Discussion Paper Series
No. ***107***

Department of Urban Engineering
University of Tokyo

Object-oriented spatial analysis: Set-based exploratory analysis of the relations among the distributions of spatial objects

Yukio Sadahiro
Department of Urban Engineering, University of Tokyo

Object-oriented spatial analysis: Set-based exploratory analysis of the relations among the distributions of spatial objects

Yukio Sadahiro

Department of Urban Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
sada@ua.t.u-tokyo.ac.jp

Keywords: Object-oriented spatial analysis, topology diagram, spatial relations

Abstract

This paper develops a new exploratory method for analyzing the relations among distributions of spatial objects. Existing methods usually focus on the distribution of one or two types of spatial objects. This paper, on the other hand, deals with the relations among more than two distributions of spatial objects. In addition, the method is applicable independent of the type of spatial objects. These advantages assure a wide applicability and flexibility of the method. Analysis starts with the evaluation of local spatial relations among objects. This paper then develops a computational algorithm for clustering distributions based on a specific spatial relationship such as spatial proximity. The result is visualized by a graph-based representation and evaluated by numerical measures. The method is applied to the analysis of the distributions of commercial facilities in Chiba City, Japan. Technical soundness of the method is discussed as well as empirical findings.

1. Introduction

Geography deals with the distribution of a wide variety of geographical entities and phenomena. Geographical information science represents them as spatial objects such as points, lines, polygons, and surfaces and assigns their attributes as nominal, categorical, and numerical variables. Geographers analyze their spatial distribution, the relationship between the distribution of objects and their attributes, and so forth.

The point is one of the most basic and fundamental spatial objects used in geographical information science. Statistical analysis of point distributions is often called point pattern analysis and numerous methods have been developed in the literature. They include quadrat method (Goodall, 1952; Greig-Smith, 1952; Pielou, 1969), nearest neighbor distance (Skellam, 1952; Clark and Evans, 1954; Diggle, 2003), K-function (Ripley, 1976, 1977, 1981), and kernel density estimation (Rosenblatt, 1956; Parzen, 1962; Silverman, 1986).

The line also plays a critical role in geographical information science. Lines represent water stream, gas pipelines, traffic, electric, and information networks. Analytical methods based on graph theory permit us to evaluate the properties of networks including their size, connectivity and density (Shimble, 1953; Haggett and Chorley, 1969). These methods are recently applied to the analysis of social and web networks (Wasserman and Faust, 1994; Carrington *et al.*, 2005; Knoke and Yang, 2008; Abraham *et al*., 2009).

The above methods treat the distribution of a single type of spatial objects. In the real world, however, a wide variety of spatial objects exists and affect with each other. It is clearly indispensable to analyze the relations among more than a single distribution of spatial objects to understand the whole picture of the real world.

In point pattern analysis, several methods are available to examine the relationship between the distributions of two types of points. They include nearest neighbor contingency table (Pielou, 1961; Dixon, 1994), bivariate J-function (van Lieshout and Baddeley, 1999), cross K-function (Ripley, 1981), and nearest neighbor measures (Lee, 1979; Okabe and Miki, 1984).

Concerning the relationship between points and other types of spatial objects, a few methods have been proposed in the literature. Okabe and Fujii (1984) and Okabe *et al*. (1988) propose a statistical method for analyzing the relationship between points and a network and that between points and polygons, respectively. Sadahiro (1999) discusses the relationship between points and a surface.

Unfortunately, the type and the number of distributions are quite limited in exploratory spatial analysis. There are few methods that treat spatial objects other than points and lines. Given more than two distributions of spatial objects, we have to evaluate every pair of two distributions separately. Since each method is tailored for a specific type of spatial objects, numerous methods have to be developed to treat a wide variety of spatial objects.

To resolve the problem, this paper proposes a new exploratory method for analyzing the relations among distributions of spatial objects. It is a generalized method based on the those proposed in Sadahiro (2010, 2011, 2012), Sadahiro and Kobayashi (2012), and Sadahiro *et al*. (2012), each of which focuses on a specific type of spatial objects: 1) point distributions on a discrete space, 2) spatial tessellations, 3) a set of single polygons, 4) trajectories on a network, and 5) spatially distributed time series data. On the basis of these papers, this paper proposes a general method that is applicable independent of the type of spatial objects. Its applications include point distributions, polygons distributions, and trajectories on a continuous space.

Section 2 proposes a general method of analyzing the relations among the distributions of spatial objects. Section 3 describes the way of applying the method to several specific types of spatial objects. Section 4 applies the method to the analysis of the distributions of commercial facilities in Chiba City, Japan. Section 5 summarizes the conclusions with a discussion.

2. Method

This paper proposes a general method for analyzing the relations among the distributions of spatial objects. The method does not assume a specific type of spatial objects. However, the description below goes with a concrete illustration of the analysis of point distributions on a continuous space. This aims to avoid an abstract explanation of the method that is not easily accessible to readers. We choose point distributions on a continuous space because they are a more general case than those discussed in earlier papers. Point distributions are suitable for explaining a broad range of applications.

The method primarily consists of four steps:

1) Preprocessing
2) Clustering of distributions
3) Visualization of the relations among distributions
4) Quantitative evaluation

The following sections describes each step successively.

2.1 Preprocessing

Analysis starts with the definition of the minimal set of elements by which all the distributions are described. This paper calls the elements *parts*. Parts are the set of largest elements that do not need further division into smaller pieces to compose all the distributions. Any distribution can be represented as a set of parts.

There are at least two ways of defining parts. If spatial objects do not overlap with each other, parts are given by the union set of all the objects. In point distributions, for instance, each point in the distributions is a part. Every distribution can be represented as a set of points. If spatial objects overlap with each other, we make intersection of all the objects to obtain fragmented pieces each of which is defined as a part. Sadahiro *et al*. (2012) adopts this definition in the analysis of trajectories on a network where parts are defined as a set of links of the network.

This paper calls a distribution of spatial objects a *body*, because our method often treats the relations among distributions each of which consists of a single object (Sadahiro, 2012; Sadahiro and Kobayashi, 2012; Sadahiro *et al*., 2012). Let $\mathbf{B}=\{B_1, B_2, ..., B_M\}$ be the set of bodies. Body B_i is represented as a set of parts $\{P_{i1}, P_{i2}, ..., P_{imi}\}$. The number of elements and the ith element in set Q are denoted by $\#(Q)$ and $e(Q, i)$, respectively.

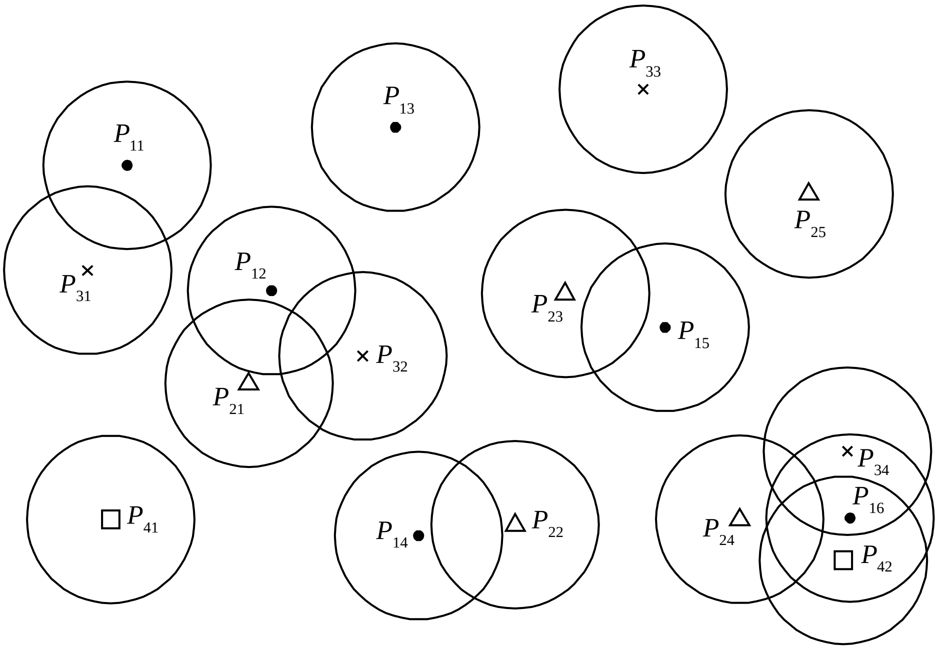
We then evaluate the spatial relations among parts contained in different bodies. To this end, we define *neighborhoods* of parts and examine the smaller regions obtained by their intersection. The definition of neighborhood depends on the type of spatial objects and the objective of analysis. A natural definition for point distributions is the buffer region of points as shown in Figure 1a.

Every region generated by the intersection of all the neighborhoods is assigned a spatial *tag* (Figure 1b). A set of tags is denoted by $\mathbf{T}=\{T_1, T_2, ..., T_K\}$. Tags provide a means of evaluating spatial relations among parts and bodies. If two parts share the same tag, the parts are considered to be spatially proximal. Two bodies are spatially similar if their parts share many tags. Every tag has its own *size* denoted by $s(T_i)$. When tags are defined on a two-dimensional space, their size is the area of their assigned regions. The size of a tag defined on a one-dimensional space is the length of its assigned line segment.

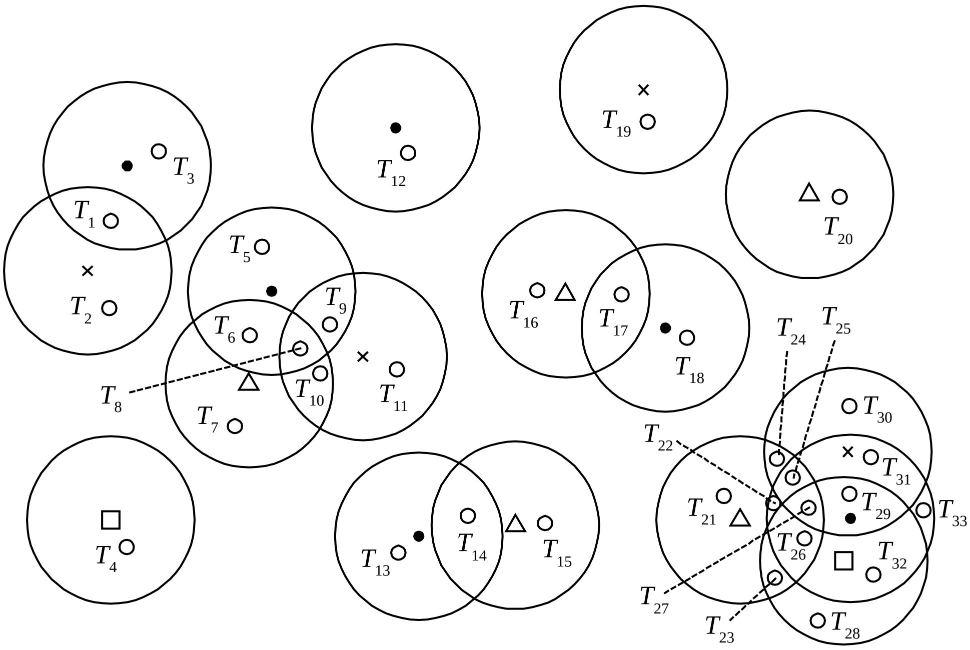
If parts are generated by intersection of overlapping objects, tags can be defined without considering the neighborhoods of parts. Every part has its own tag and only parts in different bodies that share the same tag are regarded as spatially proximal.

