

1

Introduction

This book presents statistical and computational methods for analyzing events that occur on or alongside networks. To this end, the first three chapters are concerned with preparations. This chapter shows the scope of this book, Chapter 2 fixes a general framework for spatial analysis, and Chapter 3 describes computational methods commonly used throughout the subsequent chapters. In this introductory chapter, we first describe the events under consideration, i.e., events that occur on and alongside networks, termed *network events*. Second, we show that if traditional spatial analysis assuming a plane with Euclidean distances, referred to as *planar spatial analysis*, is applied to network events, then it is likely to lead to false conclusions. Third, to overcome this shortcoming, we propose a new type of spatial analysis, namely *network spatial analysis*, which assumes a network with shortest-path distances. Fourth, we review studies on network events in the related literature and show how to apply network spatial analysis to those studies. Last, we describe the structure of the twelve chapters of the book and suggest how to read them according to the reader's interests. Note that network spatial analysis viewed from a board perspective is described in the preface of this volume.

1.1 What is network spatial analysis?

To introduce this new type of spatial analysis, we first define a key concept, *network events*, and next consider typical questions about network events to be solved by network spatial analysis. We then describe the salient features of network spatial analysis in contrast to the traditional planar spatial analysis.

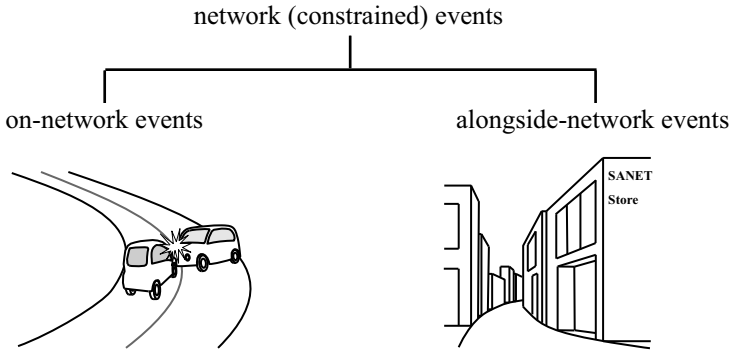


Figure 1.1 Network (constrained) events consisting of on-network events and alongside-network events.

1.1.1 Network events: events on and alongside networks

In the real world, there are numerous and various events that are strongly constrained by networks, such as car crashes on roads and fast-food shops located alongside streets. We call them *network-constrained events* (Yamada and Thill, 2007) or *network events* for short. Network events can be classified into two classes: events that occur directly on a network (e.g., car crashes on a road), and events that occur alongside a network rather than directly on it (e.g., fast-food shops located alongside a street). We refer to the former as *on-network events* and the latter as *alongside-network events*. Consequently, network events consist of on-network events and alongside-network events (Figure 1.1). Note that we sometimes use ‘along’ for both ‘on’ and ‘alongside.’

Figure 1.2 illustrates an actual example of on-network events, where each dot represents a traffic accident around Chiba station, Japan. As with this example, many types of network event have been reported in the related literature, including pedestrian and motor vehicle street accidents, roadkills of animals on forest roads, street crime sites, tree spacing along the roadside, seabirds located along a coastline, beaver lodges in watercourses, levee crevasse distribution on river banks, leakages in gas and oil pipelines, breaks in a wiring network, disconnections on the Internet, and blood clots in a vascular network (studies on network events including these examples will be reviewed in Section 1.2).

Figure 1.3 depicts an actual example of alongside-network events, where the black dots indicate advertisement agency sites alongside streets in Shibuya ward, one of the subcentral districts in Tokyo. There are many facilities that are located alongside street networks within densely inhabited areas. In fact, the entrances to almost all facilities in a city are adjacent to streets and users access amenities through these (Figure 1.1). Consequently, the locations of almost all facilities within an urbanized area can be regarded as alongside-network events.



Figure 1.2 Sites of traffic accidents around Chiba station, Japan (private roads are not shown).



Figure 1.3 The distribution of advertisement agency sites (the black points) alongside streets (the gray line segments) in Shibuya ward, one of the subcentral districts in Tokyo.

4 SPATIAL ANALYSIS ALONG NETWORKS

On- and alongside-network events such as those in the above examples are the major concern of this book. More specifically, this book primarily focuses on spatial distributions and relationships of such events on and alongside networks. Typical questions to be discussed in this volume are as follows:

- Q1: How can we obtain the catchment areas of parking lots in a downtown area including one-way streets, assuming that drivers access their nearest parking lots?
- Q2: Do boutiques tend to stand side-by-side alongside streets in a downtown area?
- Q3: Do street burglaries tend to take place near railway stations?
- Q4: Is the roadside land price of a street segment similar to those of the adjacent street segments?
- Q5: How can we locate clusters of fashionable boutiques alongside downtown streets?
- Q6: How can we estimate the density of traffic accidents and street crimes incidence, and how can we identify locations where the densities of those occurrence are high, referred to as *black spots* and *hot spots*?
- Q7: How can we spatially interpolate an unknown NO_x (nitrogen oxides) density at an arbitrary point on a road using known NO_x densities at observation points in a high-rise building district, such as Midtown Manhattan?
- Q8: How can we estimate the probability of a consumer choosing a specific fast-food shop among alternative shops located alongside streets in a downtown area?

1.1.2 Planar spatial analysis and its limitations

To answer the above types of question, we might conventionally use spatial methods that assume:

AP1: Events occur on a continuous (unbounded) plane.

AP2: If a method for analyzing the events includes distance variables, the distances are measured by Euclidean distance.

These types of spatial approach are referred to as *planar spatial methods*, and analyses made in this way are termed *planar spatial analyses*. Originally, planar spatial methods were designed for analyzing events on a plane, but in practice, as a matter of convenience, planar spatial methods are often applied to network events. However, this use is likely to lead to false conclusions, which are clearly demonstrated in Figure 1.4.

Having assessed the distribution of points in Figure 1.4a, nobody would consider that the points are randomly distributed. This view is true if the points are considered as being distributed on a plane; however, this becomes false when the points are seen to be located on a network indicated by the line segments in

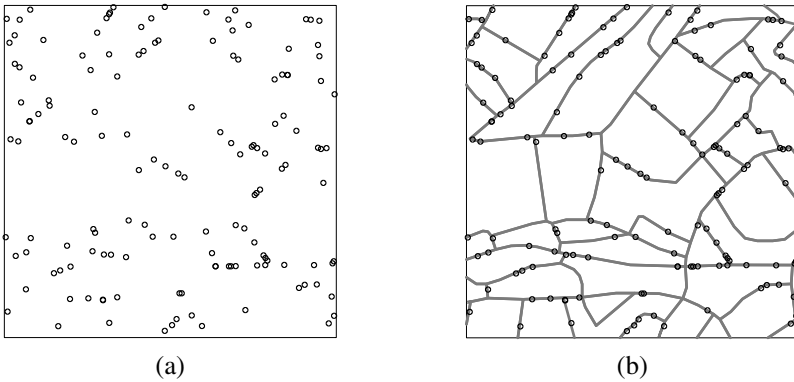


Figure 1.4 Point distributions: (a) nonrandomly distributed points on a bounded plane, (b) randomly distributed points on a network (note that the point distributions in (a) and (b) are the same).

