ATSUYUKI OKABE
KOKICHI SUGIHARA

Spatial Analysis along Networks

Statistical and Computational Methods

WILEY

STATISTICS IN PRACTICE

Spatial Analysis along Networks

Statistics in Practice

Series Advisors

Human and Biological Sciences
Stephen Senn
CRP-Santé, Luxembourg

Earth and Environmental Sciences
Marian Scott
University of Glasgow, UK

Industry, Commerce and Finance
Wolfgang Jank
University of Maryland, USA

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series can be found at www.wiley.com/go/statisticsinpractice

Spatial Analysis along Networks

Statistical and Computational Methods

Atsuyuki Okabe

School of Cultural and Creative Studies
Aoyama Gakuin University
Emeritus Professor, University of Tokyo, Japan

Kokichi Sugihara

Graduate School of Advanced Mathematical Sciences
Meiji University
Emeritus Professor, University of Tokyo, Japan

WILEY

A John Wiley & Sons, Ltd., Publication

This edition first published 2012

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex,
PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data
Okabe, Atsuyuki, 1945-
Spatial analysis along networks : statistical and computational methods / Atsuyuki Okabe and Kokichi Sugihara.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-77081-8 (cloth)
1. Spatial analysis (Statistics) 2. Spatial analysis (Statistics)–Data processing. 3. Geography–Network analysis. I. Sugihara, Kokichi, 1948- II. Title.
QA278.2.O359 2011
519.5'36–dc23

2011040047

A catalogue record for this book is available from the British Library.

ISBN: 978-0-470-77081-8

Set in 10.25/12pt Times by Thomson Digital, Noida, India

Contents

Preface xiii

Acknowledgements xvii

1 Introduction 1
- 1.1 What is network spatial analysis? 1
  - 1.1.1 Network events: events on and alongside networks 2
  - 1.1.2 Planar spatial analysis and its limitations 4
  - 1.1.3 Network spatial analysis and its salient features 6
- 1.2 Review of studies of network events 10
  - 1.2.1 Snow's study of cholera around Broad Street 10
  - 1.2.2 Traffic accidents 12
  - 1.2.3 Roadkills 14
  - 1.2.4 Street crime 16
  - 1.2.5 Events on river networks and coastlines 17
  - 1.2.6 Other events on networks 18
  - 1.2.7 Events alongside networks 19
- 1.3 Outline of the book 20
  - 1.3.1 Structure of chapters 20
  - 1.3.2 Questions solved by network spatial methods 21
  - 1.3.3 How to study this book 23

2 Modeling spatial events on and alongside networks **25**
- 2.1 Modeling the real world 26
  - 2.1.1 Object-based model 26
    - 2.1.1.1 Spatial attributes 27
    - 2.1.1.2 Nonspatial attributes 28
  - 2.1.2 Field-based model 28
  - 2.1.3 Vector data model 29
  - 2.1.4 Raster data model 30
- 2.2 Modeling networks 31
  - 2.2.1 Object-based model for networks 31
    - 2.2.1.1 Geometric networks 31
    - 2.2.1.2 Graph for a geometric network 32
  - 2.2.2 Field-based model for networks 33
  - 2.2.3 Data models for networks 34

2.3 Modeling entities on network space 34
2.3.1 Objects on and alongside networks 34
2.3.2 Field functions on network space 37
2.4 Stochastic processes on network space 37
2.4.1 Object-based model for stochastic spatial events on network space 38
2.4.2 Binomial point processes on network space 38
2.4.3 Edge effects 41
2.4.4 Uniform network transformation 42