Tags relate parts in different bodies with each other. In Figure 1, for instance, tag T_8 relates parts P_{12}, P_{21}, and P_{32}, and T_6 relates P_{12} and P_{21}. Tags also relate different bodies through parts. Tag T_8 relates bodies B_1, B_2, and B_3 through parts P_{12}, P_{21}, and P_{32}. Tag T_6 relates bodies B_1 and B_2 through parts P_{12} and P_{21}.

Tags, parts, and bodies are related with each other. This paper calls this relationship *assignment*. In Figure 1, tag T_8 is assigned to points P_{12}, P_{21}, and P_{32}, bodies B_1, B_2, and B_3, while P_{12}, P_{21}, P_{32}, B_1, B_2, and B_3 are assigned to T_8. Every body and part can be represented by a set of assigned tags, while every tag can be represented as a set of assigned parts and bodies. Part P_{12} and body B_2 can be represented as $\{T_5, T_6, T_8, T_9\}$ and $\{P_{21}, P_{22}, P_{23}, P_{24}\}=\{\{T_5, T_6, T_8, T_9\}, \{T_{14}, T_{15}\}, \{T_{16}, T_{17}\}, \{T_{21}, T_{22}, T_{23}, T_{24}, T_{25}, T_{26}, T_{27}\}\}$, respectively. The set of bodies represented by the sets of tags is denoted by $\mathbf{B_T}$.



(a)



(b)

Figure 1. Point distributions, neighborhoods, and tags. Different symbols indicate different distributions. (a) Point distributions and their neighborhoods. (b) Tags assigned to individual regions.

2.2 Clustering of distributions

Having defined parts and tags, we cluster distributions of spatial objects in a specific relation. This section focuses on the spatial proximity among distributions because it has been drawn much attention of geographers as shown by numerous applications in point pattern analysis mentioned earlier. Clustering based on other relations will be briefly described later.

We cluster distributions whose elements are spatially proximal with each other at a local scale. Suppose, for instance, the distributions of commercial facilities. Supermarkets, convenience stores, fast-food restaurants, and drug stores display different distributions at a global scale. However, their distributions are often very similar with each other at a local scale. We often find shopping malls that contain all the four types of commercial facilities. Our method clusters the distributions of these facilities since they are similar at a local scale.

The above shopping malls are called *centers* denoted by $\mathbf{C}=\{C_1, C_2, ..., C_N\}$. They are formally defined as the sets of tags each of which are shared by many parts of different bodies. Each center consists of a set of tags assigned to the same set of parts. A body is said to be *assigned* to C_i if all the tags of C_i are assigned to parts in the body. Let Γ_i be a set of bodies assigned to center C_i. The set of body sets is denoted by $\mathbf{\Gamma}=\{\Gamma_1, \Gamma_2, ..., \Gamma_N\}$.

The following is a computational algorithm that detect centers and cluster distributions:

Algorithm CB (Center detection and Body clustering)

Input:

Set of bodies represented by both $\mathbf{B}$ and $\mathbf{B_T}$

Set of tags $\mathbf{T}$

Conditions $\{\vartheta_S, \vartheta_{A1}, \vartheta_{A2}\}$

Parameters $\{\alpha, \beta\}$

Output:

Set of tags representing centers $\mathbf{C}$ and their related tags $\mathbf{C'}$

Sets of bodies assigned to centers $\mathbf{\Gamma}$ and related bodies $\mathbf{\Gamma'}$

Algorithm:

1. $\mathbf{C}=\Gamma\in\mathbf{\Gamma}=C_i\in\mathbf{C}=\emptyset$. $j=0$.
2. Repeat the steps 3-26 while $\#(\Psi)\geq\alpha$ at step 6.
3. $\Theta=\Psi=\Psi'=\emptyset$. $k=1$.
4. Choose the set of tags in **T** satisfying ϑ_{A1}.
5. Add the tags to Θ.
6. Add the bodies in **B** assigned to the chosen tag to Ψ.
7. Repeat the steps 8-21 while $\#(\Psi)>1$.
8. If $s(\Theta)\geq\beta$ then
9. If $C_j\neq\emptyset$ then
10. Move the bodies from Ψ to Ψ' that do not contain all the elements of Θ.
11. If $C_j=\emptyset$ and $\#(\Psi)\geq\alpha$ then
12. $j=j+1$.
13. $C_j=\Theta$.
14. $\Gamma_j=\Psi$.
15. If $C_j=\emptyset$ and $\#(\Psi)<\alpha$ then
16. Move the bodies from Ψ to Ψ' that do not contain all the elements of Θ.
17. If $s(\Theta)<\beta$ then
18. Move the bodies from Ψ to Ψ' that do not contain all the elements of Θ.
19. Choose the set of tags in $\mathbf{T}\backslash\Theta$ satisfying ϑ_S and ϑ_{A2}.
20. Add the tags to Θ.
21. Move the bodies from Ψ to Ψ' that are not assigned to all the elements of Θ.
22. If $C_j\neq\emptyset$ then
23. $C_j'=\Theta$.
24. $\Gamma_j'=\Psi'$.
25. Remove the tags of parts of Γ_j containing C_j from $\mathbf{B_T}$.
26. If $C_j=\emptyset$ then $k=k+1$.
27. Return **C**, **C'**, $\mathbf{\Gamma}$, and $\mathbf{\Gamma'}$.

Algorithm CB detects centers represented by **C** and clusters bodies into similar groups based on **C** represented by $\mathbf{\Gamma}$. Since it is designed as a general algorithm applicable independent of the type of spatial objects, conditions ϑ_S, ϑ_{A1}, and ϑ_{A2} and parameters α and β are not specified. They have to be defined in each application as well as tags, parts, and bodies.

Condition ϑ_S is the spatial requirement on tags in **C** while conditions ϑ_{A1} and ϑ_{A2} indicate requirements on their attributes. The former depends on whether or not each center has to be in a specific spatial form. Analysis of point distributions does not usually impose this condition. When centers have to be spatially clustered, we define ϑ_S as a tag adjacent to those in Θ.

Condition ϑ_{A1} is given to choose a tag that best represents many bodies. We usually define ϑ_{A1} as the largest tag among those assigned to the kth most bodies in $\mathbf{B_T}$. Modifying slightly condition ϑ_{A1}, we obtain Condition ϑ_{A2}: the largest set of tags in $\mathbf{B_T}$ that are assigned to the most bodies in Ψ.

Parameters α and β indicate the requirements for centers. The former is the minimum number of bodies assigned to a center while the latter is the minimum size of a center. Large α and β yield large centers assigned to many bodies. Though such centers are useful in analysis, large parameter values are restrictive conditions so that Algorithm CB may detect only a few centers. In practice, we should start with small values, say, $\alpha=0.001\times M$ and $\beta=0.0001\times s(\mathbf{T})$, and gradually increase them until a reasonable number of centers are obtained.

Figure 2 shows the process of Algorithm CB applied to the point distributions shown in Figure 1. Let us first suppose the case where $\alpha=3$ and $\beta=\beta_1$, the latter of which is indicated by the area of the dotted circle in Figure 2a. Algorithm CB chooses tag T_{27} at step 4 because it is the only tag assigned to four bodies. The sets Θ and Ψ become $\{T_{27}\}$ and $\{B_1, B_2, B_3, B_4\}$, respectively, the former of which is indicated by the bold line in Figure 2a. Since $s(T_{27})<\beta_1$, Algorithm CB proceeds to step 17 and then 19 to choose $\{T_8, T_{25}\}$ because $s(T_8)+s(T_{25})$ is largest among the sets of tags assigned to three bodies in Ψ. The sets Θ and Ψ then become $\{\{T_{27}\}, \{T_8, T_{25}\}\}$ and $\{B_1, B_2, B_3\}$, respectively. The set $\{T_8, T_{25}\}$ is indicated by thin lines in Figure 2a. Algorithm CB returns to step 8. Since $s(T_{27})+s(T_8)+s(T_{25})>\beta$, the tag set $\Theta=\{\{T_{27}\}, \{T_8, T_{25}\}\}$ is substituted to center C_1 at step 13. This center is indicated by dark gray shades in Figure 2b.

We then turn to the case where $\alpha=3$ and $\beta=\beta_2$. Since $s(T_{27})+s(T_8)+s(T_{25})<\beta$, Algorithm CB proceeds to step 17. It adds tags $\{T_6, T_{14}, T_{17}, T_{24}\}$ to Θ. Though $s(\Theta)>\beta_2$, set Θ is not a center because $\#(\Psi)<\alpha$. Algorithm CB does not detect any center because the requirements for centers are too restrictive.

If $\alpha=2$, Algorithm CB detects the center represented by $\{\{T_{27}\}, \{T_8, T_{25}\}, \{T_6, T_{14}, T_{17}, T_{22}\}\}$ indicated by light gray shades in Figure 2b. Two bodies $\{B_1, B_2\}$ are assigned to the center.