Figure 1.4b. In fact, the points in this figure are randomly generated according to the uniform distribution across the network (for details, see Section 2.4.2 in Chapter 2). Figure 1.4 provides the following warning: analysis of network events using a planar spatial method is likely to lead to false conclusions. We shall show examples in subsequent chapters.

The second assumption AP2, i.e., the Euclidean distance assumption, is also arguable. The reasons for making this assumption are:

- it is much easier to compute Euclidean distance on a plane than the shortest-path distance on a network; and
- it is believed that the shortest-path distance is approximated by Euclidean distance.

The first reason remains true, although the difficulty is nowadays reduced because the use of geographical information systems (GIS) makes it easy to manage network data and to calculate shortest-path distances (a concise introduction to GIS is provided by Okabe (2004, 2005, Chapter 1)). The second reason may be true over a large region, but the validity of this concept is questionable across a small area or within a city. For example, Maki and Okabe (2005) report that in Kokuryo, a suburb of Tokyo, the difference between shortest-path distances and their corresponding Euclidean distances is significant if the Euclidean measurement is less than 400 m (see Figure 1.5). In addition, as shown in Table 1.1, the average radii of the service areas of many types of downtown store, exemplified by Shibuya ward in Tokyo, are less than 400 m. Planar spatial methods may be inappropriate therefore for analyzing alongside-network location events affected by trip behavior (for a further discussion, see Section 6.3 in Chapter 6).

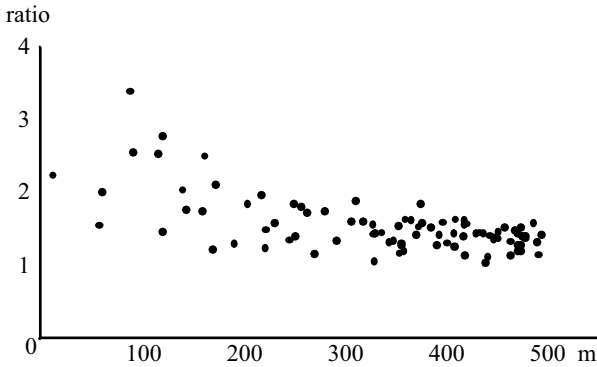


Figure 1.5 Ratio of the shortest-path distance to its corresponding Euclidean distance for the street network in Kokuryo, a suburb in Tokyo (data source: Maki and Okabe (2005)).

1.1.3 Network spatial analysis and its salient features

To overcome the above limitations of planar spatial methods, we now introduce a new type of spatial analysis that assumes:

AN1: Events occur on and alongside a network.

AN2: If a method for analyzing the events includes distance variables, the distances are shortest-path distances.

Corresponding to the planar spatial methods mentioned above (AN1 and AN2 correspond to AP1 and AP2, respectively), we call these methods *network spatial*

Table 1.1 Average radii of service areas in Shibuya ward, Tokyo.

Store type	Average radius (m)
Aromatherapy shop	282
Bag shop	271
Interior design shop	249
Daily necessities store	217
Preparatory school	216
Apartment estate agent	175
Printing store	167
Cafe	130
Japanese-style restaurant	106
Clothing store	85
Beauty shop	73

methods, and analyses that use network spatial methods, we call *network spatial analyses*. It should be noted that network spatial analysis does not imply the analysis of a network itself, such as geographical network analysis (Haggett and Chorley, 1969), communication network analysis (Kesidis, 2007), and circuit network analysis (Stanley, 2003). To avoid this confusion, we could use the terms *on-* or *alongside-network spatial analysis*, *network-constrained spatial analysis* (Yamada and Thill, 2004), *network-based spatial analysis* (Downs and Horner, 2007a, 2007b; Shiode, 2008) or more strictly, *spatial analysis on and alongside networks*. In this text, we use *network spatial analysis* for short.

We make a few remarks on the above two assumptions, AN1 and AN2. The first assumption AN1 describes places where events occur. The *on-network relation* is obvious. Events occur exactly on a network, such as traffic accidents. The *alongside-network relation* includes fairly broad spatial relations. It implies that the physical unit of an event (e.g., a store located at a site) has an access point on a network (the entrance of the store indicated by the black circle in Figure 1.6a) or the physical unit (the lot of the store) shares a common boundary line segment with a network (the bold line segment in Figure 1.6b). In addition, the alongside-network relation includes relations in which the physical unit of an event may intersect a network, for instance, a river intersects a road (Figure 1.6c) or a network goes through a forest area (Figure 1.6d). Computational treatments of these alongside-network relations are developed in Chapter 3 in detail.

The second assumption, AN2, specifies distance variables included in spatial methods. Consider, for instance, the analysis of boutique clusters in a downtown area using cluster analysis (for details, see Chapter 8). Because boutiques in clusters in a downtown area are located side-by-side alongside streets and customers access boutiques from entrances facing streets, it is natural to measure the closeness in terms of the shortest-path distance along streets. If a river separates two boutiques, it is not natural to assume that the boutiques belong to the same cluster even if the Euclidean distance between them is short. Underlying activities that result in boutique clusters are trips through streets, for example, window-shopping on sidewalks. In addition, many kinds of activities in a city are achieved through a street network, and so the configuration of activities may be influenced by trip

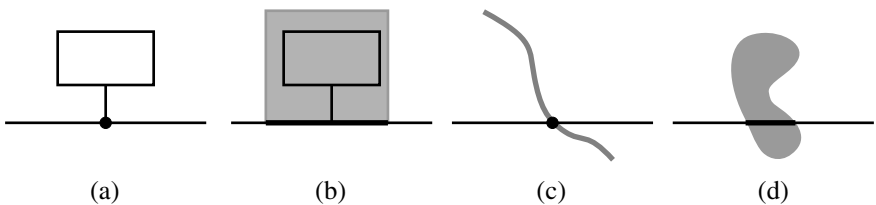


Figure 1.6 Alongside-network relations: (a) an access point (the black circle) of a polygon to a network (the horizontal line segment), (b) a boundary line segment of a polygon shared with a network (the bold line segment) (c) an intersection point of two networks (the black circles), (d) a network intersecting an area (the bold line segment).

behavior constrained by a street network. Consequently, network events may be best analyzed in terms of the shortest-path distance.

It should be noted, however, that there may be cases in which the shortest-path distance is not appropriate even if events occur on a network. For instance, consider the service area of a cell phone antenna. Although cell phone antennas stand on the edge of a street, their service areas are determined by Euclidean distance, because electric waves go straight through the air. The reader who wants to use a network spatial method should confirm whether or not the network spatial method is appropriate even when events are network events.

We notice from the above definition of network spatial analysis that it has salient features distinct from those of planar spatial analysis. First, by definition, network spatial analysis can properly analyze events occurring on and alongside a network. As a result, we can avoid the misleading conclusion illustrated by Figure 1.4. It is apparent from that figure that the selected points inevitably form clusters on a plane, because the points can exist only on a network. In fact, Yamada and Thill (2004) claimed that a planar spatial method (the K function method) overestimates clusters of traffic accidents in Buffalo (for details, see Chapter 6). Lu and Chen (2007) gave similar warning when analyzing urban crime distributed along streets. Such an overestimation is likely to happen not only for on-network events but also for alongside-network events. Therefore, clusters of stores in a city examined by planar spatial methods should be reexamined by network spatial methods.

