it successfully to the organization. This training is usually provided as a series of seminars or workshops, either by existing GIS staff or by external GIS consulting companies. Second, before the GIS vendor is selected and the project is implemented, the GIS staff needs to possess the skills necessary to make educated decisions on topics such as database management systems, project management techniques, land records and cartography, IT, and, of course, GIS. This training can be provided by universities, technical institutions, and some consulting companies. Finally, once the particular GIS vendor is selected, the training needs focus on the implementation, use, and management of the GIS software and hardware. This training is usually provided by the vendor as part of the GIS procurement contract, or it can be provided by the vendor's business partner or other training agency that specializes in the particular GIS vendor.

Other Needs

Computerization, and especially new technology such as GIS, can cause great changes to an organization. Organizational changes may be required to support changes in responsibilities such as who is responsible for the new system (i.e., should the system be placed in the IT department or the primary user department, or should it be placed in a more strategic location, such as the CEO's office?). After GIS has been implemented, data and maps move throughout the organization differently, and determining who is responsible for maintaining the newly computerized data and maps may cause changes. In many cases, the way work is done changes, and that may require changes to the organizational structure or in organizational procedures. In some cases, policy and legal changes are necessary in order for the organization to gain the fullest benefit from the technology. Some of the most common legal issues that arise when public agencies adopt GIS involve the dissemination of data to the public and the public's right to privacy. Copyrighting and licensing digital data is becoming commonplace, while researchers and nongovernmental agencies are calling for public access to the data. In addition, many surveyors and engineers use the technology, and so local governments are beginning to require them to submit subdivision plans in digital form to make the map maintenance procedures easier after the development becomes a reality.

Conclusion

There are many needs to consider when an organization begins the adoption of new technology such as GIS. Since GIS has benefits in many types of industries and organizational functions, and, since organizations differ in how they operate, the particular package of GIS hardware and software, data, people, procedures, and other considerations differ for each organization. That is why many GIS experts agree that a GIS is built—not bought.

William E. Huxhold

See also Database Design; Database Management System (DBMS); Software, GIS

Further Readings

Gilfoyle, I., & Thorpe, P. (2004). *Geographic information management in local government.* London: CRC Press.

Huxhold, W. E., & Levinsohn, A. G. (1995). *Managing geographic information systems projects.* New York: Oxford University Press.

Obermeyer, N. J., & Pinto, J. K. (1994). *Managing geographic information systems.* New York: Guilford Press.

Tomlinson, R. (2003). *Thinking about GIS.* Redlands, CA: ESRI Press.

Network Analysis

Network analysis consists of a set of techniques for modeling processes that occur on networks. A *network* is any connected set of vertices (e.g., road intersections) and edges (e.g., road segments between intersections) and can represent a transportation or communications system, a utility service mechanism, or a computer system, to name only a few network applications. Although network analysis is a broad and growing discipline within geographic information science (GISci), this entry addresses the topic by identifying and outlining three major components of network analysis: finding locations on networks, routing across networks, and network flow analysis.

Location on Networks

Both the process of locating network elements themselves and the process of locating facilities on existing networks can be structured as optimization problems, which are problems that seek to minimize or maximize a particular goal within a set of constraints. As an example, the minimum spanning tree problem seeks to find locations for new network edges such that the cost of constructing those edges is minimized yet every node or vertex is connected to the network. Both Kruskal's and Prim's algorithms solve this problem using a greedy approach that sequentially chooses the next minimum cost edge and adds it to the network until all locations are connected. Other objectives in designing networks may be to maximize the connectivity of the network under a cost constraint or to minimize the dispersion of the network nodes.

Location on networks involves selecting network locations on an existing network such that an objective is optimized. These problems can be differentiated by the objective function, by the type of network on which location occurs, or by the number of facilities to locate. Since it is not possible to outline all possible permutations of these factors in this entry, several classic location objectives on networks are presented here.

Median problems on networks seek to locate facilities such that the demand-weighted distance is minimized. That is, it is assumed that varying demand for service exists at the nodes of the network, and facilities would be best located if the total cost incurred in transporting the demand to a facility is at a minimum. When a single facility is being located, this problem is termed the *Weber problem,* due to Alfred Weber's early work on the location of industries. When a single facility is located on a Manhattan network (a rectangular network of intersecting edges such as the road network in Manhattan), the point of minimum aggregate travel can be determined by finding the median point along both of the axes of the network. When more than one facility is to be located, this problem is termed the p-median problem, where p designates the number of facilities to be located. Due to the combinatorial complexity of this problem, it is very difficult to solve large-problem instances, and therefore the p-median problem is one that demands substantial research effort.

