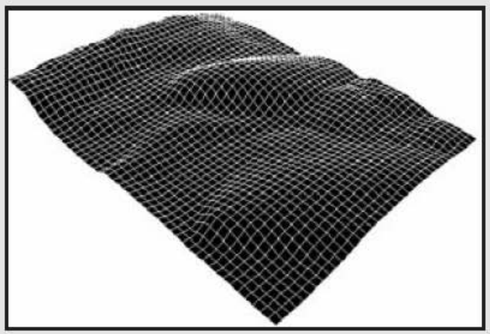
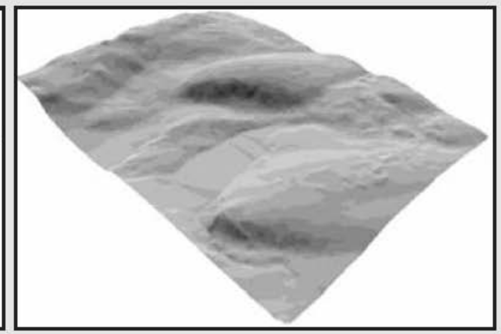
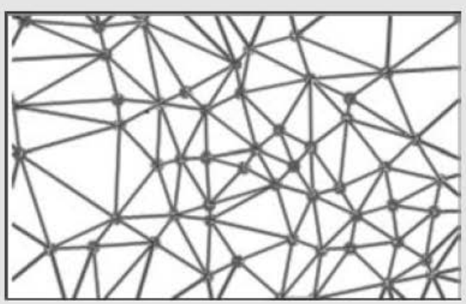
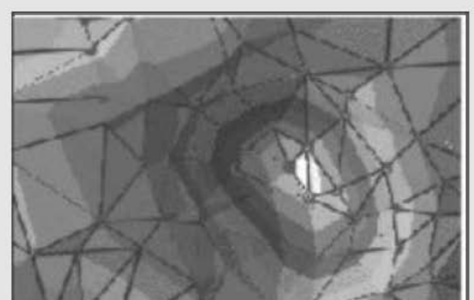
CHAPTER 8

ADVANCED DATA MODELS

SURFACE REPRESENTATION

- Several data models used in GIS can extend the real world to include the terrain surface, the time factor, and movable objects.
- The digital representation of a terrain surface is called either a digital terrain model (DTM) or a digital elevation model (DEM).

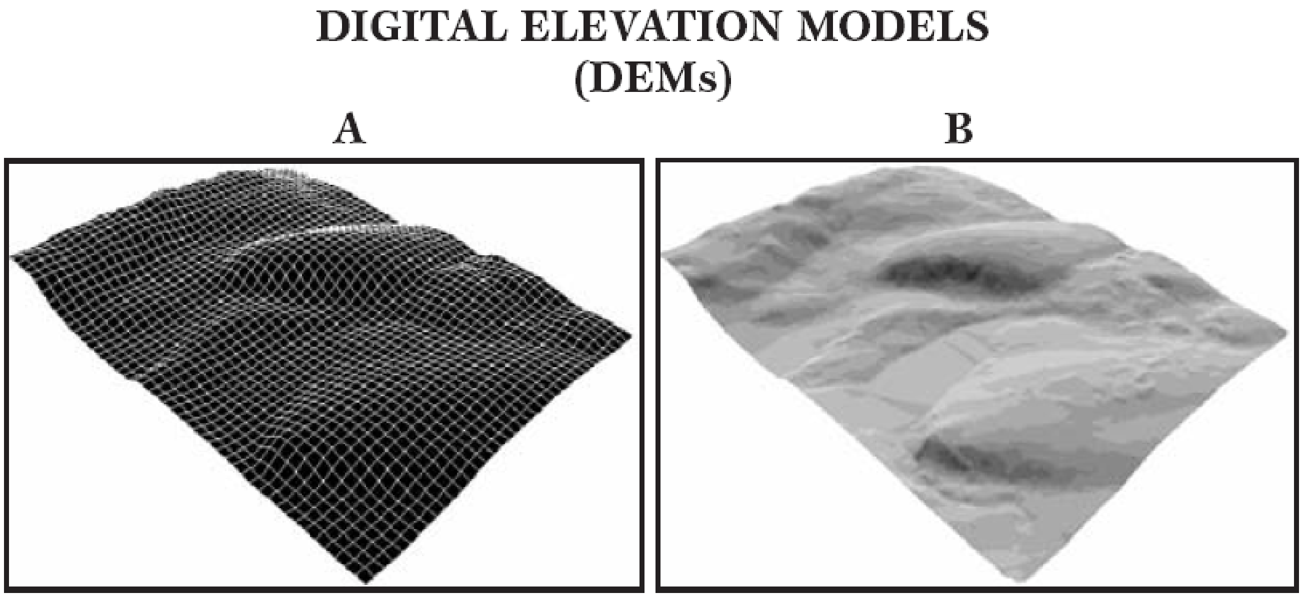


Figure 8.1: A DEM is an essential layer in the representation or analysis of any area with variable terrain.

- The terrain surface can be described as comprising two basically different elements.
- The random (stochastic) elements
  - These are the continuous surfaces with continuously varying relief.
  - These can be described in practice with a network of points.
  - It is usual to use a network that creates sloping triangles or regular quadrants.
- The systematic elements
  - The systematic part of the terrain surface are characterized either by sharp cracks in the terrain or by characteristic points.
    - sharp cracks such as the top or bottom of a road cut, or characteristic points such as spot depression and spot height
  - The systematic part is thus best represented by lines and typical single points.