Second, network spatial analysis can easily take account of directions, such as directions of current in a river and traffic flow regulation on a street network. In cities, particularly in downtown areas, many streets are one-way. In fact, about one third of streets in the downtown area of Kyoto are one-way (Okabe *et al.*, 2008). This implies that we cannot precisely estimate the delivery service areas of retail stores (e.g., pizza delivery stores) with Euclidean distances. Alternatively, we estimate the service areas in terms of the shortest-path distance on a directed network, and this estimation is investigated in detail in Chapter 4 (deterministic service areas) and Chapter 11 (probabilistic service areas).

Third, network spatial analysis can treat detailed networks using a common data structure. In a simple case, we represent a street by a line segment, but the street may consist of several components. For example, a street consists of vehicular roads (with two-way lanes), sidewalks, and crossings (Figure 1.7a). We can represent these details by a set of networks, as shown in Figure 1.7b, that share the same data structure (see Chapters 2 and 3).

Fourth, network spatial analysis can easily treat networks in three-dimensional space, such as underpaths and crossover bridges. This easy treatment is powerful when we analyze, for example, the incidence of pickpockets in a department store, egg-laying sites in an ant nest or blood clots in a vascular network. Figure 1.8 illustrates walkways, stairs, up/down escalators, and elevators in a department store, which are represented by a directed network.

Fifth, as will be shown in Section 2.3, network spatial analysis can treat nonuniform activities on a network more easily than planar spatial analysis can. While traditional spatial analysis methods are mostly designed to test the null

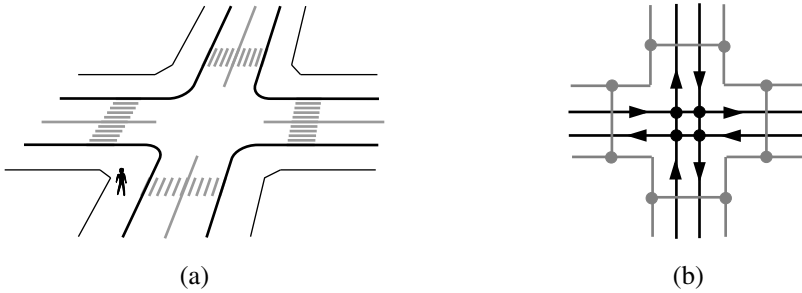


Figure 1.7 Entities represented by networks: (a) sidewalks, vehicular roads and crossing (entities), (b) the networks representing those entities.

hypothesis that events are uniformly distributed over a plane or network, this assumption is often violated in real-world phenomena. Consider, for example, traffic accidents on a road network. Obviously, traffic accidents do not occur uniformly across the network. Traffic accidents result from many factors, one of which is traffic volume (see Section 1.2.2). It is likely that the density of traffic accidents is proportional to traffic volume which naturally varies over a road network. Therefore, we cannot directly apply the traditional methods that assume uniform traffic volume to the distribution of traffic accidents resulting from nonuniform traffic volume. It is difficult to incorporate such nonuniformity in planar spatial analysis. Fortunately, however, we have good ‘magic’ that transforms a nonuniform density of an activity to a uniform density of the activity (to be shown in Section 2.4 in Chapter 2), to which we can apply traditional spatial methods assuming a uniform density. Through this transformation, we can easily analyze nonuniform activities on networks.

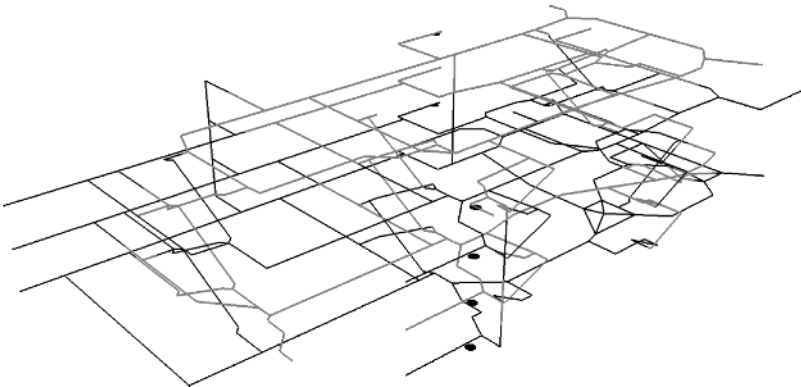


Figure 1.8 Walkways, stairs, up/down escalators, and elevators in a department store in Tokyo, where the circles indicate toilets (provided by T. Satoh). The subnetworks in different gray colors indicate the Voronoi cells of the three-dimensional network Voronoi diagram generated by the toilets (for definition, see Chapter 4).

Sixth, network spatial analysis gains analytical tractability because a network consists of one-dimensional line segments. Mathematical derivations on a one-dimensional space are more tractable than those on a two-dimensional space. For instance, to derive indexes or statistics, we often do integral computation, and single-integral computation is easier than double-integral computation. Therefore, we may obtain exact statistics for a network that could not be obtained for a plane.

Last, we should note the shortcomings of network spatial methods. On a plane, once the coordinates of points are given, we can easily compute the Euclidean distance between them. On a network, however, computation of the shortest path is not so simple and requires several steps. First, we must construct a database for managing a network. In practice, point data and network data are obtained from different sources and points that are supposed to be on a network are likely to be off the network rather than exactly on the network. Therefore, second, we must assign the points to the network. Third, we must use an algorithm for computing the shortest path on a network. In addition, we must perform many kinds of geometrical computation inherent in network spatial analysis. As a result, it is not straightforward in practice to extend statistical methods for planar spatial analysis to those for network spatial analysis. Network spatial analysis becomes practical only when its computation is possible. That is why the subtitle of this book is *Statistical and Computational Methods*. The computational methods in each chapter show how to solve difficult geometric computations encountered in network spatial methods in practice.

1.2 Review of studies of network events

As the above salient features indicate, network spatial analysis provides a suitable and powerful approach to the analysis of events occurring on and alongside networks. In fact, we can find many empirical studies of network events in various fields, although they do not always call their analyses network spatial analysis. In this subsection, we review these studies, but note that our review is not exhaustive and that our intent is merely to provide illustrative examples to be discussed in the following chapters.

