GIS

A Computing Perspective

Third Edition

Edited by Matt Duckham, Qian (Chayn) Sun and
Michael F. Worboys

First published 1995

ISBN: 978-1-4665-8719-9 (hbk)
ISBN: 978-1-032-53973-7 (pbk)
ISBN: 978-0-429-16809-3 (ebk)

Chapter 3

Fundamental Spatial Concepts

DOI: 10.1201/9780429168093-3

The funder for this chapter is Matt Duckham

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

FUNDAMENTAL SPATIAL CONCEPTS

3

SECTIONS

Many different representations of space are needed for a GIS. This chapter explores the fundamental concepts and relationships underlying these different spaces. The topics covered in this chapter provide a guide for learning to:

- manipulate **Euclidean geometry**, in which **distances**, **angles**, and **coordinates** may be defined;
- understand **topology**, and its construction from the basic concept of a **neighborhood**;
- construct simple representations of space using **sets** and **graphs**; and
- appreciate how a **metric space** formalizes the concept of distance between points in space.

SPACE is a term that is difficult to precisely define (see Box 3.1 on the next page). We all have an intuitive idea about the concrete space in which our bodies move. In the context of GIS, we normally use the term "space" to refer to "geographic space": the structure and properties of the relationships between locations at the Earth's surface. In this chapter we examine "space" more carefully by considering the different ways of representing and reasoning about geographic space.

A fundamental concept underlying these different representations is that of *geometry*. A *geometry* provides a formal representation of the abstract properties and structures within a space. Modern treatments of geometry are founded on the notion of *invariance*: geometries can be classified according to the group of transformations of space under which their propositions remain true. This idea was first proposed in 1872, by the German mathematician Felix Klein in his inaugural address to the University of Erlangen (known as the *Erlangen Program*).

geometry

invariance

To illustrate, consider a space of three dimensions and our usual notion of distance between two points. Then a geometry is formed by the set of all transformations that preserve distances (that is, for which distance is an *invariant*). Into this set would fall translations and rotations, because the distance between two points is the same before and after a translation or rotation is effected. Scalings (enlargements) would not be members because they usually change distance. Looking at this the other way around provides us with a definition of a geometry as the study of the invariants of a set of

Box 3.1: What is space?

The notion of space is not easy to define. Gatrell (1991) defines space as "a relation defined on a set of objects," which includes just about any structured collection. This definition is too general to describe the spaces of interest in a GIS. A distinction is often made between the space we can apprehend using our visual perception, termed *perceptual*; and space that is too big for humans to observe all at once, termed *transperceptual* (see Kuipers, 1978; Montello, 1993). Zubin (1989) provides more detail and distinguishes four types of space: A-space contains manipulatable everyday objects, like phones and books; B-space contains objects larger than humans, but still observable from a single perspective, like buildings and buses; C-space contains geographic scenes that are too large to apprehend at one time, like landscapes; D-space contains objects that are too large for any human to truly experience, like the solar system or the galaxy. Freundschuh & Egenhofer (1997) give a full overview and synthesis of the different classifications of space that have been proposed. In GIS, we are primarily interested in Zubin C-spaces, although there is also growing interest in integrating information from B- and even A-spaces, as well as other types of virtual spaces, such as the Internet and cyberspace (an area of study termed "cybergeography," see Dodge & Kitchin, 2002; Fabrikant & Buttenfield, 2001; Kwan, 2001).

transformations. Thus the invariants of the set of translations, rotations, and scalings include angle and parallelism, but not distance.

Coordinatized Euclidean geometry provides a view of space that is intuitive, at least in Western culture, so this is our starting point (Section 3.1). In later sections we discuss the most primitive space of all—just collections of objects with minimal structure (Section 3.2)—and proceed to build up to richer geometries (Sections 3.3–3.4). Because of the nature of the topic, the treatment will of necessity sometimes be abstract and formal, but examples are provided along the way.

3.1 Euclidean space

Euclidean space

origin

axis

Cartesian plane

Spatial phenomena are most commonly modeled as embedded in a physical 2- or 3D space, called *Euclidean space*. We assume here for simplicity a 2D model, although all the concepts in this section can be generalized to higher dimensional spaces. For the 2D Euclidean plane, we can set up a coordinate frame consisting of a fixed, distinguished point (*origin*) and a pair of orthogonal lines (*axes*), intersecting in the origin. This coordinatized model of the Euclidean plane, known as the *Cartesian coordinate plane* or simply the *Cartesian plane*, transforms spatial properties into properties of tuples of real numbers and is the primary focus of this section.

3.1.1 Point objects

point

vector

A *point* in the Cartesian plane has associated with it a unique pair of real numbers (x, y) measuring its distance from the origin in the direction of each axis, respectively. The collection of all such points is often written as $\mathbb{R}^2$. It is often useful to view Cartesian points (x, y) as *vectors*, measured from the origin to the point (x, y), having direction and magnitude and denoted by a directed line segment (Figure 3.1). Vectors may be added, subtracted, and multiplied by scalars according to the rules:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$
$$(x_1, y_1) - (x_2, y_2) = (x_1 - x_2, y_1 - y_2)$$
$$k(x, y) = (kx, ky)$$

To distinguish clearly between points as locations and vectors as displacements, we use the over-arrow notation $\vec{a}$ to identify vectors (see Figure 3.1). Given a point vector, $\vec{a} = (x, y)$, we may form its *norm*, defined as follows:

norm

$$\|\vec{a}\| = \sqrt{(x^2 + y^2)}$$

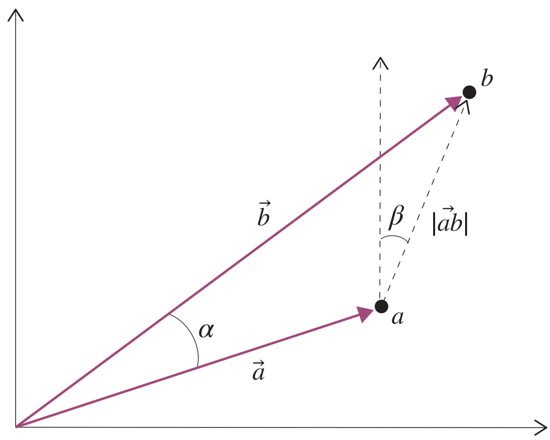


Figure 3.1: Distance $|\vec{ab}|$, angle α, and bearing β between points (vectors) $\vec{a}$ and $\vec{b}$ in the Euclidean plane

In a coordinatized system, measures of distance may be defined in a variety of ways (see Section 3.4.2 on metric spaces). A *Euclidean plane* is a Cartesian coordinate plane with the particular measures of distance and angle given below.[1] These measures form the foundation of most school geometry courses and refer to the "as the crow flies" concept of distance. Given points (vectors) $\vec{a}, \vec{b}$ in $\mathbb{R}^2$, the distance from $\vec{a}$ to $\vec{b}$, $|\vec{ab}|$, is given by:

Euclidean plane

[1] Conversely, a Cartesian plane is a Euclidean plane parameterized with coordinates.

$$|\vec{ab}| = \|\vec{b} - \vec{a}\|$$

Suppose that the points $\vec{a}, \vec{b}$ in $\mathbb{R}^2$ have coordinates (x_a, y_a) and (x_b, y_b), respectively. Then the distance $|\vec{ab}|$ is precisely the Pythagorean distance familiar from school days, given by:

$$|\vec{ab}| = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}$$

The *angle* α (see Figure 3.1) between vectors $\vec{a}$ and $\vec{b}$ is given as the solution of the trigonometrical equation:

angle

$$\cos\alpha = \frac{x_a x_b + y_a y_b}{\|\vec{a}\| \times \|\vec{b}\|}$$

bearing

The *bearing* β (see Figure 3.1) of $\vec{b}$ from $\vec{a}$ is given by the unique solution in the interval $0 \le \theta < 360$ of the simultaneous trigonometrical equations:

$$\sin\theta = \frac{x_b - x_a}{|\vec{ab}|}$$
$$\cos\theta = \frac{y_b - y_a}{|\vec{ab}|}$$

3.1.2 Line objects

Line objects are very common spatial components of a GIS, representing the spatial attributes of objects and their boundaries. The definitions of some commonly used types of line are introduced below in a parameterized form. The parameter λ is a real number constrained to vary over a given range (the range depending upon the object type). As λ varies, so the set of points constituting the linear object is defined.

line
line segment
half line

- Given two distinct points (vectors) $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^2$, the *line* incident with $\vec{a}$ and $\vec{b}$ is defined as the point set $\{\lambda\vec{a} + (1-\lambda)\vec{b}\}$.
- Given two distinct points (vectors) $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^2$, the *line segment* between $\vec{a}$ and $\vec{b}$ is defined as the point set $\{\lambda\vec{a} + (1-\lambda)\vec{b} \mid 0 \le \lambda \le 1\}$.
- Given two distinct points (vectors) $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^2$, the *half line* radiating from $\vec{b}$ and passing through $\vec{a}$ is defined as the point set $\{\lambda\vec{a} + (1-\lambda)\vec{b} \mid \lambda \ge 0\}$.

Figure 3.2 shows some examples.

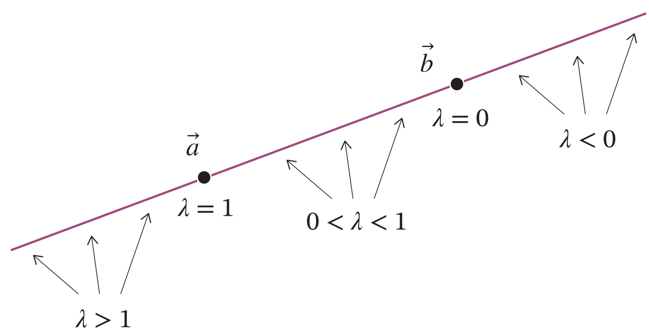


Figure 3.2: Parameterized representation of linear objects

Straight lines may also be specified by a single bivariate polynomial equation of degree one ($ax + by = k$). Not only straight lines are of interest for GIS, though. Higher-degree bivariate polynomials specify further classes of one-dimensional (1D) objects. Thus, polynomials of degree two (quadratics of the form $ax^2 + bxy + cy^2 = k$) specify conic sections, which could be circles, ellipses, hyperbolas, or parabolas.

3.1.3 Polygonal objects

polyline

A *polyline* in $\mathbb{R}^2$ is defined to be a finite set of line segments (called *edges*) such that each edge end-point is shared by exactly two edges, except possibly for two points, called the *extremes* of the polyline. If, further, no two edges intersect at any place other than possibly at their end-points, the polyline is called a *simple* polyline. A polyline is said to be *closed* if it has no extreme points. A (simple) *polygon* in $\mathbb{R}^2$ is defined to be the area enclosed by a simple closed polyline. The polyline forms the *boundary* of the polygon. Each end-point of an edge of the polyline is called a *vertex* of the polygon. Some possibilities are shown in Figure 3.3. An extension to the definition would allow a general polygon to contain holes, islands within holes, etc.

simple
polygon
boundary
vertex

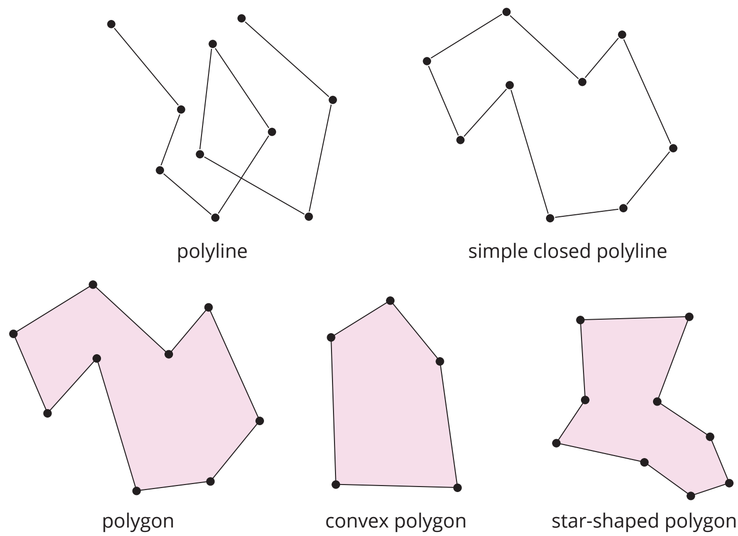


Figure 3.3: Polylines and types of polygons

Many different types of polygon have been defined in the computational geometry literature. A useful category is the set of *convex polygons*, each of which has an interior that is a convex set and thus has all its internal angles not greater than 180°. For a convex polygon, every interior point is visible from every other interior point in the sense that the line of sight lies entirely within the polygon. A *star-shaped polygon* has the weaker property that there exists at least one point which is visible from every point of the polygon (see Figure 3.3). Convexity is discussed further in Section 3.2.4.

convex polygon
star-shaped polygon

The definition of *monotone polygons* depends upon the concept of *monotone chains*. Let chain $C = [p_1, p_2, ..., p_n]$ be an ordered list of n points in the Euclidean plane. Then C is *monotone* if and only if there is some line in the Euclidean plane such that the projection of the vertices onto the line preserves the ordering of the list. Figure 3.4 shows monotone and non-monotone chains.

monotone

A polygon is a monotone polygon if its boundary can be split into two polylines, such that the chain of vertices of each polyline is a monotone chain. Clearly, every convex polygon is monotone. However, the converse of this statement is not true. It is not even true that every monotone polygon is star shaped, as may be seen from the example in Figure 3.5.

monotone polygon

A *triangulation* of a polygon is a partition of the polygon into triangles

triangulation

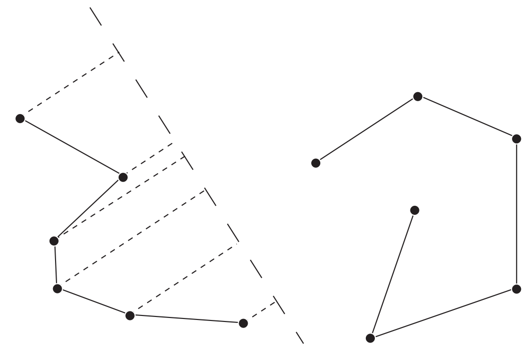


Figure 3.4: Monotone and non-monotone chains

that intersect only at their mutual boundaries. It is not too hard to show that a triangulation of a simple polygon with n vertices that introduces an extra m internal vertices (sometimes called *Steiner points*) will result in exactly $n + 2m - 2$ triangles.

Steiner point

3.1.4 Transformations of the Euclidean plane

This section describes some common transformations of the Euclidean plane. A *transformation* of $\mathbb{R}^2$ is a function from $\mathbb{R}^2$ to itself. Thus every point of the plane is transformed into another (maybe the same) point. Some transformations preserve particular properties of embedded objects:

transformation

congruences

Euclidean transformations (also termed *congruences*) preserve the shape and size of embedded objects. An example of a Euclidean transformation is a translation. For example, a translation of the point (x, y) through real constants a and b has the formula $(x + a, y + b)$.

similarity transformation

Similarity transformations preserve the shape but not necessarily the size of embedded objects. All Euclidean transformations are also similarities. An example of a similarity transformation is a scaling. Scaling a point (x, y) by real constants a and b has the formula (ax, by).

affine transformation

Affine transformations preserve the affine properties of embedded objects, such

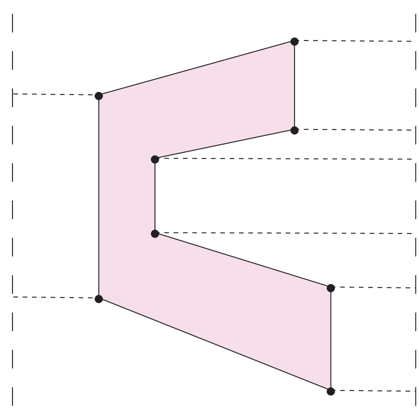


Figure 3.5: A monotone but not star-shaped polygon

as parallelism. All similarity transformations are also affine. Examples of affine transformations of the point (x, y) include *rotations*, e.g., through angle θ about the origin with the formula $(x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta)$; *reflections*, e.g., reflected in the line through the origin at angle θ to the x-axis with the formula $(x \cos 2\theta + y \sin 2\theta, x \sin 2\theta - y \cos 2\theta)$; and *shears*, e.g., parallel to the x-axis with real constant a and formula $(x + ay, y)$.

rotation

reflection

shear

Projective transformations preserve the projective properties of embedded objects. An intuitive idea of a central projection is the action of a point light source, sending a figure to its projection upon a screen. For example, the projection of a circle may result in an ellipse. All affine transformations are also projective.

projective transformation

Topological transformations (homeomorphisms or bicontinuous maps) preserve topological properties of embedded objects. We shall study this class in detail later.

