Cartography and Geographic Information Systems
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcag19

# Use of Graph Theory to Support Map Generalization

William A. Mackaness & Kate M. Beard
Published online: 14 Mar 2013.

To cite this article: William A. Mackaness & Kate M. Beard (1993) Use of Graph Theory to Support Map Generalization, Cartography and Geographic Information Systems, 20:4, 210-221

To link to this article: http://dx.doi.org/10.1559/152304093782637479

# Use of Graph Theory to Support Map Generalization

**William A. Mackaness and M. Kate Beard**

**ABSTRACT.** *In the generalization of a concept, we seek to preserve the essential characteristics and behavior of objects. In map generalization, the appropriate selection and application of procedures (such as merging, exaggeration, and selection) require information at the geometric, attribute, and topological levels. This article highlights the potential of graph theoretic representations in providing the topological information necessary for the efficient and effective application of specific generalization procedures. Besides ease of algebraic manipulation, the principal benefit of a graph theoretic approach is the ability to detect and thus preserve topological characteristics of map objects such as isolation, adjacency, and connectivity. While it is true that topologically based systems have been developed for consistency checking and error detection during editing, this article emphasizes the benefits from a map-generalization perspective. Examples are given with respect to specific generalization procedures and are summarized as a partial set of rules for potential inclusion in a cartographic knowledge-based system.*

**KEY WORDS:** *automated map generalization, graph theory, network generalization*



## Introduction

There is a tendency to view generalization as a set of procedures (for example, simplify, omit, exaggerate, and displace) that are applied depending on design constraints such as scale, aesthetics, content, and theme. But the subtleties of map design mean that generalization techniques rarely are applied wholesale to map objects. Their application varies both within classes of objects (for example, roads will be generalized differently according to their classification) and between different classes of objects (for example, roads are not generalized in the same way as rivers). This reflects the fact that objects have their own behavior, with a response that varies according to map task, and this response dictates design parameters such as scale, theme, symbology, and contextual content. The spatial interplay among objects makes map design a complex decision-making process, and as was argued by Mackaness (1991), a complete knowledge of the map objects is required for successful design; not least of these is knowledge of the role played by each feature: Does the feature provide a contextual setting, and does it connect, divide, or include other objects? Is it unique or isolated, or is it representative of other features? All these characteristics govern a feature's likely inclusion in or exclusion from a map. Inclusion is also a compromise among many competing factors, such as their relative importance, the size of the screen or paper on which the map is displayed, and the user's knowledge and powers of interpretation. To interpret a map, users bring their understanding of the world around them—the notion that mountain chains impede, continents have shape, rivers flow, airports connect, and cities congest; cartographers use this understanding when designing maps.

Though the application of automated generalization procedures is based principally on an analysis of space and attribute, a map is as much about the communication of spatial concepts. What determines good design (and the "successful" application of various generalization techniques) is the success with which the nature of space is conveyed and how a graphic helps in its interpretation: the ordering, distinction, comparison, combination, and recognition of relationships. This article seeks to highlight the role of topological concepts in map generalization and shows through the use of graph theory how concepts such as isolation, connectivity, adjacency, and neighborhood proximity influence the application of certain generalization operations. Map generalization is known to be complex, and its application to each map requires understanding and modeling at the geometric, topological, and attribute levels. For the purpose of analytic cartography, the aim of graph analysis is to characterize the objects represented by the graph.

This article begins with a brief review of the graph theory relevant to analytical cartography, and by examining some examples shows how it can be used to generalize networks without disconnecting objects, to measure content in a map, and to select information for inclusion in strip maps. The influence of topological concepts on network generalization are formalized as a partial set of rules.

*William Mackaness is a research assistant professor for the National Center for Geographic Information and Analysis (NCGIA) in the Department of Surveying Engineering, University of Maine, ME 04469. Kate Beard is an associate professor of surveying engineering also carrying out research for the NCGIA.*

## A Brief Review of Graph Theory

Graph theory enables one to characterize topological relationships among objects and has been used to solve such problems as minimum vertex coloring, modeling Markov chains, labyrinths (Konig 1990), enumeration of chemical molecules, network flows, scheduling, and critical path analysis. This section provides a review of graph theory terminology relevant to map generalization. For a complete

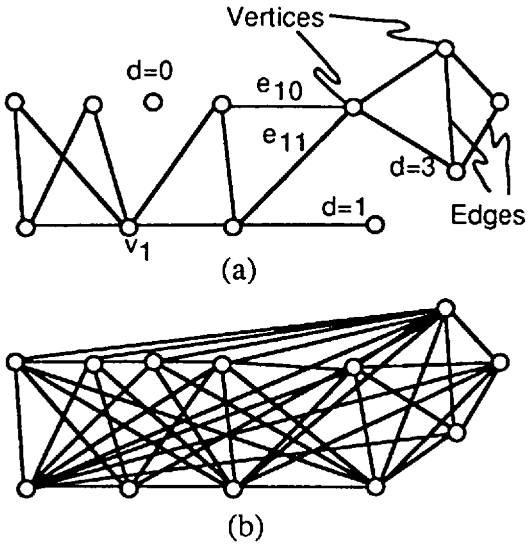


**Figure 1.** *Examples of completeness.*



discussion, see Chartrand (1977) and Wilson (1979). A graph consists of $n$ vertices and $m$ edges. The degree of a vertex is the number of incident edges (for illustration, the degrees of a few of the vertices of Figure 1 are shown). A graph is isomorphic if the vertices have the same degree and the same vertices are incident to one another (Figure 2). Thus a graph can appear visually to differ from the network it represents.

A disconnected set of a connected graph G is a set of edges ($e_i$, $i = 1, m$) whose removal disconnects a graph. In Figure 1*a*, the disconnected set is $e_{10}$ and $e_{11}$. A separating set is similar to a disconnected set but applies to vertices ($v_i$, $i = 1, n$); if the vertices of a separating set are removed, then the graph becomes disconnected. If a separating set contains only one vertex, it is called the cut-vertex or articulation vertex. In Figure 1*a*, $v_1$ is an example of an articulation vertex. A vertex with no edges incident is called an isolated vertex, and has degree zero. If a connected graph contains a cycle (it is possible to visit the same vertex from different adjacent vertices), then removing an edge from the cycle does not disconnect the graph. If all cycles are removed, the graph has a treelike structure and is acyclic.

## Directional and Weighted Graphs

A directional graph (often called a digraph) consists of a set of directed edges, an example of which is given in Figure 3*a*. Digraphs are necessary to represent the movement of transactions between departments (Chartrand 1977), or one-way roads in a city road network.

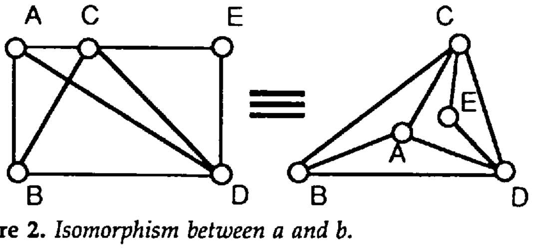


**Figure 2.** *Isomorphism between a and b.*

Graphs can be weighted to reflect the varying emphasis on the relationships between objects and sets of objects. Weights can be attributes of either the vertices or edges in either directed or undirected graphs, and this enables the modeling of networks that have capacities. In transport networks, the weights associated with the edges typically might be distance or travel time.