1.2.1 Snow's study of cholera around Broad Street

Primitive qualitative network spatial analysis might date back many centuries ago when, for instance, a Roman ruler considered the location of colony settlements along Roman roads (Hodder and Orton, 1976). As far as we know, scientific quantitative network spatial analysis originated from John Snow's study in the mid-nineteenth century (Snow, 1855, 1936). John Snow's cholera map (Figure 1.9), which he called a diagram of the *topography of the outbreak* (Snow, 1855), illustrated one of the worst outbreaks of cholera that occurred around Broad Street and Golden Square in London in the mid-nineteenth century. The black bars along streets in Figure 1.9 indicate the number of victims. To find the source of the

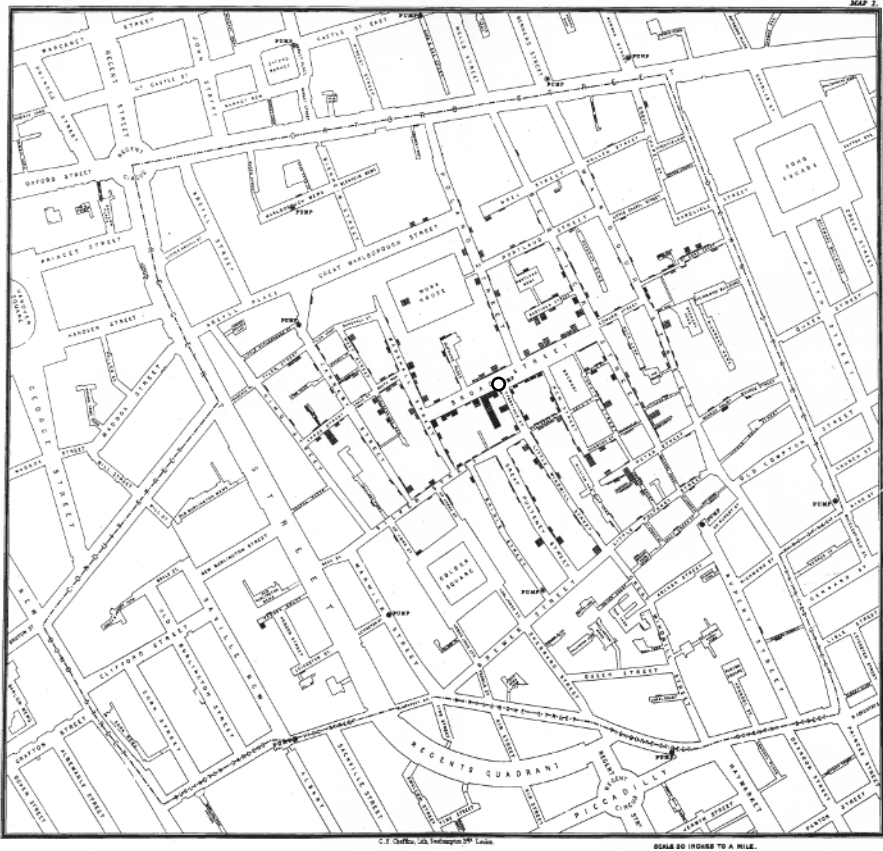


Figure 1.9 Snow's cholera map (Snow, 1855).

infections, he demarcated the area for which the Broad Street pump (the white circle in Figure 1.9) was the nearest pump among the 13 water pumps around this region in terms of the shortest-path distances along streets. Noticing that almost all victims were found in this demarcated area, he came to the conclusion that the cause of the victims' illness was contamination of the water from the Broad Street pump (his detection process was described in a book, *Ghost Map*, by Johnson (2006); see also Gilbert (1958), Smith (1982), Tufte (1997), McLeod (2000), and Koch (2004)). Snow called the demarcated area the *cholera area* (Snow, 1855). In modern terms, the area is a Voronoi subnetwork of the *network Voronoi diagram* (Okabe *et al.*, 2000, Section 3.8; Chapter 4 in this volume). Cliff and Haggett (1988, Figure 1.16D) explicitly applied the network Voronoi diagram to Snow's map, and Shiode (2012) deepened their study. Nakaya (2001) and Koch (2005) also reexamined Snow's map, although they used the planar Voronoi diagram.

Snow's study is indeed a landmark of network spatial analysis and epidemiology, but his statistical method was the descriptive statistics of the nineteenth century. Deepening his results required the 'modern' statistics (i.e., inferential statistics) created in the early twentieth century by Karl Pearson (1851–1936) and Ronald Aylmer Fisher (1890–1962), among others. Inferential statistics first developed univariate statistical methods and next multivariate statistical methods. Potentially, univariate statistical methods can be used for network spatial analysis because a univariate value can be represented by a point on a real line and the line can be regarded as a special case of a network. Using univariate distributions, Barton and David (1955, 1956a, 1956b), Pearson (1963), Pinder and Witherick (1973), and Young (1982) developed tests of randomness for points on a line. Watson (1961), Pearson and Stephens (1962), and Kuiper (1962) developed tests on a circle. A review of the early studies is provided by Selkirk and Neave (1984). In general, textbooks of spatial point processes refer to statistical methods on a line in their introductory chapters (e.g., Daley and Vere-Jones, 2003, Chapter 3). However, the methods used in those studies can be applied only to very specific examples of networks, i.e., one link between two nodes or one circle. Statistical analysis on complex networks consisting of more than one link began in the late twentieth century in various fields.

1.2.2 Traffic accidents

Researchers studying traffic accidents are concerned with the distribution of incidence spots on road networks. In particular, they want to find *black spots* or *black zones* (Maher and Mountain, 1988; Black, 1992; Flahaut, Mouchart, and Martin, 2003; Steenberghen *et al.*, 2004), where the density of traffic accidents is significantly high. In the early phase of traffic accident studies, researchers used spatially aggregated data to find black spots; specifically, the number of accidents was spatially aggregated with respect to road segments and the density of accidents on each road segment obtained from the aggregated number was used to detect black spots. The problem with that approach is how to divide a road network into road segments. Some researchers used unequal-length road segments (e.g., Ceder and Livneh, 1978; Turner and Thomas, 1986; Ng and Hauer, 1989; Stern and Zehavi, 1990; Miaou, 1994); and some researchers used equal-length road segments (e.g., Golob, Recker, and Levine, 1990; Thomas, 1996; Black, 1991; Erdogan *et al.*, 2008; Yamada and Thill, 2010). The result may vary according to the lengths of the road segments. This variation is a notorious known problem in spatial analysis, termed the *modifiable area unit problem* (abbreviated to MAUP) (Openshaw and Taylor, 1979, 1981; Openshaw, 1984).

For statistical tests, equal-length road segments are preferable, because statistical adjustments required by differences in size are not necessary, and several statistical methods assume equal-size spatial units, for instance, the *cell-count method*, i.e., the analysis using the numbers of road segments (cells in general) having certain numbers of accidents (see Section 9.1.3 in Chapter 9 for the precise

definition). Nicholson (1989) and Black (1992) used a similar method using accident counts on equal-size road segments. In practice, however, it is not easy to divide a road network into equal-length road segments, because the length of a link between nodes varies from link to link and it is impossible to find a common divisor for all the links. We discuss this problem in depth in Chapter 9, and show a few practical methods in Section 9.3.1.

Once the numbers of accidents are given with respect to road segments (whether they are equal length or not), we can regard the numbers as the attribute values of statistical data units. In this general context, various methods developed in ordinary statistics can be used for finding high-density road segments, for example, the goodness-of-fit test in Erdogan *et al.* (2008), and the Gini coefficient in Nicholson (1989) for pedestrian accidents. However, if attribute values of spatially close (or distant) units are correlated, which is referred to as *spatial autocorrelation* (see Chapter 7 for the precise definition), then statistical indexes that take spatial autocorrelation into account should be used. Black (1991, 1992), Black and Thomas (1998), and Yamada and Thill (2010) analyzed traffic accidents using such indexes. In Chapter 7, we describe a few types of spatial autocorrelation defined on networks.