As with many network location problems, the p-median problem can be formulated as a linear programming optimization model. Such a model consists of an objective function to be optimized, a set of constraints, and sets of decision variables that represent decisions about where to locate on the network. The goal is to determine the values of those variables such that the objective is optimized, while respecting the constraints. The notation for the p-median formulation consists of

i and j = indices of network node locations that serve as both demand locations (i) and potential facility sites (j),

a_i = the level of demand at network location i,

d_{ij} = the distance (or cost of travel) between network locations i and j, and

P = the number of facilities to locate.

The decision variables are defined as

$$x_j = \begin{cases} 1 \text{ if a facility is located at network location } j \\ 0 \text{ otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} 1 \text{ if demand at } i \text{ is served by a facility} \\ \quad \text{at network location } j \\ 0 \text{ otherwise} \end{cases}$$

With this notation defined, one can formulate the objective of minimizing demand-weighted distance as

$$\text{Minimize } Z = \sum_i \sum_j a_i d_{ij} y_{ij}$$

This function states that the values of the decision variables (y_{ij}) must be chosen in such a way that the sum of the products of the demands and their respective distances to facilities is minimized.

The objective function operates under several sets of constraints, including

$$\sum_j y_{ij} = 1 \text{ for all } i$$

which ensures that any demand point i is only assigned to a single facility location j, and

$$y_{ij} - x_j \leq 0 \text{ for all } i \text{ and } j$$

which ensures that for every pair of locations i and j, a demand can be assigned to a facility at j ($y_{ij} = 1$) only if a facility is located at j ($x_j = 1$); and in order to ensure that exactly P facilities are located, a single constraint is added:

$$\sum_j x_j = P$$

Finally, all of the decision variables must be either 0 or 1, since a facility cannot be located at more than one location and demand ought not be served by more than one facility:

$$x_j = 0, 1 \text{ for all } j$$

$$y_{ij} = 0, 1 \text{ for all } i, j$$

Although the p-median is a very widely used network location problem, there are a multitude of other objectives. Among these are *center problems* that seek to locate facilities such that the maximum distance between a demand point and a facility is minimized. This problem optimizes the worst-case situation on the network. Still other important problems seek to maximally cover the demand within an acceptable service distance. These problems are frequently used for the purpose of locating emergency service facilities. A variation of this type of problem seeks to locate facilities such that flow across the network is covered. These are termed *flow covering* or *flow interdiction problems*.

Routing Across Networks

Routing is the act of selecting a course of travel. The route from home to school, the path taken by a delivery truck, or the streets traversed by a transit system bus are all examples of routing across a network. Routing is the most fundamental logistical operation in network analysis. As in location on networks, the choice of a route is frequently modeled as an optimization problem.

Finding the Shortest Path

Without question, the most common objective in routing across networks is to minimize the cost of the route. Cost can be defined and measured in many ways but is frequently assumed to be a function of distance, time, or impedance in crossing the network. There are several extremely efficient algorithms for determining the optimal route, the most widely cited of which was developed by Edsgar Dijkstra. *Dijkstra's algorithm* incrementally identifies intermediate shortest paths through the network until the optimal path from the source to the destination is found. Alternative algorithms have been designed to solve this problem where negative weights exist, where all the shortest paths between nodes of the network must be determined, and where not just the shortest path but also the 2nd, 3rd, 4th, or kth shortest path must be found.

The Traveling Salesman Problem

The *traveling salesman problem (TSP)* is a network routing problem that may also be the most important problem in combinatorial optimization. This classic routing problem presumes that a hypothetical salesman must find the most cost-efficient sequence of cities in the territory, stopping once at each and returning to the initial starting location. The TSP has its origins in the Knight's Tour problem, first identified by L. Euler and A. T. Vandermonde in the mid-1700s. In the 1800s, the problem was identified as an element of graph theory and was studied by the Irish mathematician, Sir William Rowan Hamilton, whose name was subsequently used to describe the problem as the Hamiltonian cycle problem.