## Graph Completeness and Other Descriptions

A graph is said to be complete if every pair of vertices is adjacent. It is possible to calculate the number of edges of a complete graph if the number of vertices is known. For vertices, the number of edges is exactly $[n(n - 1)]/2$. Such an index can help describe the level of connectivity in a graph. Consider Figure 1*a*, which has 16 edges. If it were complete, according to the formula above, it would have 66 edges; thus it is 16/66 or 26% complete. Besides being an indicator of connective redundancy in a network, the percentage of completeness is also an indicator of network content. For example, the graph in Figure 1*b* is 66% complete, with 44 edges. Completeness is not intended as the sole measure of content in networks, but when combined with other content predicting techniques such as Topfer and Pillewizer's (1966) Radical Law, which computes the reduced number of objects based on changing scales, completeness can be used to gauge changes in ink space.

In addition to the measure of completeness, there are a number of ways of describing the attributes and characteristics (Cliff et al. 1979) and the complexity (Temperley 1981) of graphs. These descriptions can be used to identify features with important attributes and to indicate the amount of space devoted to representing network features. The number of edges in a graph is called its size, and the number of vertices is its order. The diameter of a graph is the maximum distance between any two vertices (some of these characteristics are illustrated in Figure 3*b*). The eccentricity of a vertex is the distance from that vertex to the farthest vertex. This definition can be used to describe the center of a graph as being the vertex with minimum eccentricity. It is also possible to define the radius of a graph and its capacity (Bollobas 1979). Using such descriptions, it is possible to model features and their relationships and to generalize their properties. Such descriptions afford a means of classifying objects in order to symbolize them. For example, urban centers are characterized by attributes such as population, political function, and areal extent. Graph theory can provide additional information, such as the thresholds of suburbia (Borchert 1961), degree of connectivity, and network complexity.

## Graph Theory in Geography and Cartography

Graph theory has been applied to many aspects of geographical analysis and many people are familiar with the problem of the bridges of Konigsberg (Hartsfield and Ringel 1990). Graph theory has been used by locational theorists, such as Kohl and Christaller, transport planners (James et al. 1970; Haggett 1990), in studies of contiguity (Richardson 1961), in urban growth models (Taaffe et al. 1963), and in the characterization of urban settlement patterns. For example, Borchert (1961) was able to identify the fuzzy urban-rural subdivision by measuring the characteristics of road patterns using graph theory. The relevance of topological models to editing and consistency checking in automated cartography has also been well demonstrated (White 1978). For example, ARITHMICON, a computer system intended to be a development tool for computerized cartography (Corbett 1979), was based on a topological cellular model (Switzer 1975). Indeed, according to Corbett (1979, p. 18), "every procedure of ARITHMICON depends in some way on the analysis of a characteristic graph." The TIGRIS system, also developed around a cellular topology, was designed for "interactive edit, query and spatial analysis" (Herring 1987, p. 291) of large databases in which all features derive their spatial extent through relationships to topological objects. Similarly, Williams (1985) describes a system to encode and interrogate efficiently the infrastructure over large geographic areas. However, none of these systems has been used from the perspective of automated map generalization. In the following six sections, we look at a variety of ways that topological characteristics can be used to select map objects and to preserve topological characteristics during map generalization.

## Weighted Graphs Used to Guide the Generalization of Networks

Using weighted graphs, it is possible to generate minimum trees that replicate some of the behavior of a geographic feature. For a given set of map objects, weights can be assigned to each edge of a graph to indicate factors such as distance between features or capacities. Indeed, each edge might have a number of weights depending on the number of attributes associated with the edge.

For example, Figure 4*a*, depicts part of a stream network; the numbers have been assigned on the basis of Strahler's (1960) stream-ordering principle. Alternatively, Horton's stream ordering could have been used, and some authors have expressed this ordering in preference to Strahler's (Rusak et al. 1990). Such weighting has immediate and apparent use in network generalization (Rusak Mazur and Castner 1990) as well as in symbolization (pen thickness according to stream-order number is illustrated in both parts of Figures 4). Figure 4*b* shows the same network but with only stream orders 3 and 4.

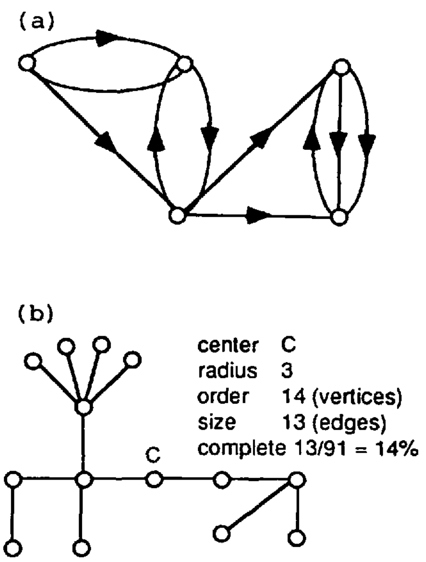


**Figure 3.** *Example of a directional graph and a description of a graph.*

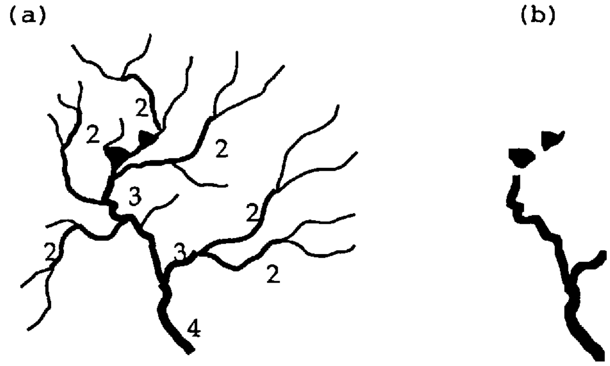


**Figure 4.** *Weights assigned to a stream network by using Strahler's stream-ordering principle.*

Numerous researchers have highlighted problems such as the one illustrated in Figure 4*b*, where the lakes have become detached from the network because part of the core path comprises a section of lower stream order. Most cartographers would extend the line to catch the lakes. This can be achieved similarly through the application of graph theory by classifying geographical objects according to graph theoretic attributes (Paiva et al. 1992). In general, it has been observed that most natural lakes are connected to river networks. From graph theory, it is known that a vertex with degree zero is an isolated vertex; thus it is possible to formulate the following graph theoretic rule: IF a lake prior to generalization has degree greater than zero THEN after generalization the degree should remain greater than zero.

By inspecting the graph before and after generalization, it is possible to identify lakes that have become isolated as well as the disconnected set (the set of edges) that isolated the lakes. Given the graph for the stream network of Figure 4*a*, the following instruction, which reflects the more subtle application of a generalization technique, can be given: Remove all edges with a weight less than 3 unless that edge is the disconnecting set for vertices labeled "lake."

Such a rule (above) applies only to a range of scales. It could be further refined to control, for example, the elimination of lakes as an alternative to retaining the disconnected edges.