3 Basic computational methods for network spatial analysis 45
3.1 Data structures for one-layer networks 46
3.1.1 Planar networks 46
3.1.2 Winged-edge data structures 47
3.1.3 Efficient access and enumeration of local information 49
3.1.4 Attribute data representation 51
3.1.5 Local modifications of a network 52
3.1.5.1 Inserting new nodes 52
3.1.5.2 New nodes resulting from overlying two networks 52
3.1.5.3 Deleting existing nodes 53
3.2 Data structures for nonplanar networks 54
3.2.1 Multiple-layer networks 54
3.2.2 General nonplanar networks 56
3.3 Basic geometric computations 57
3.3.1 Computational methods for line segments 57
3.3.1.1 Right-turn test 57
3.3.1.2 Intersection test for two line segments 58
3.3.1.3 Enumeration of line segment intersections 58
3.3.2 Time complexity as a measure of efficiency 59
3.3.3 Computational methods for polygons 60
3.3.3.1 Area of a polygon 60
3.3.3.2 Center of gravity of a polygon 61
3.3.3.3 Inclusion test of a point with respect to a polygon 61
3.3.3.4 Polygon-line intersection 62
3.3.3.5 Polygon intersection test 62
3.3.3.6 Extraction of a subnetwork inside a polygon 63
3.3.3.7 Set-theoretic computations 64
3.3.3.8 Nearest point on the edges of a polygon from a point in the polygon 65
3.3.3.9 Frontage interval 66
3.4 Basic computational methods on networks 66
3.4.1 Single-source shortest paths 67
3.4.1.1 Network connectivity test 70
3.4.1.2 Shortest-path tree on a network 71

3.4.1.3 Extended shortest-path tree on a network 71
3.4.1.4 All nodes within a prespecified distance 72
3.4.1.5 Center of a network 72
3.4.1.6 Heap data structure 73
3.4.2 Shortest path between two nodes 77
3.4.3 Minimum spanning tree on a network 78
3.4.4 Monte Carlo simulation for generating random points on a network 79

4 Network Voronoi diagrams 81
4.1 Ordinary network Voronoi diagram 82
4.1.1 Planar versus network Voronoi diagrams 82
4.1.2 Geometric properties of the ordinary network Voronoi diagram 83
4.2 Generalized network Voronoi diagrams 85
4.2.1 Directed network Voronoi diagram 86
4.2.2 Weighted network Voronoi diagram 88
4.2.3 *k*-th nearest point network Voronoi diagram 89
4.2.4 Line and polygon network Voronoi diagrams 91
4.2.5 Point-set network Voronoi diagram 93
4.3 Computational methods for network Voronoi diagrams 93
4.3.1 Multisource Dijkstra method 94
4.3.2 Computational method for the ordinary network Voronoi diagram 95
4.3.3 Computational method for the directed network Voronoi diagram 96
4.3.4 Computational method for the weighted network Voronoi diagram 97
4.3.5 Computational method for the *k*-th nearest point network Voronoi diagram 98
4.3.6 Computational methods for the line and polygon network Voronoi diagrams 99
4.3.7 Computational method for the point-set network Voronoi diagram 100

5 Network nearest-neighbor distance methods 101
5.1 Network auto nearest-neighbor distance methods 102
5.1.1 Network local auto nearest-neighbor distance method 103
5.1.2 Network global auto nearest-neighbor distance method 104
5.2 Network cross nearest-neighbor distance methods 106
5.2.1 Network local cross nearest-neighbor distance method 106
5.2.2 Network global cross nearest-neighbor distance method 108
5.3 Network nearest-neighbor distance method for lines 111

5.4 Computational methods for the network nearest-neighbor distance methods 112
5.4.1 Computational methods for the network auto nearest-neighbor distance methods 112
5.4.1.1 Computational methods for the network local auto nearest-neighbor distance method 113
5.4.1.2 Computational methods for the network global auto nearest-neighbor distance method 116
5.4.2 Computational methods for the network cross nearest-neighbor distance methods 116
5.4.2.1 Computational methods for the network local cross nearest-neighbor distance method 116
5.4.2.2 Computational methods for the network global cross nearest-neighbor distance method 117

6 Network *K* function methods 119
6.1 Network auto *K* function methods 120
6.1.1 Network local auto *K* function method 121
6.1.2 Network global auto *K* function method 122
6.2 Network cross *K* function methods 122
6.2.1 Network local cross *K* function method 123
6.2.2 Network global cross *K* function method 124
6.2.3 Network global Voronoi cross *K* function method 126
6.3 Network *K* function methods in relation to geometric characteristics of a network 127
6.3.1 Relationship between the shortest-path distance and the Euclidean distance 127
6.3.2 Network global auto *K* function in relation to the level-of-detail of a network 129
6.4 Computational methods for the network *K* function methods 131
6.4.1 Computational methods for the network auto *K* function methods 131
6.4.1.1 Computational methods for the network local auto *K* function method 132
6.4.1.2 Computational methods for the network global auto *K* function method 133
6.4.2 Computational methods for the network cross *K* function methods 133
6.4.2.1 Computational methods for the network local cross *K* function method 133
6.4.2.2 Computational methods for the network global cross *K* function method 134
6.4.2.3 Computational methods for the network global Voronoi cross *K* function method 136