- A terrain model can be realized by linking height as an attribute to each point *(x, y).*
  - This type of elevation model can only describe a surface and cannot handle more *z* values to the same point.
- The term 2.5 – dimensional is often used to describe the DTM dimension.
  - This model is most suited to visualization.
- In a three-dimensional elevation model, elevation is an integral part of position *(x, y, z)* and the model can handle several *z* values for the same *x, y* pair.
  - It can handle different geological layers, roof heights on buildings, roads that cross each other, together with the terrain surface.
  - A three-dimensional model is also suited to volume calculations.

- The *z* value of a new point is calculated by interpolation from the z value to the closest existing points.
- If the points are stored in an unstructured way, it is usual to use data structures which also describe the contiguity between the points.
  - This is achieved by using data structures based on single points in a raster (grid) or triangles covering a surface.

GRID MODEL

- A systematic grid, or raster, of spot heights at fixed mutual spaces is often used to describe terrain (Figure 8.1 A).
- Elevation is assumed constant within each cell of the grid and the size of cells is constant in a model.
  - Small cells detail terrain more accurately than large cells.
  - Areas with a greater variation of terrain may be described less accurately than those with less variation.
- The grid model is most suitable for describing random variations in the terrain, while the systematic linear structures can easily disappear or be deformed.
  - One possible solution can be to store the data as individual points and generate grids of varying density as required.

- Elevation values are stored in a matrix and the contiguity between points is thus expressed through the column and line number.
  - When the data points are dispersed, the averages of the elevations of those closest to grid points, within a given circle or square, are assigned to the grid points with inverse weighting in proportion to the intervening distances involved.
  - In the contours, grid point elevations are interpolated from the elevations at the intersections of the original data lines and the lines of the grid.

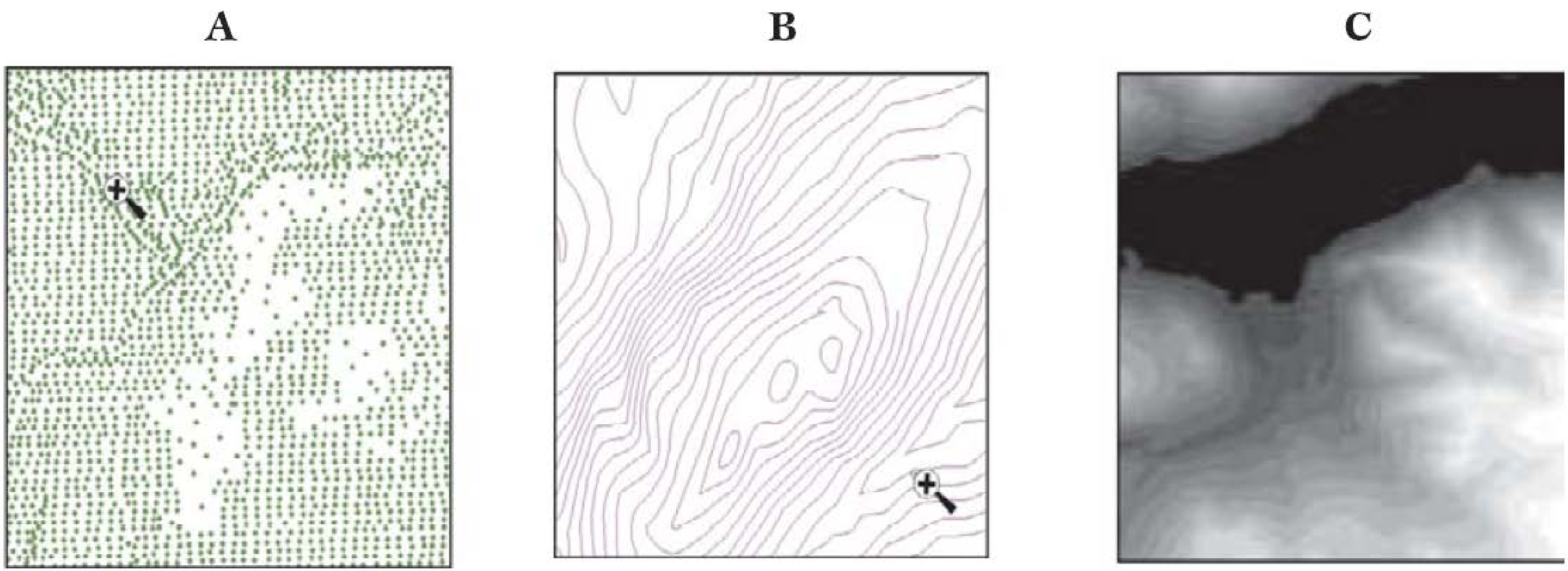


Figure 8.2: Elevation data are acquired through:

a. Mass points (lattices)

b. Contour lines

c. GRIDs (interpolated from points or lines; or created currently from digital imagery)

- Terrain may also be described in terms of chosen or arbitrarily selected individual points (*i.e.,* a point cloud).
  - Point densities should be greatest in areas where terrain features vary the most.
  - The amount of memory storage required for the point arrangements between the regular grid and the cloud points differs.
- For describing abrupt terrain variations, such as the top and bottom of a road cut, point models are inferior.