## Hierarchical Generalization of Road Networks

Figure 5*a* shows a weighted (nondirectional) finite graph of an urban area at 1:50,000 (Figure 5*b*). Weights were assigned hierarchically based on the road classification, although traffic loadings or other road attributes could be used as the basis of the hierarchical assignment. The graph (which in this example was derived from the map at the 1:50,000 scale) has a similar orientation and shape to help the reader interpret the weight assignment for each section of road. Some vertices have been added at the periphery to make the graph finite. The weights given to both the edges and the vertices (indicated in the key) enable one to

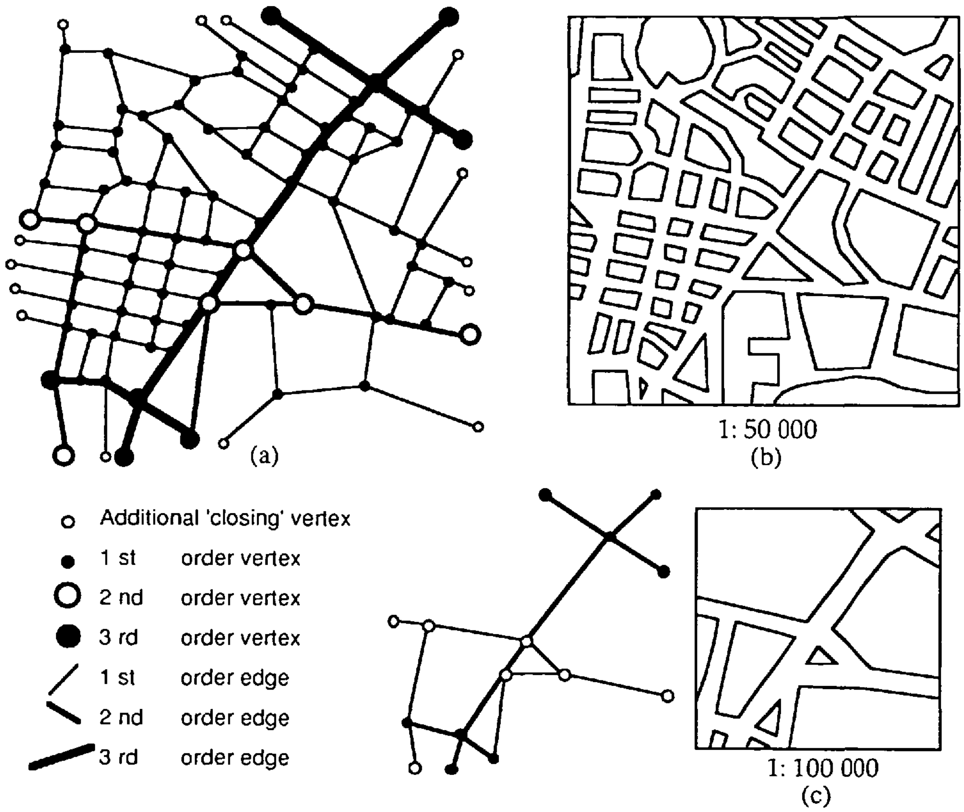


**Figure 5.** *Graph theoretical representations of the example by Keates (1989, p. 46).*

subsequently generalize the network to 1:100,000 (Figure 5c). Observe that both the vertices and edges have been hierarchically weighted. This provides two ways of selecting content at different scales, which are discussed below.

The first method of generalizing this network would be to identify key intersections, to derive the disconnected set of edges that join those vertices, and to have those edges drawn on the map. In this example, because of the high degree of completeness, the disconnected set would present alternative routes (cycles) between any two vertices. Therefore it would still be necessary to weight the edges in order to select from a choice of edges connecting the same vertices. A second solution consists of first selecting the key edges (rather than vertices) that one wishes to retain (for example, a major road bisecting a city) and then removing auxiliary vertices. An auxiliary vertex (having in-degree and out-degree equal to 1) is where just one road joins another. In graph theory, an auxiliary vertex can be defined as one having a degree of exactly 2 with the edges incident to a vertex of the same type. If this is true, the vertex can be removed, and the edges can be seamed together, as demonstrated by White (1978) in ARITHMICON. Therefore the rule base might contain this: IF a vertex has degree 2 AND the incident edges are of the same type, THEN remove the vertex.

Since in this example the vertices merely represent road intersections, the latter choice is more appropriate. As a final observation, it is interesting to look at the degree of completeness for the two maps in Figure 5. Table 1 shows that the more generalized maps are more complete. This is due to the sharp reduction in the number of road intersections at the decreased scale. There are, therefore, proportionally fewer vertices and consequently fewer edges required to make the graph complete (shown in the fourth column of Table 1).

## Applying Minimum Spanning Trees to Generalize Networks

Minimum spanning tree (MST) algorithms can be used to generalize networks represented in the form of a finite weighted graph and are, therefore, applicable to any set of geographical objects comprising a network. There are numerous articles on minimum spanning tree algorithms (Tarjan 1983) as well as articles that specifically discuss and implement Kruskal's algorithm (Sedgewick 1992). The minimum weight spanning tree of a connected graph connects all vertices by using the least number of edges of the lowest weights (Aho et al. 1974); thus the minimum spanning tree is acyclic. Figure 6*b* shows the result of applying Kruskal's algorithm to the graph in Figure 6*a*.

A cartographic assumption is that at small scales, there is not enough space to show every class of road. Never-

**Table 1.** Degree of Completeness of the Two Maps in Figure 5.

| Scale | Vertices | Edges | $1/2\, n\, (n - 1)$ | Complete |
|---|---|---|---|---|
| 1:50,000 | 100 | 139 | 4,950 | 3% |
| 1:100,000 | 15 | 16 | 105 | 16% |

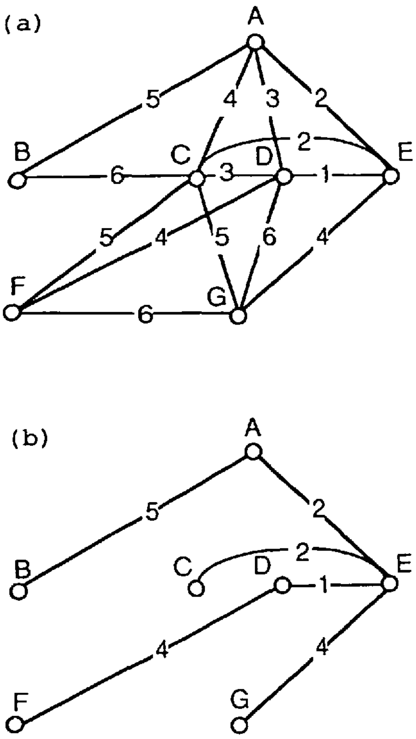


**Figure 6.** *Figure b is the minimum weight spanning tree of the connected graph in figure a. Figures a and b are 38% and 18% complete, respectively.*

theless, the need to convey the notion of connectivity between features, such as cities, remains a very important objective. For example, in a graph representing a road network, the edges might be weighted proportionally to travel time between major cities. The shorter the travel time, the lower the weight. The advantage of reducing map content using an MST approach is that irrespective of the magnitude or variance in weights across the tree, no vertex becomes isolated (of degree zero). If a town is connected solely by a minor road (therefore, it has a very high relative weight), then that minor road will be retained when using this algorithm, thereby preserving the connectivity of that town with other towns. A connected graph in Figure 7*a* for part of a road network (Figure 7*c*) has been created and weighted according to Euclidean distance between towns. The minimum spanning tree has been derived (Figure 7*b*) and used to select road network content for a smaller map scale (Figure 7*d*). Figure 7*e* was generated using the same algorithm but with travel time rather than distance as the weight. A marginally different result is obtained, one that reflects travel via faster roads (e.g., Interstate 44 between Springfield and St. Louis, which is faster than Highway 60 connecting Poplar Bluff with Springfield). The source of both distance and time data is Rand McNally (1991).