3.2 Set-based spaces

The Euclidean plane is a highly organized kind of space, with many well-defined operations and relationships that can act upon objects within it. This section retreats to the much more rarefied realm of set-based space, and then gradually builds up more structure with topological and metric spaces.

3.2.1 Sets

The set-based model of space does not have the rich set of constructs of the Euclidean plane. The set-based model simply involves:

- The constituent objects to be modeled, called *elements* or *members*.
- Collections of elements, called *sets*. For computer-based models, such collections are usually finite, or at least countable.
- The relationship between the elements and the sets to which they belong, termed *membership*. We write $s \in S$ to indicate that an element s is a member of the set S.

element

set

membership

The set-based model is abstract and provides little in the way of constructions for modeling spatial properties and relationships. Nevertheless, sets are a rich source of modeling constructs (see "Russell's paradox," Box 3.2 on the following page) and are fundamental to the modeling of any spatial information. For example, relationships between different base units of spatial reference may be modeled using set theory. States or counties may be contained within (i.e., members of) a country; countries may themselves be members of continents; continents may be elements of the set of landmasses. Such hierarchical relationships are adequately modeled using set theory. Sometimes, areal units are not so easily handled. For example, in the US there is no simple set-based relationship between zip codes, used for the distribution of mail, and

Box 3.2: Russell's paradox

Even though the set concept expresses the most rudimentary relationships and structuring of space, it is surprisingly difficult to capture the essence of a set in a few words. A first attempt might be "a set is any collection of objects," but this lays itself open to Russell's paradox (after the British mathematician Bertrand Russell). Russell's paradox is sometimes explained in terms of a library index. Imagine that an assiduous librarian wishes to index every book in the library that does not contain a reference to itself (unlike this book, Duckham, Sun, & Worboys, 2023). The librarian can write down the names of books that do not refer to themselves in a new index book. To be complete, this index should list itself, as it does not contain a self-reference. But as soon as the librarian adds a self-reference to the index, that reference should be removed, because now the index book does cite itself! We can express this tortuous problem mathematically by considering the set S of all sets that are not members of themselves. If S is a member of itself, then by definition it is not a member of itself; on the other hand, if S is not a member of itself, then it must be a member of itself. Either way, we arrive at a contradiction. This paradox spurred considerable efforts in the 20th century to find a more adequate definition of a set, but no definition has proved completely satisfactory.

the administrative units such as city, county, and state (as already noted in Box 2.4 on page 58). This mismatch causes considerable problems when data referenced to one set of spatial units is compared or combined with spatial data referenced to a different set (discussed further in Box 8.5 on page 321).

In classical set theory an object is either an element of a particular set or it is not. There is no halfway house or degree of membership. If the binary on-off nature of the membership condition is relaxed, then it is possible to arrive at some more expressive models. Some of these models are important for modeling uncertainty in spatial information, a topic tackled in Chapter 10.

From the basic constructs of element, set, and membership, a large number of modeling tools may be constructed. We shall consider just a few here.

- equality — *Equality* is defined as a relationship between two sets that holds when the sets contain precisely the same members.
- subset — The relationship between two sets where every member of one set is a member of the second is termed *subset*. The relationship that set S is a subset of set T is denoted $S \subseteq T$.
- power set — A *power set* is the set of all subsets of a set. The power set of set S is denoted $\mathcal{P}(S)$.
- partition — A *partition* of a set is a set of subsets such that each and every member of the original set appears in exactly one subset in the partition.
- empty set — The *empty set* is the set containing no members, denoted $\emptyset$.
- cardinality — The number of members in a set is termed *cardinality*. The cardinality of set S is denoted $\#S$ or $|S|$.
- intersection — The *intersection* operation is a binary operation that takes two sets and returns the set of elements that are members of both the original sets. The intersection of sets S and T is denoted $S \cap T$.
- union — The *union* operation is a binary operation that takes two sets and returns the set of elements that are members of at least one of the original sets. The union of sets S and T is denoted $S \cup T$.
- difference — The *difference* operation is a binary operation that takes two sets and re-

turns the set of elements that are members of the first set but not the second set. The difference of sets S and T is denoted $S \backslash T$.

- The *complement* operation is a unary operation that when applied to a set returns the set of elements that are not in the set. The complement is always taken with reference to an (implicit) universal set. The complement of set S is denoted S'.

complement

Figure 3.6 shows by shading on set diagrams some of the Boolean set operations described above.

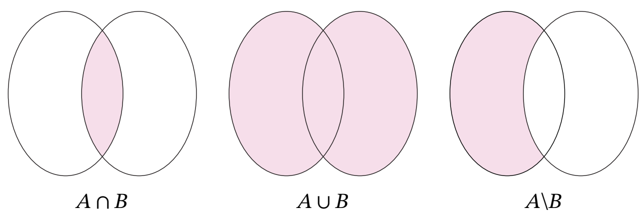


Figure 3.6: Set intersection, union, and difference

Some sets, particularly sets of numbers, are used so often that they have a special name and symbol. Some of these are listed in Table 3.1. The Boolean set $\mathbb{B}$ is used whenever there is a two-way choice to be made. This choice may be viewed as between on and off, true and false, or one and zero, depending upon the context. The set of integers $\mathbb{Z}$ is used in discrete models. Sometimes, only the positive integers are needed, written $\mathbb{Z}^+$.

For continuous models, the real numbers $\mathbb{R}$ are required. In fact, it is provably impossible to capture the reals completely using a computer; therefore, rational numbers are used in practice. Rational numbers are real numbers that can be expressed as a ratio of integers, e.g., 123/46. Like real numbers, rational numbers have the property that they are *dense*: that is, between any two rationals a and b, where a is greater than b, no matter how close a and b are, it is always possible to find a third rational c such that $a < c < b$ (see Figure 3.7). In this way, the rationals are a useful approximation for modeling continuous processes. Any particular computer implementation will place a restriction upon the precision of the rationals.

dense set

Name	Symbol	Description
Booleans	$\mathbb{B}$	Two-valued set of true/false, 1/0, or on/off
Integers	$\mathbb{Z}$	Positive and negative numbers, including zero
Rationals	$\mathbb{Q}$	Ratios of two integers
Reals	$\mathbb{R}$	Measurements on the number line
Real plane	$\mathbb{R}^2$	Ordered pairs of reals, the Cartesian plane
Closed interval	$[a, b]$	All reals between a and b (including a and b)
Open interval	$]a, b[$	All reals between a and b (excluding a and b)
Semi-open interval	$[a, b[$	All reals between a and b (including a and excluding b)

Table 3.1: Some distinguished sets

Some subsets of the reals are particularly useful. *Intervals* are connected sets of real numbers. They may or may not contain their end-points and are then called *closed* or *open*, respectively. It is also possible for intervals to be

interval

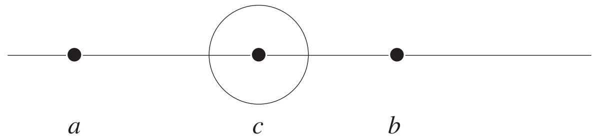


Figure 3.7: The rationals are dense

closed at one end and open at the other. Such intervals are called *semi-open* (or *semi-closed*). Thus the closed interval $[2, 5]$ denotes the set of all real numbers not less than two and not greater than five. The semi-open interval $]2, 5]$ denotes the set of all real numbers greater than two but not greater than five.

3.2.2 Relations

Sets on their own are limited in their application to modeling. Life becomes more interesting when relationships between two or more sets are modeled. In order to provide these tools, a further set-based operation is defined.

product

- The binary operation *product* returns the set of ordered pairs, whose first element is a member of the first set and second element is a member of the second set. The product of sets S and T is denoted $S \times T$.

An example of a product set is the set of points in the Cartesian coordinate plane, introduced in Section 3.1.1. Each point in the Cartesian plane is represented as an ordered pair of real numbers, measuring the point's distance from a given origin in the direction of the two axes. In set-theoretic terms, the collection of all such points is a product set, being the product of the set of real numbers with itself. This set is denoted by $\mathbb{R} \times \mathbb{R}$ or $\mathbb{R}^2$. This notion may be generalized to Cartesian coordinate 3-space $\mathbb{R}^3$ or indeed Cartesian n-space $\mathbb{R}^n$. A second example would be the set of points in the *unit square*, the square with vertices $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$ in the Cartesian plane (Figure 3.8). This set is the product of two intervals, $[0, 1] \times [0, 1]$. It is the case that $[0, 1] \times [0, 1] \subseteq \mathbb{R}^2$.

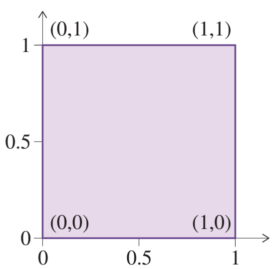


Figure 3.8: The unit square

Product spaces provide a means of defining relationships between objects. Hence the following construction:

binary relation

- A *binary relation* is a subset of the product of two sets, whose ordered pairs show the relationships between members of the first set and members of the second set.

Suppose that $S = \{\text{Fred}, \text{Mary}\}$ and $T = \{\text{apples}, \text{oranges}, \text{bananas}\}$. Then the formal way of expressing the relationship "likes" between people and fruit is by means of a relation, constructed as a set of ordered pairs $\{(\text{Fred}, \text{apples}), (\text{Fred}, \text{bananas}), (\text{Mary}, \text{apples})\}$. This set of ordered pairs is a subset of the entire product space $S \times T$, containing six pairs.

Relations may in general apply to the product of more than two sets (as in the relations in the relational database model). However, even binary relations have many modeling applications. For example, suppose that we are given two sets: the places of interest in The Potteries P and town centers in The Potteries T. A relation $R_1 \subseteq P \times T$ might provide for each a place of interest $p \in P$ the nearest town center $t \in T$ to p. Figure 3.9 shows the relation R_1 diagrammatically. The fact that Hanley is the nearest town center to the City Museum is indicated in Figure 3.9 with an arrow from the City Museum to Hanley, and would be written (City Museum, Hanley) $\in R_1$.

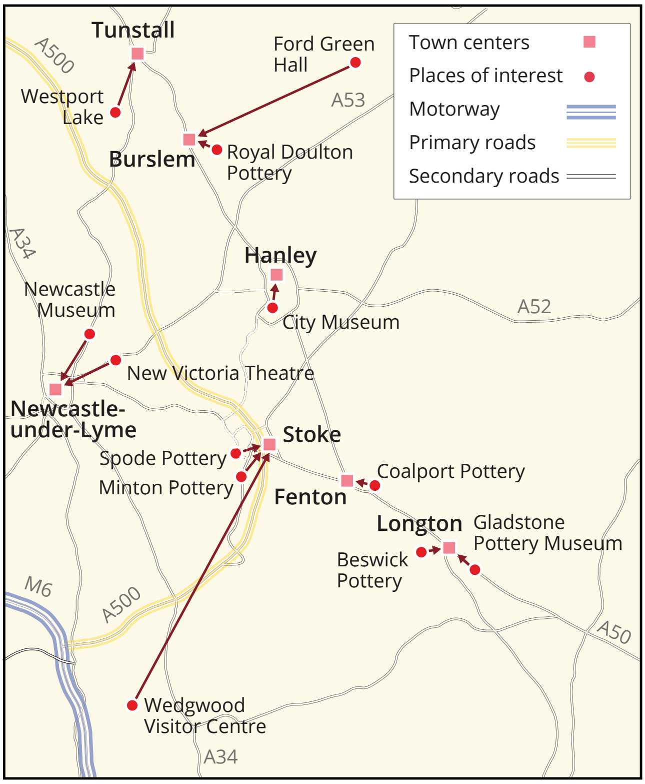


Figure 3.9: Binary relation on the sets of Potteries places of interest and town centers

As we have seen with the relation $\mathbb{R}^2$, a relation may act upon a single set. Suppose for the set of places P, a place of interest $p \in P$ is related to $p' \in P$ if p' is the nearest place of interest to p. This relation $R_2 \subseteq P \times P$ is shown in Figure 3.10. The fact that Spode Pottery is the nearest place of interest to the City Museum is indicated with an arrow from the City Museum to Spode Pottery, and written (City Museum, Spode Pottery) $\in R_2$.

Some binary relations between objects of the same set have special properties.

- A binary relation R on X where every element of the set $x \in X$ is related to itself, $(x, x) \in R$, is termed a *reflexive* relation. A relation is *irreflexive* if every element of the set is not related to itself, $(x, x) \notin R$ for all $x \in X$.
- A binary relation R where if x is related to y, then y is related to x is termed a *symmetric* relation. A relation is *antisymmetric* if for every pair of elements, x and y in X, if x related to y then y is not related to x.
- A binary relation R where if x is related to y and y is related to z, then x is related to z is termed a *transitive* relation.

reflexive

irreflexive

symmetric

antisymmetric

transitive

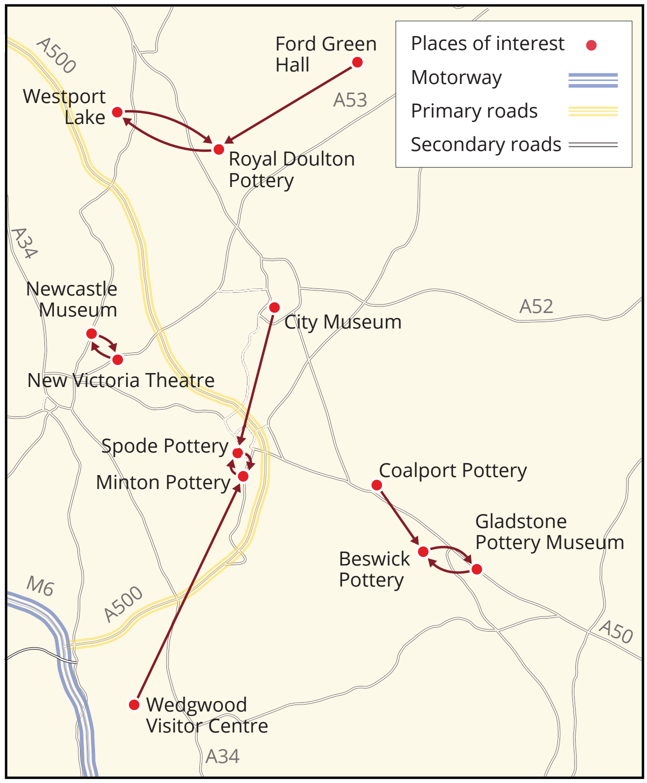


Figure 3.10: Binary relation on the sets of Potteries places of interest

In the example of Figure 3.10, there is an implicit assumption that a place cannot be related to itself; therefore, the relation is not reflexive and is also irreflexive. It is less obvious whether or not the relation "has nearest place of interest" is symmetric. For example, Spode Pottery is nearest to Minton Pottery and vice versa. However, the nearest place to the City Museum is also Spode Pottery while the converse is not true, so the relation "has nearest place of interest" cannot be symmetric. Also, the relation is not transitive because if p has nearest place of interest q and q has nearest place of interest r, then p cannot have nearest place of interest r unless $q = r$ or $q = p$ (and we have already disallowed a place being nearest to itself).

equivalence relation

A binary relation that is reflexive, symmetric, and transitive is termed an *equivalence relation*. Another useful class of relations is order relations, which satisfy the transitive property and are also irreflexive and antisymmetric. These further properties and partial orders are discussed in Section 3.4.1.