TIN MODEL

- The triangulated irregular network (TIN) model is an array of triangular areas with their comers stationed at selected points of most importance, for which the elevations are known.
  - The inclination of the terrain is assumed to be constant within each triangle.
  - The area of the triangles may vary, with the smallest representing those areas in which the terrain varies the most.
- Insofar as possible, small equilateral triangles are preferable.
- In the TIN model, the $x – y – z$ coordinates of all points, as well as the triangle attributes of inclination and direction, are stored.

Box 16: Make your own TIN model

A simple experiment using a piece of paper can give you an idea of TIN model. If you take a sheet of paper squeeze it in your palm. Now if open it again you would find many irregular sized creases. *This is a good and simple example of* TIN MODEL. The creases are the ridges and valleys and the intersections of the creases are the peaks, depressions and passes. The areas of flat paper between the creases are the irregular triangles of the TIN model, which may be assigned area slope and aspect values. Certainly, this is not a perfect model as not all of the facets on your piece of paper will be triangles, but it gives a rough idea to illustrate TIN principle. Further if you hold the paper level with a light source you could get a fair idea of a miniature terrain of peaks and valleys casting shadows in front of you. The tighter you squeeze the paper more complex terrain you produce. This experiment also demonstrates the *'two and a half'* dimensional nature of terrain models in GIS. This model is the surface with no depth.

- The Delaunay triangle (Figure 8.3)
  - Delaunay triangulation is a method used to fit triangles in a point cloud.
  - Triangles can be formed by laying circles through three points and testing whether there are other points within the circle.
  - If other points are not available, a new triangle will be formed.
  - This method produces triangles with a low variance in length.

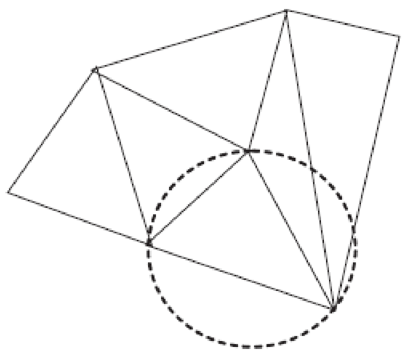
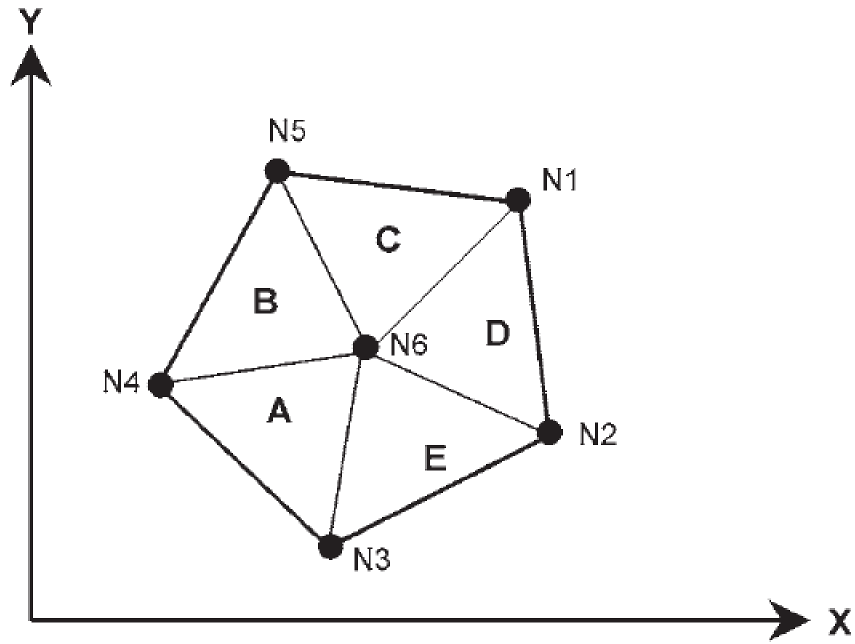


Figure 8.3: Delaunay triangulation is a method used to fit triangles in a point cloud. The circle described ensures that the triangles have good geometry with least possible variation in page lengths.

- Compared to the grid model, the TIN model (Figure 8.4) is cumbersome to establish but more efficient to store.
- However, the TIN model normally requires considerably larger storage capacity than the grid model.
- TIN models are good for describing terrain because the sharp breaks of slope between uniform-slope facets fit certain types of terrain well.



Triangle Table and Node

Triangle	Adjacent	Nodes
A	B, E	N3, N4, N6
B	A, C	N4, N5, N6
C	B, D	N1, N5, N6
D	C, E	N1, N2, N6
E	D, A	N2, N3, N6

Coordinate Table

Node	Coordinates
N1	X1, Y1, Z1
N2	X2, Y2, Z2
N3	X3, Y3, Z3
N4	X4, Y4, Z4
N5	X5, Y5, Z5
N6	X6, Y6, Z6

Figure 8.4: TIN model: the triangles are stored in a topological structure.