Note that the degree of completeness also has been calculated. In this instance, the number of vertices has not changed, so the completeness values of Figures 7*c* and 7*d* are directly comparable (from 38% to 18% complete). A large percentage difference could indicate that the notion of connectivity had been partially lost. The graph could be revisited, and edges of lower weights could be added until the percentage difference fell within some predefined tolerance. As stated previously, feature inclusion on a map is a compromise between many competing factors, and it is not a simple decision based solely on scale, or a decision that one feature type stays at the expense of another. In the example above, it is possible to vary the content from 38% complete to 18% with the knowledge that connectivity has been retained. Furthermore, a need to reduce the degree of completeness below 18% indicates a need also to reduce the number of vertex features (in this case, cities or towns). Thus content reduction becomes a bipartisan affair and reflects the interdependence that exists among features of different types in the real world. The application of MSTs provides a method of selecting components of networks for removal while retaining the connectivity of map objects. It thus overcomes one of the criticisms of the Radical Law. The Radical Law specifies amounts of data according to scale, but it fails to identify which elements should be discarded (Maling 1966 and Robinson et al. 1984).

## Using Directed Graph Information to Generalize Networks

Information associated with directed graphs also can be used to finely tune the generalization of networks. For example, it is possible to identify the shortest paths through directed graphs (Tarjan 1983), and such an approach can be used to create strip maps between specified vertices. It also is possible to generalize directed graphs through edge removal. For example, Figure 8 illustrates how graph theory can be used to generalize a network that has direction. It is known that if a connected graph contains a cycle, removing an edge from the cycle will not disconnect the graph. Thus a rule can be formalized that allows the removal of less important edges (such as the cul-de-sac in Figure 8*a*). But deciding which edge of a cycle to remove may not be easy if, as in this case, the edges are of equal weight. However, having identified the disconnected set and having assumed the direction of the edges is known, a test for connectivity can be performed after any one edge is removed, thereby reducing the number of edges but maintaining connectivity.

Graph theory can be used to help arrive at Figure 8*c* by collapsing edges. In graph theory, the collapse of edges is referred to as the contraction of a line (Temperley 1981), and the equivalent process in map generalization is referred to as merging (McMaster and Shea 1992). Generating maps from contracted graphs would not be a problem if the data were stored at high resolution (Figure 8*a*) and lower levels of detail were thus derived. It would still be necessary to have an algorithm that calculated the position of the new edge; this may not be as simple as taking the mean position of two edges. Furthermore, the algorithm would need to be able to generalize the topology of each of the two edges to one.

Finally, note that in this instance, the small cul-de-sac (which can be identified as the disconnected set of a vertex of degree 1) is not critical to the preservation of overall connectivity and has, therefore, been removed. In a later example, a similar identification technique is used to preserve features of this type rather than delete them; in

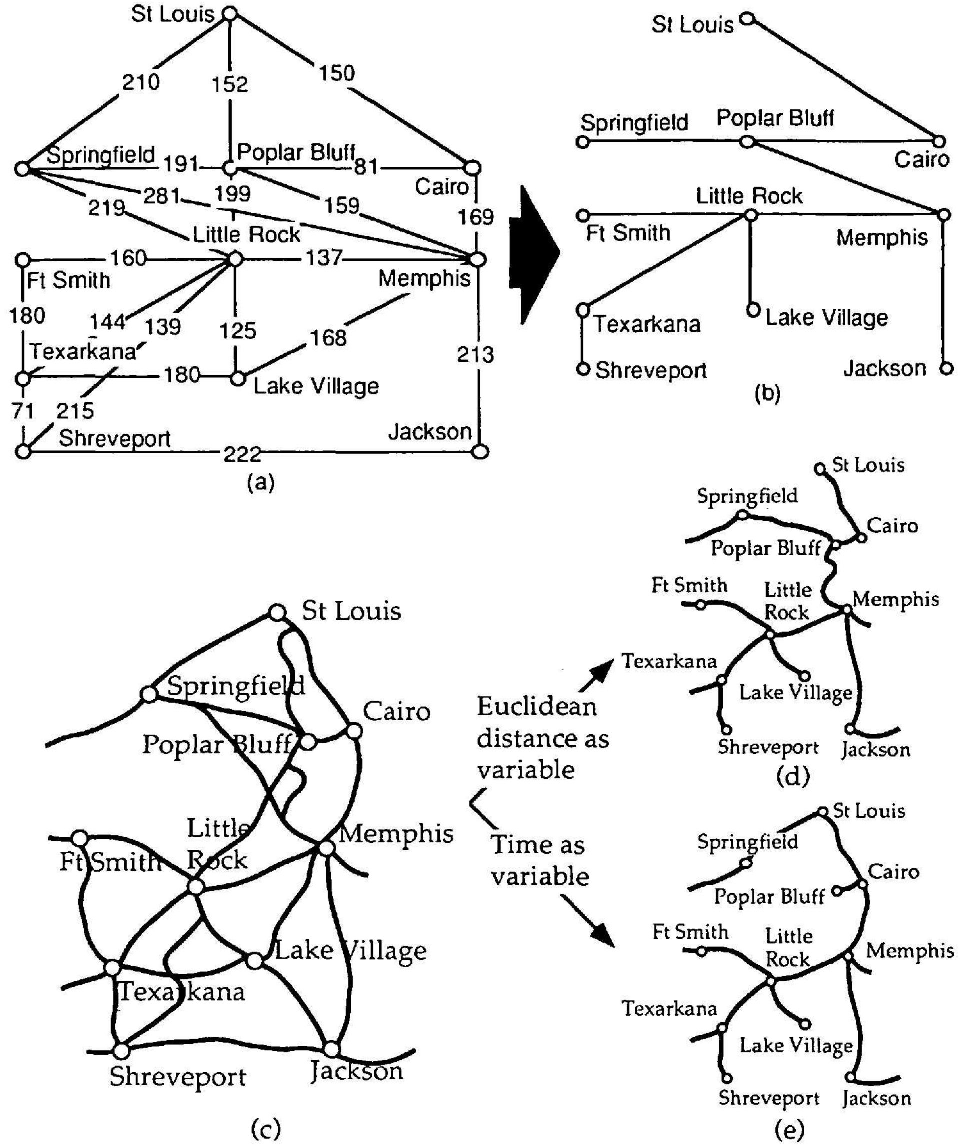


**Figure 7.** *Graph (a) of part of a road network (c), and the minimum spanning tree (b) used to create a reduced road network (d).*

either event, the topological properties of the graph combined with other geometric and attribute information can help in defining the most appropriate map generalization procedure.

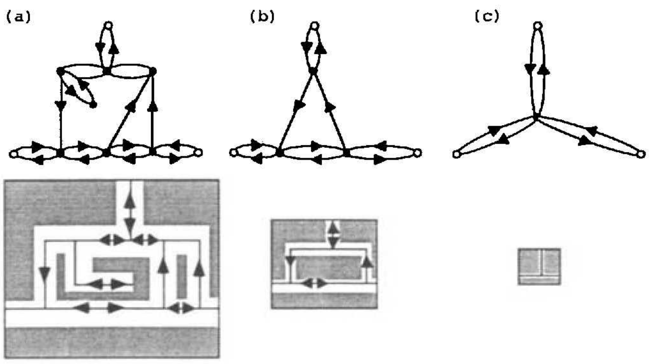


**Figure 8.** *Three states of generalization of a road network and their graphs.*

## Connectivity Influencing the Likelihood of Inclusion in a Map

Context information on a map is an important component of the recognition process and provides a setting for the message the designer wishes to convey. It provides a framework that facilitates the understanding of the relationship between known and unknown objects. For example, contextual information may be an inset, a section of coast, or a capital. Part of that context is the representation of the connectivity that exists between features, and this information can be indicative of its remoteness or centrality. Certain types of maps might not show edges of low weight, such as footpaths and train routes. However, in remote regions with features characterized by a low degree (the vertex is connected to relatively few edges in the graph), it becomes pertinent to show the features' connection to other features. Thus a town might be connected to the sea by a low-order stream that normally would not be shown. But such a section of river might be included if the town had a low degree. The following example (Figure 9) illustrates this point.