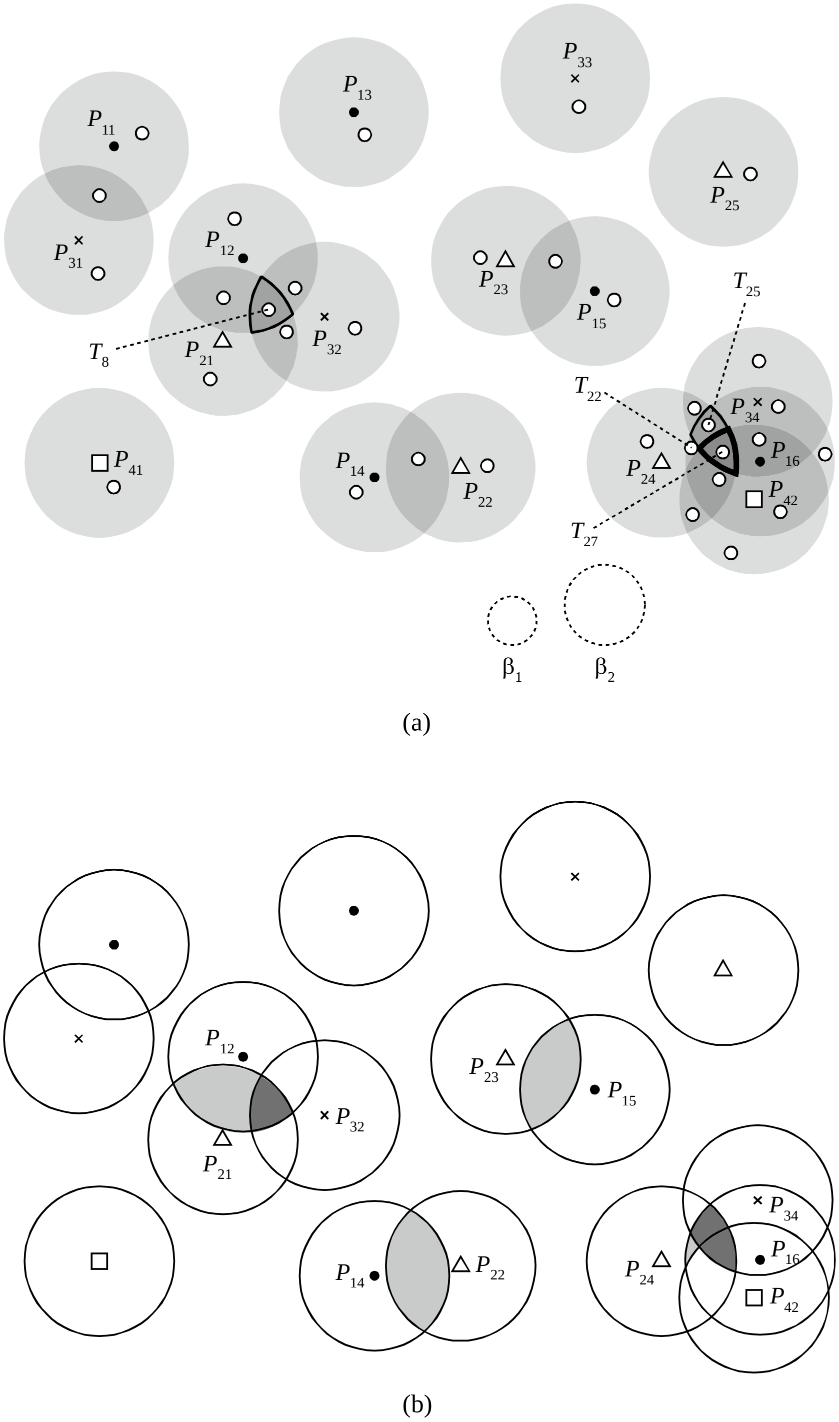


Figure 2. Center detection and body clustering in point distributions. Parameter β is indicated as the area of the dotted circles. (a) Tags chosen by Algorithm CB when α=3. Bold and thin lines indicate the first and second set of tags, respectively. (b) Centers detected by Algorithm CB. Dark and light gray shades indicate the centers when α=2 and α=3, respectively.

Algorithm CB is a generalized algorithm of Algorithm TC proposed by Sadahiro and Kobayashi (2012) that was originally developed by Kharrat *et al*. (2008). Though existing algorithms focus on a specific type of spatial objects under limited circumstances, Algorithm CB does not depend on the type of spatial objects.

Though only a single center is detected in Figure 2, many centers are usually detected when a number of distributions are analyzed. Each body can be assigned to more than one center, and centers sharing the same tags spatially overlap with each other.

Algorithm CB generates two ordered sets of tags and bodies **C'** and **Γ'** for clustering bodies. In Figure 2, $\mathbf{C'}=\{\{T_{27}\}, \{T_8, T_{25}\}, \{T_6, T_{14}, T_{17}, T_{22}\}, \{T_1, T_3, T_5, T_9, T_{12}, T_{13}, T_{18}, T_{26}, T_{29}, T_{31}, T_{32}, T_{33}\}\}$ and $\mathbf{\Gamma'}=\{B_4, B_3, B_2, B_1\}$. The tag sets are arranged in the order of addition while the bodies are arranged in the order of removal. Since the orders reflect that of similarity of elements, **C'** and **Γ'** provide a means of classify tags and bodies. In $\mathbf{\Gamma'}=\{B_4, B_3, B_2, B_1\}$, for instance, bodies $\{B_3, B_2, B_1\}$ are more similar with each other than B_4. Consequently, we may classify the bodies as $\{\{B_4\}, \{B_3, B_2, B_1\}\}$, $\{\{B_4\}, \{B_3\}, \{B_2, B_1\}\}$, or $\{\{B_4\}, \{B_3\}, \{B_2\}, \{B_1\}\}$. Classifications such as $\{\{B_4, B_2\}, \{B_3, B_1\}\}$ and $\{\{B_4, B_1\}, \{B_3, B_2\}\}$ are not permissible because they are inconsistent with the order of bodies in **Γ'**.

This classification scheme permits each body or tag contained in more than one group when multiple centers are detected. To avoid this, we simply remove all the tags of Γ_j in $\mathbf{B_T}$ at step 25. This enables the classification of all the bodies and tags without overlap.

Algorithm CB adds tags to Θ while removes bodies from Ψ under given conditions. Tags and bodies are interchangeable because they are both represented as the sets of the others. Exchanging tags and bodies, we obtain a dual algorithm of Algorithm CB:

Algorithm CT (Center detection and Tag clustering)

Input:

Set of bodies represented by both **B** and $\mathbf{B_T}$

Set of tags **T**

Conditions $\{\vartheta_{A1}, \vartheta_{A2}\}$

Parameters $\{\alpha, \beta\}$

Output:

Set of tags representing centers **C** and their related tags **C’**

Sets of bodies assigned to centers **Γ** and related bodies **Γ’**

Algorithm:

1. $\mathbf{C}=\Gamma\in\mathbf{\Gamma}=C_i\in\mathbf{C}=\emptyset$. $j=0$.
2. Repeat the steps 3-26 while $\#(\Psi)\geq\alpha$ at step 6.
3. $\Theta=\Theta'=\Psi=\emptyset$. $k=1$.
4. Choose the set of bodies in **B** satisfying ϑ_{A1}.
5. Add the bodies to Ψ.
6. Add the tags in **T** assigned to the chosen body to Θ.
7. Repeat the steps 8-21 while $\#(\Theta)>1$.
8. If $s(\Psi)\geq\beta$ then
9. If $C_j\neq\emptyset$ then
10. Move the tags from Θ to Θ' that do not contain all the elements of Ψ.
11. If $C_j=\emptyset$ and $\#(\Theta)\geq\alpha$ then
12. $j=j+1$.
13. $C_j=\Theta$.
14. $\Gamma_j=\Psi$.
15. If $C_j=\emptyset$ and $\#(\Theta)<\alpha$ then
16. Move the tags from Θ to Θ' that do not contain all the elements of Ψ.
17. If $s(\Psi)<\beta$ then
18. Move the tags from Θ to Θ' that do not contain all the elements of Ψ.
19. Choose the set of bodies in $\mathbf{B}\backslash\Psi$ satisfying ϑ_{A2}.
20. Add the body to Ψ.
21. Move the tags from Θ to Θ' that are not assigned to all the elements of Ψ.
22. If $C_j\neq\emptyset$ then
23. $C_j'=\Theta$.
24. $\Gamma_j'=\Psi'$.
25. Remove the tags of parts of Γ_j containing C_j from $\mathbf{B_T}$.
26. If $C_j=\emptyset$ then $k=k+1$.
27. Return **C**, **C’**, **Γ**, and **Γ’**.

Conditions CT

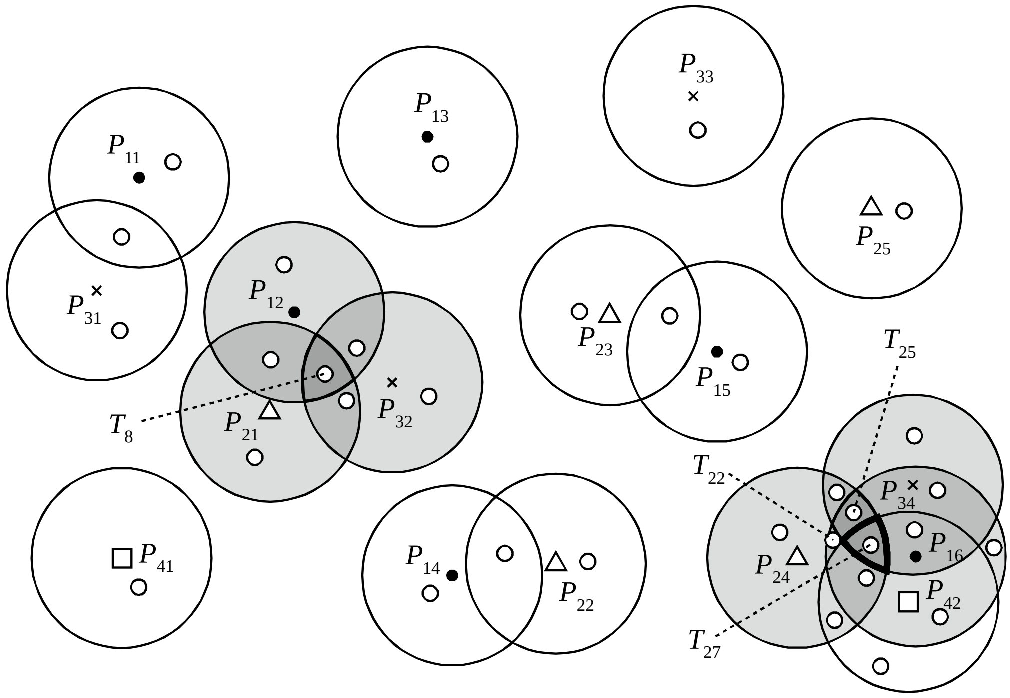
ϑ_{A1}: Largest among those assigned to the kth most tags in $\mathbf{B_T}$

ϑ_{A2}: Largest among those assigned to the most tags in $\mathbf{B_T}$

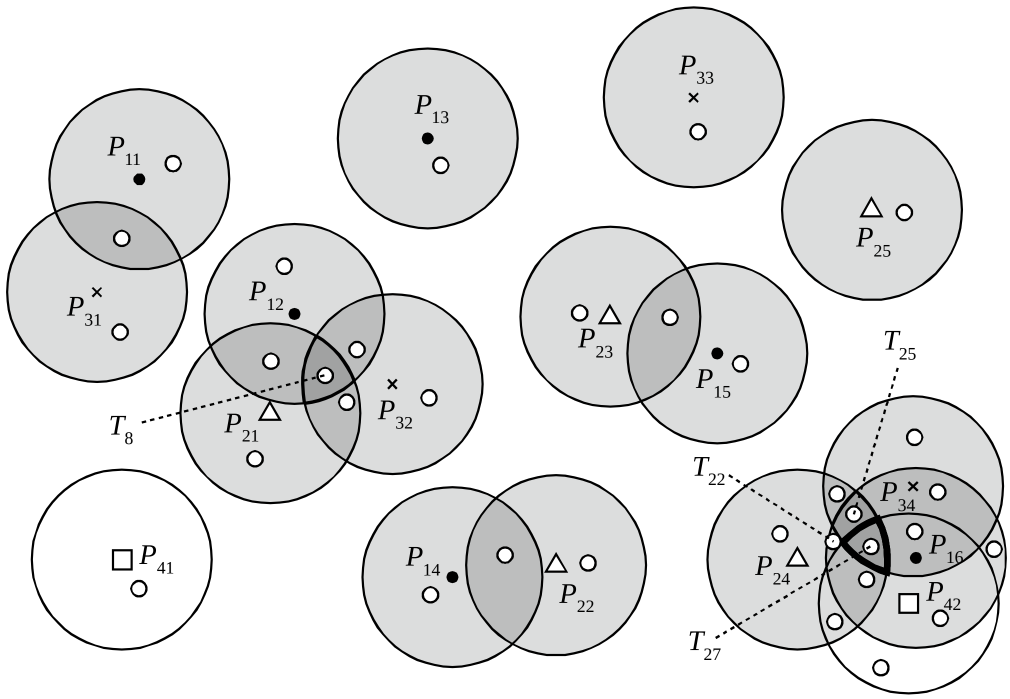
The size of a body is the sum of the size of tags assigned to the body. Algorithm CT is a kind of reverse process of Algorithm CB, that is, it adds bodies to Ψ while removes tags from Θ. Though both algorithms cluster bodies and tags, they usually generate different results because Algorithm CB focuses on the coverage of bodies represented by $\#(\Psi)$ while Algorithm CT focuses on that of tags represented by $\#(\Theta)$. Another difference is that Algorithm CB can consider spatial condition at step 19. Algorithm CT does not impose spatial conditions on centers.