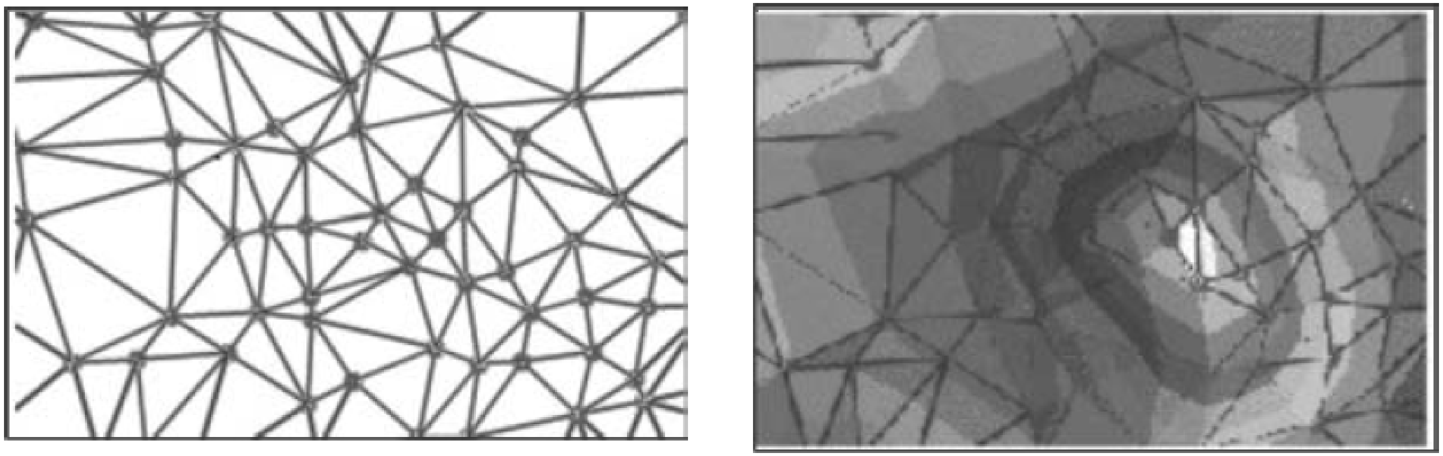


Figure 8.5: An example of topological structure in TIN model.

OTHER MODELS

- Isolines – continuous lines connecting points of the same elevation may represent terrain in much the same way as contour lines depict terrain on conventional maps (Figure 8.6).

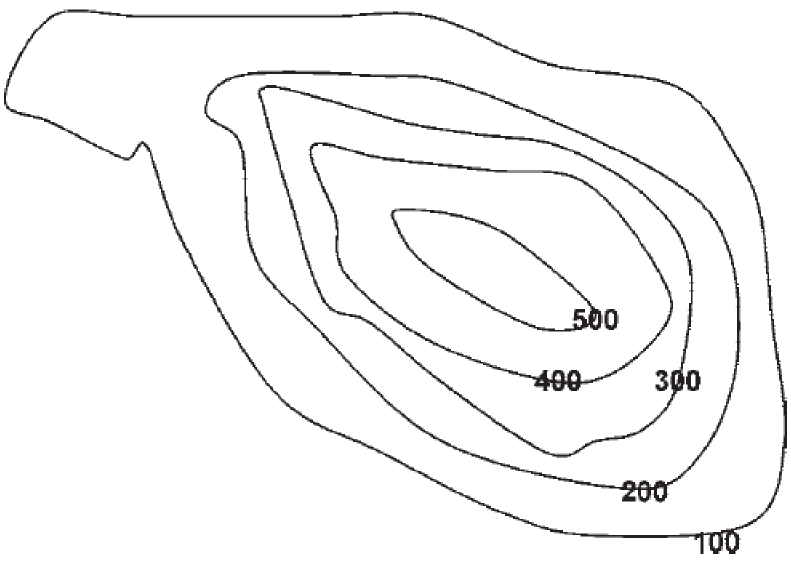


Figure 8.6: Lines that link points with the same terrain height are used to represent terrain surface, which corresponds to traditional elevation contours. However, this structure is poorly suited to the calculation of elevation values for new points.

  - The point densities should be greatest in those areas in which the terrain varies the most.
  - Smaller elevation increments between isolines result in greater accuracy of description.
- Parallel profile lines connecting points of varying elevation may be used to describe terrain.
  - The density of points along profile lines should be increased in areas where there are major variations in the terrain.
  - The closer the lines, the greater the accuracy of description.
- A combination of isolines and individual points may also be used to describe terrain,
  - especially when specifying such point features as peaks and valley floors, or vital terrain lines, such as the top and bottom of a fill.
- As mentioned previously, the grid and TIN models are best suited for calculation of the *z* value of new points.

PRACTICAL OBSERVATIONS

- GIS based on vector models can easily manipulate elevations stored as spaghetti data, but can handle elevation grid data less easily.
- Only a GIS based on a topological model can manipulate TIN data.
- Terrain data are usually compiled from survey point elevations, from isolines digitized from existing maps, or from photogrammetric point and/or line (contour or profile) registration.
- The ways in which data are represented and stored are decisive in determining the type and efficiency of the computations.
- The methods used to describe terrain surfaces may also be used to describe other continuously varying phenomena.
  - For example, population density, prevailing temperatures, or biomass production can be described quite simply by assigning the parameter involved to the z axis.

ACCURACY

- The accuracy of terrain descriptions is determined primarily by;
  - random variations in the terrain, spreading of measured points, distance between measured points
  - accuracy of points by the method of generating the model grid and triangular surfaces
  - the method used to interpolate between points in the model.
- For a grid model, the following degrees of accuracy are typical:

Source	Accuracy in elevation and ground plan
Surveying	±5cm
Photogrammetric data from	
1: 6000 images	±20 to 30 cm
Digitized 1: 1000 maps	±50 cm

- In the models in which cells and profiles are recreated from a point cloud, accuracy depends on the cell or profile density.