Kuujjuaq is a remote coastal town of Labrador, with a population of approximately 500. It has no road connection, but is connected to the sea by a bay, and can be reached overland by driving to Sept-îlse by taking the train to Schefferville and by canoeing the river Caniaspiscau (represented by the graph in Figure 9*b*). Immediately the edges of the disconnecting set can be identified (such as the railway and the river). Additionally, members of the separation set can also be identified (such as Schefferville). Because Schefferville is an articulation vertex it may be assigned greater importance. While railways and minor rivers normally might not be included on maps of certain scales and themes, Figure 9*a* represents the minimum information needed to convey Kuujjuaq's links with the rest of the world. There is sufficient space to do this because of the low denasity of cultural features. Thus graph theory can be used to secure the connectivity context of features when generalization techniques are being applied.

## Use of Graph Theory to Select Features that Provide Localized Context

Graph theory also can be used to select features based on their adjacency to salient features. The emphasis on a map is directly affected by which types of features are shown. The example in Figure 10 contains a pedestrian trail (part of the Appalachian Trail), a road vehicle network, and a flight network for float planes for part of the eastern seaboard of the United States. As demonstrated in Figure 11 (which shows the graph of Figure 10), graph theory does not preclude mixing edges and vertices of a different type. The use of similar line types as in the key of Figure 10 makes it easier to compare Figure 10 with Figure 11. Figure 11 can be used for a variety of applications, such as determining sightseeing tours using different modes of transport. It also is possible to determine the likely functionality of vertex features; in the example, some towns are termini (having degree 1) while others are centers (vertex with minimum eccentricity) for a variety of activities.

Graphs also can be used to derive corridor maps. For example, assuming an intended route has been identified through the graph, it is possible to identify features local to the route by identifying adjacent edge and vertex features. Figure 12 was derived from the graph of Figure 11 by first selecting the salient corridor feature (in this case, the Appalachian Trail) and then identifying the roads and towns in local proximity. This was done by searching for edges and their vertices adjacent to the path. Thus graph theory can be applied to derive corridor maps (or strip maps)

