If an intersection check has been performed already, then this island will fit entirely within a single polygon of the main component. The resolution itself is made by resolving a single point of the region to be within one polygon of the main component. This resolution is done starting with any polygon and chaining adjoining polygons to find the correct one. Wherever this point falls will have the entire outer boundary recoded to it. Island resolution is only necessary during the insertion mode. Its edit mode equivalent ensures that disconnected complexes are labeled to agree with the surrounding polygon. This is more convenient to do in the polygon analysis phase.

EULER EDIT

The final edit done on the file before the program terminates in both the insertion and the edit mode is a modified Euler edit. The Euler edit is based upon simple principles. If a single line chain is drawn and examined, it is obvious that there is a single chain, two endpoints and one surrounding region. In this example, the number of nodes and regions equals the number of chains, plus two. The Euler Theorem states that this numerical relationship is always true for a connected set of chains on a plane. This may be illustrated by iterative chain building. When a new chain is joined to an existing network of chains, one endpoint must begin at an existing node. If the other end of the added chain goes to a node that already exists, it will create a new bounded polygon inside an existing polygon or in the outer region. If the other end of an added chain does not link up with an existing node, it creates a new endpoint.

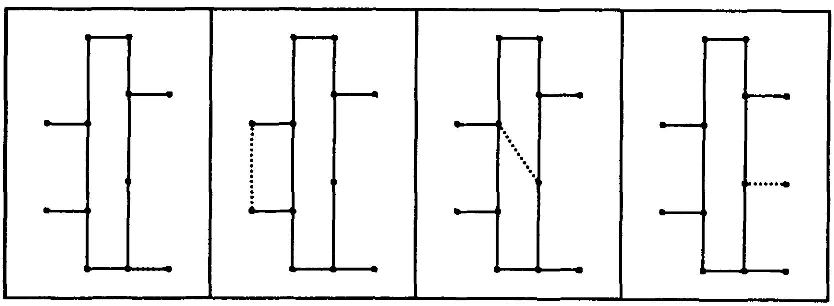

FIGURE 5 OPTIONS FOR ADDING A CHAIN
TO AN EXISITING CONNECTED NETWORK

Thus, after the basic relationships among the node counts, chain counts, and polygon counts are established, that relationship will remain constant as long as new chains are attached to the existing network of chains.

However, a cartographic database need not have all chains connected. This leads to the necessity to modify the basic Euler formula. It is obvious that each disconnected network of chains will obey the basic Euler formula. However, in such a complex, the outer boundary of each disconnected network must be properly related to some region of another graph. Either the network will be entirely internal to one polygon of the other graph (its chains can not cross those of the other graph) or it will share in the outer region of that other graph. The basic Euler formula is modified to take this into account. The modified Euler formula is stated as, "the total number of nodes and polygons must equal the total number of discrete components in the graph plus the number of chains in the graph plus one."

The modified Euler edit can be used on any planar graph. Although it does not ensure overall accuracy of the file, it is a good test of consistency among the counts of nodes, chains and polygons of the file. A failure of the modified Euler test is a guarantee that the file is not topologically consistent. However, passing this edit is no guarantee that the file is sound. Compensating errors can still exist that would allow the counts to balance.

CONCLUSION

This program is used in the exchange of digital map files with the U.S. Geological Survey. The polygonizing of the files is one of the first processes done to a file when it is received. After the Census Bureau performs the internal updates to the file, the TOPOLY program is run in edit mode. In this mode, it is one of the last processes run before returning the updated files to the U.S. Geological Survey. To date, the program has been run on over half of the 50,000 7.5 minute quadrangles involved in building the National Digital Cartographic Data Base. The program has been converted successfully to run on the most recent TIGER file structure of the Census Bureau and it is anticipated to be run in the edit mode for years to come.

REFERENCES

Corbett, James P., 1979, Topological Principles of Cartography, Technical Paper 48, Bureau of the Census, Department of Commerce

Lefschetz, Solomon, 1975, Applications of Algebraic Topology , Springer-Verlag, New York

Nievergelt, A. and Preparata, F.P., 1982, "Plane-Sweep Algorithms for Intersecting Geometric Figures", Communications of the ACM, Vol 25 No 10.

Prather, Ronald E., 1976, Discrete Mathematical Structures for Computer Science, Houghton Mifflin Company, Boston

White, Marvin, 1984, "Technical Requirements and Standards for a Multipurpose Geographic Data System", The American Cartographer, Vol 11 No 1.