THREE-DIMENSIONAL OBJECTS

- A complete data model should be based on the three dimensions: ground, position, and elevation.
- The realization of three-dimensional objects in GIS still has theoretical and practical limitations.
  - It is a theoretical and mathematical problem to establish topology for three-dimensional objects.
  - It is also difficult to establish satisfactory routines for checking whether declared data exist in three-dimensional topology.
  - Specification of all objects in three dimensions can easily increase the amount of data.
  - It may also influence the techniques used to collect data.
- Even though relational databases support binary large objects (BLOBs) for storage of texture (building facades or similar), this type of data cannot be searched for as with other data.
  - Models can be constructed, but should in this case be carried out in systems for computer-aided design (CAD).

REPRESENTATION OF TIME

- The most usual way of handling the time factor in GIS is to look on time as an attribute to the objects in the same way as for other attributes.
  - However, this simple approach will not necessarily create a logical connection between the various time layers.
- Possible practical solutions will therefore be:
- 1. The attributes of the objects will be changed.
  - a. Historical data are stored only in fixed or variable time intervals (*e.g.,* every second year, every fifth year, etc.). The attribute values between these intervals may have to be interpolated.
  - b. All changes are registered and stored for selected types of objects (*i.e.,* historical data have to be preserved by date stamping).

- 2. The geometry of objects is changed.
  - a. Historical data are stored only in certain time intervals (*e.g.,* every second or tenth year). The geometry between these intervals may have to be calculated. As in the case of attribute values, the object type will decide where it is possible to interpolate new geometry.
  - b. All changes are registered and stored for selected object types (*i.e.,* all historical data have to be preserved by date stamping).

Box 17: Dimension of time in GIS

Integrating the dimension of time into GIS presents challenges. The main reasons for it is that, data about spatial object are not easily available for a continuous period, or data models and structures that allow us to record store and visualize in different temporal states are in infancy. This problem is bad enough when the geographic entity under investigation is fixed with respect to location, but it is more complex when the object is either mobile or changes its entity type through time. There are four type of temporal event. This provides an indication of the types of changes that may affect an entity:

- Continuous – these events go on throughout some interval of time.
- Majorative – these events go on most of the time.
- Sporadic – these events occur some of the time.
- Unique – events that occur only once.

Handling time in GIS: In a raster or vector layer based GIS, one option for handling time is to store multiple layers for the theme under investigation. The problem with this approach is that it generates lots of duplicate data. One solution is to store only information that changes to reduce the data storage requirements.

In a GIS using an object-oriented data model, a different approach is used. The various elements and attributes which make up an object can each be assigned a time tag.

- Registration in time intervals is more of a practical solution, where the main aim is to maintain rapid access to data and limited data volumes rather than realization of a basic data model, of which time is an integral part (Figure 8.7).

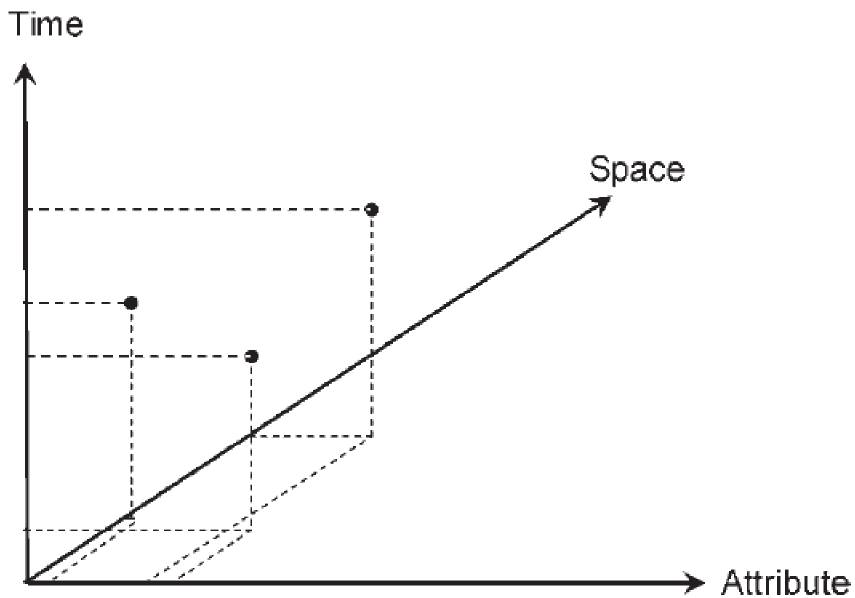


Figure 8.7: The real world changes occur in time, attribute and space, of which time is an integral part for realization of a basic data model.

- Even though a data model can be created that can handle the time factor satisfactorily, we are, in practice, reliant on changes being registered and stored in the database within a reasonable time.

MODELS FOR MOVEABLE OBJECTS

- A considerable part of the real world consists of moveable objects:
  - vehicles on a road network that carry passengers or goods or water running over the terrain surface.
- Special models have therefore been developed to handle these conditions.

NETWORK MODEL

- The network model comprises road systems, power grids, water supply, sewerage systems, and the like, all of which transport movable resources.
  - The most usual type of network is road systems.
- For most purposes, reality can be simplified to a model that can handle two different situations:
  - I. Displacement of resources or objects from one place to another.
  - II. Allocation of resources or objects from or to a center.