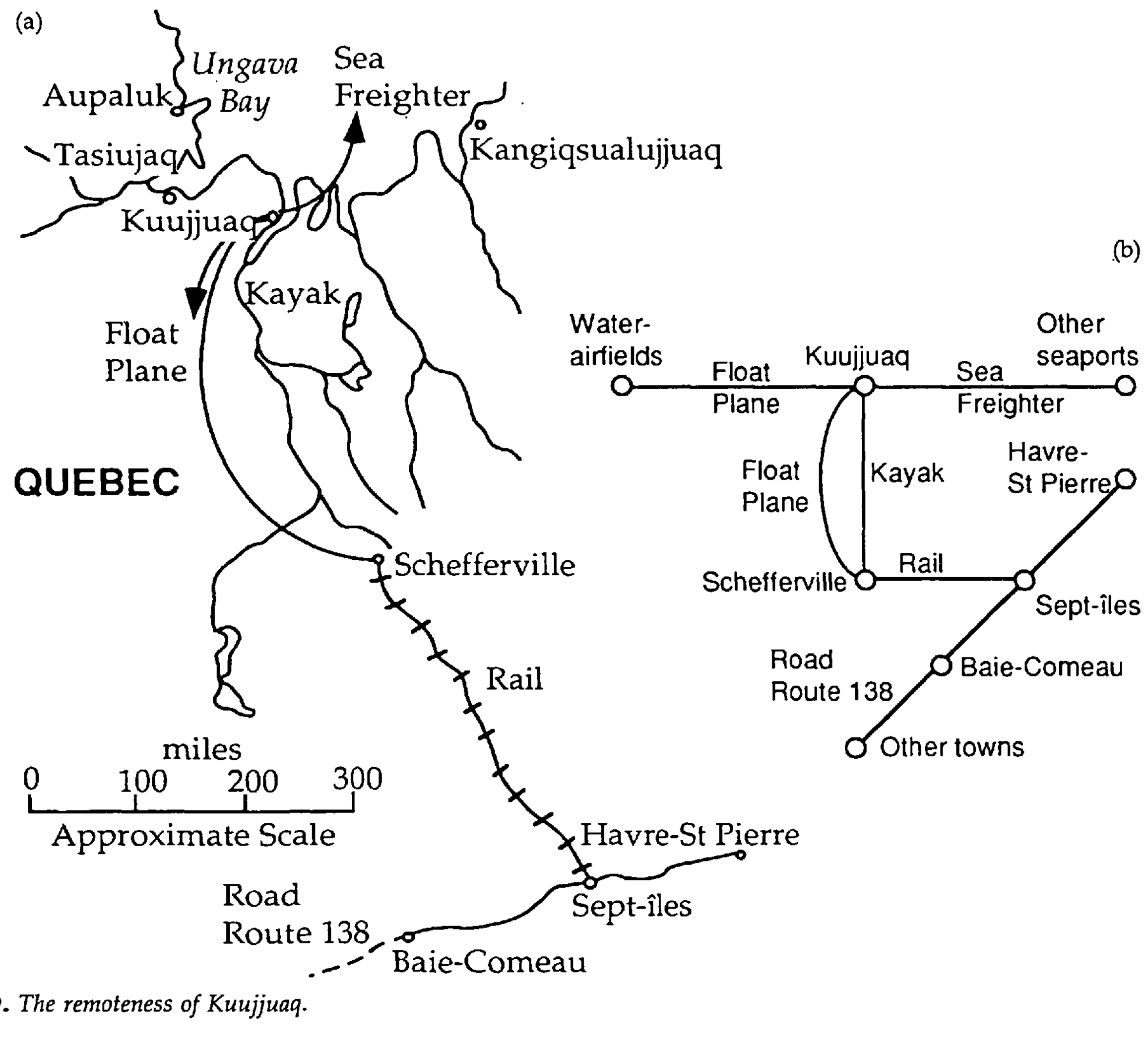


**Figure 9.** *The remoteness of Kuujjuaq.*

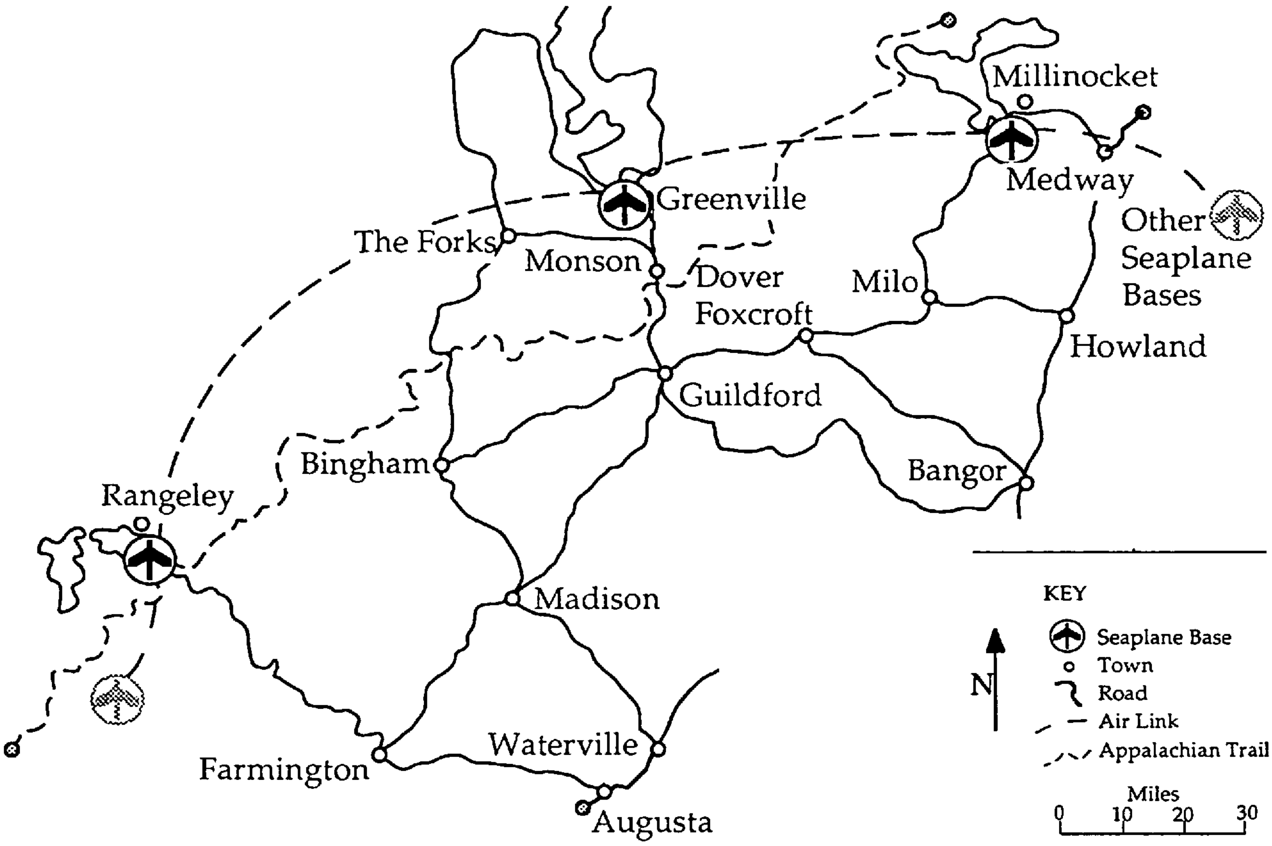


**Figure 10.** *Roads, trails, and planes in part of the Eastern Seaboard.*

showing local towns and amenities (Figure 13). Figure 13 is a simple example. A more sophisticated version of a corridor map could be produced by weighting the graph. For example, it is likely that second-order towns or large towns that provide a context (such as Bangor) would also be included. Thus a compilation rule could have the following form: Include the disconnecting set of all vertices with a context weight greater than 6 and all vertices adjacent to the path of interest and their disconnecting set.

## Some Preliminary Rules

The preceding examples demonstrate how rules on graphs can be formulated to help in the generalization process. Some examples of rules are set out in Appendix B (the order of the rules is not significant). Observe that some of them are compound and would require additional analysis through the application of theories other than just graph theory.

Appendix B is not an exhaustive list of rules formalized from graph theory, and the rules are not exhaustive in their definition. The interpretation of these rules depends on context. As has been illustrated, some terminus vertex features may be retained (rule 3), but in other situations they are removed (as in the cul-de-sac in Figure 8). This does not negate the use of graph theory in map generalization, but it does demonstrate how graph theory can be used in conjunction with other assessments in order to make intelligent design decisions. These rules provide additional information to help select features based on their attributes and their role in the map. For example, assuming many of the spatial and nonspatial attributes of a set of towns are known, it should be possible to prioritize them based on which of the rules applied.