3.2.3 Functions

function

domain

codomain

A *function* is a special type of relation which has the property that each member of the first set relates to exactly one member of the second set. Thus, a function provides a rule that transforms each member of the first set, called the *domain*, into a member of the second set, called the *codomain*. We use the notation:

$$f : S \to T$$

to mean that f is a function, S is the domain, and T is the codomain. If the result of applying function f to element x of S is y, we write $y = f(x)$ (read "y equals f of x") or $f : x \mapsto y$ (read "f maps x to y"). Figure 3.11 shows schematically the relationship between function, domain, codomain, and image (defined shortly).

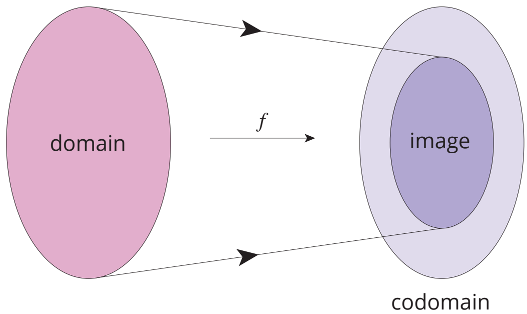


Figure 3.11: An abstract function f

For example, suppose that S is the set of points on a spheroid and T the set of points in the plane. A map projection is essentially a function with domain S and codomain T. Usually, map projection functions are constructed to leave some properties of the sets invariant, for example, lengths, angles, or areas. The familiar Universal Transverse Mercator (UTM) projection is a function that preserves angles. The UTM projection has a further property that any two different points in the domain are transformed to two distinct points in the codomain. Such a function is said to be an *injection*. Map projections are discussed further in the context of cartography, in Section 8.2.3.

injection

Not all functions are injections. Consider, for example, the function that computes the square of any integer. Then, because both 2 and −2 square to the same number, 4, the square function is not an injection. The two relations described in the previous section (Figures 3.9 and 3.10) are also functions (i.e., each member in the first set relates to exactly one member in the second set) but not injections.

The set of outputs from the application of a function to elements in its domain will form a subset of the codomain. Thus the UTM map projection will project a spheroid onto a finite subset of the plane. The set of all possible outputs is called the *image* or *range* of the function. If the image actually equals the codomain, then the function is termed a *surjection*. The UTM projection is not a surjection because the spheroid projects onto only a finite portion of the plane and not the whole plane. A function that is both a surjection and an injection is termed a *bijection*.

image

surjection

bijection

Injective functions have the special property that they have inverse functions. Consider again the UTM projection from the spheroid to the plane. Given a point in the plane that is part of the image of the transformation, it is possible to reconstruct the point on the spheroid from which it came. This reversal of the process allows us to form a new function whose domain is the image of the UTM, and which maps the image back to the spheroid. This is called the *inverse* function.

inverse function

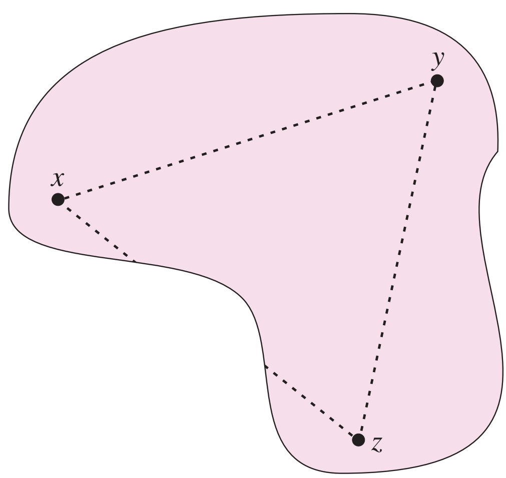
3.2.4 Convexity

Convexity has already been discussed for polygons and is now generalized to the same property for arbitrary point-sets in the Euclidean plane. The same notion is also meaningful in Euclidean 3-space, and the definition is easily extensible to this case. The essential idea is that a set is convex if every point is visible from every other point within the set. To make this idea precise, we define visibility and then convexity. Let S be a set of points in the Euclidean plane. Then:

- the point x in S is *visible* from point y in S if either $x = y$ or it is possible to draw a straight-line segment between x and y that consists entirely of points of S;
- the point x in S is an *observation point* for S if every point of S is visible from x;
- semi-convex — the set S is *semi-convex* (*star-shaped* if S is a polygonal region) if there is some observation point for S; and
- convex — the set S is *convex* if every point of S is an observation point for S.

Figure 3.12 shows the visibility relation within a set between three points x, y, and z. Points x and y are visible from each other, as are points y and z. But, points x and z are not visible from each other. The visibility relation is reflexive and symmetric, but not transitive. Also, observe that any convex set must be semi-convex (although the converse is not necessarily true). Figure 3.13 gives some examples of sets that are not semi-convex, semi-convex but not convex, and convex, respectively.

Figure 3.12: Visibility between points x, y, and z



The intersection of a collection of convex sets is also convex, and therefore any collection of convex sets closed under intersection has a minimum member. This leads to the definition of a *convex hull* of a set of points S in $\mathbb{R}^2$ as the intersection of all convex sets containing S. From above, the convex hull must be the unique smallest convex set that contains S (Figure 3.14). A convex hull of a finite set of points is always a polygonal region.

convex hull

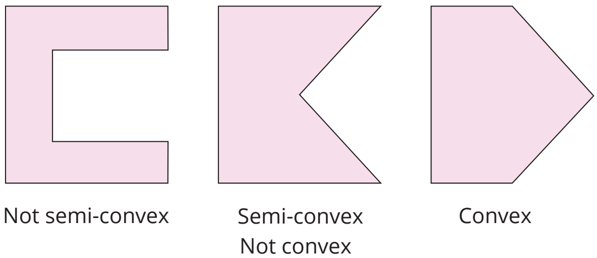


Figure 3.13: Degrees of convexity in point sets

3.3 Topology of space

The word "topology" derives from Greek and literally translates to "study of form." This may be contrasted with the word "geometry," which translates to "measurement of the Earth." *Topology* is a branch of geometry concerned with a particular set of geometrical properties—those that remain invariant under topological transformations. So, what is a topological transformation? In fact, topology is a general notion that can be applied to many different kinds of space.

topology

3.3.1 Topological spaces

To gain some topological intuition, imagine the Euclidean plane to be an unbounded sheet of fine-quality rubber that has the ability to stretch and contract to any desired degree. Imagine a figure drawn upon this rubber sheet. Allow this sheet to be stretched but not torn or folded. Certain properties of the original figure will remain, while others will be lost. For example, if a polygon were drawn upon the sheet, and a point was drawn inside the polygon, then after any amount of stretching the point would still be inside the polygon; on the other hand, the area of the polygon may well have changed. The property of "insideness" is a topological property (because it is invariant under rubber sheet transformation), while "area" is not a topological property. The transformation induced by stretching a rubber sheet is called a *topological transformation* or *homeomorphism*. Thus we have the following definitions:

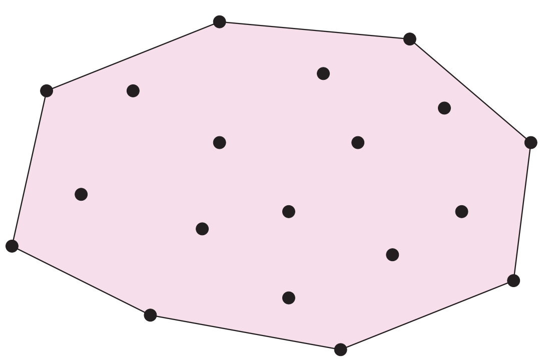


Figure 3.14: Convex hull of a point set

Topological properties properties that are preserved by topological transformations of the space.

Topology the study of topological transformations and the properties that are left invariant by them.

Table 3.2 lists some topological and non-topological properties of objects embedded in the Euclidean plane. For object types we take points, arcs, and area. Later, these types will be defined more carefully; for now, assume the obvious meanings to arc (possibly curved linear object) and area (2D piece of plane, possibly with holes and islands).

Table 3.2: Topological and non-topological properties of objects in the Euclidean plane with the usual topology

Topological	A point is at an end-point of an arc
	A point is on the boundary of an area
	A point is in the interior/exterior of an area
	An arc is simple
	An area is open/closed/simple
	An area is connected
Non-topological	Distance between two points
	Bearing of one point from another point
	Length of an arc
	Perimeter of an area

point-set topology

combinatorial topology

The discussion will cover two branches of topology: point-set (or analytic) topology and combinatorial (or algebraic) topology. In *point-set topology*, the focus, as one would expect, is on sets of points and in particular on the concepts of neighborhood, nearness, and open sets. We shall see that several important spatial relationships, such as connectedness and boundary may be expressed in point-set topological terms. The other important branch of topology, which has been applied to spatial data modeling, is *combinatorial topology*, in particular the theory of simplicial complexes. Even though these ideas may at times seem rarefied and far removed from spatial databases, in fact they do form the basis of several prominent conceptual models used in GIS. It is certainly true that the construction of sound and lasting generic spatial models relies on knowledge of the material that is introduced here.

3.3.2 General point-set topology

It is possible to define a topological space in several different ways. The definition below is based upon a single primary notion, that of *neighborhood*. A set upon which a well-defined notion of neighborhood is provided is then a topological space. It turns out that all the familiar topological properties are definable in terms of the single concept of neighborhood. Given any set, the approach is to define a collection of its subsets, constituting the neighborhoods, and thus provide a neighborhood topology on the set. The formal definition is now given.

topological space

neighborhood

- Let S be a given set of points. A *topological space* is a collection of subsets of S, called *neighborhood*, that satisfy the following two conditions.

$T1$ Every point in S is in some neighborhood.
$T2$ The intersection of any two neighborhoods of any point x in S contains a neighborhood of x.

Figure 3.15 shows the two conditions of a topological space in action. Neighborhoods are shown surrounding each point in the set, and two neighborhoods are shown overlapping and containing in their intersection another neighborhood.

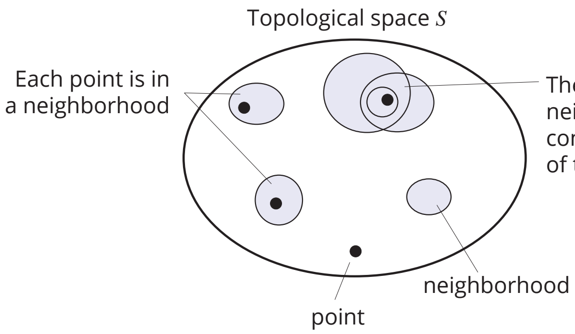


Figure 3.15: Points and neighborhoods in a topological space

By far, the most important example of a topological space, for our purposes, is the *usual topology* for the Euclidean plane. The usual topology is so called because it is the topology that naturally comes to mind with the Euclidean plane, and it corresponds to the rubber-sheet topology introduced earlier. It is possible to define other (unusual?) topologies on the Euclidean plane (see Box 3.3 on the next page), and one important example of these, travel time topology, follows the usual topology.

The usual topology of the Euclidean plane Define an *open disk* to be a set of points bounded by a circle in the Euclidean plane, but not including the boundary. An example is given in Figure 3.16. The convention is that a hatched line at the boundary indicates that the boundary points are excluded, whereas a continuous line indicates that boundary points are included.

open disk

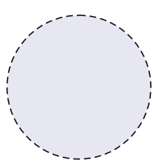


Figure 3.16: An open disk in the Euclidean plane with the usual topology

Define a neighborhood of a point x in $\mathbb{R}^2$ to be any open disk that has x within it (see Figure 3.17). We now show that, under this definition of neighborhood, $\mathbb{R}^2$ is a topological space. To check that condition $T1$ for a topological space holds, it is sufficient to observe that every point in $\mathbb{R}^2$ can certainly be surrounded by an open disk. For $T2$, take any point x in $\mathbb{R}^2$ and surround it by two of its neighborhoods (open disks with x inside), N_1 and

Box 3.3: Topological spaces

There are many common topological spaces other than the usual topology. For example, let S be any set, and define the neighborhoods to be *all* the subsets of S. It is easy to confirm that this neighborhood structure defines a topological space (by checking conditions $T1$ and $T2$). The space is called the *discrete topology*, because the smallest neighborhood of each point x in S is $\{x\}$, so each point in S is separated by a neighborhood from every other point. Another example of an extreme topology occurs if we let S be any set, and define the only neighborhood to be the set S itself. Again, this may easily be verified to be a topological space, called the *indiscrete topology*. The usual topology of the Euclidean plane may be scaled up or down to Euclidean space of any dimension. This is illustrated with the usual topology on the Euclidean line. For any real number $x \in \mathbb{R}$, define a neighborhood of x to be any open interval containing x. This is the 1D equivalent of the usual topology on the Euclidean plane, and it can be shown to satisfy properties $T1$ and $T2$ in a similar way.

N_2. Now, x will lie in the intersection of these two neighborhoods, and it is always possible to surround x with an open disk entirely within this intersection. To see this, let d_1 be the minimum distance of x from the boundary of N_1 and d_2 be the minimum distance of x from the boundary of N_2. Then the open disk with center x and radius the minimum of d_1 and d_2 will contain x and lie entirely in the intersection of N_1 and N_2 (see Figure 3.18). Thus $T1$ and $T2$ are satisfied, and under this definition of neighborhood, $\mathbb{R}^2$ is a topological space, called the *usual topology* for $\mathbb{R}^2$.

usual topology

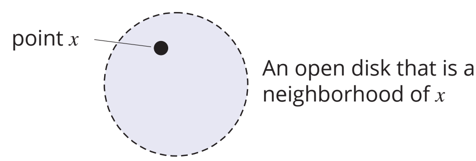


Figure 3.17: A neighborhood of x in $\mathbb{R}^2$ with the usual topology

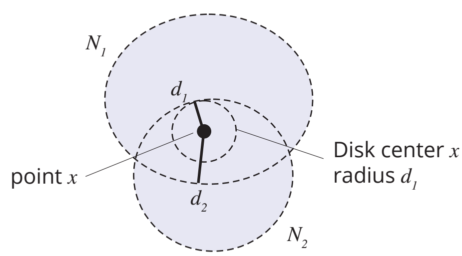


Figure 3.18: Condition $T2$ is satisfied for $\mathbb{R}^2$ with the usual topology

travel time topology

Travel time topology Another important topology for spatial information is *travel time topology*. Let S be the set of points in a region of the plane. Suppose that the region contains a transportation network and that we know the average travel time between any two points in the region using the network, following the optimal route. For the purposes of this example, we need to assume that the travel time relation is symmetric: that is, it must always be the case that the travel time from x to y is equal to the travel time from y to x. For each time t greater than zero, define a t-zone around point x to be the set of all points reachable from x in less than time t. As an illustration, Figure 3.19 shows a 5-zone, 10-zone, and 15-zone around the Spode Pottery.

Let the neighborhoods be all t-zones (for all times t) around all points. Then, clearly $T1$ is satisfied, since each point will have some t-zones surrounding it. The argument that $T2$ is satisfied is similar to that used for the usual topology of the Euclidean plane and is omitted. However, the symmetry of the travel time relation is required at a critical stage in this argument.

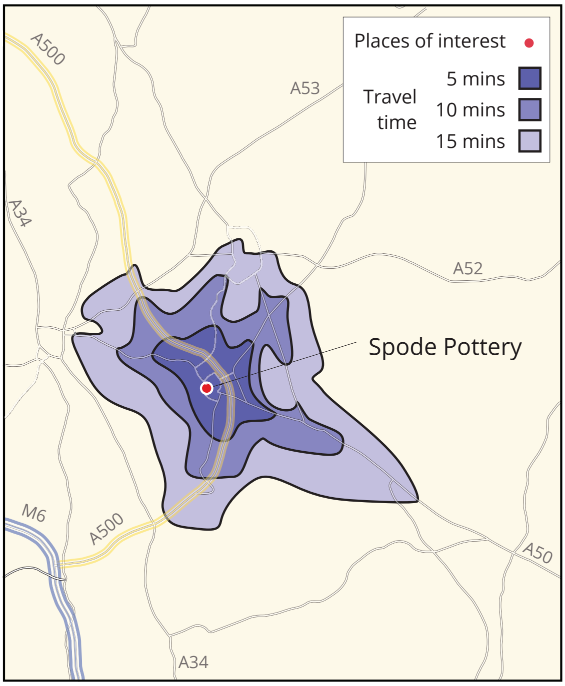


Figure 3.19: Travel-time topology example, showing some neighborhoods (t-zones)

The symmetric travel time measure between two points is an example of a *metric* (discussed later in Section 3.4.2). The travel time topology is a special case of the topology that can be induced by any metric on a space.