2.3 Visualization of the relations among distributions

We then visualize the result of the clustering of distributions. One method is a map representation such as shown in Figure 3. Figure 3a and Figure 3b show the parts and bodies assigned to the center, respectively. Comparing the two figures, we notice that some parts such as P_{11}, P_{23}, and P_{31} are assigned to the center indirectly through the tags of the same bodies assigned to the center, say, P_{12}, P_{21}, and P_{32}.



(a)



(b)

Figure 3. Tags and bodies assigned to the center detected by Algorithm CB when α=2 and β=β_1. (a) Tags, (b) bodies.

Figure 3, unfortunately, does not clearly present the relations among tags, bodies, and centers. To complement this representation, this paper proposes a graph-based representation called *topology diagram*. It has been originally developed in Sadahiro (2010, 2011, 2012), Sadahiro and Kobayashi (2012), and Sadahiro *et al*. (2012). The power set of tags $\mathbf{T}$ and Boolean operations $\{\cap, \cup\}$ form a lattice (Anderson, 2002; Pemmaraju and Skiena, 2003), where the least and greatest elements are $\varnothing$ and the union set of all the tags. A lattice is a poset (partially ordered set), and consequently, visualized by Hasse diagram (Birkhoff, 1979; Davey and Priestley, 2002). Topology diagram is a modified subset of Hasse diagram that represents the topological relations among spatial objects. Nodes indicate tags, bodies, and centers, while edges indicate their topological relations. Parts are represented implicitly by their composing tags. The vertical axis indicates the size of spatial objects.

Topology diagram is not uniquely defined. One method to create a topology diagram is to trace the entire process of detecting a single center in Algorithm CB. Addition of tags to Θ and removal of bodies from Ψ can be represented as a graph whose nodes indicate a center and its related bodies and tags.

Figure 4 shows the topology diagram representing the process of detecting the center in Figure 1 when α=2. Bold lines represent the growth of Θ. Algorithm CB adds tag sets $\{T_8, T_{25}\}$, $\{T_6, T_{14}, T_{17}, T_{22}\}$, $\{T_1, T_3, T_5, T_9, T_{12}, T_{13}, T_{18}, T_{26}, T_{29}, T_{31}, T_{32}, T_{33}\}$ to Θ while removes bodies B_4, B_3, and B_2 from Ψ. The former process is represented as a tree indicated by bold and thin solid lines while the latter is a tree consisting of bold solid and dotted lines in Figure 4.

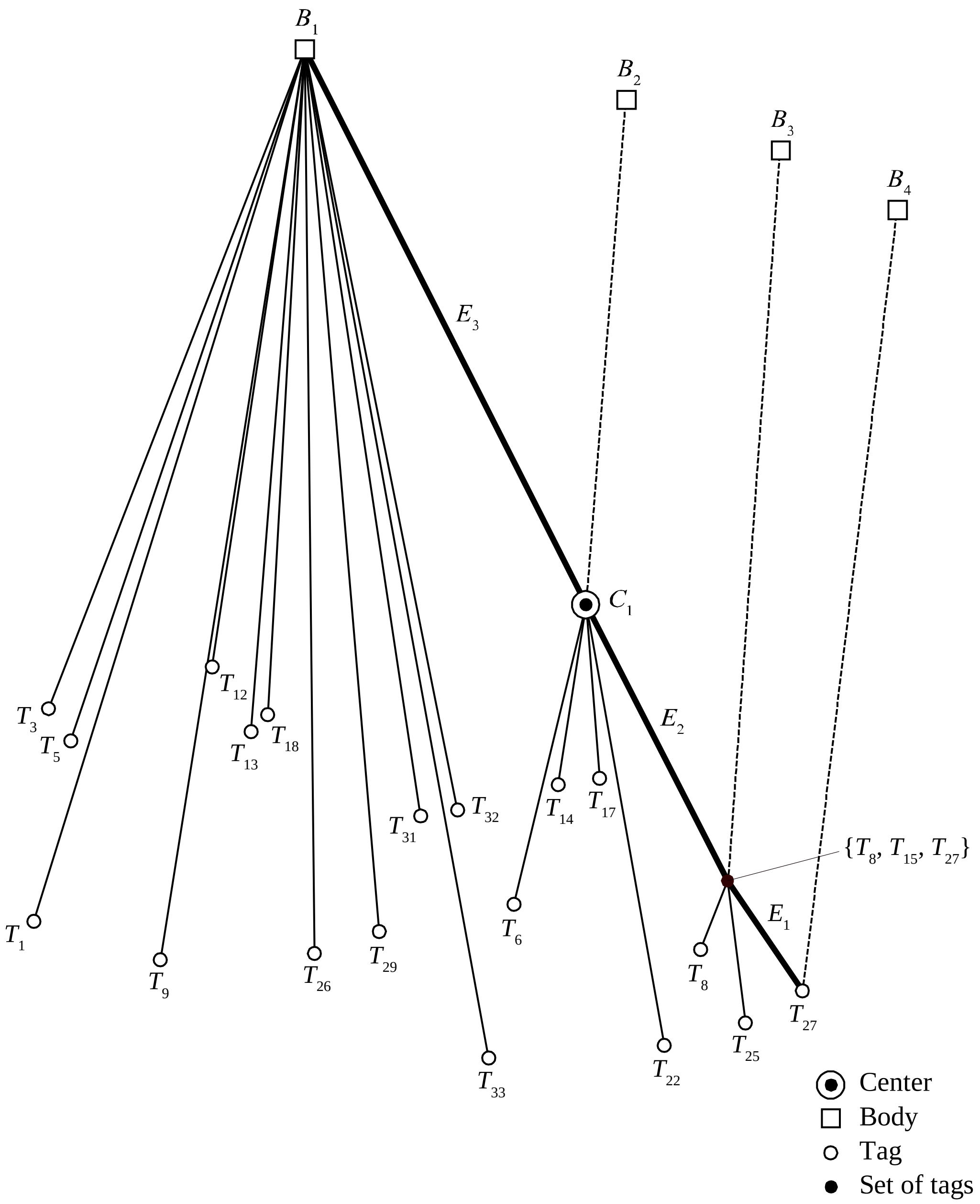


Figure 4. Topology diagram indicating the process of detecting a center in Algorithm CB. Bold and thin solid lines indicate the addition of tags to Θ. Bold solid lines and dotted lines indicate the removal of bodies from Ψ. Edges E_1-E_3 represent the growth of Θ.

Topology diagram permits us to grasp visually the entire process of Algorithm CB. In addition, it is convenient for classifying bodies and tags without inconsistency mentioned earlier. Cutting some edges connecting intermediate sets of tags in Θ, we obtain groups of bodies and tags. In Figure 4, for instance, we cut E_2 to obtain two sets of bodies {B_1, B_2}, {B_3, B_4} and two sets of tags {T_{27}, {T_8, T_{25}}}, {{T_6, T_{14}, T_{17}, T_{22}}, {T_1, T_3, T_5, T_9, T_{12}, T_{13}, T_{18}, T_{26}, T_{29}, T_{31}, T_{32}, T_{33}}}. Cutting E_2 and E_3, we obtain three sets of bodies. Topology diagram inherently prohibits classifications inconsistent with the order of bodies in **Γ'** such as {{B_4, B_2}, {B_3, B_1}} and {{B_4, B_1}, {B_3, B_2}}.

Topology diagram can also visualize the relations among multiple centers. To generate a topology diagram for multiple centers, we employ Algorithm CB by regarding centers as bodies. Tracing the entire process of Algorithm CB, we obtain a topology diagram that indicates the topological relations among centers.

2.4 Quantitative evaluation

Algorithm CB yields the sets of centers and bodies. The result is informative and useful when many centers are detected and many bodies are clustered into groups. To evaluate the result, this section proposes quantitative measures.

A basic measure is N, the number of centers detected by Algorithm CB. In addition, the ratios of the bodies and tags assigned to centers are also useful:

$$R_{NB} = \frac{\#\left(\bigcap_{\Gamma_i \in \mathbf{\Gamma}} \Gamma_i\right)}{M} \tag{1}$$

and

$$R_{NT} = \frac{\#\left(\bigcap_{C_i \in \mathbf{C}} C_i\right)}{K} . \tag{2}$$

The size of centers is also a critical measure of evaluating the effectiveness of analysis. Its standardized form is defined by

$$R_S = \frac{\sum_i \#(\Gamma_i)\, s(C_i)}{s(\mathbf{B_T})} . \tag{3}$$

The above measures reflect the similarity among the distributions of spatial objects. A high similarity permits Algorithm CB to detect large and many centers, and consequently, yields large values. Measures become small when a wide variety exists among the distribution of spatial objects.

Measures can also be defined for individual centers. A center represents its assigned bodies and it is more representative and informative if it is large and assigned many bodies. The representativeness of a center can be evaluated by its size and the ratios of assigned bodies and tags:

$$r_{SN}\left(C_i\right)=\frac{\#\left(\Gamma_i\right)s\left(C_i\right)}{\sum_{B_j\in\Gamma_i}s\left(B_j\right)}, \tag{4}$$

$$r_{NB}\left(C_i\right)=\frac{\#\left(\Gamma_i\right)}{M}, \tag{5}$$

and

$$r_{NT}\left(C_i\right)=\frac{\#\left(C_i\right)}{K}. \tag{6}$$

These measures are all ratio variables ranging from 0 to 1. Large values indicate the high representativeness of C_i.

2.5 Discretization

Intersection of spatial objects in preprocessing often generates numerous parts and tags (Sadahiro, 1999). Since the computing time of Algorithm CB heavily depends on the number of parts and tags, calculation may not terminate within a reasonable time.

A practical solution to this problem is to discretize the space on which spatial objects are distributed into small units. When the space is two-dimensional, a square lattice is a reasonable solution where cells serve as basic units in discretization. Parts, neighborhoods, and bodies are approximated by sets of cells and tags are assigned to the cells. If the space is one-dimensional such as a network space, we can use edges as basic units in discretization (Sadahiro *et al.*, 2012).

Computational complexity of Algorithm CB after discretization is O(Mm), where M and m are the numbers of bodies and basic units in discretization, respectively. Since it is a linear function of M and m, computation time is practically acceptable.