The use of aggregated data in ordinary statistical analyses produces the MAUP. Alternatively, researchers can use disaggregated data, i.e., the data of individual incidence spots on a network. There are several statistical methods based on disaggregated data in the traffic accident literature. One such method is the *kernel density estimation method* (for the precise definition, see Chapter 9). Intuitively speaking, this method estimates the density of accidents by generating ‘mole hills’ (bell-shaped density functions with small domains; see Figure 9.3a in Chapter 9) at incidence spots and estimates the density by heaping those ‘hills’ (Figure 9.3b). Sabel *et al.* (2006) and Erdogan *et al.* (2008) employed this method to study traffic accidents in Christchurch in New Zealand and Afyokarahisar in Turkey, respectively.

A second method is the *nearest-neighbor distance method*, i.e., a statistic defined in terms of the distance from each point to its nearest neighbor (see Chapter 5 for the precise definition). Black (1992) applied this method to traffic accidents in a time–space context. Modifying the nearest-neighbor distance method, Steenberghen *et al.* (2004) used a statistic defined in terms of the number of accidents within a certain radius centered at each incidence spot. A third method is an extension of their method, called the *K function method*, in which the radius varies from zero to infinity (see Chapter 6 for the precise definition). Applications of the *K function method* to traffic accidents are found in Jones, Langford, and Bentham (1996), Yamada and Thill (2004), and Erdogan *et al.* (2008). A fourth method is *cluster analysis*, which finds sets of points that are spatially clustered on a network. Levine, Kim, and Nitz (1995) applied this method to traffic accidents in Honolulu. We develop cluster analysis for points on networks in Chapter 8.

Once black zones are detected by the above methods, the next problem is to identify the factors that produce those zones. Examples of influential factors reported in the literature are: traffic density (Pfundt, 1969; Ceder and Livneh, 1978, 1982; McGuigan, 1981; Ceder, 1982; Brodsky and Hakkert, 1983;

Lassarre, 1986; Ng and Hauer, 1989; Golias, 1992; Steenberghen *et al.*, 2004); geometry of junctions (Tanner, 1953; McGuigan, 1981; Nicholson, 1989), signal control (McGuigan, 1981; Braddock *et al.*, 1994), points of access and egress (Black, 1991), trees (Mok, Landphair, and Naderi, 2006), access to the nearest emergency services (Brodsky and Hakkert, 1983), population density along roads (Levine, Kim, and Nitz, 1995), lane and shoulder width (Ng and Hauer, 1989; Klop and Khattak, 1999), median strips (Bellis and Graves, 1971), speed limits (Lassarre, 1986; Abdel-Aty, Cunningham, and Gayah, 2008), curvature (Fink and Krammes, 1995), horizontal and vertical alignments (Shankar, Mannering, and Barfield, 1995; Kim and Boski, 2001), light conditions (Aerts *et al.*, 2006), line-of-sight distances (Van Kirk, 2000; Kim and Boski, 2001), bicycles traveling with or against traffic (Wachtel and Lewiston, 1994), and natural environmental factors, for example, fog (Black, 1991).

Analysis of the spatial relationship between black spots (more generally, network events) and the above influential factors differs according to whether data are aggregated or disaggregated. When the number of traffic accidents and the values of explanatory variables (e.g., traffic volume, speed limit) are spatially aggregated with respect to road segments, we can simply use a regression model, for example, the numbers of traffic accidents of road segments are regressed on the traffic volumes and speed limits of these road segments. In this case, as mentioned above, we may encounter the MAUP. To avoid it and to deepen the results of aggregated data analysis, we can alternatively use the disaggregated data of accident incidence spots on a road network, and treat the above influential factors in the following disaggregated manner (full developments are shown in Chapter 2).

The factors (in the form of entities) are classified in four types according to their geometrical forms. First, junctions, points of access/egress, traffic signs, traffic lights, street lights and other elements are represented by points on a network with their attribute values (the black circles in Figure 1.10a). Second, road segments with specific discrete or categorical characteristics, such as 50 m width roads and national roads, are represented by line segments with their attribute values (e.g., 50 m and national; the bold line segment in Figure 1.10b). Third, areas with specific characteristics, say foggy areas are represented by polygons, and the foggy areas are treated as foggy road segments that are included in the polygons or the minimum distance between a point in the polygon and a point on the road (the bold line segments or the broken line segment in Figure 1.10c). Last, traffic density, population density, illumination intensity, curvature, line-of-sight distance, access to the nearest emergency services and other attribute values that vary continuously along roads are represented by a function with a network as its domain (called a *field function* defined in Section 2.1.2 in Chapter 2, as in Figure 1.10d). Geometrical formulations and computational methods are developed in Chapters 2 and 3.

1.2.3 Roadkills

Animals can be traffic accident victims as well as humans. The former accidents are referred to as *roadkills of animals* or *roadkills* for short. It is reported in the

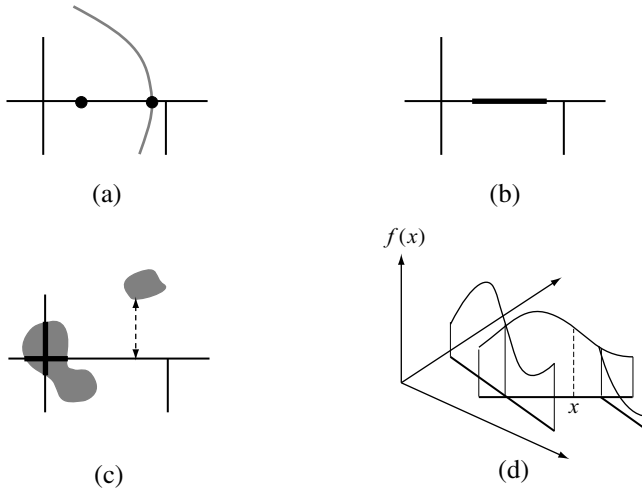


Figure 1.10 Influential factors classified according to geometrical forms: (a) points on a network, (b) line segments on a network, (c) line segments on a network included in a polygon and the distance from a polygon to the nearest point on a network, (d) a field function with a network as its domain.

literature that many kinds of animals can be victims of roadkills, including deer (e.g., Bellis and Graves, 1971), moose (Seiler, 2005), raccoon dogs (Saeki and Macdonald, 2004), wild boars (Malo, Suarez, and Diez, 2004), panthers (Forman and Alexander, 1998), otters (Philcox, Grogan, and Macdonald, 1999), rodents, voles, hedgehogs, rats, moles, shrews, and beech martens (Orlowski and Nowak, 2006), frogs and snakes (Forman and Alexander, 1998), and butterflies and burnets (Munguira and Thomas, 1992).

Statistical methods used in roadkill analysis are almost the same as those for traffic accident analysis. Differences appear in the factors influencing roadkills, although some are common. As in Section 1.2.2, the factors (entities) may be categorized by four types according to their geometrical forms. The first type of influential factor (entity) includes buildings alongside roads (Bashore, Tzilkowski, and Bellis, 1985; Malo, Suarez, and Diez, 2004), warning devices (Reed, Beck, and Woodard, 1982), swareflex reflectors (Schafer and Penland, 1985; Reeve and Anderson, 1993), underpasses (Ward, 1982; Foster and Humphrey, 1995; Malo, Suarez, and Diez, 2004), overpasses (Forman and Alexander, 1998), river crossings (Philcox, Grogan, and Macdonald, 1999), and crossroads or intersections (Malo, Suarez, and Diez, 2004; Seiler, 2005). These factors are represented by points on roads (Figure 1.10a).