3.3.3 Properties of a topological space

It is surprising that out of the single primitive notion of neighborhood it is possible to construct all the features and properties of a topological space. This section describes some of these constructions, in a similar way to Henle (1979), beginning with the definition of "nearness." Many topologists use the phrase "limit point" to replace our use of "near point."

- Let S be a topological space. Then S has a set of neighborhoods associated with it. Let X be a subset of points in S and x an individual point in S. Define x to be *near* X if every neighborhood of x contains some point of X.

For example, in the Euclidean plane with the usual topology, let C be the open *unit disk*, centered on the origin, $C = \{(x, y)|x^2 + y^2 < 1\}$. Then the point $(1, 0)$, although not a member of set C, is near to C, because any open disk (no matter how small) that surrounds $(1, 0)$ must impinge into C. In fact, any point on the circumference of C is near to C, as indeed is any point inside C.

However, any point exterior to C and not on the circumference is not near C, since it will always be possible to surround it with a neighborhood that separates it from C (see Figure 3.20).

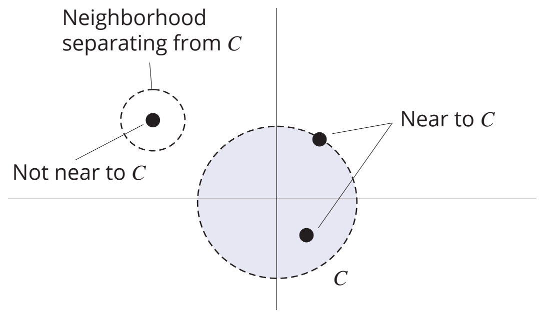


Figure 3.20: Points near and not near to the open unit disk C in the Euclidean plane with the usual topology

A fundamental topological invariant of any set is its boundary. This notion can be constructed out of our primitives as follows. First, open and closed sets are introduced using the neighborhood idea. It will emerge that an *open set* is a set that does not contain its boundary, whereas a *closed set* is a set that contains all its boundary.

open set

closed set

- Let S be a topological space and X be a subset of points of S. Then X is *open* if every point of X can be surrounded by a neighborhood that is entirely within X.
- Let S be a topological space and X be a subset of points of S. Then X is *closed* if it contains all its near points.

The open unit disk C above is (obviously) open, because any point, no matter how close to the circumference, may be surrounded by a neighborhood made sufficiently small that it is entirely within C. C is not closed, because points on the circumference are near points to C but not contained in C. For C to be closed, it would have to include its circumference. Note that it is possible for a topological space to contain sets that are both open and closed or neither open nor closed (see Box 3.4 on the facing page).

This leads us to the related definition of *closure*.

closure

- Let S be a topological space and X be a subset of points of S. Then the *closure* of X is the union of S with the set of all its near points. The closure of X is denoted X^-.

Clearly the closure of a set is itself closed; in fact the closure of set X is the smallest closed set containing X. In our example, the closure of the open unit disk C is formed by annexing its circumference. Thus $C = \{(x, y)|x^2 + y^2 \leq 1\}$. This set is called the *closed unit disk*. We may also force a set to be open by stripping away unwanted near points, constructing the *interior* of a set as follows:

interior

- Let S be a topological space and X be a subset of points of S. Then the *interior* of X consists of all points which belong to X and are not near points of X', the complement of X. The interior of set X is denoted X°.

Box 3.4: Topological spaces (open and closed sets)

Recall from the previous "Topological spaces" Box 3.3 on page 94, that the discrete topology on a space S defined the neighborhoods to be all the subsets of S. Let X be any subset of S, then the only near points of X are the points of X itself (since any point not in X may be surrounded by the neighborhood containing just that point). Thus X is closed. X is also open, because we can surround any point of X with a neighborhood entirely in X containing just that point. The discrete topology is therefore odd in that every set within it is both open and closed. This is not the case for the indiscrete topology.

Let S be a set upon which the indiscrete topology is defined and let X be any subset of S. Then, the only neighborhood in S is S itself. Therefore, every point in S is a near point of X. Thus, unless X is either empty or equal to S, it is neither open nor closed. For the travel time topology, an example open set is the set of all points less than 1 hour's traveling time from a specified point, say, Spode Pottery. This set has as its boundary the set of all points that are exactly 1 hour from Spode Pottery, and as its closure the set of all points having travel time from Spode Pottery not greater than 1 hour.

Notice that for a point x to be near to the complement of set X, it must be the case that each neighborhood of x impinges upon X'. Therefore, a point in X which is not a near point of X' has at least one neighborhood of it that is entirely within X. Thus, the interior of a set is open. In fact, the interior of a set X is the largest open set contained in X. As an example, if D is the closed unit disk, then D° is the open unit disk. Figure 3.21 shows a further example of an open and closed set in the Euclidean plane with the usual topology.

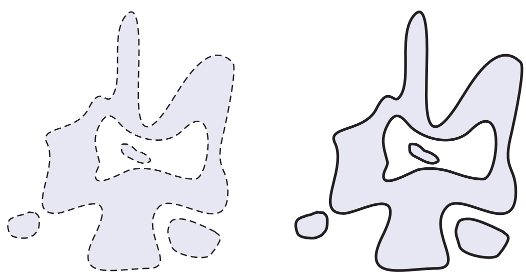


Figure 3.21: Open and closed sets in the Euclidean plane with the usual topology

boundary

We are now sufficiently prepared to define the *boundary* of a set in purely topological terms.

- Let S be a topological space and X be a subset of points of S. Then the boundary of X consists of all points which are near to both X and X'. The boundary of set X is denoted ∂X.

Let point x be a member of ∂X. Since x is near to X, then x must be in X^-. Since x is near to X', then x cannot be in X°. Thus ∂X is the set difference of X^- and X°. In the case of the unit disk C in the Euclidean plane with the usual topology, ∂X is the circumference of C, as we would expect. As a further example using the Euclidean plane with the usual topology, suppose that S is the connected region of the plane containing a single hole shown in Figure 3.22. We see that the outer boundary of S is excluded from S, but the inner boundary of S (i.e., the boundary of the hole) is contained in S. The interior, closure, and boundary of S are as shown in Figure 3.22. S° contains all the points of S excluding its inner boundary. S^- includes both inner and outer boundaries. ∂S is the union of the inner and outer boundaries of S.

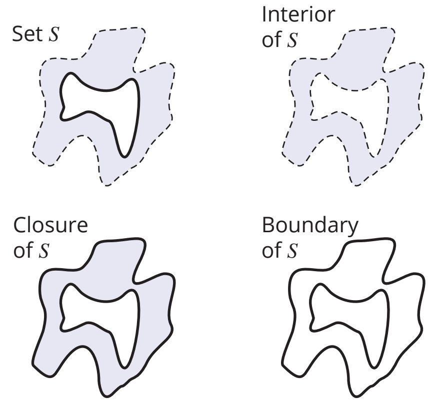


Figure 3.22: Interior, closure, and boundary of a region in the Euclidean plane with the usual topology

It is important to realize that it is not possible to consider the topological properties of sets in exclusion from the larger spaces in which they are embedded. To illustrate this point, consider a finite length of straight line. If we take the line to be embedded in a 2D (or higher-dimensional) Euclidean space, as shown in Figure 3.23a, then its interior is the empty set, and its boundary and closure are both the line itself. On the other hand, if the same line is embedded in a 1D Euclidean space, that is, in the real number line, as shown in Figure 3.23b, then its interior is the line excluding its end-points, its closure is the line itself, and its boundary consists of the end-points of the line.

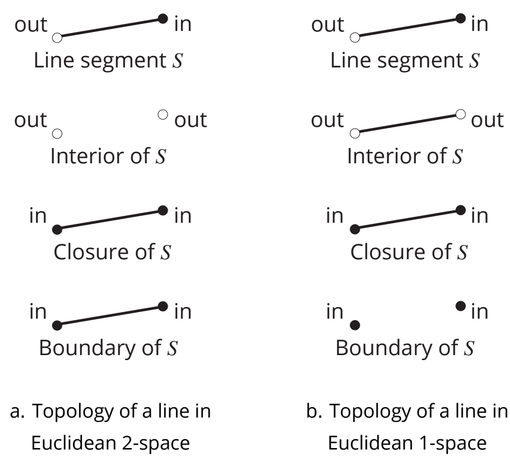


Figure 3.23: Interior, closure, and boundary of a line segment in Euclidean 2-space and 1-space with the usual topology

We conclude this brief excursion into point-set topology by considering the notion of *connectedness*. In fact, point-set topology recognizes several different kinds of connectedness. This section defines a simple form based directly upon the neighborhood properties of a topological space. The next section includes a description of further forms, including weak, strong, and path-connectedness.

- Let S be a topological space and X be a subset of points of S. Then X is *connected* if whenever it is partitioned into two non-empty disjoint subsets, A and B, then either A contains a point near B, or B contains a point near A, or both.

connected

Consider the three sets shown in Figure 3.24. In the case of Figures 3.24a and b, no matter how we choose to divide them into two, the partition will always satisfy the condition of the definition of connectedness above. Even if the set in Figure 3.24a is partitioned into its upper and lower disks, with the point of intersection included in the upper disk, then this point is certainly a near point of the lower disk. Therefore, the sets in Figures 3.24a and b are connected. Figure 3.24c shows a set that is not connected. To see this, partition the set in Figure 3.24c into its upper and lower component disks. Then no point of the upper disk is a near point of the lower disk, and no point of the lower disk is a near point of the upper disk. This example shows that the topological definition of connectedness accords with our intuition.

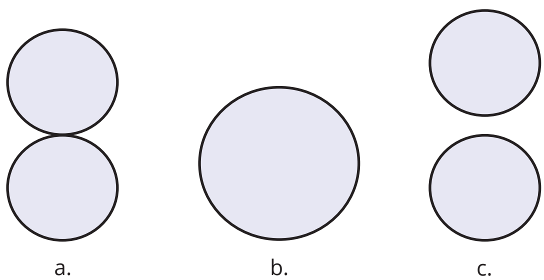


Figure 3.24: Connected (a, b) and disconnected (c) sets in the Euclidean plane with the usual topology

3.3.4 *Point-set topology of the Euclidean plane*

The Euclidean plane with the usual topology provides by far the most important example of a topological space, for the purposes of GIS. A *homeomorphism* (or *topological transformation*) of $\mathbb{R}^2$ is a bijection of the plane that transforms each neighborhood in the domain to a neighborhood in the image. Furthermore, any neighborhood in the image must be the result of the application of the transformation to a neighborhood in the domain. Put more simply and intuitively, a homeomorphism corresponds to the notion of a rubber sheet transformation, which stretches and distorts the plane without folding or tearing.

homeomorphism

If the result of applying a homeomorphism to a point-set X is point-set Y, we say that X and Y are *topologically equivalent*. Thus, in Figure 3.25, the disk S and the area T are topologically equivalent, but neither is topologically equivalent to the area U. From an intuitive point of view, it is clearly possible (at least, in the mind's eye, with a lot of stretching) to transform S to T (and back again) by stretching and contracting a rubber sheet. However, the only option to arrive at the area U from S is to tear the sheet, so as to form the hole in U. Tearing is not allowed and therefore S and U are not topologically

topological equivalence

equivalent. Set V, which is formed by gluing the disk to itself at a single point, is homeomorphic to none of the sets S, T, or U.

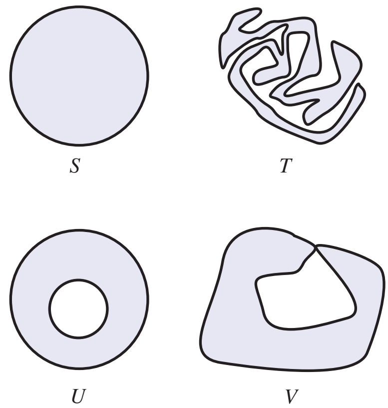


Figure 3.25: Topologically equivalent and inequivalent planar objects

Many mapping ideas are based upon the idea of homeomorphism. For example, Figures 3.26 and 3.27 show two maps of The Potteries' bus routes, first (Figure 3.26) as a similarity transformation of the actual routings (neglecting the Earth's curvature), and second (Figure 3.27) as a topological transformation of the actual routings. The two bus route maps are topologically equivalent.

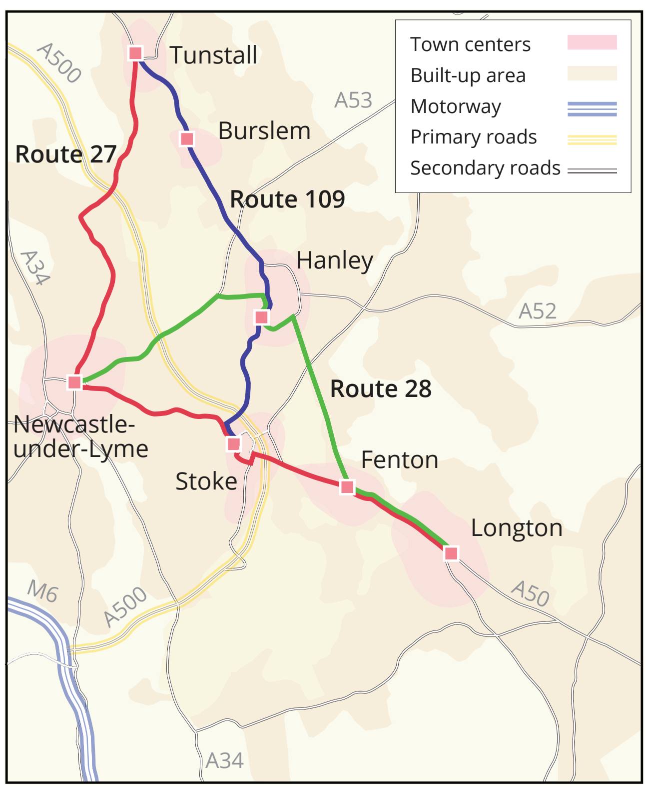


Figure 3.26: Map (similarity transformation) of Potteries' bus routes

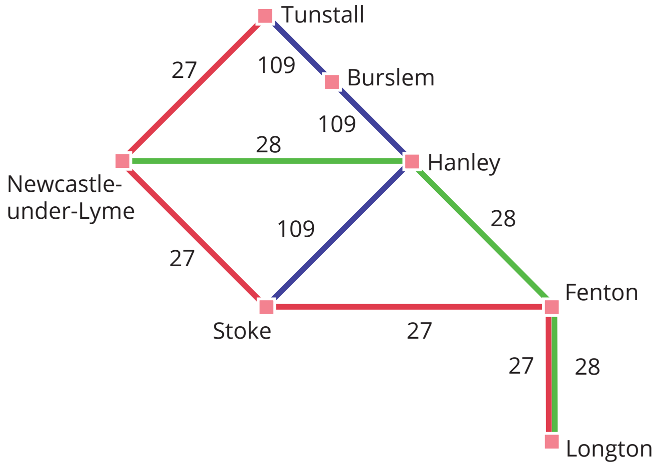


Figure 3.27: Map of Potteries' bus routes, topologically equivalent to Figure 3.26

Properties that are preserved by homeomorphisms are called *topological invariants*. From the definition of a homeomorphism, it is clear that the configuration of neighborhoods of a space is a topological invariant. Other constructs, such as open set, closed set, and boundary, are also topological invariants because they are defined purely in terms of neighborhoods.

topological invariance

The paradigm for all open sets in the Euclidean plane is the unit open disk, with its center at the origin. Therefore, all topologically equivalent areas to this are open as well. Any set homeomorphic to the unit open disk we term an *open cell*. Any set homeomorphic to the closed unit disk is a *closed cell* (see also Box 3.5 on the following page). An open cell is clearly an open set, and similarly a closed cell is clearly a closed set. In general, if a set of points owns its entire boundary then it is closed, and if it owns none of its boundary then it is open.