7 Network spatial autocorrelation **137**
7.1 Classification of autocorrelations 139
7.2 Spatial randomness of the attribute values of network cells 145
7.2.1 Permutation spatial randomness 145
7.2.2 Normal variate spatial randomness 146
7.3 Network Moran's *I* statistics 146
7.3.1 Network local Moran's *I* statistic 147
7.3.2 Network global Moran's *I* statistic 148
7.4 Computational methods for Moran's *I* statistics 150

8 Network point cluster analysis and clumping method **153**
8.1 Network point cluster analysis 155
8.1.1 General hierarchical point cluster analysis 155
8.1.2 Hierarchical point clustering methods with specific intercluster distances 160
8.1.2.1 Network closest-pair point clustering method 160
8.1.2.2 Network farthest-pair point clustering method 161
8.1.2.3 Network average-pair point clustering method 161
8.1.2.4 Network point clustering methods with other intercluster distances 162
8.2 Network clumping method 162
8.2.1 Relation to network point cluster analysis 162
8.2.2 Statistical test with respect to the number of clumps 162
8.3 Computational methods for the network point cluster analysis and clumping method 164
8.3.1 General computational framework 164
8.3.2 Computational methods for individual intercluster distances 166
8.3.2.1 Computational methods for the network closest-pair point clustering method 166
8.3.2.2 Computational methods for the network farthest-pair point clustering method 168
8.3.2.3 Computational methods for the network average-pair point clustering method 169
8.3.3 Computational aspects of the network clumping method 170

9 Network point density estimation methods **171**
9.1 Network histograms 172
9.1.1 Network cell histograms 172
9.1.2 Network Voronoi cell histograms 174
9.1.3 Network cell-count method 175
9.2 Network kernel density estimation methods 177
9.2.1 Network kernel density functions 178
9.2.2 Equal-split discontinuous kernel density functions 181
9.2.3 Equal-split continuous kernel density functions 183

9.3 Computational methods for network point density estimation 184
9.3.1 Computational methods for network cell histograms with equal-length network cells 184
9.3.2 Computational methods for equal-split discontinuous kernel density functions 186
9.3.3 Computational methods for equal-split continuous kernel density functions 190

10 Network spatial interpolation 195
10.1 Network inverse-distance weighting 197
10.1.1 Concepts of neighborhoods on a network 197
10.1.2 Network inverse-distance weighting predictor 198
10.2 Network kriging 199
10.2.1 Network kriging models 200
10.2.2 Concepts of stationary processes on a network 201
10.2.3 Network variogram models 203
10.2.4 Network kriging predictors 206
10.3 Computational methods for network spatial interpolation 209
10.3.1 Computational methods for network inverse-distance weighting 209
10.3.2 Computational methods for network kriging 210

11 Network Huff model 213
11.1 Concepts of the network Huff model 214
11.1.1 Huff models 214
11.1.2 Dominant market subnetworks 215
11.1.3 Huff-based demand estimation 216
11.1.4 Huff-based locational optimization 217
11.2 Computational methods for the Huff-based demand estimation 217
11.2.1 Shortest-path tree distance 218
11.2.2 Choice probabilities in terms of shortest-path tree distances 220
11.2.3 Analytical formula for the Huff-based demand estimation 220
11.2.4 Computational tasks and their time complexities for the Huff-based demand estimation 221
11.3 Computational methods for the Huff-based locational optimization 222
11.3.1 Demand function for a newly entering store 223
11.3.2 Topologically invariant shortest-path trees 224
11.3.3 Topologically invariant link sets 225
11.3.4 Numerical method for the Huff-based locational optimization 227
11.3.5 Computational tasks and their time complexities for the Huff-based locational optimization 230