- Network model is based on geometry and attributes (Figure 8.8).
  - The geometry in the network is represented by lines consisting of connected lines of vector data.
  - The model is based on three basic relationships:
    - I. Continuous, connected networks
    - II. Rules for displacements in a network
    - III. The possibility of attribute value accumulations due to displacements
  - Attributes are connected to links and nodes and consist of two main categories.
    - One sets conditions for transfer in the network, while the other specifies which resistance occurs at different locations in the network (Table 8.1).

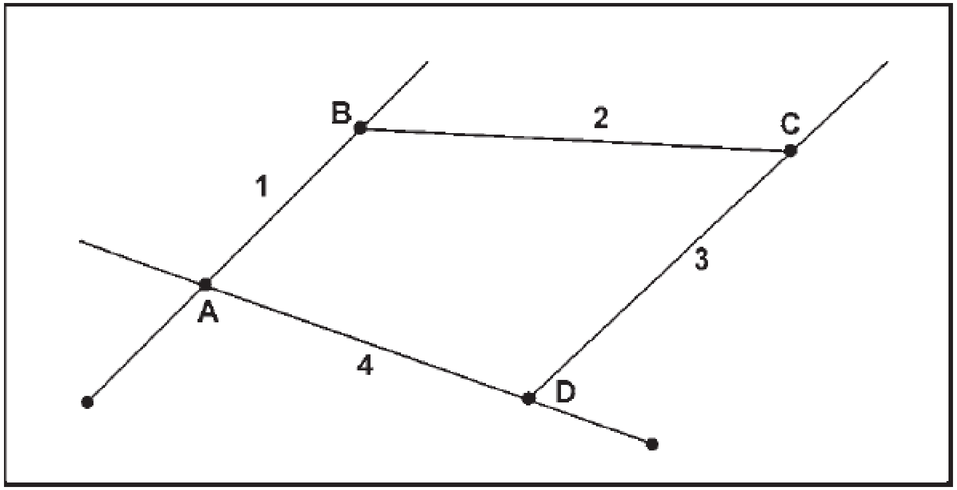


Figure 8.8: A road network splits into links and nodes, here link is a line without logical intermediate intersection and a node is an intersection point where two or more links meet or a start/end point.

Table 8.1: Attributes are attached to links and nodes and the resulting data are displayed in tables.

Link	Distance	Restriction	Node	Resistance
1	5 kilometers	25 (km/h)	1	2 minutes
2	4 kilometers	60 (km/h)	2	1 minute
3	5 kilometers	40 (km/h)	3	1.5 minutes
4	3 kilometers	30 (km/h)	4	2 minutes

- Once the model has been constructed, it is possible to simulate the quickest and/or the shortest route between points A and B based on the route with lowest accumulated resistance.
- Certain relationships need to be taken into consideration when establishing network models.
  - The initial step, how a road is divided into links, determines the nature of the nodes.
  - If all intersections, events, and features along the road result in nodes, the number of links may be enormous, resulting in the need for large storage capacity and slow data retrieval.

MODEL FOR MOVEMENT OVER SURFACES

- The free flow of resources in the terrain can also be modeled by using geometry and attributes.
  - such as water that flows on the surface.
- In this case, it is practical to use the full-coverage raster model instead of vector data.
- The geometry is thus represented with regular cells, and attributes are represented with coded values for each cell (Figure 8.9).
- Connective models of raster data may also be used to determine travel distances, to identify areas of given shapes and sizes, and so on.

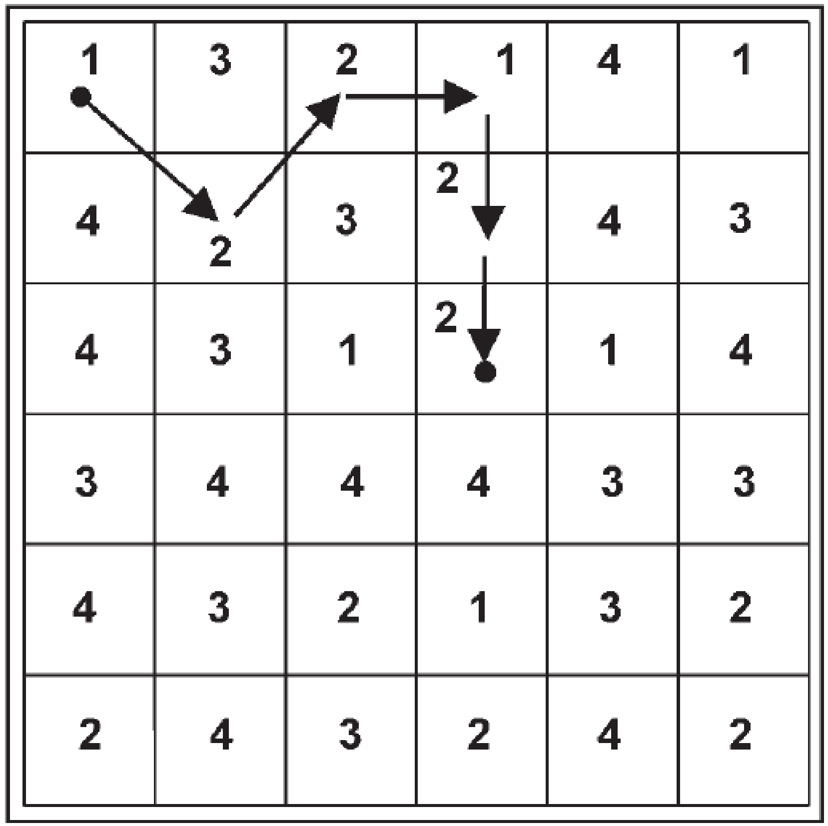


Figure 8.9: An example of optimizing route location on raster data.