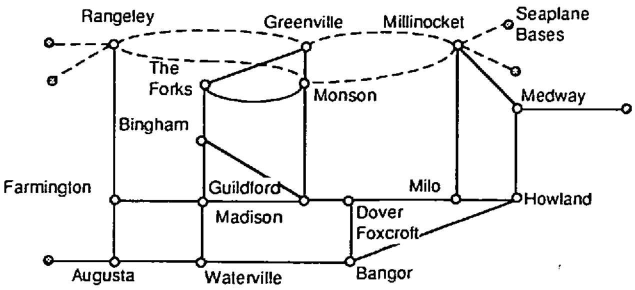


**Figure 11.** *Graph of part of the Eastern Seaboard.*

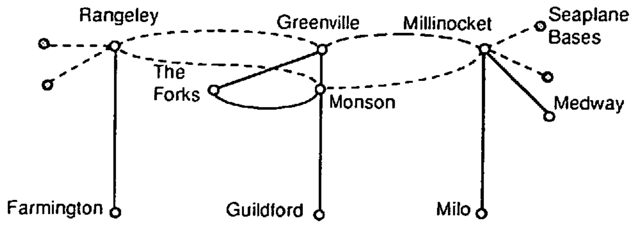


**Figure 12.** *Graph of vertices in proximity to edge activities used as a basis of corridor maps.*

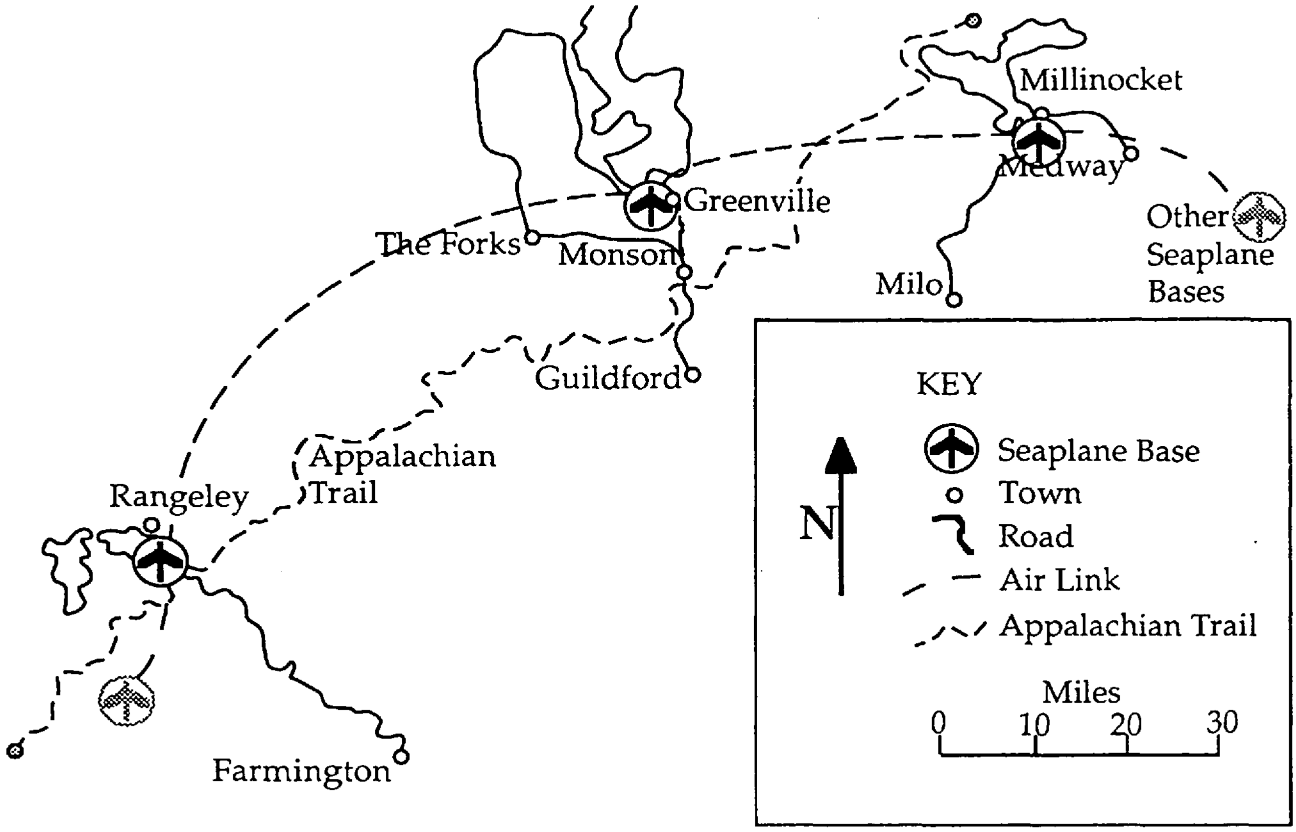


**Figure 13.** *Map derived from a finite, unweighted graph of vertices in proximity to specified edges.*

## Discussion

Of the various elementary procedures of cartographic generalization, such as those listed by Lichtner (1979) and McMaster and Shea (1992), graph theory can be used in the application of six of them (Appendix A). In addition, it has been demonstrated that graph theory can provide a measure of content in networks using the completeness parameter. Graph theory emphasizes topological properties and can be used to identify remote features and critical links among features. During the process of generalization, graph theory can be applied to ensure connectivity among features by preventing features from falling below degree minima. While this would have the effect of de-emphasizing the difference in urban street complexity between different vertex features (for example, towns), it may be just as important to represent the connectivity between features as their difference in size. Graph theory can also be used to help select features in general proximity to other features. Thus it is possible to create strip maps, for example, when creating a map showing towns and historical sites of interest along the Appalachian Trail (Figure 13) or when making a map showing views from a train, on a train map. While the application of certain generalization techniques affords graphical control within resolution bands, there are transitional resolutions at which major changes in behavior (therefore map content) take place (Ratajski 1967 and Buttenfield et al. 1991). For example, in Figure 5, simplification, enlargement, displacement, and selective omission can be used to reduce the road network through scales 1:50,000 to 1:100,000. But at a smaller scale, this network would become a symbol representing a town. These critical transition points are what Muller refers to as "conceptual cusps" (Muller 1989). It is important to appreciate that each object has its own unique conceptual cusp signature, which varies with map task.

For reasons of clarity, the examples in this paper have been in large part literal ones, but graph theory also can be used in more abstract situations to show, for example, associations or the physical or logical dependence between features. For example, a graph could be made of the different types of features shown on a map (cultural and natural). The edges linking those features would indicate associations and the weights would indicate the degree of association. Such a graph could then be used to select clusters of features (groups of vertices with high association) appropriate to a particular map type.

Similarly, graphs also can be used to model association through process. For example, the predominant geological unit in the surficial geological map of Maine is glacial till. The same geomorphic processes produced other geological units that are associated with the till. These include swamp marsh and bog deposits, stream alluvium, moraine, eskers, and other glaciofluvial deposits, all of which occur in small splatterings across the till. Due to their small size, at reduced scales these small geological units typically are generalized out of the map, and the associative process that links them with glacial till is consequently lost. A typical solution would have two stages: First, select a small representative subset of these small geological units, and second, exaggerate their size in order to make them visible. A graph can be used to model the association of these smaller units with the predominant unit and thus support the first stage of this generalization process. An acyclic graph (tree-like in shape) would represent the association, the edges would represent "belonging," and their weight would represent their degree of association. Each occurrence of a small geological unit would be represented by a vertex connected by an edge to the predominant unit. The representative set would be selected by pruning the graph based on the number of vertices of a particular geological unit and the weight of association. In this way it is possible to preserve both the association and the existence of small geological units at smaller scales.

## Conclusion

The application of graph theory in map generalization is seen as part of a wider use of other models that will help to formalize knowledge required to make intelligent and aesthetic decisions in map design. This article has demonstrated how graph theory can be used to provide and to disseminate the information required in this decision-making process.

The interpretation of a feature depends on a context provided by other features. Thus a feature's inclusion in or exclusion from a map graphic is not based solely on theme and scale but on its interaction with other features and the competition for space. If generalization techniques are to be applied in the same subtle way that they are applied in human map design (in order to convey ordering, distinction, comparison, combination, and recognition of relations), then additional information relating to context is required. From a design perspective, that information includes the spatial and behavioral attributes of data and their interdependence. The authors have, therefore, argued for a course that requires a rich database, one in which these additional attributes are collected at the first abstraction (i.e., when the data are first captured), as no amount of second-level processing will bring back what was not collected. The application of graph theory enables us to formalize some of the relationships among features and thereby gain a better insight into how features interact; in turn, this will help us to make better decisions concerning map design.



### ACKNOWLEDGMENTS

Both authors are grateful for the reviewers' comments and for support for the NCGIA provided by the National Science Foundation under grant SES 88-10917. Thanks are also due to Todd Rowell, graduate student at the University of Maine, for an implementation of the Kruskal algorithm.

### REFERENCES

Aho, A.V., J.E. Hopcroft, and J.E. Ullman. 1974. *The Design and Analysis of Computer Algorithms.* Reading, Massachusetts: Addison Wesley.

Bollobas, B. 1979. *Graph Theory: An Introductory Course.* New York: Springer Verlag.

Borchert, J.R. 1961. The Twin Cities Urbanized Area: Past, Present, Future. *Geographical Review,* vol. 51, pp. 47-70.

Buttenfield, B.P., C.R. Weber, M. Leitner, J.J. Phelan, D.M. Rasmussen, and G.R. Wright. 1991. How Does a Cartographic Object Behave? Computer Inventory of Topographic Maps. *Proceedings of GIS/LIS '91,* vol. 2, pp. 891-900.

Chartrand, G. 1977. *Graphs as Mathematical Models.* Boston: Prindle, Weber and Schmidt.

Chou, Y.-H. 1991. A Network Analysis of Domestic Airline Transportation in the United States. In *Proceedings of GIS/LIS '91,* vol. 2, pp. 518-527.

Cliff, A.D., P. Haggett, and J.K. Ord. 1979. Graph Theory and Geography. In *Applications of Graph Theory,* R.J. Wilson and L.W. Beineke (eds.). London: Academic Press.

Corbett, J.P. 1979. Topological Principles in Cartography. *U.S. Department of Commerce and the Bureau of the Census, Technical Paper 48; GPO stock no. 1980-311-046/1142.* Washington, D.C.: U.S. Government Printing Office.

Haggett, P. 1990. *Geography: A Modern Synthesis.* New York: Harper Collins.

Hartsfield, N., and G. Ringel. 1990. *Pearls in Graph Theory.* Boston: Academic Press.

Herring, J.R. 1987. TIGRIS: Topologically Integrated Geographic Information System. In *Proceedings of Auto Carto 8,* Baltimore, vol. 1, pp. 282-291.

James, G.A., A.D. Cliff, and P. Haggett. 1970. Some Discrete Distributions for Graphs with Applications to Regional Transport Networks. *Geografiska Annaler,* vol. 52B, pp. 14-21.

Keates, J.S. 1989. *Cartographic Design and Production,* second edition. London: Longman.