12 GIS-based tools for spatial analysis along networks and their application 231
12.1 Preprocessing tools in SANET 232
12.1.1 Tools for testing network connectedness 233
12.1.2 Tool for assigning points to the nearest points on a network 233
12.1.3 Tools for computing the shortest-path distances between points 234
12.1.4 Tool for generating random points on a network 234
12.2 Statistical tools in SANET and their application 235
12.2.1 Tools for network Voronoi diagrams and their application 236
12.2.2 Tools for network nearest-neighbor distance methods and their application 237
12.2.2.1 Network global auto nearest-neighbor distance method 238
12.2.2.2 Network global cross nearest-neighbor distance method 239
12.2.3 Tools for network *K* function methods and their application 240
12.2.3.1 Network global auto *K* function method 241
12.2.3.2 Network global cross *K* function method 241
12.2.3.3 Network global Voronoi cross *K* function method 243
12.2.3.4 Network local cross *K* function method 244
12.2.4 Tools for network point cluster analysis and their application 245
12.2.5 Tools for network kernel density estimation methods and their application 246
12.2.6 Tools for network spatial interpolation methods and their application 247

References 249

Index 271

Preface

As its title indicates, this book is devoted to spatial analysis along networks, referred to as *network spatial analysis*, or more explicitly, statistical and computational methods for analyzing events occurring on and alongside networks. Network spatial analysis is of practical use for analyzing, among other things, the occurrence of traffic accidents on highways, the incidence of crime on streets, the location of stores alongside roads, and the contamination of rivers (Chapter 1 introduces many applications). This usefulness is the main reason we focus on network spatial analysis in this volume. However, there is also a more general and somewhat more ambitious justification for this work. That is, when viewed from a broader perspective, we expect that network spatial analysis will prove to be a first step toward next-generation spatial analysis.

Having reviewed the extant literature on spatial analysis, we note that most empirical studies incorporate spatially aggregated data across subareas, such as administrative districts, census tracts, and postal zones. We refer to this type of spatial analysis as *subarea-based spatial analysis* or *meso-scale spatial analysis*. One of the earliest and most notable examples of this type of spatial analysis is included in a compilation titled *The City* (Park, Burgess, and McKenzie, 1925), written by sociologists at the Chicago School (sometimes described as the Ecological School). More specifically, Burgess (1925) surveyed land use of subareas in Chicago and formulated the concentric-zone model, subsequently followed by Hoyt's (1939) sector model and the Harris–Ullman multiple nuclei model (Harris and Ullman, 1945).

Since then, subarea-based spatial analysis has become one of the most important approaches to empirical spatial analysis. Even today, we frequently employ subarea-based spatial analysis for empirical studies because subarea data, including population and other census-related data, are widely available and because it is generally straightforward to apply ordinary statistical techniques, including regression analysis, to the attribute values of subareas. Unlike the empirical literature, we find that the development of most theoretical work on spatial analysis has assumed an 'ideal space', that is, real space is represented by unbounded homogeneous space with Euclidean distance. This ideal space is convenient for developing pure theories of spatial analysis or spatial stochastic processes; indeed, the derivations of many useful theorems employ this assumption (see, e.g., Illian *et al.*, (2008)). However, ideal space is far from the real world.

In the late twentieth century, the availability of detailed spatial data increased dramatically thanks to rapid progress in data acquisition technologies, such as the global positioning system (GPS) and many kinds of geosensors. Better data availability potentially enables us to analyze spatial events in detail by representing individual entities in the real world in terms of geometric objects in two- or three-dimensional Cartesian space instead of aggregating them into subareas (see Chapter 2 for this representation). We describe this possible form of spatial analysis as *object-based spatial analysis* or *micro-scale spatial analysis*, in contrast to the well-established subarea-based spatial analysis or meso-scale spatial analysis. At present, however, the methods for micro-scale spatial analysis are at an early stage. We believe that one clue to micro-scale spatial analysis would be to represent real space by networks embedded in two- or three-dimensional Cartesian space. This is because many kinds of events or activities in the real world are constrained by networks, such as streets, railways, water and gas pipe lines, rivers, electric wires, and communication networks. A first step toward micro-spatial analysis would thus appear to be network-constrained spatial analysis, which is the main concern of this volume.