A fine discretization sounds desirable because it provides a good approximation of spatial objects. However, a high resolution is not always effective because it often generates many small centers that are not significant in analysis. Considering the initial value of β mentioned earlier, this paper recommends to set m from ten thousands to a million. Sadahiro and Kobayashi (2012) reports that lattices of resolution 200×200 to 1000×1000 yielded almost similar results.

2.6 Clustering of distributions based on relations other than spatial proximity

As mentioned in Section 2.2, distributions can be clustered based on not only spatial proximity but also other relations. This section briefly illustrates the clustering of distributions based on other relations, that is, complete, exclusive, and complementary relations discussed in Sadahiro (2010).

A set of bodies are *complete* if every tag is assigned to at least one body in the set. Bodies are *exclusive* if no tag is assigned to more than one body in the set. Complete and exclusive set of bodies are *complementary*. A set of bodies in a specific relation are denoted by G_i, and their set is denoted by $\mathbf{\Lambda}=\{G_{1,} G_{2, \ldots,} G_N\}$.

The three relations can be handled with in the same algorithm only by using different conditions in choosing bodies.

Algorithm BC (Body Clustering)

Input:

Set of bodies **B**

Set of tags **T**

Conditions $\{\vartheta_R, \vartheta_S, \vartheta_{A1}, \vartheta_{A2}\}$

Parameters $\{K, \gamma, \mu\}$

Output:

Sets of bodies $\mathbf{\Lambda}$

Algorithm:

1. Λ=G$_i\in\Lambda=\varnothing$. i=0. k=1
2. Repeat the steps 3-12 while $k \leq K$.
3. $\Theta=\Psi=\varnothing$.
4. Add B_k to Ψ.
5. Add the tags composing B_k to Θ.
6. Repeat the steps 7-9 until ϑ_R is satisfied.

7. Choose the body in **B** satisfying ϑ_S and ϑ_{A1}.
8. Add the body to Ψ.
9. Add the tags composing the body to Θ.
10. If ϑ_{A2} is satisfied then
11. i=i+1.
12. G_i=Ψ.
12. k=k+1.
13. Return $\boldsymbol{\Lambda}$.

Conditions $\mathbf{BC_C}$

ϑ_R: #(Ψ)=M

ϑ_{A1}: Largest among those assigned to the most tags in $\mathbf{T}\backslash\Theta$

ϑ_{A2}: #(Ψ)<γ

Conditions $\mathbf{BC_E}$

ϑ_R: No body can be further added to Ψ without overlap of tags in Θ

ϑ_{A1}: Largest among those assigned to the most tags in $\mathbf{T}\backslash\Theta$ without overlap of tags in Θ

ϑ_{A2}: #(Ψ)<γ and μ<$s(\Theta)$

Conditions BC_C and BC_E are used for detecting sets of complete and exclusive bodies, respectively. We can obtain complementary bodies by applying both conditions simultaneously. Spatial condition ϑ_S depends on the type of spatial objects. Sadahiro (2010) uses the Delaunay triangulation generated from all the points and chooses points adjacent on the network.

Parameters K, γ, and μ give the number of sets of bodies, the maximum number of bodies in a set, and the minimum size of a set, respectively. Parameter μ is necessary only for detecting sets of exclusive bodies. Exchanging bodies and tags in Algorithm BC, we can also detect complete, exclusive, and complementary tags (Sadahiro, 2010).

2.7 Formal concept analysis

The power set of tags **T** and Boolean operations {$\cap$, $\cup$} form a lattice as mentioned earlier. This enables us to analyze the relations among tags, bodies, and centers by using formal concept analysis (Ganter et al., 2005; Kwuida and Sertkaya, 2010). Once we perform preprocessing, we can employ analytical methods developed in formal concept analysis with extensions and modifications.

For instance, computational algorithms are available to generate a concept lattice that visualizes the relations among objects and attributes. They are useful for efficiently visualizing the relations among tags, centers, and bodies. Methods for classifying objects and attributes have also been developed in the literature (Arévalo *et al*., 2009; Janowitz, 2010). We can extend them to classify bodies and tags.

3. Distributions of spatial objects other than point distributions

The method proposed in the previous section does not depend on the type of spatial objects. Its applications require the specification of parts, neighborhoods, and tags, and the definition of conditions and parameters in Algorithm BC. This section illustrates several applications discussed in earlier papers (Sadahiro, 2010, 2011, 2012, Sadahiro and Kobayashi, 2012, and Sadahiro *et al*., 2012).

3.1 A set of single polygons

Sadahiro (2012) analyzes a set of single polygons where each body is defined as a single polygon. If polygons overlap with each other, we make intersection of all the polygons to obtain smaller fragments as parts. If some polygons do not overlap, we may consider buffer regions of polygons as neighborhoods and make their intersections to define parts and tags.

Definition of ϑ_S depends on whether or not each center needs to be spatially connected. To obtain connected centers, we define ϑ_S as a tag adjacent to those in Θ. If we permit disconnected centers, this condition is not necessary.

3.2 Single paths on a network

Sadahiro *et al*. (2012) analyzes paths on a network space on which each body is defined as a single path. When paths overlap with each other, we make intersection of all the paths to obtain parts defined by sets of edges. If some paths do not overlap, we may consider their neighborhoods on the network to define parts and tags. Spatial condition ϑ_S can be represented as spatial adjacency on the network.

3.3 Time series data

Sadahiro and Kobayashi (2012) applies the method to the analysis of spatially distributed time series data. The paper treats graphs representing the data as curved lines on a two-dimensional space. The treatment of two dimensions, however, is different because the horizontal axis represents time while the vertical axis indicates attribute value.

Neighborhoods are buffer regions of the graphs extended only in the vertical direction. This definition is different from that used in GIS since the latter is isotropic with respect to generators. A body is the neighborhood of a time series graph. Overlaying neighborhoods, we obtain small polygons in each of which a tag is defined. A body is split into parts by vertical lines at the earliest and latest times of every polygon composing the body. The size of a tag is its length measured in the horizontal direction.

Condition ϑ_S is given by the spatial adjacency of polygons assigned to tags. Condition ϑ_{A1} is defined similarly as that for polygons and paths. Condition ϑ_{A2}, on the other hand, is defined as the largest tag outside the time period of Θ among those assigned to the kth most bodies in $\mathbf{B_T}$. This aims to detect centers long enough in the horizontal rather than vertical direction.

3.4 Point distributions on a discrete space

Section 2 describes a method of analyzing point distributions on a continuous space. Since discrete space is a special case of continuous ones, the method is applicable as it is to the analysis of point distributions on a discrete space.

However, when points can be distributed only on a limited number of locations, points may overlap with each other. In such a case, we can define parts as individual locations without considering neighborhoods (Sadahiro, 2010). In this case, condition ϑ_S is not naturally given by spatial proximity. We have to define explicitly a spatial structure among locations such as by using a Delaunay triangulation.

3.5 Polygon distributions

When polygons overlap with each other, the method discussed in Section 3.1 is directly applicable. If polygons do not overlap, we define neighborhoods and make their intersection to obtain parts. Though condition ϑ_S is not imposed in usual, we can define it as a tag adjacent to those in Θ if each center has to be spatially connected.

3.6 Spatial tessellations

Since spatial tessellations are a special case of polygon distributions, the method proposed in the previous section is applicable as it is. However, as seen in Sadahiro (2011), consideration of the special properties of spatial tessellations permits us to find a wider variety of relations among tessellations, especially when they are defined in the same region.

A difference in Sadahiro (2011) lies in the definition of the relations among tessellations. Given two tessellations, Sadahiro (2011) considers the topological relation between every pair of regions each of which is in a different tessellation. Two tessellations are hierarchical if every region of one tessellation is fully contained in a single region of the other. We can adopt this definition in Algorithm CT with a slight modification.

The definition of topology diagram is also different. In Sadahiro (2011), the vertical axis of the diagram indicates the granularity of tessellations. The granularity of a tessellation Ω is defined by

$$g(\Omega) = \frac{\sum_i \{a(\omega_i)\}^2}{\left\{\sum_i a(\omega_i)\right\}^2}, \tag{7}$$

where ω_i is the ith region of Ω and $a(\omega_i)$ is its area. Nodes indicate spatial tessellations. Two nodes are connected by an edge if the tessellations are hierarchical. This representation is useful if a focus is on the hierarchical relation among spatial tessellations.

4. Empirical application

To validate the method proposed in the previous section, this section applies it to the analysis of the distributions of commercial facilities in Chiba City, Japan. Chiba is located 30 kilometers away from the east of Tokyo. We generated spatial data of commercial facilities by geocoding of their addresses in NTT telephone directory. Chiba has 16,311 commercial facilities of 235 categories.

Figure 5 shows the distribution of commercial facilities in Chiba. Downtown of Chiba is located in the southeast of Chiba station indicated by the largest cluster of darker cells. The northern and western parts of Chiba serve as residential areas for people working in Tokyo. The population density is higher in these areas. The southern and eastern parts are residential areas for people working in the downtown of Chiba. Commercial facilities are concentrated mainly around railway stations. Chiba station has the largest cluster and stations in subcenters such as Inage and Tsuga also draw many facilities.

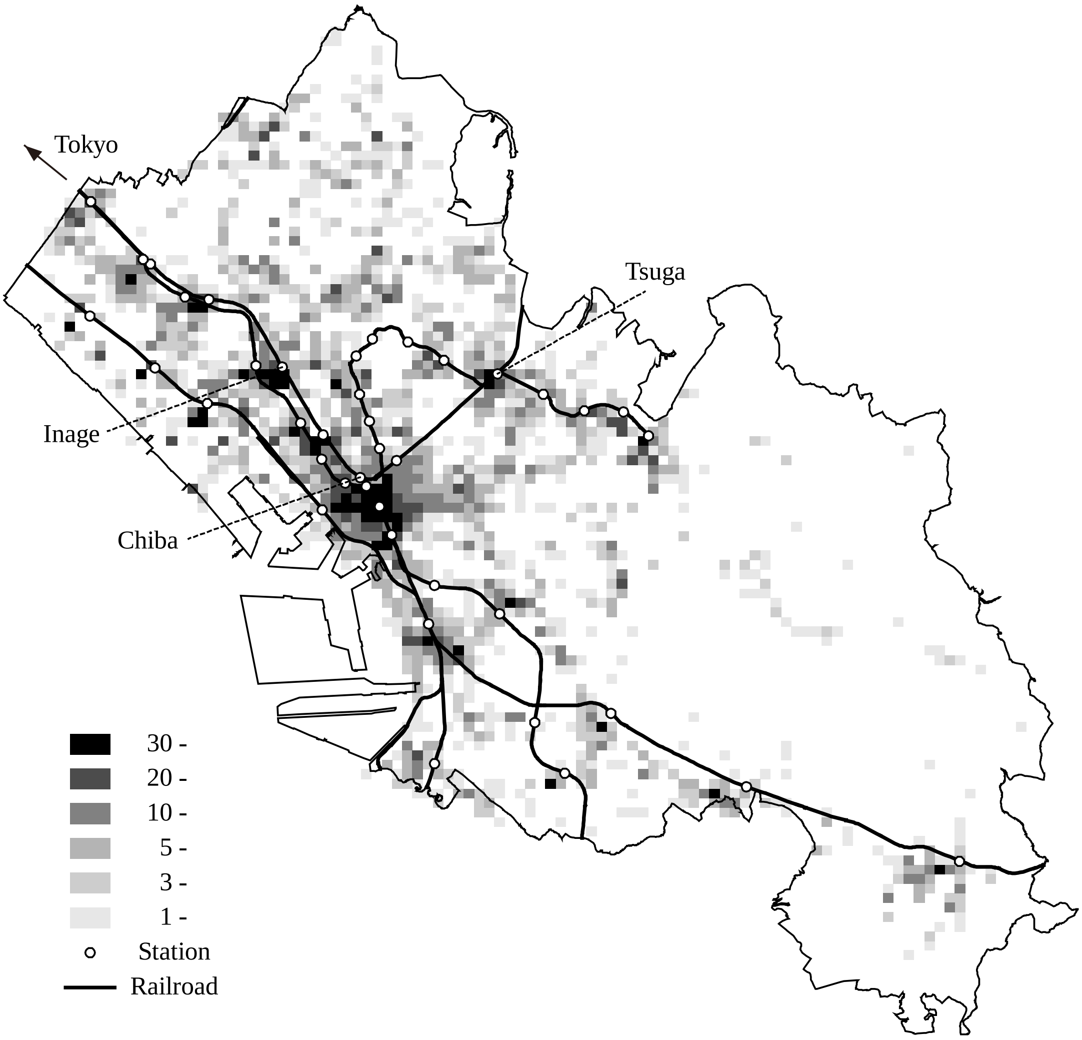


Figure 5. Distribution of commercial facilities in Chiba City, Japan.