The second type of influential factor (entity) includes bridges (Hubbard, Danielson, and Schmitz, 2000), fences (McKnight, 1969; Ward, 1982; Bashore, Tzilkowski, and Bellis, 1985; Philcox, Grogan, and Macdonald, 1999; Malo, Suarez, and Diez, 2004; Seiler, 2005), hedges (Malo, Suarez, and Diez, 2004), watercourses (Philcox, Grogan, and Macdonald, 1999), median strips (Bellis and Graves, 1971), gullies (Finder, Roseberry, and Woolf, 1999), the number of lanes (Hubbard,

Danielson, and Schmitz, 2000), road width (Forman and Alexander, 1998) and speed limit (Bashore, Tzilkowski, and Bellis, 1985; Forman and Alexander, 1998). These factors can be represented by line segments on roads with their attributes (Figure 1.10b). The third type of influential factor (entity) is represented by polygons characterized by land-use categories, for example, areas covered with woods (Bellis and Graves, 1971; Puglisi, Lindzey, and Bellis, 1974; Bashore, Tzilkowski, and Bellis, 1985; Finder, Roseberry, and Woolf, 1999; Hubbard, Danielson, and Schmitz, 2000; Saeki and Macdonald, 2004; Malo, Suarez, and Diez, 2004; Seiler, 2005) or grass or crops (Hubbard, Danielson, and Schmitz, 2000). These factors can be analyzed by the distance from an incidence spot to the nearest point in those polygons or road segments intersected by those polygons (Figure 1.9c). The last type of influential factor (entity) includes traffic volumes (Saeki and Macdonald, 2004; Seiler, 2005; Orłowski and Nowak, 2006) and in-line visibility (Case, 1978; Bashore, Tzilkowski, and Bellis, 1985). These factors can be represented by field functions defined on roads (Figure 1.10d). Geometrical formulations of these four types of influential factor are shown in Chapter 2 and computational methods are illustrated in Chapter 3.

1.2.4 Street crime

According to Sherman, Gartin, and Bürger (1989), 3.3% of street addresses and intersections in Minneapolis generated 50.4% of all dispatched calls for police service. These concentrated places are called *hot spots* (Brantingham and Brantingham, 1982; Sherman, Gartin, and Bürger, 1989; Sherman, 1995; Sherman and Weisburd, 1995; Bürger, Cohn, and Petrosino, 1995) or sometimes referred to as *mean streets* (Celik *et al.*, 2007) or *deviant streets* (Cohen, 1980). LaFree (1998) reported in the book subtitled *Street Crime and the Decline of Social Institutions in America* that in the early 1990s, nearly 25 000 Americans were being killed each year; in 1994, an estimated 620 000 Americans were robbed, 102 000 were raped, and 1.1 million were assaulted. In addition to these crimes, street crime (in a broad sense) includes car thefts, threats by street gangs, illegal drug trades, and graffiti and vandalism of public properties on and alongside streets. Street crime is a major subject of network spatial analysis.

Einstadter and Henry (2006) wrote that a French lawyer and statistician, Guerry, and a Belgian mathematician and astronomer, Quetelet first noted the uneven distribution of crime. This tendency was intensively surveyed in the 1920s by the Chicago School (e.g., Park, Burgess, and McKenzie, 1925); these are often referred to as *ecological crime studies*, and most of the research has been performed since then. Until the 1980s, studies mainly employed ordinary statistics without considering geographical proximity among data units, but since the 1990s, spatial statistical methods have been introduced to consider the configuration of crime spots (a comprehensive review is provided by Anselin *et al.* (2000), and Anselin, Griffiths, and Tita (2008)). The use of those methods was accelerated in practice by GIS-based tools, such as CrimeStat (Levine, 2004, 2006) and Spatial and Temporal Analysis of Crime (STAC) (Block, 1995). However, most methods and tools assume that crimes occur on the continuum of a plane (not restricted to streets) except

for Shiode (2011). This assumption is, as noticed by Levine (2004), hard to accept for street crimes. We show proper statistical methods for analyzing crimes on streets in the following chapters.

The statistical methods for street crimes are common to those for traffic accidents and roadkills, but factors affecting street crimes are different from those studies. We classify those influential factors (entities) according to the types defined in Section 1.2.2. The first type (Figure 1.10a) includes street lights (MacDonald and Gifford, 1989; Perkins, Meeks, and Taylor, 1992; Perkins *et al.*, 1993; Painter, 1996; Loukaitou-Sideris, 1999), trees, security/watch signs (Perkins *et al.*, 1993), liquor shops (Block and Block, 1995; Buslik and Maltz, 1998; Loukaitou-Sideris, 1999), taverns (Roncek and Maier, 1991), bars (Sherman, Gartin, and Bürger, 1989), convenience stores (Duffala, 1976; Brantingham and Brantingham, 1982), elevated stations (Block, 1998), abandoned dilapidated housing (Perkins *et al.*, 1993; Buslik and Maltz, 1998), schools (Buslik and Maltz, 1998; Block, 1998), bus stops, adult bookstores and pawn shops (Loukaitou-Sideris, 1999), and graffiti and litter (Perkins *et al.*, 1993; Loukaitou-Sideris, 1999). These factors are represented by points. The second type (Figure 1.10b) includes arterial streets (Eck, 1998), wide streets and unlit streets (Cohen, 1980), which are represented by line segments. The third type (Figure 1.10d) includes pedestrian flows (Nasar and Fisher, 1993), represented by a filed function across a network. Geometrical treatments of those factors are shown in Chapter 2.

1.2.5 Events on river networks and coastlines

Networks are not limited to road networks. Rivers and coastlines also form networks. Roline (1988) examined the effects of heavy metal pollution of the upper Arkansas River on the distribution of aquatic macroinvertebrates (see also Soares *et al.* (1999) and Elbaz-Poulichet *et al.* (1999)). Anderson, Brown, and Rappleye (1968) surveyed water quality and plant distributions along the Upper Patuxent River, Maryland (Milovanovic, 2007; Perona, Bonilla, and Mateo, 1999). Garcia (1999) investigated the spawning distribution of fall Chinook salmon in the Snake River (Heise *et al.*, 2004). Hayes, Leathwick, and Hanchet (1989) studied fish distribution patterns with environmental factors in the Mokau River, New Zealand (also see Nicola and Almodovar (2004)). Fustec, Cormier, and Lode (2003) surveyed the beaver lodge locations in the upstream Loire River. Cabrero *et al.* (1997) studied the geographical distribution of B chromosomes in the grasshopper *Eyprepocnemis plorans* along rivers in the Sugura River Basin. Moss *et al.* (1989) studied the phytoplankton distribution in a temperate flood-plain lake and river system. Richter *et al.* (1998) made a spatial assessment of hydrologic alteration within a river network in the Upper Colorado Basin. Paul and Pillai (1986) investigated the distribution of radium in the Periar River from Karala to Moolampalli (El-Gamal, Nasr, and El-Taher, 2007). Gregory, Davis, and Tooth (1993) studied the spatial distribution of coarse woody debris dams in the Lymington Basin (also see Evans, Hungr, and Clague (2001)). In practice, tools for analyzing events along rivers are useful. Ries *et al.*

(2008) provided a software package, called *StreamStats*, for analyzing stream networks consisting of rivers, water bodies and constructed channels.