In network spatial analysis, we measure the shortest-path distance. Unfortunately, its computation is much more difficult than that of Euclidean distance because it requires the management of network topology. Therefore, network spatial analysis becomes practical only when efficient computational methods are available. Dijkstra (1959) developed a key algorithm for this purpose in the middle of last century. Since then, there has been extensive study of location problems on networks by a variety of researchers, mainly in operations research (Handler and Mirchandani, 1979; Daskin, 1995; for a review, see Labbe, Peeters, and Thisse (1995)). We should note that the focus in these studies has been locational optimization or the computing of network characteristics (e.g., Kansky, 1963; Haggett and Chorley, 1969), with rather less attention paid to the statistical analysis of events on networks.

To fill this gap in the literature, we develop statistical and computational methods for network spatial analysis by introducing computational methods originally developed for operations research and computational geometry (Preparata and Shamos, 1985; Chapter 3 in this volume presents some basic computational methods). In this sense, the network spatial analysis presented in this volume is a first step toward micro-scale spatial analysis. However, we cannot present real world space by either network or Euclidean space alone as it is a complex hybrid system with elements of both. The next step, then, would be object-based spatial analysis in a hybrid space consisting of a discrete network space with shortest-path distance and a continuous space with Euclidean, or more generally, geodesic distance. An initial attempt is Cressie *et al.* (2006).

We are now in the midst of an ongoing revolution brought about by information and communication technologies. In the future, microcomputer tips, tags, and geosensors will be embedded in almost every entity (including moving objects) in our environment, and the integration of these devices with communication systems (e.g., the Internet) will establish an intellectual system joining the virtual world of computers and the global real world. This system will then realize a society we refer to as the *ubiquitous computing society*, in which at any time and in any place, people can receive the most appropriate personalized information for action given their particular circumstances in time and space (Sakamura and Koshizuka, 2005). To construct this system, micro-scale spatial analysis is expected to extend to *real-time spatial analysis*, that is, spatial analysis in which the circumstances of an acting body (including a person, a group of persons, a company, or possibly a robot) are analyzed and appropriate personalized information for action is derived almost instantaneously (Okabe, 2009a, 2009b). We intend that this volume, in presenting state-of-the-art methodology for network spatial analysis, will contribute a first step toward micro-scale spatial analysis and encourage our readers to further develop micro-scale spatial analysis and, from there, tackle the challenge of real-time spatial analysis.

Atsuyuki Okabe
Kokichi Sugihara
March 2012

Acknowledgements

When we first thought of the concept underlying this book in June 2007, we consulted Noel Cressie on possible publication. In turn, he was kind enough to introduce our proposal to the statistics and mathematics section at John Wiley. A positive response meant that our long project could begin in September 2007. Since then, so very many people have helped us in different ways in developing and presenting this book that it would be impossible to acknowledge all of them individually.

To start with, we are very grateful to those who have read our drafts and offered useful comments, particularly Ikuho Yamada on the general concepts underpinning network spatial analysis (Chapters 1 and 2) and spatial autocorrelation (Chapter 8), Toshiaki Satoh on kernel density estimation (Chapter 9) and GIS-based tools (Chapter 12), and Kei-ichi Okunuki on the Huff model (Chapter 11). Our special thanks also go to those with whom we discussed related subjects and who in turn provided us with inspiration. These especially include Mike Tiefelsdorf and Barry Boots on spatial autocorrelation, Yuzo Maruyama and Yonghe Li on kriging, Shino Shiode on inverse-distance weighting and cell counting, and Hisamoto Hiyoshi on spatial interpolation. They also include Atsuo Suzuki, Takehiro Furuta, and Shinji Imahori on equal cell splitting, Kei-ichi Okunuki and Masatoshi Morita on the K function method, and Yasushi Asami and Yukio Sadahiro on urban analysis.

We would also like to express our thanks to those who helped us to run the necessary programs, particularly Toshiaki Satoh, Kayo Okabe, Akiko Takahashi, and the staff at the Center for Spatial Information Science (CSIS) at the University of Tokyo and the Information Science Research Center at Aoyama Gakuin University. We are also indebted to Ayako Teranishi for collecting the more than 500 related papers, entering them in our database, and editing the references and compiling the index, and to Tsukasa Takenaka for constructing the online database with which we could develop our book while we were away. We also thank Masako Yoshida for the retrieval program used for the references, Tetsuo Kobayashi for collecting research articles, and Aya Okabe for designing the website, along with the web crew members involved in its management at CSIS, through which we received many practical comments on the GIS-based toolbox known as SANET from users across 51 countries.