COMBINATION OF MODELS

- Models can be better in combination than singly.
- The technique of multimedia integrates several types of models: vector, raster, 3D, time, and so on.
  - It helps the user to develop complete mental models of spatial problems and gives the user the ability to navigate in a GIS-derived information Space.
- An example network is given in figure 8.10.
  - nodes : are represented by small circles
  - arcs : are represented by the lines connecting the nodes.
  - node identifier (ID) : the number in each circle.
  - the distance (time) of traversing that arc : the numerical value next to each arc
  - A direction is also given for each arc.

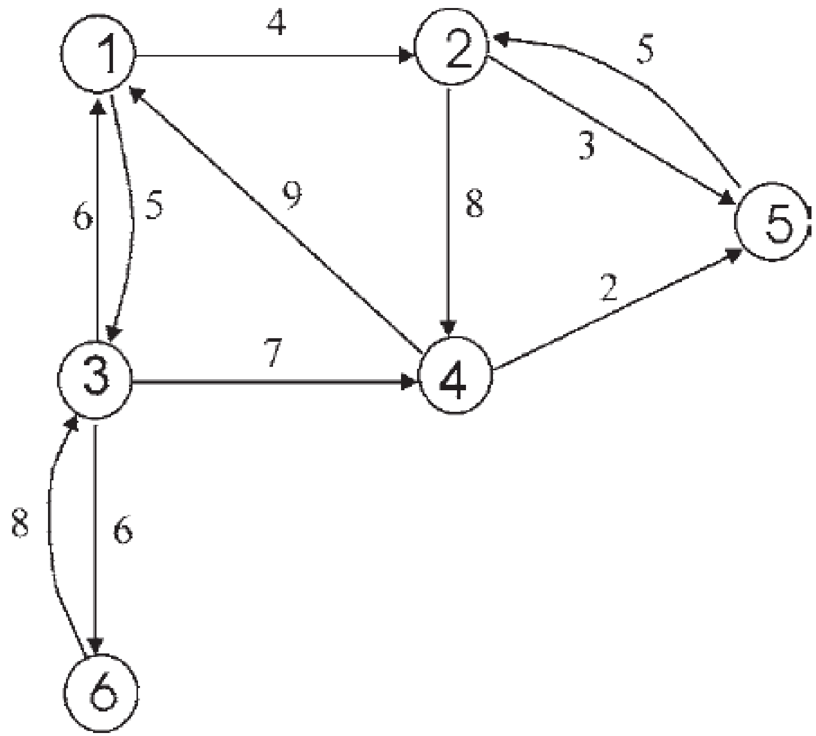


Figure 8.10: An example network.

THE REPRESENTATION OF NETWORKS

- The tradeoff of choosing a particular data structure is often between speed and storage space.
- In network analysis, commonly used representations of a network include:
  - ⇒ Node-Arc Incidence Matrix
  - ⇒ Node-Node Adjacency Matrix
  - ⇒ Adjacency Lists
  - ⇒ Forward and Reverse Star Representation
- The Node-Node Adjacency Matrix is the most basic form of representing network topology.
- The Forward and Reverse Star representation is the most efficient.

NODE-NODE ADJACENCY MATRIX

- Following is the Node-Node Adjacency Matrix representation of the network given in table 8.2:
  - The rows and columns in the matrix correspond to the nodes on the network.
  - A non-zero element in the ith row and jth column in the matrix represents the numerical value associated with arc *(i, j).*
  - A zero element in the matrix in the ith row and jth column in the matrix indicates that there exists no arc going from node *i to node j.*

Table 8.2: Node-Node Adjacency Matrix of the example network

0	4	5	0	0	0
0	0	0	8	3	0
6	0	0	7	0	6
9	0	0	0	2	0
0	5	0	0	0	0
0	0	8	0	0	0

- The storage space required is an^2 for a network with n nodes, where a is a constant.
- Node-Node Adjacency Matrix is very easy to implement and is suitable for dense networks.

FORWARD AND REVERSE STAR REPRESENTATION

- The Forward and Reverse Star representation stores the arcs eminating from the nodes in a single array.
- The Forward Star representation of a network,
  - A unique sequence number is assigned to each arc to obtain the ordering of the arc list.
  - Arcs are numbered in the following order:
    - ⇒ First arcs eminating from node 1 are numbered, then those from node 2, and so forth.
    - ⇒ Arcs eminating from the same node are numbered in an arbitrary fashion.
  - Once this list of ordered arcs is obtained, data associated with the arcs are stored in single arrays sequentially.
  - In addition to the list of ordered arcs, a pointer is also maintained for each node *i,* denoted by *pointer(i).*
    - – The numerical value associated with *pointer(i)* is the smallest-numbered arc eminating from node *i.*
    - – *pointer(1) = 1* and *pointer(n+1) = m+1.*

- Tables 8.3 and 8.4 shows the Forward Star Representation of the network example given in Figure 8.10.

Table 8.3: A list of order arcs in the Forward Star Representation.

Arc No.	Starting-node	Ending-node	Arc-length
1	1	2	4
2	1	3	5
3	2	4	8
4	2	5	3
5	3	1	6
6	3	4	7
7	3	6	6
8	4	5	2
9	4	1	9
10	5	2	5
11	6	3	8

Table 8.4: Pointer to each node in the Forward Star Representation.