Concerning parameter values, we have tried various values to evaluate the relationship between parameters and result. Parameters α and β range from 1 to 10 and 10 to 500, respectively. We finally set the former to 5 because this yields a reasonable number of centers. Neighborhood of tags is defined as the circle of radius r ranging from 100 to 1000 meters.

Discretization used lattices of a resolution ranging from 100×100 to 1000×1000. Since the results are almost consistent, we only discuss the result of 100×100 lattice in the following.

Figure 6 shows the relationship between parameter values and the result of analysis. In general, Algorithm CB detects more and larger centers with an increase in r and a decrease in β. Only the number of centers N decreases for large r as shown in

Figure 6a because an increase in the size of centers is often accompanied by a decrease in the number of centers.

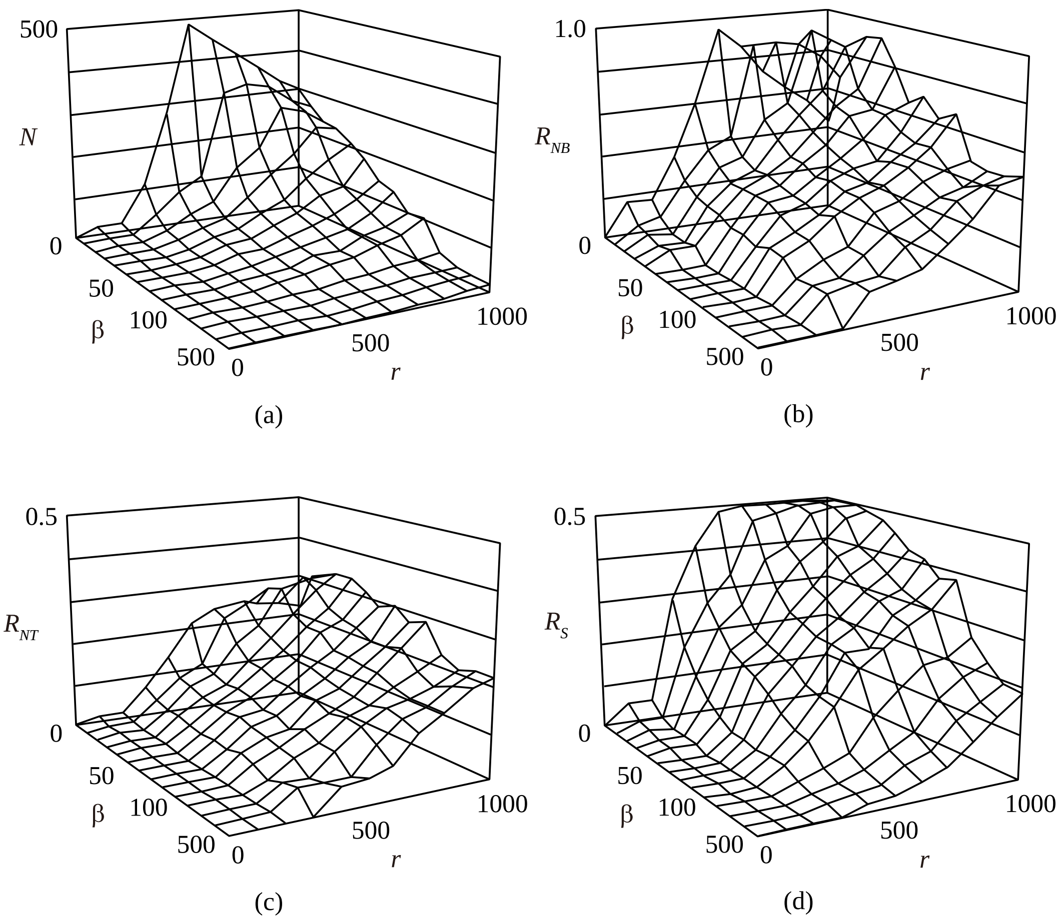


Figure 6. The relationship between parameter values and the result of analysis. (a) The number of centers, (b) the ratio of bodies assigned to centers, (c) the ratio of tags assigned to centers, (d) the standardized size of centers.

We then move to the details of the result. Table 1 shows the commercial facilities assigned to centers detected when r=300 and β=50. Nine centers C_1-C_9 are detected whose size is all close to 50 given by β.

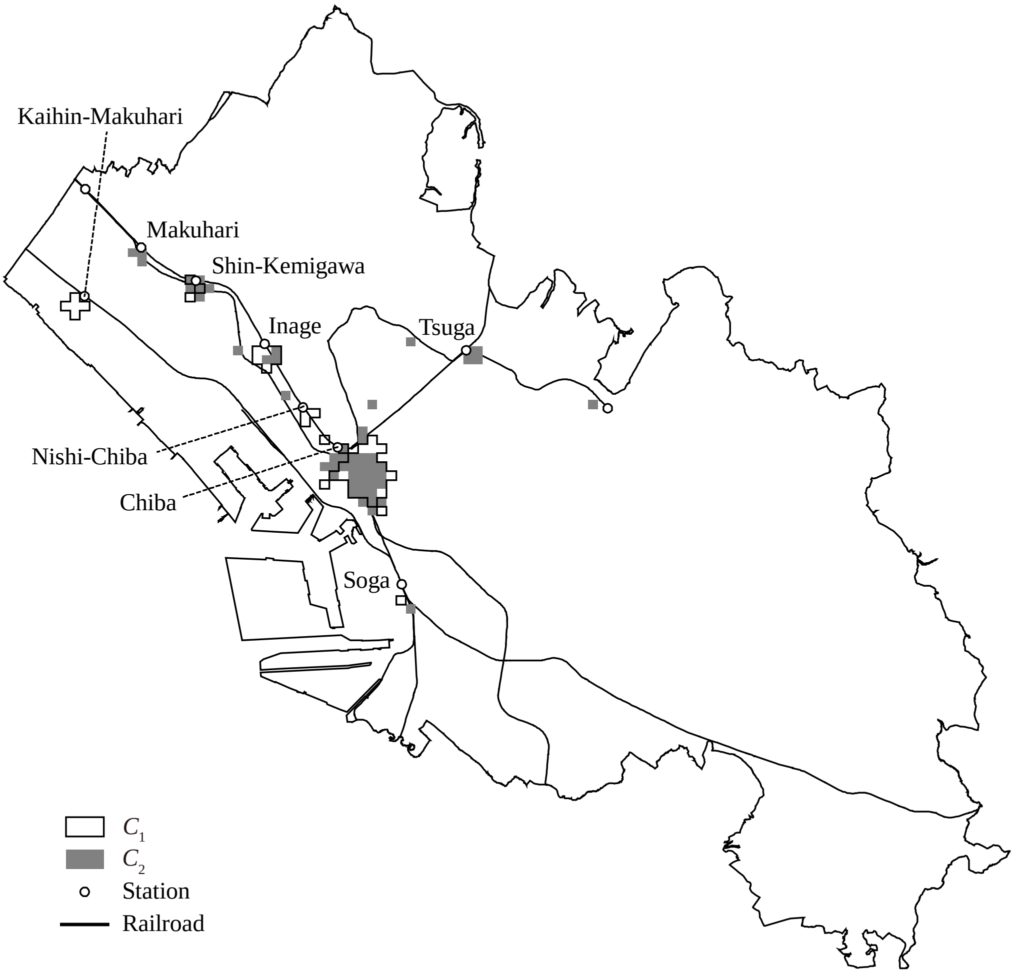
A wide variation exists in the location of centers as seen in Figure 7. Tags assigned to C_1 and C_2 are found only around railway stations such as Chiba, Nishi-Chiba, Inage, and Kaihin-Makuhari. These centers represent large shopping malls and districts. Center C_1 is characterized by bodies B_1-B_4: Japanese fast-food restaurants, banks, cosmetic stores, and coffee shops. This result is reasonable because these facilities are found mainly around railway stations in suburban areas of Japan. Center C_2

is characterized by kimono and flower shops, whose distribution is quite similar to that of C_1. Since kimono shops are usually located in traditional shopping malls and department stores, center C_2 is considered to represent the traditional shopping districts around Chiba, Inage, and Tsuga. Kaihin-Makuhari has only C_1 tags because the station opened in 1986 and the shopping district around the station was developed in late 1980s.