We are thankful to the staff at John Wiley, particularly Richard Davies, Ilaria Meliconi, Heather Kay, Susan Barclay, Kathryn Sharples, and Prachi Sinha Sahay for their helpful assistance. We also acknowledge a grant-in-aid by the Japan

Society for the Promotion of Science for a project entitled ‘Development of methods, algorithms, and GIS-based tools for statistical spatial analysis on networks’ (#20300098), and data provision by the Chiba Prefectural Police, NTT Data, and CSIS. Finally, we thank our respective partners, Kayo Okabe and Keiko Sugihara, for their lifelong encouragement and invaluable support before and during the writing of this book.

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.

Spatial Analysis along Networks: Statistical and Computational Methods, First Edition.
Atsuyuki Okabe and Kokichi Sugihara.

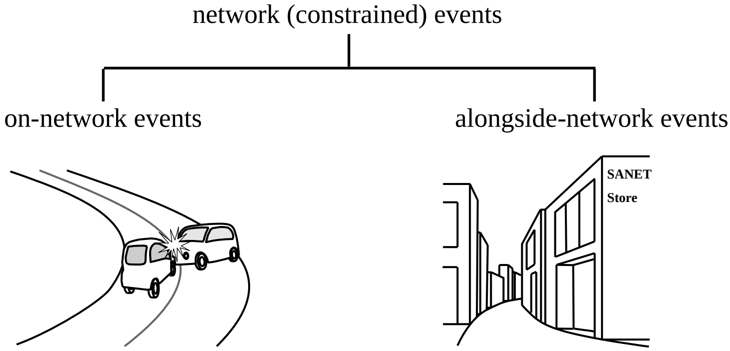


Figure 1.1 Network (constrained) events consisting of on-network events and alongside-network events.

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.

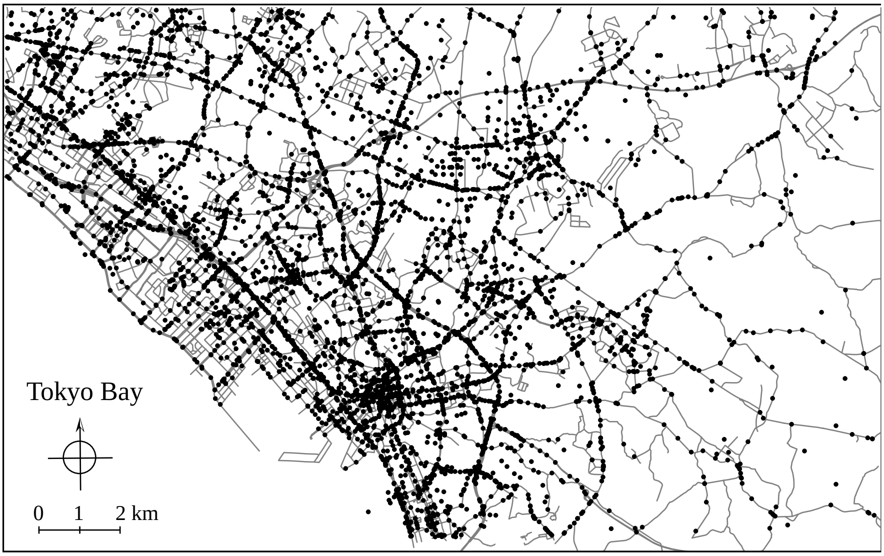


Figure 1.2 Sites of traffic accidents around Chiba station, Japan (private roads are not shown).

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.

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

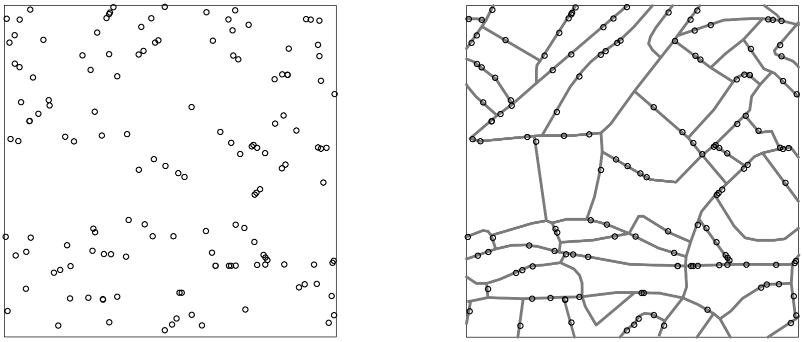
To answer the above types of question, we might conventionally use spatial methods that assume:

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



(a) (b)

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).