Corresponding node	Element value
1	1
2	3
3	5
4	8
5	10
6	11
(7)	12

- The Reverse Star representation,
  - It is a data structure that provides an efficient means to determine the set of incoming arcs for any node.
    - The Forward Star representation can be used to efficiently determine the set of arcs outgoing from any node.
  - It can be constructed in a manner similar to the Forward Star representation, but the only difference is that *incoming arcs* at each node are numbered sequentially.

- Tables 8.5 and 8.6 are the Reverse Star Representation of the example network shown in Figure 8.10.

Table 8.5: A list of order arcs in the Reverse Star Representation.

Arc No.	Starting-node	Ending-node	Arc-length
1	3	1	6
2	4	1	9
3	1	2	4
4	5	2	5
5	1	3	5
6	6	3	8
7	2	4	8
8	3	4	7
9	4	5	2
10	2	5	3
11	3	6	7

Table 8.6: Pointer to each node in the Reverse Star Representation.

Corresponding node	Element value
1	1
2	3
3	5
4	7
5	9
6	11
(7)	12

- There is a significant amount of duplicate information when both the forward star and reverse star representations are stored in a computer.
  - To avoid the duplication, we only maintain a single array called *trace* which stores the arc numbers in the forward star representation.
- A compact forward and reverse star representation of the example network is given in Figure 8.11.
- The storage space required for the Forward and Reverse Star Representation is $an+bm$ for a network with n nodes and m arcs, where a and b are constants.
- The advantages of the Forward and Reverse Star Representation are that it saves space, it is efficient to manipulate, and it is suited for dense as well as sparse networks.

	Pointer
1	1
2	3
3	5
4	8
5	10
6	11
7	12

Arc No.	Starting Node	Ending Node	Arc Length
1	1	2	4
2	1	3	5
3	2	4	8
4	2	5	3
5	3	1	6
6	3	4	7
7	3	6	6
8	4	5	2
9	4	1	9
10	5	2	5
11	6	3	8

	Trace
1	5
2	9
3	1
4	10
5	2
6	11
7	3
8	6
9	8
10	4
11	7

	R-Pointer
1	1
2	3
3	5
4	7
5	9
6	11
7	12

Figure 8.11: Compact forward and reverse star representation of the example network.

REPRESENTATION OF NETWORK ATTRIBUTES

- The key to network representation is to represent nodes, arcs and network topology efficiently.
- Once the nodes, arcs, and network topology are efficiently represented, other data and information associated with nodes, arcs, stops, centers, and turns can be represented as attributes either associated with nodes or arcs.

COMPUTATION OF SHORTEST PATHS ON A NETWORK

- The computation of shortest path algorithms is a vital component of any network analysis task.
  - Shortest paths are often needed as input to ‘higher level’ models.
- Existing shortest path algorithms can be categorized into two groups: *label-setting* and *label-correcting.*
  - Both groups of algorithms are iterative and both employ the *labeling method* in computing one-to-all (one node to all other nodes) shortest paths.
  - They differ in the ways in which they update the estimate (*i.e.,* upper bound) of the shortest path distance and in the ways in which they converge to the final optimal one-to-all shortest paths.

- label-setting algorithm
  - the final optimal shortest path distance from the source node to the destination node is determined once the destination node is scanned.
- label-correcting algorithm
  - It treats the shortest path distance estimates of all nodes as temporary and converges to the final one-to-all optimal shortest path distances until its final step when the shortest paths from the source node to all other nodes are determined.
- A key operation in many shortest path algorithms is the *labelling method.*

Common network operations

- Pathfinding
  - the process to find the shortest, least cost, or most efficient path or tour on a network.
- Tracing
  - the process to determine a connected portion of a network that are either flow from this connected portion of the network to a given node or flow from a given node to this connected portion of the network.
- Allocation
  - the process to assign portions of a network to a location (*e.g., a center)* based on some given criteria.

Common network applications

- Geocoding
  - the process for building a relationship between locational data in a database and street address data that are normally in a tabular format.
  - *e.g.,* in retail analysis, customers' addresses can be used to create maps showing locations of different customers with different shopping behaviors.
- Location-allocation
  - the process of determining the optimal locations for a given number of facilities based on some criteria and simultaneously assigning the population to the facilities.
  - *e.g.*, the determination of locations for retail stores, restaurants, banks, factories, and warehouses or the choice of locations for libraries, hospitals, post offices, and schools can be supported by analysis results from location-allocation models.

- Business logistics
  - The optimization of vehicle routing and delivery scheduling
  - The combined power of GIS and network analysis makes GIS an ideal environment for analyses related to business logistics.
- Spatial interaction and gravity modelling
  - The interaction between different locations in geographic space and the mathematical modelling of the interaction are important in application areas such as transportation and retail analyses.
  - Gravity models are commonly used to support these analyses. Gravity modelling can be conveniently supported through network analysis in a GIS environment.

- Dynamic segmentation
  - Dynamic segmentation is a particular network model used to represent, analyze, query, and display linear features.
  - Dynamic segmentation vs. the network representations
    - dynamic segmentation has the flexibility to associate an attribute to a portion of an arc or several arcs (*e.g.,* through the definition of a route).
  - Dynamic segmentation is commonly used to model linear features such as highways, river networks, power lines, city streets, and telephone lines.