The problem was introduced to researchers (including Merrill Flood) in the United States in the early 20th century. Flood went on to popularize the TSP at the RAND Corporation, in Santa Monica, California, in late 1940s. In 1956, Flood mentioned a number of connections of the TSP with Hamiltonian paths and cycles in graphs. Since that time, the TSP has been considered one of the classic models in combinatorial optimization and is used as a test case for virtually all advancements in solution procedures.

There are many mathematical formulations for the TSP, with a variety of constraints that can be used to enforce variations of the requirements described above. The most difficult problem in formulating the TSP involves eliminating subtours, which are essentially smaller tours among a subset of the "cities" to be visited. These subtour elimination constraints can substantially increase the size of the problem instance and therefore make solution more difficult.

Other Vehicle Routing Problems

The shortest-path problem and the TSP are two of many different possible *vehicle routing problems (VRPs)*. Most of the VRPs in the literature are cost minimization problems, although there are others that seek to maximize consumer surplus, maximize the number of passengers, seek equity among travelers, or seek to minimize transfers while encouraging route directness and demand coverage. A substantial subset of the literature posits that multiple objectives should be considered. Among the proposed multiobjective models are those that trade off maximal covering of demand against minimizing cost, those that seek to both minimize cost and maximize accessibility, and those that trade off access with service efficiency.

Network Flow Analysis

Most networks are designed to support the flow of some objects across them. The flow may be water through a river network, traffic across a road network, or current through electricity transmission lines. To model flow, networks must be able to support the concepts of capacity and flow direction. In the context of geographic information systems (GIS), *network capacity* is implemented as an attribute value associated with features. The concept of *flow direction* can be assigned with an attribute value but more accurately is a function of the topological connections to sources and destinations of flow (sinks). Using flow direction, GIS can solve problems such as tracing up- or downstream.

Beyond simply modeling the concept of flow, there is a large family of problems known as *network flow problems.* The focus of these is on finding the optimal flow of some objects across the network. It may be that the optimal solution is the one that determines the maximal flow through the network from a source to a sink without violating capacities. Another problem tries to find the minimal cost flow of commodities across the network from a set of sources to a set of destinations. This problem is sometimes referred to as the *transportation problem,* and there are efficient algorithms (such as the network simplex algorithm) for optimal solution under certain conditions. When congestion on the network causes the cost of traversing its edges to vary with the amount of flow moving across them, the network is said to have *convex costs,* and advanced methods allow for the solution of such problems.

Challenges for Network Analysis in GISci

There are two primary challenges for researchers interested in network analysis and GISci. First, the implementations of network analysis in current GIS software are in their infancy. In the most recent network analysis software package from the industry-leading software developer, there are only four primary network analysis functions. There are many network analytical techniques and methods that have not yet been integrated into GIS.

Second, many network analysis problems are extremely difficult to solve optimally. These are so difficult that even modestly sized instances of these problems cannot be solved by enumeration or by linear programming methods. The GISci community must accept the challenge of reformulating problems, developing new solution techniques, and, when necessary, developing good heuristic or approximate methods to quickly find near-optimal solutions. Last, there has recently been increased interest in the use of simulation methods, such as agent-based modeling and cellular automaton models to generate optimal solutions to network problems. While addressing these issues, network analysis will continue to be one of the most rapidly growing elements of GISci. It has a deep body of theory behind it and a great diversity of application that encourages continued research and development.

Kevin M. Curtin

See also Geographic Information Science (GISci); Geographic Information Systems (GIS); Geometric Primitives; Network Data Structures; Optimization; Spatial Analysis; Topology

Further Readings

Current, J. R., & Marsh, M. (1993). Multiobjective transportation network design and routing problems: Taxonomy and annotation. *European Journal of Operations Research, 65,* 4–19.

Dantzig, G. B., & Ramser, J. H. (1959). The truck dispatching problem. *Management Science, 6,* 80–91.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1,* 269–271.

Evans, J. R., & Minieka, E. (1992). *Optimization algorithms for networks and graphs* (2nd ed.). New York: Marcel Dekker.

Flood, M. M. (1956). The traveling salesman problem. *Journal of the Operations Research Society, 4,* 61–75.

Magnanti, T. L., & Wong, R. T. (1984). Network design and transportation planning: Models and algorithms. *Transportation Science, 18,* 1–55.