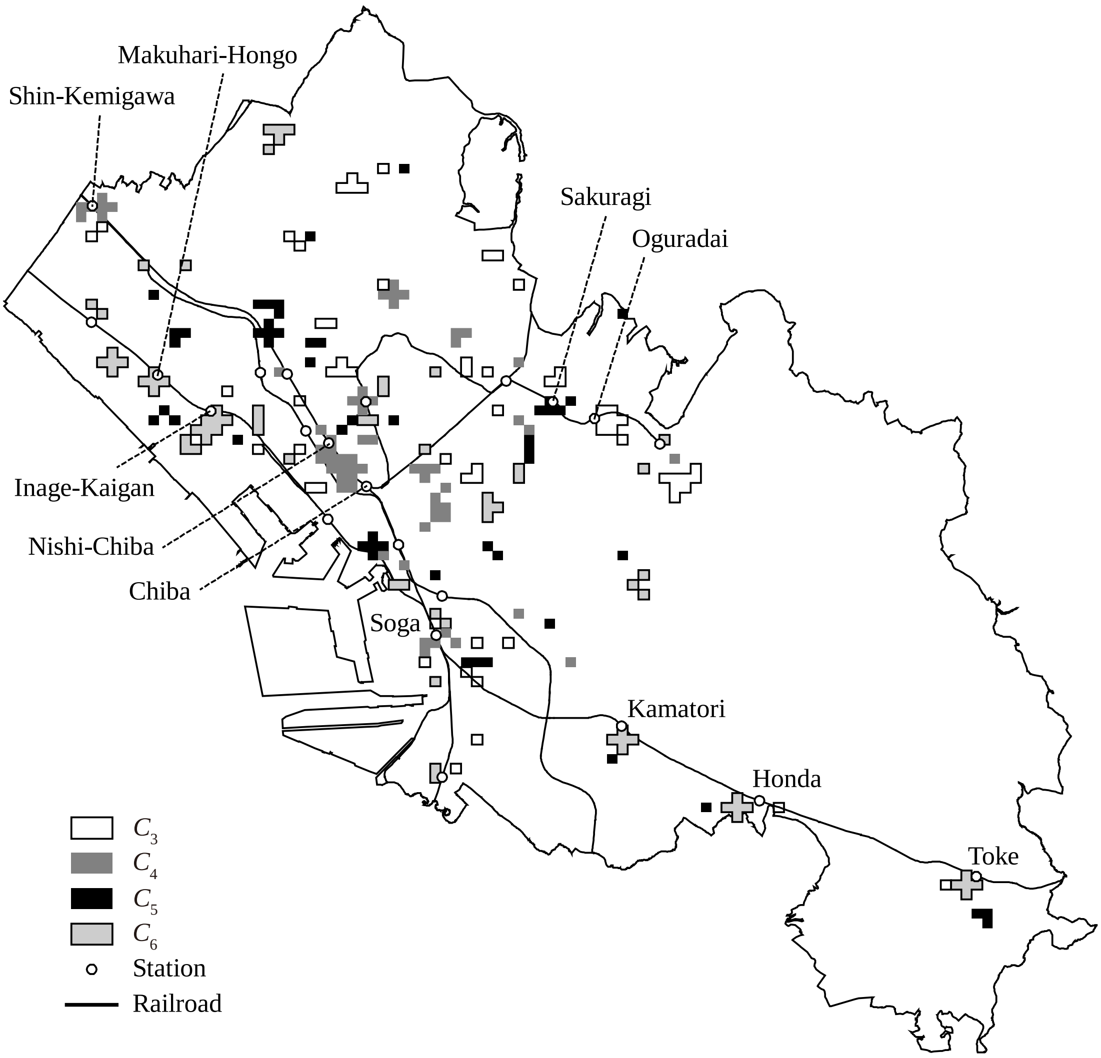
Tags assigned to centers C_3-C_9 are dispersed all over Chiba City. They represent small shopping districts in and around residential areas. Table 1 suggests that these centers can be classified into three groups: $\{C_3, C_4, C_5, C_6\}$, $\{C_7, C_8\}$, and $\{C_9\}$. Centers $\{C_3, C_4, C_5, C_6\}$ are characterized by convenience stores, beauty shops, and laundry shops while $\{C_7, C_8\}$ are characterized by sushi restaurants and barber shops. The former represents relatively new while the latter is traditional local shopping malls and districts in residential areas. The former can be further classified into $\{C_3, C_4\}$ and $\{C_5, C_6\}$ in Table 1, because only $\{C_3, C_4\}$ is assigned Japanese and American pubs. This reflects that local shopping districts often contain local pubs for residents in Japan. Center C_9 is characterized by supermarkets, noodle restaurants, and fast-food restaurants. It is a typical combination of commercial facilities found in shopping centers in suburban and rural areas of Japan. It is confirmed in Figure 7c, though some C_9 tags are found around railway stations.

Table 1 Commercial facilities assigned to centers detected when r=300 and β=50.

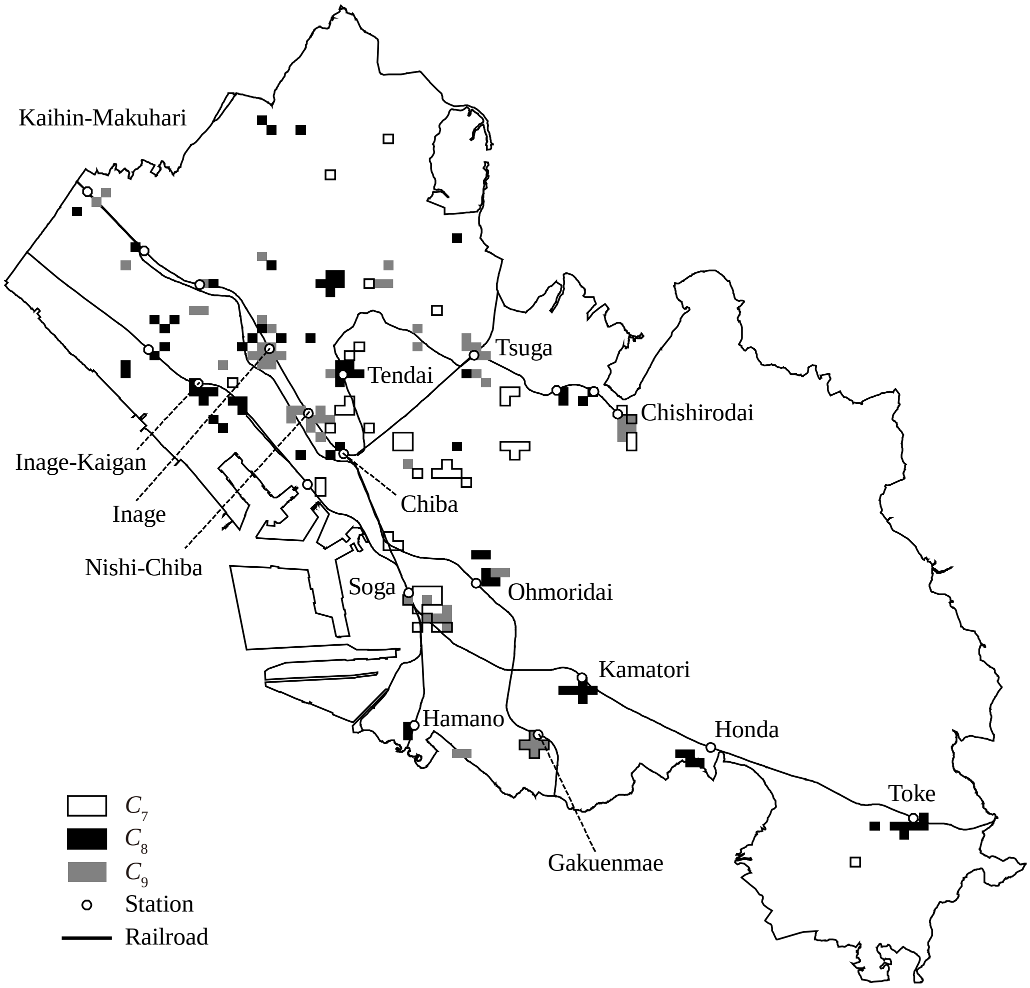
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
B_1	Japanese fast-food restaurants	○								
B_2	Banks	○								
B_3	Cosmetic stores	○								
B_4	Coffee shops	○								
B_5	Liquor shops	○				○			○	
B_6	Book stores	○					○			
B_7	Chinese restaurants	○						○		
B_8	Kimono shops		○							
B_9	Flower shops		○							
B_{10}	Fruits and vegetable shops	○	○							
B_{11}	Japanese noodle restaurants		○							○
B_{12}	Chinese noodle restaurants									○
B_{13}	Fast-food restaurants	○	○	○						○
B_{14}	Japanese pubs	○	○	○	○					
B_{15}	American pubs	○	○	○	○					
B_{16}	Convenience stores	○	○		○	○	○			
B_{17}	Beauty shops		○	○	○	○	○			
B_{18}	Laundry shops		○	○	○	○	○			
B_{19}	Sushi restaurants	○	○					○	○	
B_{20}	Pharmacies	○	○					○	○	
B_{21}	Barber shops		○			○		○	○	
B_{22}	Gas stations							○		
B_{23}	Supermarkets						○			○
B_{24}	Cram schools								○	○
B_{25}	Japanese pub-style restaurants				○					


Kaihin-Makuhari
Makuhari
Shin-Kemigawa
Inage
Tsuga
Nishi-Chiba
Chiba
Soga
C_1
C_2
Station
Railroad

(a)


Makuhari-Hongo
Shin-Kemigawa
Sakuragi
Oguradai
Inage-Kaigan
Nishi-Chiba
Chiba
Soga
Kamatori
Honda
Toke
C_3
C_4
C_5
C_6
Station
Railroad

(b)

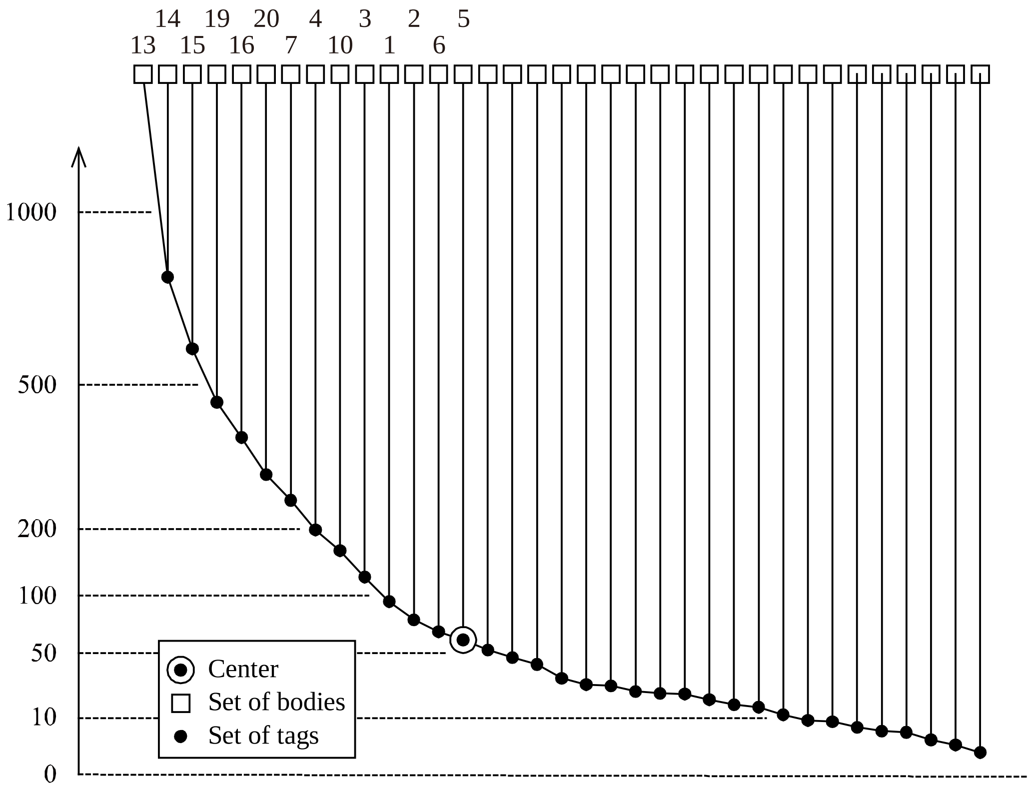


(c)