open cell

closed cell

Connectedness has already been defined in the context of a general topological space. Connectedness is defined purely in terms of the topology, so it is a topological invariant. Essentially, a connected set is "all of a piece," such as a line, a cell, or an annulus (a ring). Connected sets are not necessarily homeomorphic to each other; a cell is not homeomorphic to an annulus because the rubber sheet would need to be torn to make the hole. Connected sets that do not have holes are called *simply connected*. Being simply connected is a topological invariant. Cells are simply connected; annuli are not simply connected.

simply connected

Two basic 1D object types are the straight-line segment and the circle (a circle is the boundary of a disk). Using the notion of homeomorphism, we may generalize these two basic object types as follows. A simple arc is topologically equivalent to a straight-line segment: it is clearly connected. The homeomorphism from a straight-line segment to a simple arc is a bijection, so it cannot be possible for a simple arc to cross over itself or for its end-points to be coincident. If the condition on no self-crossings is relaxed, then the resulting 1D object is termed an *arc*. If the end-points are coincident and self-crossings are not allowed, then the object is termed a *simple loop*. If the

simple arc

arc

simple loop

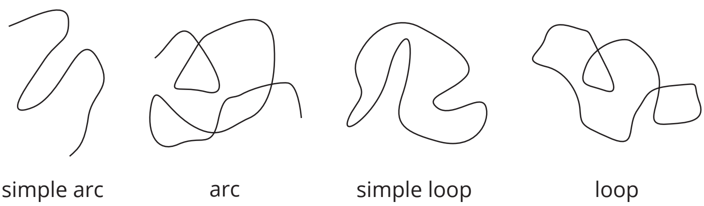
Box 3.5: Brouwer's fixed point theorem

Cells, though basic to topological work, are the subject of one of the most intriguing of results in introductory topology. Imagine a cell laid out as a rubber shape on the plane. Now deform the cell by twisting, stretching, even folding, but not tearing, and replace the cell on the plane in such a way that it is entirely within its original outline. Then, there will always be at least one point that has not moved from its original position! Such a point is called a *fixed point* and the result is known as *Brouwer's fixed point theorem*. Brouwer's fixed point theorem holds for Euclidean n-space and has many intriguing and counter-intuitive implications. As a 3D example, assume the water within a lake is homeomorphic to the closed unit sphere. Assume further that the currents within the lake operate as a continuous rubber sheet transformation of the water. Then according to Brouwer's fixed point theorem, for any two points in time there must always be some water that is in exactly the same location within the lake at both points in time (although it may have moved in between). Fixed point theorems have played an important role in computer science, in particular underpinning the theory of recursion.

loop

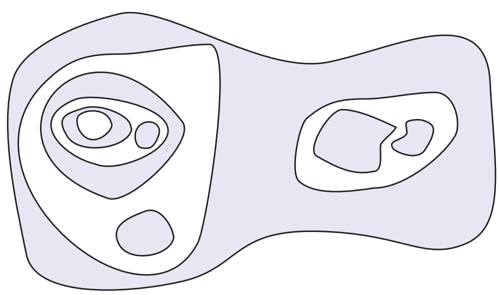
end-points are coincident and self-crossings are allowed, then the object is termed a *loop*. A simple loop is topologically equivalent to a circle. Examples of some 1D planar objects are shown in Figure 3.28.

Figure 3.28: Some 1D planar objects



Moving up a dimension, we have already defined the cell as the primary 2D topological object, topologically equivalent to the disk and simply connected. Another notable class of areal objects is the annuli (cells with a single hole). We can also allow the holes to be occupied by further objects (islands) and so on (see Figure 3.29).

Figure 3.29: Holes within holes within holes



The next paragraphs explore further the concept of connectedness, particularly as it relates to areal objects. The topological property of connectedness was defined earlier, in Section 3.3.2. We begin by providing a different definition.

path-connected

- A set in a topological space is *path-connected* if any two points in the set can be joined by a path that lies wholly in the set.

A natural first question is: Given that there are two notions of connectedness, namely, connectedness (as defined earlier) and path-connectedness, do these notions define the same property? The answer to this question is, in general, no. Although it is possible to show that every path-connected set is connected, there are sets that are connected but not path-connected. However, in the special case of the Euclidean plane with the usual topology, each example of a connected but not path-connected set is pathological, involving an infinite number of twists and turns. Therefore, for practical purposes, we may identify notions of connectedness and path-connectedness, certainly for the areal objects that we will define shortly. Path-connectedness is a more intuitive notion than pure connectedness, and can therefore be used as a test for connectedness in practical cases. To summarize, test for connectedness by asking the question: Given any two points in the set, is it possible to move from one point to the other along a path entirely within the set?

Many applications of spatial analysis require classes of planar objects that are purely areal, that is, not mixtures of points, lines, and areas. Also they do not have isolated missing points (*punctures*) or arcs (*cuts*). Interestingly, it is possible to define the notion of a purely areal object using only topological notions.

- Let X be a set of points in the Euclidean plane under the usual topology. Then define the *regularization* of X to be the closure of the interior of X, that is, $reg(X) = X^{\circ -}$.

regularization

The regularization process has the effect of eliminating from a set any pathological and non-areal features. Consider the example shown in Figure 3.30a, which is an amalgamation of a punctured and cut cell with some arcs and isolated points. The regularization of the set in Figure 3.30b removes all cuts, punctures, extraneous arcs, and isolated points. Regularization first finds the interior of the object, which will remove exterior arcs and points. Taking the closure will then remove cuts and punctures. What remains is always a closed, purely areal object. In our example, shown in Figure 3.30b, the result is a cell.

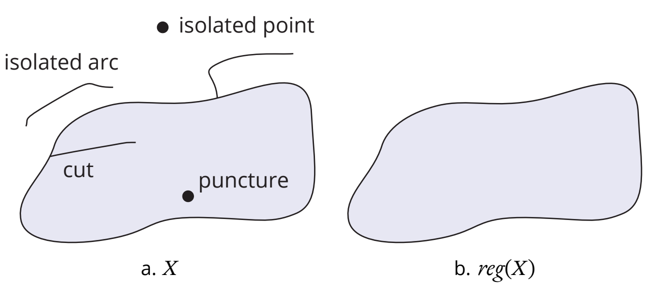


Figure 3.30: Spatial object X comprising an area with cut, puncture, arcs, and points, and $reg(X)$, its regularization

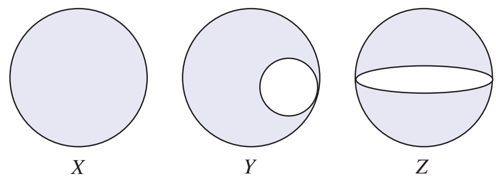
The regularization concept can now be used to characterize pure area. If an object is already purely areal, then regularizing it will have no effect. An object for which regularization has no effect is termed *regular closed*. The regular closed sets are exactly the purely areal objects that we require. The formal definition follows.

regular closed

- Let X be a set of points in the Euclidean plane under the usual topology. Then X is *regular closed* if and only if $X^{\circ -} = X$.

Having finally arrived at a topological characterization of objects that are purely areal, we now reconsider in more detail the notion of connectedness. Figure 3.31 shows three connected sets.

Figure 3.31: Three connected sets



There are some clear differences in the kinds of connectedness here. In the first two cases X and Y, given any two points in each set, there are few constraints upon the path of connection from the first point to the second. All that is required is that the path starts at the first point, stays within the set, and ends at the second point. However, in the case of set Z, if the two points are in the upper portion and lower portion, respectively, then the path is constrained to pass through one of the two points on the horizontal diameter. This difference is expressed by saying that X and Y are *strongly connected*, but that Z is *weakly connected*. To arrive at a formal definition, note that Z may be made disconnected by removing a finite number of points (in fact, the two points on its horizontal diameter). However, no matter how large a finite number of points we remove from X and Y, they will remain connected.

weakly connected

strongly connected

- A connected set X in the Euclidean plane with the usual topology is *weakly connected* if it is possible to transform X into an unconnected set by the removal of a finite number of points.
- A connected set X in the Euclidean plane with the usual topology is *strongly connected* if it is not weakly connected.

Figure 3.32 shows some more strongly and weakly connected sets. While the sense of strongly and weakly connected sets will resonate with the definitions of strongly and weakly connected directed graphs introduced in the previous chapter (Section 2.4.1), the two ideas are quite distinct as can be seen from their definitions. The notions of strong and weak connectedness in sets play an important role later in the categorization of planar objects in the object-based approach to spatial modeling.

The 4-intersection model (4IM) We have seen above how point-set topology leads directly to formal of *boundary*, *interior*, and *closure*. An important characterization of the topological relationships between spatial objects—more widely cited in the literature than any other—is based on the question: What can be deduced about the topological relationship between two spa-

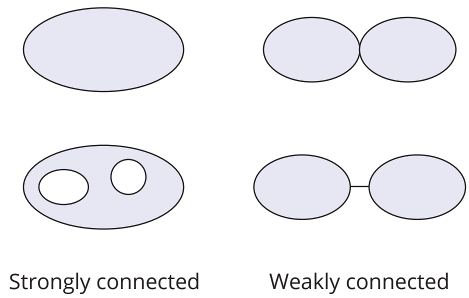


Figure 3.32: More strongly and weakly connected sets

tial regions from the set-theoretic relationship of intersection between their interiors and boundaries?

To be more precise, let X and Y be the spatial cells, and assume that the boundaries, ∂X and ∂Y, and interiors, X° and Y°, are known. In order to find the topological relationship between X and Y, compute the set-oriented relationships between ∂X, X°, ∂Y, and Y°. In fact, for a first-pass determination of the topological relationships, we consider the following four sets:

$$\partial X \cap \partial Y \qquad \partial X \cap Y^\circ$$
$$X^\circ \cap Y^\circ \qquad X^\circ \cap \partial Y$$

Each of these point sets may either be empty (written $\varnothing$, the empty set) or non-empty ($\neg\varnothing$). There are $4 \times 4 = 16$ different mutually exclusive combinations of possibilities. Each possibility is a condition on the boundaries and interiors of sets, so each possibility will lead to a relationship between the sets that is preserved under topological transformations (homeomorphisms). For general sets in a point-set topology, each of the 16 combinations can exist and lead to a distinct topological relationship between the two sets X and Y. We are, however, concerned only with spatial cells embedded in the Euclidean plane. In this case, only *eight* of the 16 combinations can occur. Table 3.3 shows the eight possibilities, together with their common names.

$\partial X \cap \partial Y$	$X^\circ \cap Y^\circ$	$\partial X \cap Y^\circ$	$X^\circ \cap \partial Y$	Name	Alternative
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	X disjoint Y	
$\neg\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	X meets Y	X touches Y
$\neg\varnothing$	$\neg\varnothing$	$\varnothing$	$\varnothing$	X equals Y	
$\varnothing$	$\neg\varnothing$	$\neg\varnothing$	$\varnothing$	X inside Y	Y contains X
$\neg\varnothing$	$\neg\varnothing$	$\neg\varnothing$	$\varnothing$	Y covers X	X covered by Y
$\varnothing$	$\neg\varnothing$	$\varnothing$	$\neg\varnothing$	Y inside X	X contains Y
$\neg\varnothing$	$\neg\varnothing$	$\varnothing$	$\neg\varnothing$	X covers Y	Y covered by X
$\neg\varnothing$	$\neg\varnothing$	$\neg\varnothing$	$\neg\varnothing$	X overlaps Y	

Table 3.3: Eight relations between cells in the Euclidean plane

Any pair of spatial cells in the Euclidean plane must satisfy one and only one of these possibilities. In technical language, the eight topological relationships are JEPD: jointly exhaustive (i.e., no other possibilities) and pairwise disjoint (i.e., a pair of regions cannot satisfy two possibilities simultaneously). These eight possibilities are summarized graphically in Figure 3.33. Because four intersections are tested for empty in this method, it is usually known as the *4-intersection* model (4IM).

4-intersection model

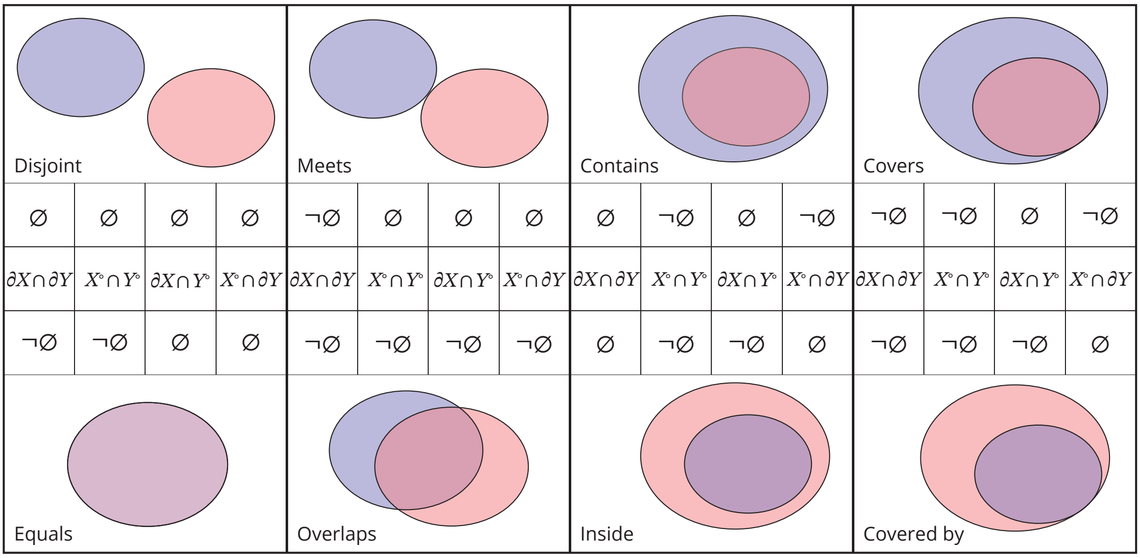


Figure 3.33: The eight JEPD topological relationships of the 4-intersection model (4IM)

It can now be seen that 4IM relations meets and overlaps are topological refinements of the set-oriented operation *intersection*, while inside and covers are refinements of the *subset of* relationship.

This work has been extended for higher-dimensional spaces, and where the *co-dimension* is non-zero (i.e., where the dimension of the spatial objects is less than the dimension of the space in which they are embedded). In this case, three topological operations are used: boundary, interior, and set complement. The binary spatial relation between two spatial objects X and Y is classified by checking for emptiness/non-emptiness of the nine combinations of the operations applied to X and Y, known as the *9-intersection* model (9IM). The 9 combinations are[2]:

co-dimension

9-intersection model

[2] Recall that X' is the set complement or *exterior* of X.

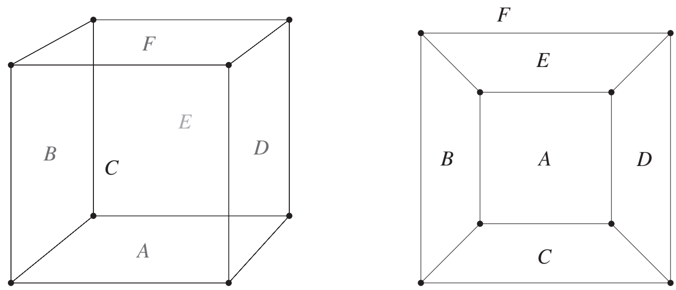
$$
\begin{array}{ccc}
\partial X \cap \partial Y & \partial X \cap Y^\circ & \partial X \cap Y' \\
X^\circ \cap \partial Y & X^\circ \cap Y^\circ & X^\circ \cap Y' \\
X' \cap \partial Y & X' \cap Y^\circ & X' \cap Y'
\end{array}
$$

3.3.5 Combinatorial topology of the Euclidean plane

The other topological area considered in this chapter is combinatorial topology. In some ways, combinatorial topology is more pertinent to computer-based models than point-set topology, because the often finite and discrete structures that arise in combinatorial topology are highly suitable for representation in computer-based data structures. A typical result of combinatorial topology is Euler's famous formula:

Euler's formula

- Given a polyhedron with f faces, e edges, and v vertices, then $f - e + v = 2$.



a. Cube

b. Planar configuration homeomorphic to a cube

Figure 3.34: Example polyhedron and 3-space homeomorphism

For example, Figure 3.34a shows a cube (six faces, 12 edges, and eight vertices). A very similar formula applies to an arrangement of cells in the plane. Remove a single face from a polyhedron (for example, in Figure 3.34a remove face F) and apply a 3-space homeomorphism to flatten the shape onto the plane. What results is a configuration of cells, with arcs forming common boundaries and nodes forming the intersection points of the arcs. Flattening the cube in Figure 3.34a results in the planar configuration in Figure 3.34b. Since we have removed a face from the polyhedron (it has actually become the exterior to the cellular configuration, as in Figure 3.34b), we may simply modify Euler's formula for the sphere to derive Euler's formula for the plane.