Konig, D. 1990. *Theory of Finite and Infinite Graphs.* Boston: Birkhauser.

Lichtner, W. 1979. Computer-Assisted Processes of Cartographic Generalization in Topographic Maps. *Geo-Processing,* vol. 1, pp. 183-199.

Mackaness, W.A. 1991. Integration and Evaluation of Map Generalization. In *Map Generalization: Making Rules for Knowledge Representation,* B.P. Buttenfield and R.B. McMaster (eds.). London: Longman Group, pp. 217-226.

Maling, D.H. 1966. Foreword to Topfer, F., and W. Pillewitzer. 1966. The Principle of Selection. *The Cartographic Journal,* vol. 3, no. 1, pp. 10.

McMaster, R.B., and K.S. Shea. 1992. *Generalization in Digital Cartography.* Resource Publications in Geography. Washington: Association of American Geographers.

Muller, J.C. 1989. Theoretical Considerations for Automated Map Generalization. *ITC Journal,* vol. 3, no. 4, pp. 200-204.

Paiva, J.A.C., M.J. Egenhofer, and A.U. Frank. 1992. Spatial Reasoning about Flow Directions: Towards an Ontology for River Networks. In *Proceedings of the 17th International Society for Photogrammetry and Remote Sensing Congress Meeting, XXIX, Part B3 Commission III,* pp. 318-324. Washington: International Archives of Photogrammetry and Remote Sensing.

Rand McNally. 1991. *Business Traveler's Road Atlas.* Boston: Rand McNally.

Ratajski, L. 1967. Phénomènes des Points de Généralization. *International Yearbook of Cartography,* vol. 7, pp. 143-51.

Richardson, L.F. 1961. The Problem of Contiguity: An Appendix to the Statistics of Deadly Quarrels. *General System Yearbook,* vol. 6, pp. 139-187.

Robinson, A., R. Sale, J. Morison, and P.C. Muehrcke. 1984. *Elements of Cartography,* fifth edition. New York: John Wiley and Sons.

Rusak Mazur, E., and H.W. Castner. 1990. Horton's Ordering Scheme and the Generalization of River Networks. *The Cartographic Journal,* vol. 27, pp. 104-112.

Sedgewick, R. 1992. *Algorithms in C++.* Reading, Massachusetts: Addison Wesley.

Strahler, A.N. 1960. *Physical Geography.* New York: John Wiley and Sons.

Switzer, R.M. 1975. *Algebraic Topology—Homotopy and Homology.* Berlin: Springer-Verlag.
Taaffe, E.J., R.L. Morrill, and P.R. Gould. 1963. Transport Expansion in Underdeveloped Countries: A Comparative Analysis. *Geographical Review,* vol. 53, pp. 503-529.
Tarjan, R.E. 1983. *Data Structures and Network Algorithms.* Philadelphia: Society for Industrial and Applied Mathematics.
Topfer, F., and W. Pillewizer. 1966. The Principles of Selection. *Cartographic Journal,* vol. 3, no. 1, pp. 10-16.
Temperley, H.N.V. 1981. *Graph Theory and Applications.* Chichester: Ellis Horwood.
White, M.S. 1978. A Geometrical Model for Error Detection and Correction. In *Proceedings of Auto Carto 3,* vol. 1, pp. 439-456.
Williams, R.J. 1985. Enquiry Systems for the Interrogation of Infrastructure. In *Proceedings of Auto Carto 7,* vol. 1, pp. 572-581.
Wilson, R.J. 1979. *Introduction to Graph Theory.* London: Longman.

## Appendix A

### Summary of Potential Use of Graph Theory in Map Generalization

Graph theory can be used to aid in applying the following generalization operators:

1. *Simplification:* Removal of auxiliary vertices while maintaining connectivity.
2. *Aggregation:* Identification of the fuzzy group boundaries of features of the same type, facilitating the coalescence of features.
3. *Merging-refinement:* Selection of the essential components of a network and maintenance of connectivity during refinement.
4. *Displacement:* Identification of dense areas of networks, and checking the consistency after the application of a displacement algorithm.
5. *Selection:* Identification of information that is contextual due to the local proximity to salient features.
6. *Exaggeration*: Identification of features in isolation or in possession of topological characteristics requiring additional cartographic emphasis.

## Appendix B

### Cartographic Generalization Procedures

*Rule 1: A feature with degree zero is an isolated feature.* Often features have prominence by virtue of their isolation or because they are important as way points in navigation (for example, where both ethnographic networks and features are sparse). Thus remote wooden mountain huts are often included in topographic maps, but large city buildings are not; such a rule in combination with a graph can be used to identify these features. A rule might be of the following form: IF a feature is isolated and feature density is low, THEN it is retained.

*Rule 2. Where a cycle exists, edges can be removed on a hierarchical basis without isolating a feature.* This is the basic rule that is applied when generalizing networks. This rule is implicit to the idea of using minimum spanning trees and was used in Figure 6. A rule might be of the following form: IF two vertices share the same edges and the edges are of the same type, THEN the number of edges is reduced to one.

*Rule 3. A feature with degree 1 is a terminus.* Features that have the attribute of being a terminus may have more importance depending on the context. For example, a terminus town may be the last opportunity to stock up with food, or a terminus may be in a region of sparse isolated features. Thus a rule might be of the following form: IF a feature is a terminus, THEN it is retained.

*Rule 4. A feature with proportionately high degree compared with the mean degree of the graph is a hub.* When a feature is common to a large number of adjacent features, that feature may become important because objects traveling through the network will travel via that feature (Chou 1991). Thus it is known that Boston and Chicago are large aviation hubs, and of course towns and cities act as hubs to local networks of roads. Thus a rule might be of the following type: IF a feature is a hub and the map type is network, THEN that feature is included in the map.

*Rule 5. A feature that is the sole link between other features is more important than features not serving this task but of a similar type.* The following rule can apply equally to vertex type features and gives us rule 6: IF an edge feature is the sole member of a disconnecting set, THEN it is retained in a network type map.

*Rule 6. A feature that is the sole member of the separation set is more important than features not serving this task but of a similar type.* A feature that is an edge connecting two vertex features has importance (rule 5). If a vertex stands on that edge (part of the separation set), then some of that importance is imparted into the vertex feature. Thus the rule might be of the following form: IF a feature is the sole member of a separation set AND the map type is network, THEN it is retained.

*Rule 7. An articulate feature displays hublike qualities and also provides unique connectivity between features.* As with rule 4, those traveling across a network will be forced to travel by an articulate feature; therefore that feature will have raised importance. For example, Calais is a small fishing port in France, but it has elevated importance as a main cross channel ferry terminus connecting the United Kingdom with France. Thus the rule might be of the following type: IF a feature is articulate and the map type is network, THEN it should be retained in a network-type map.

*Rule 8. Where edges of different types meet, the vertex has a raised importance.* Where there is a change in the mode of transport (for example, from ferry to road or from rail to float plane), the feature (such as a town) has the quality of connecting edges of different types and may have a raised importance. Again this describes one of the discerning attributes of a town, where typically you find a rail, bus, and river terminus. So the rule might take the following form: IF a group of features (all of type edge and adjacent to a common vertex) number three or more and the map type is network, THEN that feature that is the vertex of those edge types is a translator vertex.

*Rule 9. A vertex feature can be omitted if it has degree 2 and the edges are of the same type.* The application of this rule was used in creating Figure 5. By example, the following rule might exist: IF a feature has degree 2 AND edge types that are the same and the density of features of this type is high, THEN they are removed.