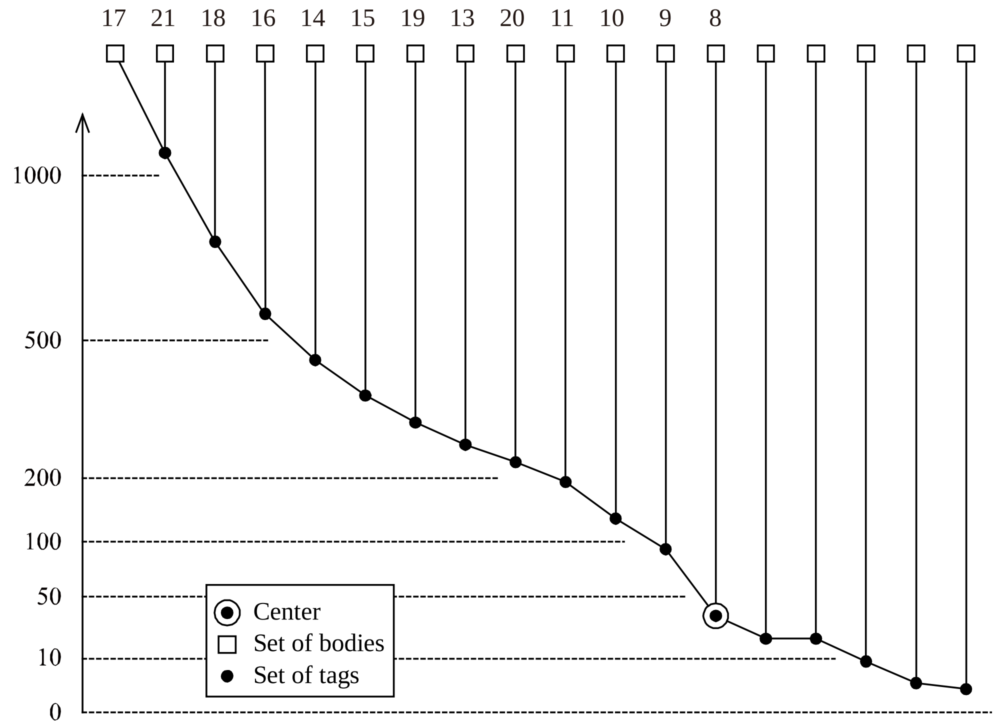
Figure 7. The location of tags assigned to centers detected when r=300 and β=50. (a) {C_1, C_2}, (b) {C_3, C_4, C_5, C_6}, (c) {C_7, C_8, C_9}.

Figure 8 shows the upper half of topology diagrams of centers C_1 and C_2. Bodies B_{13}, B_{14}, B_{15}, B_{16}, and B_{19} are closely located in both diagrams. This implies that many shopping districts represented by C_1 and C_2 contain fast-food restaurants, Japanese and American pubs, convenience stores, and sushi restaurants. Bodies B_{10} and B_{20}, on the other hand, are located relatively far from the above bodies in both diagrams. Only some shopping districts represented by C_1 and C_2 have fruit and vegetable shops and pharmacies.

Bodies located on the left hand side of topology diagram are more similar in their distributions. Bodies B_{13}, B_{14}, and B_{15} are similar in C_1 while B_{17}, B_{21}, and B_{18} are similar in C_2. The former represent fast-food restaurants, Japanese and American pubs while the latter are beauty shops, barber shops, and laundry shops. It is consistent with earlier discussion on centers C_1-C_6.



(a)



(b)

Figure 8. Topology diagrams of centers (a) C_1 and (b) C_2. The numbers over white squares indicate the suffix of bodies. The vertical axis indicates the size of spatial objects represented by the number of tags.

5. Conclusion

This paper proposes a new general method for analyzing the relations among distributions of spatial objects. The method has at least three advantages over existing ones. First, it can deal with more than two distributions of spatial objects. Second, it is applicable independent of the type of spatial objects. Third, it considers not only global but also local similarity among distributions. These advantages assure a wide applicability and flexibility of the method.

An emphasis of the method is on the relations among spatial objects rather than the relationship between the space and objects. The space is used mainly in preprocessing to define the local relations among spatial objects. Having defined the relations, analysis focuses on the relations among tags, bodies, and parts without considering the space. Though the space appears in condition ϑ_S, the its role is quite limited. Structure of space is rather defined based on the relations among spatial objects.

We thus call the method an object-oriented spatial analysis. This treatment of space sounds reasonable because space does not directly determine the status and properties of spatial objects in the real world. Space is a medium through which spatial objects affect with each other. Our focus should be on the relations among spatial objects rather than the relationship between the space and spatial objects.

We finally discuss some limitations and extensions of the paper for future research.

First, the method should be extended to treat spatiotemporal distributions. Though Sadahiro and Kobayashi (2012) discusses spatially distributed time-series data, the method cannot be applicable directly to general spatiotemporal distributions. The method proposed in this paper needs further extension and improvement to deal with a wide variety of spatiotemporal distributions.

Second, the visualization of the relations among spatial objects requires further development. Topology diagram proposed in section 2.3 is suitable for visualizing the topological structure among objects evaluated based on spatial proximity. However, as discussed in Section 2.6, a wide variety of relations can be defined among the distributions of spatial objects. New methods should be developed for visualizing the topological structure of objects evaluated from different perspectives.

Third, the framework of the method should be extended to confirmatory spatial analysis. This paper proposes the method for the use in exploratory spatial analysis, where we find research hypotheses that are tested in confirmatory spatial analysis. Though a gap usually exists between exploratory and confirmatory spatial analyses, it is desirable to perform both analyses within the same framework. Spatial modeling based on spatial relations should be further discussed.

References

Abraham, A., Hassanien, A. E., and Snasel, V., 2009. *Computational Social Network Analysis: Trends, Tools and Research Advances.* New York: Springer.

Anderson, I., 2002. *Combinatorics of Finite Sets.* New York: Dover Publication.

Arévalo, G., Desnos, N., Huchard, M., Urtado, C., and Vauttier, S., 2009. Formal concept analysis-based service classification to dynamically build efficient software component directories. *International Journal of General Systems*, **38**, 427-453.

Birkhoff, G., 1979. *Lattice Theory (3rd Ed.).* Providence, RI: American Mathematical Society.

Carrington, P. J., Scott, J., and Wasserman, S., 2005. *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press.

Clark, P. and Evans, F., 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35**, 445-453.

Davey, B. A. and Priestley, H. A., 2002. *Introduction to Lattice and Order*. Cambridge: Cambridge University Press.

Diggle, P. J., 2003. *Statistical Analysis of Spatial Point Patterns.* New York: Oxford University Press Inc.

Dixon, P., 1994. Testing spatial segregation using a nearest-neighbor contingency table. *Ecology*, **75**, 1940-1948.

Ganter B., Stumme, G., and Wille, R., 2005. Formal Concept Analysis: Foundations and Applications. *Lecture Notes in Artificial Intelligence, 3626*.

Goodall, D. W., 1952. Quantitative aspects of plant distribution. *Biological Reviews*, **27**, 194-245.

Greig-Smith, P., 1952. The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany*, **16**, 293-316.

Haggett, P. and Chorley, R. J., 1969. *Network Analysis in Geography*. London: Edward Arnold.

Janowitz, M. F., 2010. *Ordinal and Relational Clustering*. Singapore: World Scientific Publishing Co.

Kharrat, A., Popa, I. S., Zeitouni, K., and Faiz, S., 2008. Clustering algorithm for network constraint trajectories. In Ruas, A. and Gold, C.M. (eds) *Headway in Spatial Data Handling, 13th International Symposium on Spatial Data Handling, Lecture Notes in Geoinformation and Cartography*. Berlin: Springer, 631-647.

Knoke, D. and Yang, S., 2008. *Social Network Analysis*. London: Sage.

Kwuida, L. and Sertkaya, B., 2010. Formal Concept Analysis. *Proceedings of the 8th International Conference on Formal Concept Analysis, Lecture Notes in Artificial*

Intelligence, **5986**.

Lee, Y., 1979. A nearest-neighbor spatial association measure for the analysis of conditional locational interdependence. *Environment and Planning A*, **11**, 169-176.

van Lieshout, M. N. M. and Baddeley, A. J. 1999. Indices of dependence between types in multivariate point patterns. *Scandinavian Journal of Statistics*, **26**, 511-532.

Okabe, A. and Fujii, A., 1984. The statistical analysis through a computational method of a distribution of points in relation to its surrounding network. *Environment and Planning A*, **16**, 163-171.

Okabe, A. and Miki, F., 1984. A conditional nearest-neighbor spatial-association measure for the analysis of conditional locational interdependence. *Environment and Planning A*, **16**, 163-171.

Okabe, A., Yoshikawa, T., Fujii, A., and Oikawa, K., 1988. The statistical analysis of a distribution of activity points in relation to surface-like elements. *Environment and Planning A*, **20**, 609-620.

Parzen, E., 1962. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065-1076.

Pemmaraju, S. and Skiena, S., 2003. *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*. Cambridge: Cambridge University Press.

Pielou, E. C., 1961. Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. *Journal of Ecology*, **49**, 255-269.

Pielou, E. C., 1969. *An Introduction to Mathematical Ecology*. New York: Wiley.

Ripley, B. D., 1976. The second-order analysis of stationary point process. *Journal of Applied Probability*, **13**, 255-266.

Ripley, B. D., 1977. Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B*, **41**, 172-192.

Ripley, B. D., 1981. *Spatial Statistics*. New York: Wiley.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832-837.

Sadahiro, Y., 1999. Statistical methods for analyzing the distribution of spatial objects in relation to a surface. *Journal of Geographical Systems*, **1**, 107-136.

Sadahiro, Y., 2001. Number of polygons generated by map overlay: the case of convex polygons. *Transactions in GIS*, **5**, 345-353.

Sadahiro, Y., 2010. Analysis of the spatial relations among point distributions on a discrete space. *International Journal of Geographical Information Science*, **24**, 997-1014.

Sadahiro, Y., 2011. Analysis of the relations among spatial tessellations. *Journal of Geographical Systems*, **13**, 373-391.

Sadahiro, Y., Lay, R., and Kobayashi, T., 2012. Trajectories of moving objects on a network: detection of similarities, visualization of relations, and classification of trajectories. *Transactions in GIS*, to appear (draft version is available from http://ua.t.u-tokyo.ac.jp/pub/due-dp/105.pdf).

Sadahiro, Y., 2012. Exploratory analysis of polygons distributed with overlap. *Geographical Analysis,* to appear (draft version is available from http://ua.t.u-tokyo.ac.jp/pub/due-dp/102.pdf).

Sadahiro, Y. and Kobayashi, T., 2012. Exploratory analysis of spatially distributed time series data: Detection of similarities, clustering and visualization of mutual relations. *Discussion Paper Series, No. 108, Department of Urban Engineering, University of Tokyo* (available from http://ua.t.u-tokyo.ac.jp/pub/due-dp/108.pdf).

Shimble, A., 1953. Structural parameters of communication networks, *Bulletin of Mathematical Biophysics*, **15**, 501-507.

Silverman, B. W., 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Skellam, J. G., 1952. Studies in statistical ecology. I. Spatial pattern, *Biometrica*, **39**, 346-362.

Wasserman, S. and Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.