- Given a cellular arrangement in the plane, with f cells, e edges, and v vertices, $f - e + v = 1$.

Figure 3.35 shows an example of a planar configuration of cells, with 25 faces, 72 edges, and 48 nodes (inspired by the configuration of the 25 wards of the City of London).

The topological content of these results becomes clear when we observe that no matter how the surface of the sphere is divided into polyhedral arrangements, the result of $f - e + v$ is always two, and for planes the result is always one. Thus the number two characterizes a sphere and distinguishes it from a plane. If we were to perform the same exercise on the surface of a torus (doughnut shape), the result is always zero. The result of $f - e + v$ is called the *Euler characteristic* of a surface.

Euler characteristic

Simplexes and complexes As we will see in later chapters, many of the models of space found in a GIS use variations of the planar cellular arrangements described above. The most fundamental formal model of such cellular arrangements uses the notion of a *simplicial complex*. In the two-dimensional case, simplicial complexes are simple triangular network structures in the Euclidean plane.

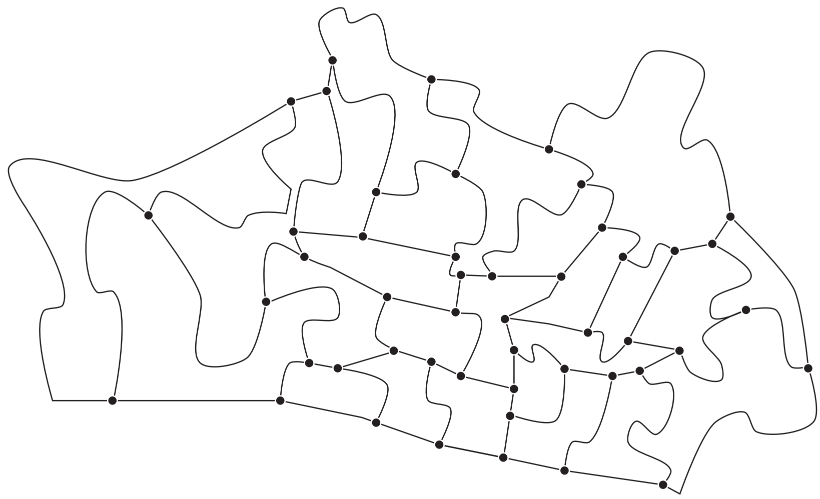


Figure 3.35: Planar cellular arrangement where $f = 25$, $e = 72$, and $n = 48$

- A *0-simplex* is a set consisting of a single point in the Euclidean plane.
- 1-simplex A *1-simplex* is a closed finite straight-line segment.
- 2-simplex A *2-simplex* is a set consisting of all the points on the boundary and in the interior of a triangle whose vertices are not collinear.

The *vertices* of a simplex are defined as: for a 0-simplex the point itself; for a 1-simplex its end-points; and for a 2-simplex the vertices of the triangle. A *face* of a simplex S is a simplex whose vertices form a proper subset of the vertices of S. Figure 3.36 shows examples of 0-, 1-, and 2-simplexes. The *boundary* of a simplex S, written ∂S, is the union of all its faces. For example, suppose that a 2-simplex S has vertices x, y, and z. Then the faces of S are the three 1-simplexes xy, xz, yz and the three 0-simplexes x, y, z. The boundary of S is the union of these faces, thus corresponding to the usual point-set topological definition of boundary.

Figure 3.36: Examples of 0-, 1- and 2-simplexes

Simplexes are the building blocks of larger structures, called simplicial complexes. Complexes are built out of simplexes in a way that is now made precise. simplicial complex A *simplicial complex* C is a finite set of simplexes satisfying the properties:

1. A face of a simplex in C is also in C.
2. The intersection of two simplexes in C is either empty or is also in C.

Figure 3.37 shows examples of configurations, two of which are simplicial complexes and two are not. In fact, the complexes on the right are formed by adding sufficient nodes and edges to the configurations on the left to make them satisfy the simplicial complex formation rules, a form of "completing the topology." In the case of the 2-simplexes abc and def, their intersection is not a face of either simplex. This is rectified by adding nodes k and l and decomposing the original simplexes abd and def into simplexes akf, afb, bfc, cfl, fkl, dkl, and del. These simplexes, along with their faces ab, af, ak, bc, bf, cf, cl, dk, dl, de, el, fl, fk, kl and vertices a, b, c, d, e, f, k, l, form a simplicial complex. The 1-simplexes in the lower part of the figure are enhanced in a similar way by adding vertex m.

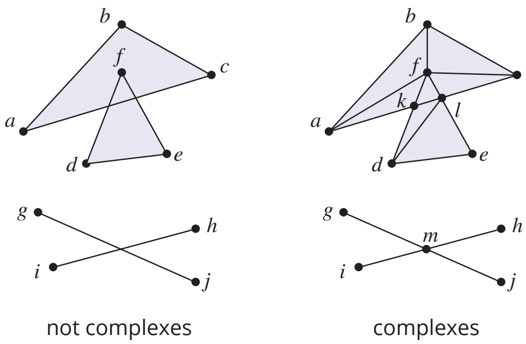


Figure 3.37: Aggregations of simplexes

The set of points contained in the constituent simplexes of a simplicial complex is a set in the Euclidean plane, called the *planar embedding* of the complex. Figure 3.38 shows the planar embedding of the upper complex in Figure 3.37.

planar embedding

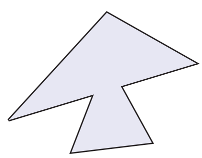


Figure 3.38: Planar embedding of a simplicial complex from Figure 3.37

Even though we have given an abstract presentation of simplicial complexes, such structures are common in GI science. An n-simplex is said to have *dimension* n. The dimension of a simplicial complex is the maximum dimension of its constituent simplexes. The constructions performed in this section are all planar, but the ideas can be generalized to higher dimensional structures. Thus a 3-simplex would be a tetrahedron.

For an n-complex, C, the boundary of C, ∂C, is a simplicial complex of dimension $n - 1$. 1D complexes are *graphs*, covered previously in Sections 2.4.1 and 3.4.1. 2D complexes may be used to model the triangulated irregular networks (TINs) used in terrain modeling, or indeed any areal objects.

Combinatorial map A different approach to combinatorial topology is known as the *combinatorial map*. Assume that the boundary of a cellular arrangement is decomposed into simple arcs and nodes, as in Figure 3.35. Next, distinguish the two directions of each arc, akin to a bidirected graph structure. A combinatorial map is then:

combinatorial map

- The set of directed arcs D, called *darts*. In Figure 3.39, the set of darts $D = \{ab, ag, ah, ba, bc, bi, cb, ci, cd, dc, de, di, ...\}$ and so on.
- A function $\sigma : D \to D$ that for each dart specifies the next dart in a counterclockwise direction at each node. For example, in Figure 3.39 $\sigma(ab) \mapsto ag$, $\sigma(ag) \mapsto ah$, $\sigma(ah) \mapsto ab$ and so on. Written as a relation, $\sigma = \{(ab, ag), (ag, ah), (ah, ab), (ba, bi), (bi, bc), (bc, ba), ...\}$. The function σ is a bijection (an injection and a surjection).
- A function $\alpha : D \to D$ that for each dart specifies the other dart that makes up that arc. For example, in Figure 3.39 $\alpha(ab) \mapsto ba$, $\alpha(ba) \mapsto ab$, $\alpha(ag) \mapsto ga$ and so on. The function α is also a bijection, as well as a symmetric, irreflexive function. Written as a relation, $\alpha = \{(ab, ba), (ag, ga), (ah, ha), (ba, ab), (bc, cb), (bi, ib), (cb, bc), ...\}$.

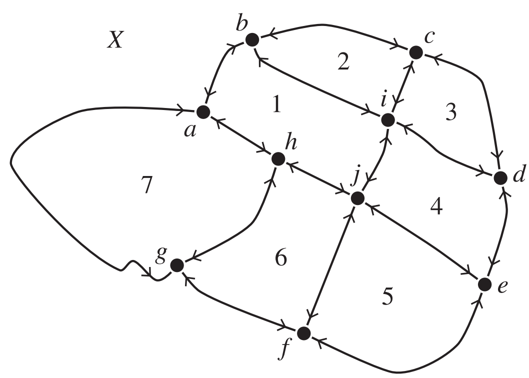


Figure 3.39: A planar cellular arrangement, augmented with a set of darts D as the basis for a combinatorial map

Intriguingly, this abstract structure of darts and functions provides a unique and unambiguous representation of a cellular arrangement, such as that in Figure 3.39. Adopting the rule that a dart bounds the face on its right as we traverse it, we can use the functions σ and α to reconstruct each face. For example, let us reconstruct the boundary of face 7 in Figure 3.39. We start with any one of its bounding darts, say ah. Next, alternating the functions α and σ will allow us to trace around face 7: $\alpha(ah) \mapsto ha$, $\sigma(ha) \mapsto hg$, $\alpha(hg) \mapsto gh$, $\sigma(gh) \mapsto ga$, $\alpha(ga) \mapsto ag$, and finally back to the start of the cycle with $\sigma(ag) \mapsto ah$. The choice of starting dart was arbitrary; we might have picked any of the darts that bound 7 on their right: ah, hg, or ga. Each of the other faces may be similarly traversed. Dart ha, for example, bounds face 1: $\alpha(ha) \mapsto ah$, $\sigma(ah) \mapsto ab$, $\alpha(ab) \mapsto ba$, $\sigma(ba) \mapsto bi$, ... and so on. For completeness, we must also include the external face, X, bounded by ba, $\alpha(ba) \mapsto ab$, $\sigma(ab) \mapsto ag$, $\alpha(ag) \mapsto ga$, $\sigma(ga) \mapsto gf$, $\alpha(gf) \mapsto fg$, $\sigma(fg) \mapsto fe$, ... and so on.

The combinatorial map is the foundation of the NAA (node-arc-area) and DCEL representations of planar configurations, discussed in more detail in Chapter 5.

3.4 Further spaces

The picture emerging over this chapter is that space is not just one thing, one reality; rather it is a family of related representations each of which foregrounds particular properties of space and offers particular characteristics and operations. This final section visits three further types of spaces: network spaces build on the foundations of abstract graphs introduced in the previous chapter; metric spaces formalize the concept of distance between objects in the space; and fractal geometry tackles spatial objects that exhibit self-similarity and scale dependence.

3.4.1 Network spaces

A great many spatial problems can be represented using a network. For example, a system of roads or rail links is often thought of as a network; the task of deciding what route to take when traveling by car is essentially the task of finding a suitable route through a network. Several famous spatial problems have been solved by representing the problem as a network, such as the Königsberg bridge problem (see Box 3.6 on the next page).

We have already encountered *abstract graphs* as a formal model of the network, in connection with graph databases in the previous chapter (Section 2.4.1). Building on these foundations, this section introduces some further graph concepts important for modeling geographic spaces, including *trees* and *planar graphs*.

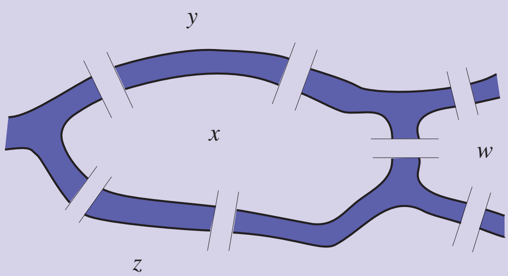
Abstract graphs revisited Recall from Section 2.4.1 that a graph is a set of nodes together with a set of edges that connect pairs of nodes. Armed with our knowledge of sets, relations, and functions gained in the previous section, we can now more precisely define the following graph structures already encountered:

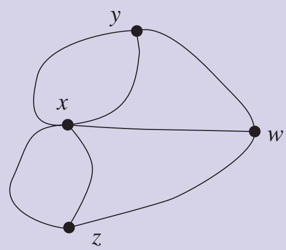
- An undirected graph $G = (N, E)$ is a set of nodes N together with a set E of unordered pairs of nodes, called edges.
- A directed graph $G = (N, E)$ is a set of nodes N together with a set of edges drawn from a relation on the set of nodes, $E \subseteq N \times N$.
- An (edge) labeled graph $G = (N, E, w)$ is a set of nodes N, edges E, together with an edge-labeling function $w : E \to L$ mapping each edge to a label in the set L.
- An (edge) weighted graph is a labeled graph $G = (N, E, w)$ where the labels are drawn from positive numbers, e.g., $w : E \to \mathbb{Q}^+$.
- A (directed) multigraph $G = (N, E, s, t)$ is a set of nodes N, a set of edges E, and two functions $s : E \to N$ and $t : E \to N$ mapping each edge to its start and end node, respectively.

Recall also from Section 2.4.1 that a path is an ordered sequence of nodes in a graph, where each consecutive pair of nodes is connected by an edge; a

Box 3.6: Königsberg bridge problem

The Königsberg bridge problem addresses the question of whether it is possible to walk a circuit that crosses each of seven bridges (shown below left) in the city of Königsberg once and only once. You might try to find such a route yourself. However, don't try for too long as in 1786 the mathematician Leonard Euler succeeded in proving that the task is impossible. Euler's proof is based on a model of the topological relationships between the Königsberg bridges (shown below right). The nodes, labeled w, x, y, and z, are abstractions of the regions of dry land. The edges between nodes are abstractions of the bridges connecting regions of dry land. Euler noted that apart from the start and end nodes, any path through a node must come in along one edge and out along another edge. So, if the problem is to be solvable, then the number of edges incident with each intermediate node must be even. However, in Königsberg *none* of the nodes is incident with an even number of edges. Thus, Euler proved that it was impossible to cross each bridge just once. A path through a graph that visits each edge exactly once is still termed an *Eulerian trail*.





cycle is a path from a node to itself traversing at least one edge; a graph is connected if there exists a path between every pair of nodes.

tree

Trees and DAGs An especially useful class of graphs is the *tree*. A tree is a connected acyclic graph. Figure 3.40 shows the three non-isomorphic trees, each with five nodes.

rooted tree

A *rooted tree* is a tree that has one of its nodes, the *root*, distinguished from the others. Rooted trees (often we omit the word "rooted") are conventionally drawn with the root at the top and nodes occupying successive levels down, depending upon their distance (in terms of path) from the root. Nodes immediately below the root are termed *immediate descendants* of the root; they themselves may have descendants, and so on. A node with no descendants is termed a *leaf*. Figure 3.41 shows an example of a layered rooted tree. Trees

leaf

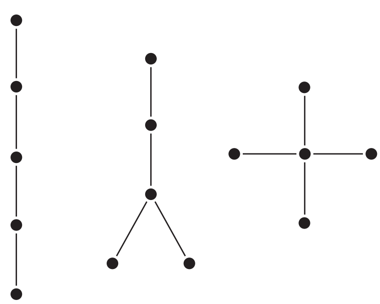


Figure 3.40: The three non-isomorphic trees with five nodes

provide some useful data structures for computational purposes, to which we return in later chapters.

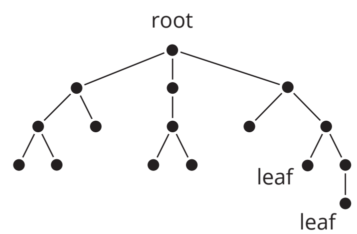


Figure 3.41: Rooted tree with five levels and eight leaves

The class of directed graphs that have no cycles is particularly useful for a wide range of applications. These graphs are termed *directed acyclic graphs* or DAGs. A DAG defines a *partial order* on its nodes. As introduced above, a (strict) partial order is a special form of relation on a set, which is irreflexive, antisymmetric, and transitive.