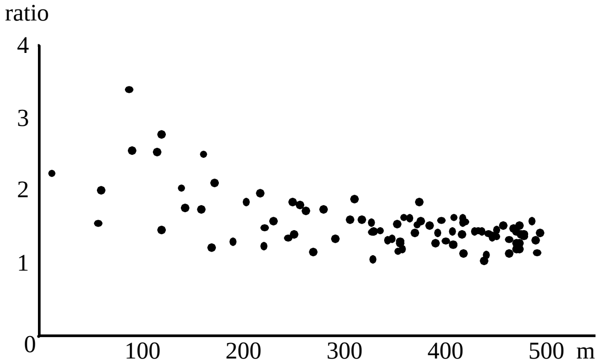


Figure 1.5 Ratio of the shortest-path distance to its corresponding Euclidean distance for the street network in Kokuryo, a suburb in Tokyo (data source: Maki and Okabe (2005)).

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*

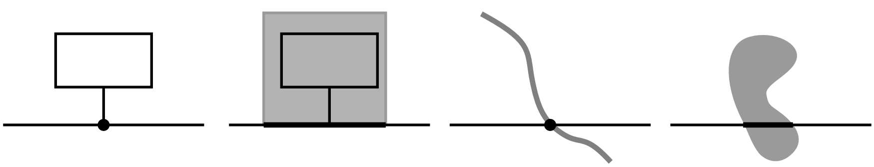
Table 1.1 Average radii of service areas in Shibuya ward, Tokyo.

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



(a) (b) (c) (d)

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

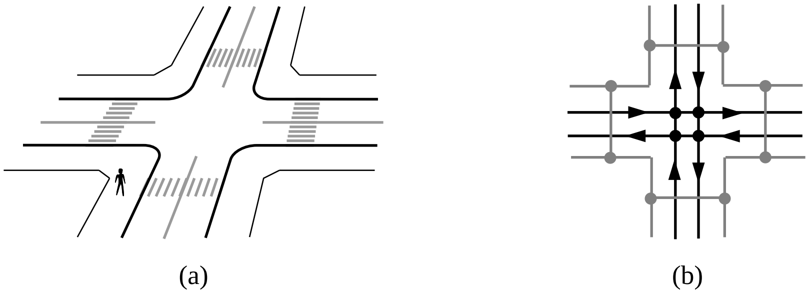


Figure 1.7 Entities represented by networks: (a) sidewalks, vehicular roads and crossing (entities), (b) the networks representing those entities.

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.

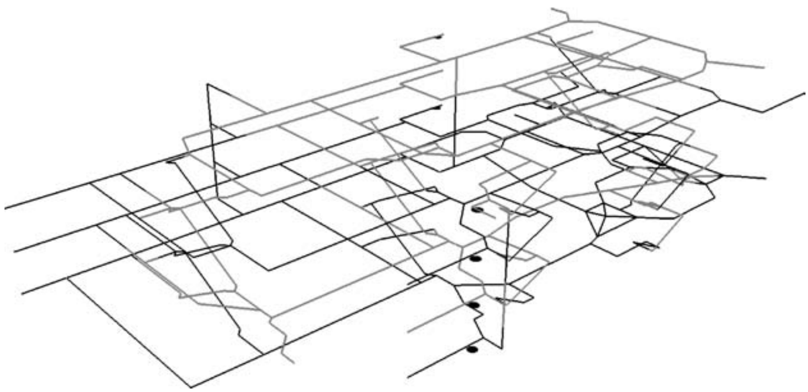


Figure 1.8 Walkways, stairs, up/down escalators, and elevators in a department store in Tokyo, where the circles indicate toilets (provided by T. Satoh). The subnetworks in different gray colors indicate the Voronoi cells of the three-dimensional network Voronoi diagram generated by the toilets (for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