It should be noted that most of these studies employed ordinary statistics without explicitly considering spatial proximity among data units. A few exceptions are, however, found in the literature. Dacey (1960b) applied the reflexive nearest-neighbor method proposed by Clark (1956) (see Chapter 5) to the spacing of cities along rivers in the central lowland USA. Lewis (1977) applied the run test (Wald and Wolfowitz, 1940) to the distribution of biochemical oxygen demand along the River Trent. Cressie and Majure (1997) interpolate nitrate concentration on streams in the upper North Bask watershed (also see the studies in Section 9.2).

In addition to rivers, water lines include coastlines. Griggs and McCrory (1975) surveyed waste-water discharges along the Pacific Coast, California. Sivertsen (1997) investigated the distribution of kelp beds along the Norwegian coast. O'Driscoll (1998) examined the distribution of sea birds along a coastline. Storlazzi and Field (2000) studied sediment distribution along a rocky, embayed coast in the Carmel Bay, California. Vila *et al.* (2001) analyzed the distribution of algal blooms along the Catalan Coast.

1.2.6 Other events on networks

Studies of events on road networks are not restricted to the traffic accidents, roadkills and street crimes reviewed above. We next refer to other events that occur on road networks, together with events on other types of network.

Spoooner *et al.* (2004) applied the network K function method (see Chapter 6) to *Acacia* populations on a road network in Lockhart Shire, Australia. Suzuki (1987), and Furth and Rahbee (2000) analyzed bus stop spacing along bus routes. Hillier (1984, 1996) analyzed straight-line visibility on roads using the concept of 'space syntax.' Using this concept, many researchers (e.g., Penn *et al.*, 1998; Jiang and Claramunt, 2002; Penn, 2003; Porta *et al.*, 2006a,b) have examined the structures of street networks in cities.

In a broad sense, road networks include passageways in complex facilities and corridors in buildings, and events on these networks are also within the scope of network spatial analysis. After the September 2001 terrorist attacks at the World Trade Center in New York City, Kwan and Lee (2005) drew attention to sites of accidents in corridor networks in high-rise buildings. Castle and Longley (2005) inspected interrupted routes at London's King's Cross Station (see also Pu and Zlatanova (2005)).

In addition, we find many kinds of network events that occur on networks other than road and river networks. Deckers *et al.* (2005) investigated the effect of landscape structure on the invasive spread of black cherry in the network of hedgerows in Flanders using the K function method (see Chapter 6) on a network. Maheu-Giroux, and Blois (2006) conducted landscape-scale analysis of *Phragmites australis* invasion patterns in two preurban areas in Quebec, focusing on the interaction between the network of wetlands and the adjacent land uses

using the K function method. Pipelines also form a network. Tucciarelli, Criminisi, and Termini (1999) analyzed reduction of leakages in a water pipe network by means of optimal valve regulation and Oliveira *et al.* (2008) studied breakages in water pipelines in Monroeville. Cevik and Topal (2003) surveyed damaged places on natural gas pipelines (also see Hwang *et al.* (2004)).

1.2.7 Events alongside networks

As illustrated in Figure 1.1, network events consist of on-network events and alongside-network events. The network events in the preceding subsections are all on-network events. However, in the real world, alongside-network events are also common, particularly location events in urbanized areas, i.e., various kinds of facilities located side by side alongside streets. In the literature, these location events are conventionally examined using planar spatial analysis, but the results may lead, as noted in Figure 1.4, to potentially false conclusions. We suggest reexamining these types of study with network spatial methods.

Of course, some studies do pay attention to alongside-network events. For example, Hodder and Orton (1976) investigated the distribution of late Iron Age coins alongside Roman roads, while Stark and Young (1981) applied a distance-based method to the distribution of archaeological sites alongside a trail in the Cabeza de la Vaca Arroyo in Mexico. Similarly, Li *et al.* (1990) surveyed the occurrence of abnormal hemoglobin in the populations of cities alongside the Silk Road in northwest China, and Xie *et al.* (2007) studied the distribution of ancient cities alongside the midsection of the Silk Road's He–Xi Corridor.

In related work, several studies more explicitly examine alongside-network events using network spatial methods. For instance, Okabe, Yomono, and Kitamura (1995) studied the distribution of liquor shops and car dealers in relation to railway stations in Nishinomiya, Japan using the nearest-neighbor distance method defined on a network (see Chapter 5). Using the same method, Sevtsuk (2010) considered the distribution of retail and food establishments in Cambridge and Somerville, Massachusetts. Moreover, because the service areas of facilities are sensitive to network patterns, the configuration of facilities is examined using Voronoi diagrams defined on a network (see Chapter 4). For example, Furuta, Uchida, and Suzuki (2005) estimated the service areas of ambulance stations in Seto, Japan, using the k -th nearest point Voronoi diagram defined on a network (Section 4.2.3 in Chapter 4).

Similarly, Okabe *et al.* (2008) determined the catchment areas of parking lots in Kyoto using the Voronoi diagram on a network. The K function method defined on a network (see Chapter 6) is also useful for analyzing facilities in urbanized areas. Using this particular network spatial method, Myint (2008) examined the distributions of banks, fast-food restaurants, schools, and churches in Norman City, Oklahoma. Besides these network spatial methods, Shiode (2008) examined the distribution of commercial facilities in Shibuya ward, Tokyo, using the cell-count method based on equal-size subnetworks (see Section 9.1.3 in Chapter 9), while Shiode and Shiode (2009b) found clumps of restaurants in Shibuya ward by using the clumping method defined on a network (see Section 8.2 in Chapter 8).

Finally, Borruzo (2008) estimated the density of banks and insurance companies in Trieste, Italy, applying a kernel density estimation method to a network embedded in a plane (see Section 9.2 in Chapter 9).

1.3 Outline of the book

As the above review indicates, a wide range of subjects can be studied by network spatial analysis. This book intends to provide statistical and computational methods for such studies, with the expectation that the methods provide more suitable approaches than the existing methods. We outline those methods in this last subsection by referring to questions solved by the methods.

1.3.1 Structure of chapters

The book consists of twelve chapters including this introductory chapter (Chapter 1), and the structure of these chapters is illustrated in Figure 1.11.