DAG

partial order

An example of a partial ordering is the relation "greater than," between two real numbers. No number can be greater than itself (irreflexive); if x is greater than y, then it cannot be that y is greater than x (antisymmetric); and x greater than y greater than z implies x greater than z (transitive). For a DAG, if we define a relation R where $(a, b) \in R$ means "there exists a path from a to b," then it can be shown that this relation is a partial order.

Planar graphs A further level of information may be added to the graph-theoretic model by considering the embedding of the graph in the Euclidean plane. A planar graph is a graph that can be embedded in the plane in a way that preserves its structure. In particular, a planar graph's edges are embedded as arcs that may only intersect at nodes of the graph. Figure 3.42a shows a planar graph while Figure 3.42b shows a non-planar graph. For the non-planar graph in Figure 3.42b, no rearrangement of arcs in the plane will preserve the original connectivity without leading to edges crossing somewhere other than at a node.

planar graph

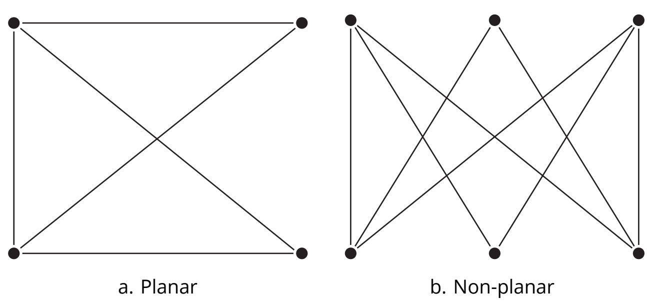


Figure 3.42: Planar and non-planar graphs

In general, there are many topologically inequivalent planar embeddings of a planar graph in the plane. A planar graph together with an embedding into the plane such that arcs intersect only at nodes of the graph is called a *plane graph*.

plane graph

Figure 3.43 shows three embeddings of the planar graph of Figure 3.42. The upper two embeddings are homeomorphic, but not homeomorphic to the lowest, with respect to the usual topology of the Euclidean plane. There is no way, using topological transformations (rubber-sheet geometry), to move the node of degree two inside the triangle to a position outside the triangle.

This last discussion raises an interesting consideration. Configurations that are taken to be equivalent in one model may very well be inequivalent in another. The second and third configurations are equivalent in a graph-theoretic sense, both being identically connected. However, when viewed as embeddings in the plane, they are inequivalent in the topological sense (and certainly in the metric sense). It all depends on the level of abstraction.

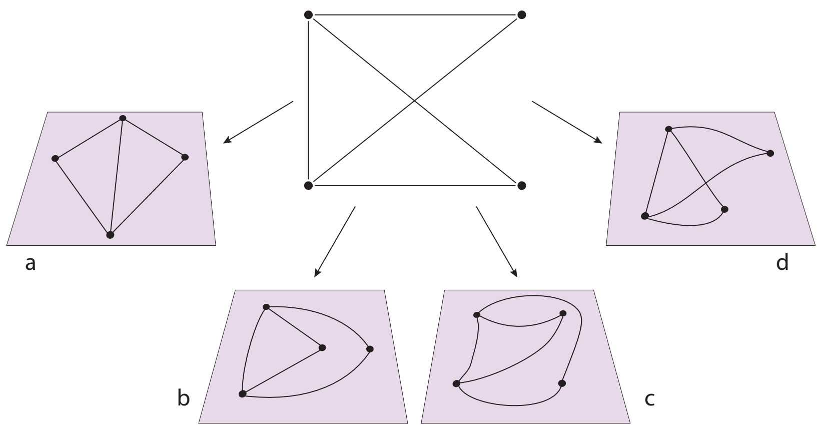


Figure 3.43: Three planar embeddings of a planar graph (plane graphs a, b, and c, where b and c are also homeomorphic) and one non-planar embedding of a planar graph (d)

A planar embedding of a planar graph determines a subdivision of the plane into faces or regions. A simple integrity constraint upon a planar graph discussed in an earlier section is given by the Euler formula for the plane, which provides a relationship between the number of nodes n, arcs a, and faces f, as $f - a + n = 1$.

A useful concept associated with planar embedded (i.e., plane) graphs is that of *duality*. The *dual* G^* of a plane graph G is obtained by associating a node in G^* with each face in G. Two nodes in G^* are connected by an edge if and only if their corresponding faces in G are adjacent. Given an edge e in G, the dual edge e^* joins the nodes in G^* corresponding to the two faces in G incident with e. Figure 3.44 shows a plane graph (nodes filled in black, edges marked with continuous lines) and its dual (nodes shown filled in color, edges marked with dotted lines).

dual graph

When the planar graph G is a diagonal triangulation of a polygon (with no Steiner points), then the dual graph G^* has the properties that the degree of each node is no more than three (because triangles have three sides), and that G^* is acyclic and connected: it is a tree.

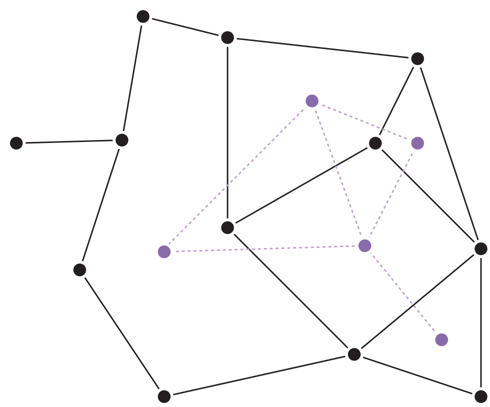


Figure 3.44: A planar graph and its dual

3.4.2 Metric spaces

This final section explores the kind of properties a model of space should have if it is to include the concept of distance between objects in the space. Such a space is called a *metric space*. The formal definition follows, but it should be noted that it does not always accord with our commonsense notion of distance.

A point-set S is said to be a *metric space* if there exists a distance function d that takes ordered pairs (s, t) of elements of S and returns a distance which satisfies the following three conditions:

metric space

1. for each pair s, t in S, $d(s, t) > 0$ if s and t are distinct points and $d(s, t) = 0$ if s and t are identical;
2. for each pair s, t in S, the distance from s to t is equal to the distance from t to s, $d(s, t) = d(t, s)$;
3. for each triple s, t, u in S, the sum of the distances from s to t and from t to u is always at least as large as the distance from s to u, that is: $d(s, t) + d(t, u) \geq d(s, u)$.

Put into more informal language, the first condition stipulates that the distance between points must be a positive number unless the points are the same, in which case the distance will be zero and the points will be *indiscernible* from each other. The second condition ensures that the distance between two points is independent of which way around it is measured, i.e., its distance measure is *symmetric*. The third condition states that it must always be at least as far to travel between two points via a third point rather than to travel directly. This third property is termed the *triangle inequality* (i.e., any two sides of a triangle must together be longer than the third side) due to the configuration shown in Figure 3.45.

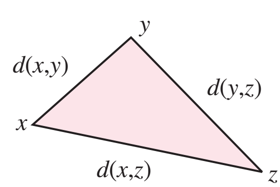


Figure 3.45: Triangle inequality

triangle inequality

These three properties of a distance measure—indiscernibility, symmetry, and the triangle inequality—together define a *metric*. In order to motivate this definition, we give below some possible distance functions and consider them with respect to properties 1–3 above. Let S be a set of cities on the globe and distance between two cities in S defined as follows (see Figure 3.46):

metric

geodesic distance

Geodesic distance The distance "as the crow flies" is termed the *geodesic distance*. In our example, it is the distance along the great circle of the Earth passing through the two city centers.

Manhattan distance

Spherical Manhattan distance The name "Manhattan distance" arises because in Manhattan many of the streets are arranged in a grid-like configuration. In the plane, the *Manhattan distance* is simply the sum of the difference in x and y coordinates. Similarly, the spherical Manhattan distance is the difference in latitudes plus the difference in longitudes (see Box 3.7 on the facing page).

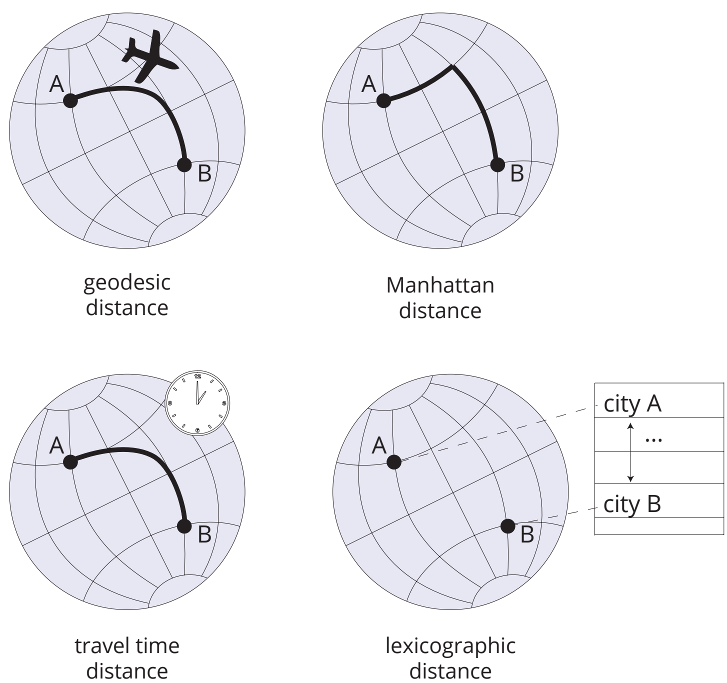
travel time distance

Travel time distance The minimum time required to travel from one city to the other (for example, using a sequence of scheduled airline flights) is the travel time distance.

lexicographic distance

Lexicographic distance The absolute value of the difference between the positions of cities in a fixed list of place names (*gazetteer*) is termed the *lexicographic distance*.

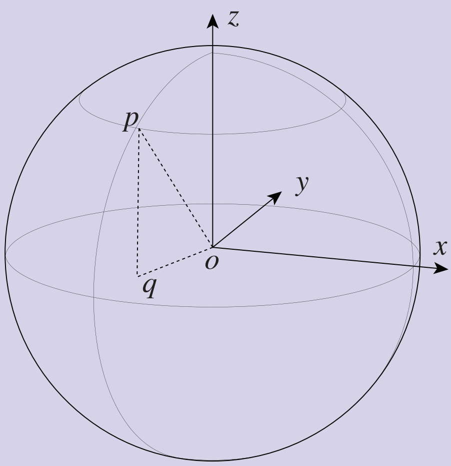
Figure 3.46: Distances defined on the globe

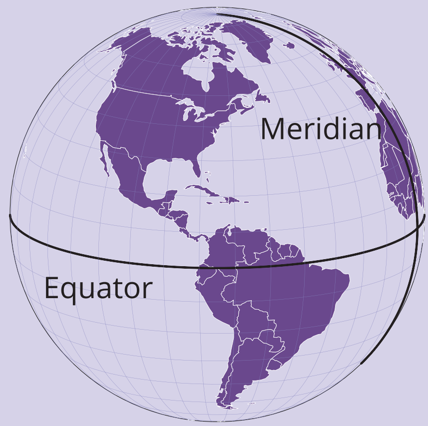


The first property of a metric space is quite uncontroversial, and satisfied by any self-respecting distance function. Sensible distances cannot be allowed to be negative. Also, the distance between an element and itself is always zero, whereas the distance between two distinct elements is always greater than zero. All the distance functions in our city example possess these properties. The third property, the triangle inequality, would also be a surprise if it did not hold for our distance measure. In plain English, the triangle inequal-

Box 3.7: Latitude and longitude

Many of the models in this chapter have been founded upon the plane. The plane is a useful approximation to the Earth's surface over small distances, but the Earth is not flat, so over larger distances other approximations are needed. The most common such approximation is the sphere. The surface of the sphere, although embedded in Euclidean 3-space, is 2D. Thus, any point on it may be uniquely specified by two numbers. A familiar system of coordinates for points on the surface of the sphere is latitude and longitude. In the diagram below, o is origin of the coordinate system, p is an arbitrary point on the sphere, and q is the projection of p onto the xy-plane. The angle between op and oq is the latitude of the point p. The angle between oq and the x-axis is the longitude of p. The great circle of the sphere in the xy-plane is the *equator* and the great circle in the xz-plane is the *meridian*. The usual topology of the surface of the sphere is similar to the usual topology of the Euclidean plane. Neighborhoods are sets of points with a constant distance from a fixed point, measured along a geodesic (great circle). Computing distances, angles, and areas on the sphere requires spherical trigonometry, with substantial differences to the planar geometry introduced in this chapter. On the sphere, straight lines meet at two points and the internal angles of a triangle sum to more than 180°, for example. A more detailed introduction to latitude and longitude may be found in Longley et al. (2015).





ity implies that it is never any farther to go by a direct route than an indirect route. All the above examples obey the triangle inequality.

Only the second property—of symmetry, that the distance from a to b is always the same as the distance from b to a—is a little more controversial. The geodesic, Manhattan, and lexicographic distances do all satisfy this condition. However, the travel time distance is not necessarily symmetric. For example, it is perfectly possible (and indeed usual due to prevailing winds) for the flight times between two cities to be different in each direction. A distance function that obeys properties 1 and 3 but not 2 is called a *quasimetric*. quasimetric

So, the collection of cities together with the geodesic distance function, the Manhattan distance function, and the lexicographic distance function are all metric spaces. The collection of cities with the travel time function is a quasimetric but not a metric space. The archetypal example of a metric space is Euclidean space, where the distance between two points is defined by the Pythagorean formula given earlier (Section 3.1.1). This distance function can easily be extended to higher dimensions.

Topology of metric spaces It turns out that a metric space has a natural topology. Let S be a metric space with distance function d. For each point, x in S and each real number r, define the *open ball* $B(x,r)$ to be the set of points whose distance from the point x is less than the number r. Expressed formally:

$$B(x,r) = \{y | d(y,x) < r\}$$

Define the set of neighborhoods to be the set of open balls. It is not hard to verify that this defines a topology for S. In the case of the Euclidean metric, this example reduces to the usual topology of the Euclidean plane. In the case of the Manhattan metric, then the open ball $B(x,r)$ will contain all points for which the sum of the horizontal and vertical distances from x is less than r. In the case of the Manhattan metric applied to a flat planar space, the open balls $B(x,r)$ will be squares of diagonal $2r$, as shown in Figure 3.47.

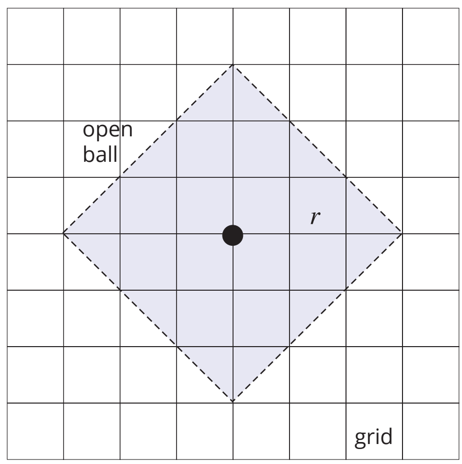


Figure 3.47: An open ball induced by the Manhattan metric

Travel time measures have some interesting properties. We have seen that, in general, they do not lead to metric spaces because they are not necessarily symmetric (for example, in a one-way traffic system). Let us for the moment make the simplifying assumption that distances are symmetric, so a topology may be defined as above. This is the travel-time topology introduced in Section 3.3.2. Computing topological neighborhoods can be illuminating in this case. Figure 3.48 shows a travel-time neighborhood (t-zone) of Liège, computed by Dussart and redrawn and considered by Tobler (1993).

The shaded region shows the area within 1 hour's travel time of the center of Liège in 1958 by the common means of travel available at that time. This "neighborhood" is not even connected, being the disjoint union of a finite number of cells. This type of pattern arises because of the discontinuous nature of travel on public transport, boarding and alighting at fixed and discrete points. As Tobler points out, such travel-time configurations are not amenable to modeling within the framework of Euclidean space. Such problems force us to use network models. Later chapters further develop this theme.

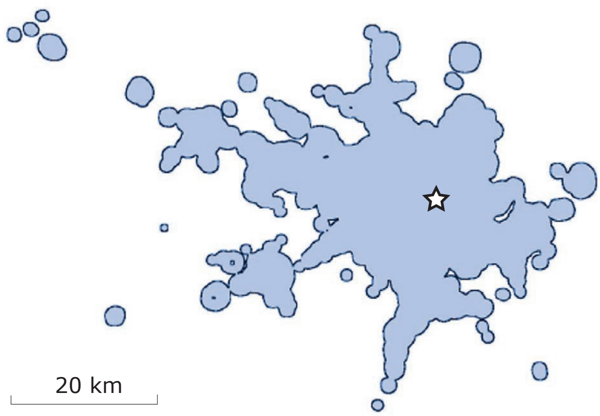


Figure 3.48: A travel-time neighborhood of Liège, computed by Dussart (Tobler, 1993). Shading indicates areas attainable within 1 hour from Liège center in 1958 by a combination of tramway, autobus, chemin de fer, and walking (5 km/h)

3.4.3 Fractal geometry

The appearance and characteristics of many geographic and natural phenomena, such as landscapes and coastlines, depend on the scale at which these phenomena are observed, termed *scale dependence*. For example, Figure 3.49 shows a satellite image of the Ganges River delta, Bangladesh, at three different scales. More detail is revealed at each finer scale, with detail revealed at finer scales tending to resemble details at coarser scales, termed *self-similarity*. The shape of the coastline and sinuosity of the rivers are similar at each scale, making it difficult to gauge the size of land forms in Figure 3.49. The scale has been deliberately omitted from Figure 3.49 to emphasize the effect; in fact, the first image is of an area about 60 km across, the final image approximately 15 km across. We can imagine repeating the process of zooming in over and over, perhaps until we reach the microscopic level. At every scale, new and self-similar detail would be revealed.

scale dependence

self-similar

The straight lines and smooth curves of Euclidean geometry are not well suited to modeling self-similarity and scale dependence. In a classic book, the Polish mathematician Benoit Mandelbrot (1982) argued that a fundamentally different type of geometry, which he dubbed *fractal geometry*, provides a more faithful representation of such natural and geographic phenomena. Fractal geometry concerns the study of shapes, called *fractals*, that are self-similar across all scales. Although the term "fractal" is relatively recent, fractal shapes and the problems they cause mathematicians have been known for hundreds of years (see Box 3.8 on page 121).

fractal geometry

fractal

True fractals are self-similar because they are defined recursively, rather than by describing their shape directly. For example, Figure 3.50 shows the first four stages in the construction of a famous fractal, the *Koch snowflake*, named after a Swedish mathematician, Helge von Koch. Building a Koch snowflake starts with an equilateral triangle, termed the *initiator* step. Then each straight line is divided into three equal parts. The middle part of each line is replaced with a new equilateral triangle, correctly scaled and with no base, termed the *generator* step. The Koch snowflake is the result of iterating

Koch snowflake

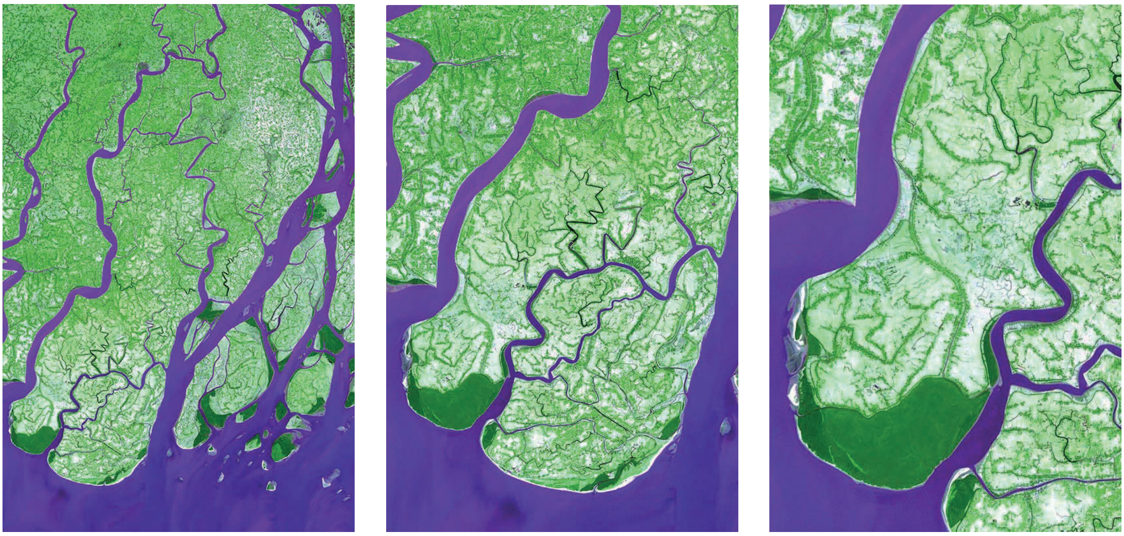


Figure 3.49: False-color composite using green, infrared, and blue wavelengths of Ganges River delta at three levels of detail (Source: NASA Landsat-7 image, February 2000)

the generator step an infinite number of times. Irrespective of the scale at which the Koch snowflake is viewed, it always exhibits the same level of detail.

Simple fractals like the Koch snowflake form a useful analogy to natural self-similar phenomena. More complex fractals can produce more "realistic" looking shapes. Figure 3.51 shows a synthetic "fractal landscape." Like the Koch snowflake, this landscape (including the clouds and water) is the result of recursively applying a feedback generator. The generator step of the Koch snowflake is essentially a similarity transformation, introducing at each iteration scaled copies of the original shape at one-third of the size of the previous iteration. In addition to being self-similar, fractals like those in Figure 3.51 are *self-affine*. Self-affine fractals can be constructed using affine transformations within the generator, so rotations, reflections, and shears can be used in addition to scaling (recall from Section 3.1.4 that all affine transformations are also similarity transformations).

self-affine

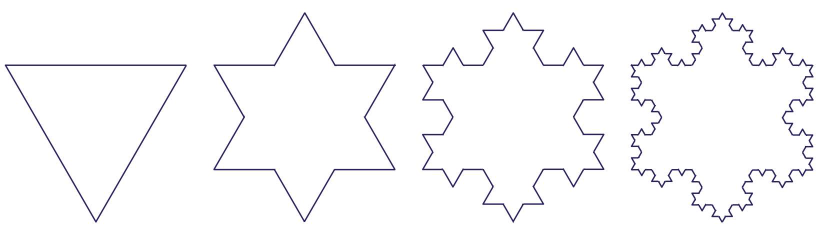


Figure 3.50: First four stages in the construction of the Koch snowflake

Aside from producing attractive pictures, fractal geometry has a number of more serious uses in GIS, for example, for map generalization (discussed further in Chapter 8). In order to produce maps at a range of different scales, it is often important to be able to decrease, and occasionally increase, the level of detail in the representation of spatial data while still retaining the essential characteristics of that line (for example, its "wiggliness"). Simplifica-

Box 3.8: Jordan curve theorem

In 1887, the French mathematician Camille Jordan formulated a famous theorem about simple loops, now known as the *Jordan curve theorem*. The theorem states that, given any simple loop, then the complement of the simple loop is not connected, but is partitioned into two connected components, one of which is bounded (called the *inside* of the loop) and one not bounded (called the *outside* of the loop). This proposition may seem so blindingly obvious as to be the sort of thing that gives mathematicians a bad name. In fact, it is quite difficult to prove convincingly. The problem lies in the wide variety of shapes, including some fractals, that qualify as loops. For example, the Koch snowflake in Figure 3.50 is a loop, but does not have a defined slope at any point on it. Technically, it is nowhere differentiable (see Chapter 4). The effort involved in proving the Jordan curve theorem resulted in the development of techniques that were instrumental in the birth of topology as a major branch of mathematics. The theorem has a practical application in the field of GIS, being the foundation of the point-in-polygon operation, discussed in more detail in Chapter 5.

tion requires less detail; enhancement requires more detail. By approximating the shape of a river, for example, as a fractal we can easily generate representations of that river at arbitrary levels of detail.

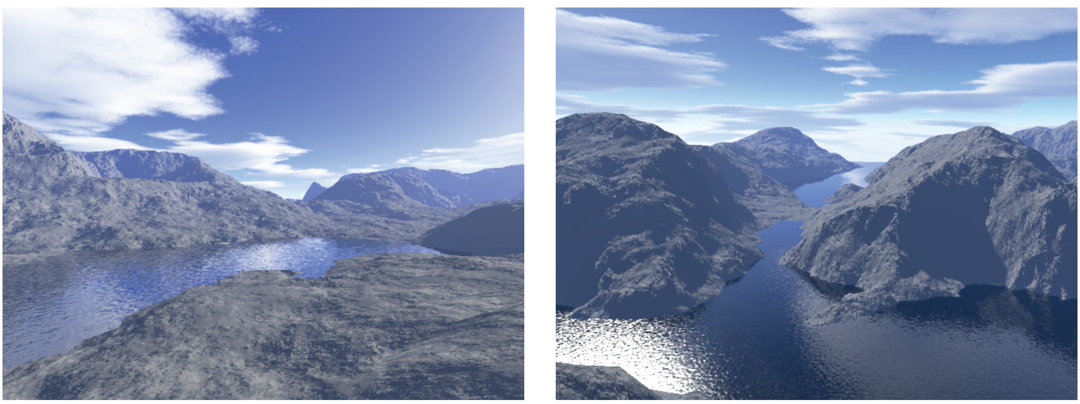


Figure 3.51: Entirely synthetic fractal landscapes

Other uses of fractals are often based on the concept of *fractal dimension*. Fractal dimension is an important property of fractals, which provides a measure of the degree to which new detail is revealed at different scales. The fractal dimension of a shape lies somewhere between the Euclidean dimensions of the shape and its embedding space. For example, the fractal dimension of the Koch snowflake in Figure 3.50 lies between 1, the dimension of a line, and 2, the dimension of the Euclidean plane.[3] Fractal dimension is an indicator of shape complexity: a shape with a high fractal dimension is complex enough to nearly fill its embedding space (e.g., a curve with a fractal dimension of 1.8 almost fills the plane). As a result, fractals are often referred to as *space-filling*. The space-filling characteristics of fractals are useful for indexing spatial information, as we shall see in Chapter 6.

fractal dimension

[3] The Koch snowflake, for example, has a fractal dimension of 1.26.

space-filling

Fractal dimension can also be a useful descriptor of a geographic shape. The fractal dimension of a river may be used as an indicator of the underlying geomorphological and hydrological processes involved in river formation. Similarly, fractal dimension analysis is used in landscape ecology to assess the complexity of geographic patches, such as plant or animal habitats. The fractal dimension of a "true" fractal shape, like the Koch snowflake, can be determined using theoretical analysis of the fractal generator. However, the fractal dimension of natural geographic phenomena, such as rivers, terrain surfaces, and animal habitats, must be determined empirically. Such empirical measurements of fractal dimension are notoriously unreliable and require careful analysis.

3.4.4 Summary of operations

This chapter has set out the fundamental spatial concepts that underpin most GIS. While other important spatial concepts exist, the Euclidean, geometric, topological, metric, and network spaces and concepts in this chapter are those most commonly encountered. We close this chapter by taking a slightly different perspective, summarizing some of the most important *operations* covered in this chapter in Table 3.4.

Table 3.4: Catalog of spatial operations

Group	Operation	Symbol	Operand(s)	Output
Set-oriented	equals	$=$	region, region	Boolean
	member of	$\in$	point, region	Boolean
	subset of	$\subseteq$	region, region	Boolean
	disjoint from		region, region	Boolean
	intersection	$\cap$	region, region	region
	union	$\cup$	region, region	region
	difference	$\backslash$	region, region	region
Network-oriented	adjacent		node, node	Boolean
	incident		node, edge	Boolean
	connected		node, node	Boolean
	degree		node	integer
	connected		graph	Boolean
	semi-connected		directed graph	Boolean
Topological	boundary	∂	cell	loop
	interior	$^\circ$	cell	open cell
	closure	$^-$	cell	closed cell
	meets		cell, cell	Boolean
	overlaps		cell, cell	Boolean
	inside		cell, cell	Boolean
	covers		cell, cell	Boolean
	connected		region	Boolean
	extremes		arc	set(points)
	within		point, loop	Boolean
Euclidean, metric	distance	$\vert\vert$	point, point	real
	bearing/angle	$\angle$	point, point	$[0, 2\pi[$
	length	$\vert\vert$	arc	real
	area		cell	real
	perimeter		cell	real

The operations in Table 3.4 are grouped into general, set-oriented, network-oriented, topological, and Euclidean and metric. The types for inputs to each operation (called *operands*) are shown along with the resulting output type. The spatial types used are *points*, of course, the basis of so many spatial structures; *regions* (arbitrary one- or two-dimensional point sets); *cells* (point sets homeomorphic to a disk); *arcs* (one-dimensional curves); *loops* (closed arcs with no self-intersections); as well as network types *nodes*, *edges*, and *graphs*, introduced in the previous chapter.

Some operations are unary (applying to a single operand); others are binary (applying to two operands). In addition to operations we have already seen in this chapter, the operation extremes acts on an object of type arc and returns the pair of points of the arc that constitute its end-points. Operation within provides a relationship between a point and a loop, returning true if the point is enclosed by the loop. This relationship is the often-used point-in-polygon operation, discussed in detail in Chapter 5.

The operations in Table 3.4 are a summary and guide, but by no means provide a complete typology.[4] For example, the operations distance, bearing, and angle are defined between the point elements of the space. In practice, it is often important to measure distances and angles between objects of different dimensions. For example, we might wish to know the distance of a town from a motorway. There are several ambiguities here. Do we mean the town center or the town as an area? Are we measuring distance along roads as the crow flies, or by some other means? If, for example, we are measuring as the crow flies, do we want the distance to the nearest point on the motorway or to the nearest intersection? These ambiguities must be resolved before the question can be answered properly. We return to some of the options in Chapter 5.

[4] It should be noted that not all operations in Table 3.4 are independent. For set theoretic operations, the usual set-theoretic constraints hold (for example, De Morgan's laws, $X \backslash (Y \cap Z) = (X \backslash Y) \cup (X \backslash Z)$ and $X \backslash (Y \cup Z) = (X \backslash Y) \cap (X \backslash Z)$). Other examples of interdependencies exist between the topological boundary, interior, closure, and set complement operations: for example, ∂X is the set difference of X^- and X°; and X^- is the set complement of $(X')^\circ$, the interior of the complement of X.

Reflections

Space is so much more than coordinates! While Euclidean geometry and the Cartesian plane are the most familiar and arguably most intuitive representations of space, a GIS employs multiple different representations in tandem. Geographic information scientists are experts in selecting the representation of space appropriate to the problem at hand, and they are also adept at switching between representations as required.

A wealth of mathematical texts exists to support those who wish to deepen their knowledge of the different types of spaces. Readers already more mathematically inclined may enjoy Coxeter (1961), which gives a fascinating overview of the variety that geometry offers, from basic work with triangles and polygons, through tessellations and 2D crystallography, to the platonic solids and golden section. A similarly stimulating tour of some of the most engaging ideas in graph theory can be found in Hartsfield & Ringel (2003), with a more systematic introduction to the topic in Saoub (2021).

Algebraic or combinatorial topology can be quite inaccessible to all but the mathematically gifted. A book that has withstood the test of time as a readable elementary introduction is Giblin (1977). Combinatorial maps were introduced in the Master's dissertation of distinguished computer scientist

Jack Edmonds (1960). Ohori, Ledoux, & Stoter (2015) explore the application of combinatorial structures, including simplicial complexes, to 3D and higher dimensional GIS. Zlatanova, Rahman, & Shi (2004) give a readable overview of some of the key concepts and challenges in capturing the topology in 3D GIS.

In connection with fractal geometry, Mandelbrot (1982) has become a modern classic. Burrough (1981) and Goodchild (1988) are two early examples of fractal geometry applied to spatial information, with further examples of fractal geometry applied to spatial information and GIS found in Lam & De Cola (1993) and Duckham, Drummond, & Forrest (2000).