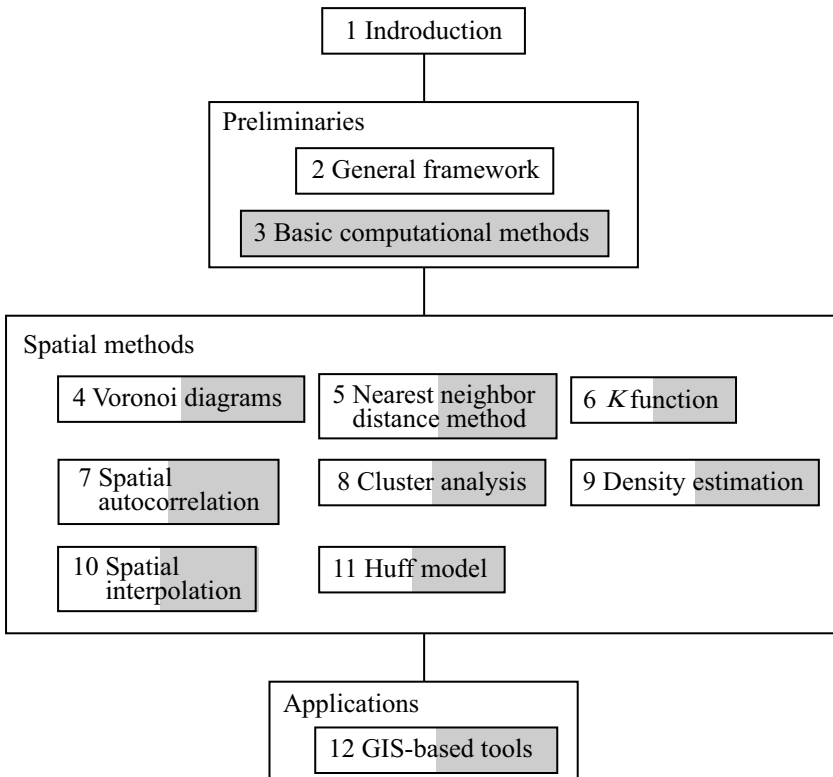


Figure 1.11 Structure of chapters (the shaded sections indicate computational parts).

The eleven remaining chapters constitute three parts: preliminaries, network spatial methods and applications.

The first part, consisting of Chapters 2 and 3 is concerned with preliminaries. Chapter 2 introduces fundamental concepts for network spatial analysis that are commonly used throughout this volume. In the chapter, we first show how conceptually to model real world networks and events on or alongside networks, and then illustrate how to numerically represent those conceptual models as data models. Third, we formulate a basic probabilistic model for events on a network, which is used as a basic null hypothesis throughout this book. In Chapter 3, we describe the elements of computational methods that are commonly used for carrying out the statistical calculations formulated in Chapters 4–11. We first show data structures for spatial analysis on one-layer and multiple-layer networks. Second, we illustrate geometric computational methods in a general way and then specify them for networks. Basic algorithms are illustrated with examples.

The main body of this book is the second part, consisting of Chapters 4–11. This part shows statistical methods for network spatial analysis and their efficient computational methods. The contents are described in the next subsection.

The last part, Chapter 12, shows tools for network spatial analysis and their applications to actual data. In practice, implementing the computational methods in each chapter in operational software requires much time and cost, particularly for those who are not skilled in programming. To simplify this task, the last chapter, Chapter 12, introduces GIS-based tools for network spatial analysis, in particular, a free software package, called *SANET* (Spatial Analysis along NETworks; Okabe, Okunuki, and Shiode, 2006a, 2006b; Okabe and Satoh, 2009). The book ends with references and an index.

1.3.2 Questions solved by network spatial methods

The second part of the book shows eight network spatial methods. We explain each method in terms of the questions it can solve.

Chapter 4 explains a family of *network Voronoi diagrams*, with which we can answer the question:

Q1': For a given set of n generators, which may be points (e.g., fast-food shops), line segments (e.g., arterial streets), or circuits enclosing polygons (e.g., parks) on a network, how can we tessellate the network into n subnetworks associated with the n generators in such a way that the generator nearest from every point in a subnetwork is the generator assigned to the subnetwork?

A specific example of this general question is question Q1 in Section 1.1.1 (the catchment areas of parking lots). The resulting set of subnetworks is referred to as the *network Voronoi diagram*. The diagram can include inward and outward

network Voronoi diagrams, weighted network Voronoi diagrams and k -th nearest point network Voronoi diagrams.

Chapter 5 formulates the *network nearest-neighbor distance method*, which gives answers to the questions:

- Q2': Given a set of points on a network, how can we test whether or not the shortest-path distance from each point to the next nearest point is significantly short (or long)?
- Q3': Given two sets of points (of types A and B) on a network, how can we test whether the shortest-path distance from each point of type A to the nearest point of type B is significantly short (or long)?

Specific examples are shown in questions Q2 (boutiques located side by side) and Q3 (burglaries occurring around stations) in Section 1.1.1.

Chapter 6 explains the *network K function method*. This method is similar to the above nearest-neighbor distance method, but different in that the method considers not only the nearest point but also points further away. This method is an alternative method to answer questions Q2 (boutiques located side by side) and Q3 (burglaries occurring around stations) in Section 1.1.1. More generally, the method can answer the questions:

- Q2'': Given a set of points on a network, how can we test whether or not the number of points within a shortest-path distance from each point is significantly many (or few)?
- Q3'': Given two sets of points (type A points and type B points) on a network, how can we test whether the number of type A points within a shortest-path distance from each point of type B is significantly many (or few)?

Chapter 7 discusses the *network spatial autocorrelation*, which can answer question Q4 (land value similarity along streets) in Section 1.1.1. This specific question can be generalized as:

- Q4': Given a set of attribute values of spatial units on a network (which may be represented by points, line segments, subnetworks) with the degrees of closeness between those spatial units (which may be categorical or numerical), are attribute values similar if their spatial units are close each other?

Chapter 8 illustrates *network spatial cluster analysis*, which gives an answer, for example, to question Q5 (clusters of fashionable boutiques) in Section 1.1.1, or more generally to the question:

- Q5': For a given set of points on a network, how can we find sets of clusters within which points are close to each other but between which the points are apart?

Chapter 9 shows the *network kernel density estimation method*. As in the example of question Q6 (detection of ‘black spots’ and ‘hot spots’) in Section 1.1.1, this method is useful for answering the question:

Q6': For a given set of points on a network, how can we estimate the density of points along the network and detect high-density areas on the network?

Chapter 10 discusses *spatial interpolation* on a network, which examines, for instance, question Q7 (interpolating NO_x densities) in Section 1.1.1. More generally:

Q7': Given known attribute values at a finite number of sample points on a network, how can we interpolate or predict an unknown attribute value at an arbitrary point on the network?

Chapter 11 formulates the *network Huff model*, a network version of the Huff model often used in marketing. The model can answer question Q8 (the choice probability of a fast-food shop), or the question:

Q8': Given facilities located alongside a network and suppose that a user accesses the facilities through the network, how can we estimate the probability of the user choosing a facility among alternative facilities?

In each chapter, statistical and computational methods for answering these questions are described in detail.

1.3.3 How to study this book

This book can be read in several ways according to the reader's interest in network events. It is recommended that readers who wish to understand network spatial analysis fully should read Chapters 1, 2 and 3 in that order; next Chapters 4–11 (the order depends on the reader's preference) and finally Chapter 12. It is recommended that readers who want to understand one of the network spatial methods shown in Chapters 4–11 should read Chapters 1, 2 and 3 in that order, and then proceed to the chapter containing the method that the reader wants to understand. The reader who wants to analyze data of network events with a specific network spatial method can first see whether a tool for that spatial method is provided in Chapter 12; if so, read the chapter containing that method in Chapters 4–11. As indicated by the subtitle of this book, *Statistical and Computational Methods*, the book shows each method in a separate subsection in every chapter. The reader who is mainly interested in statistical methods can read Chapters 1, 2 and the subsections on statistical methods in each chapter; the reader who is mainly interested in computational methods can read Chapters 1–3, the subsections on computational methods in each chapter and Chapter 12. However readers approach the book, we believe that it provides readers with theoretically as well as practically useful methods for network spatial analysis.

