Discussion Paper Series
No. 96
Department of Urban Engineering
University of Tokyo

Analysis of the relationship among spatial tessellations

Yukio Sadahiro and Toshinori Sasaya
Department of Urban Engineering, University of Tokyo

Analysis of the relationship among spatial tessellations

Yukio Sadahiro and Toshinori Sasaya
Department of Urban Engineering, University of Tokyo

Abstract

Spatial tessellation is one of the most important spatial structures in geography. It represents census tracts, postal zones, market areas, land cover, soil patterns, and so forth. Spatial tessellations in the same region are often closely related with each other. School districts and postal zones are sometimes based on administrative units, while market areas of different categories of shops affect with each other. This paper proposes a method for analyzing the relationship among spatial tessellations. Similarity between tessellations is evaluated by granularity and hierarchy. Relationship among tessellations is represented by a diagram and visualized as two trees. The method is applied to the analysis of two sets of spatial tessellations in Japan: candidate plans for the new administrative system called Doshusei system and areas covered by branches of private companies. This application reveals the properties of the method and its measures as well as empirical findings.

1. Introduction

Spatial tessellation is one of the most important spatial structures in geography. Some are defined for administrative purposes such as census tracts, postal zones, electoral and school districts. Others are based on natural phenomena which include land cover, vegetation and soil patterns. Market areas of retail stores, say, drug stores and gas stations, are often approximated by a set of polygons such as Voronoi diagrams.

Spatial tessellations in the same region are often closely related with each other (Okabe et al, 2000; Sadahiro, 2002). School districts and postal zones are sometimes based on administrative units. TAZ (Transportation Analysis Zones) are determined by census tracts and administrative units. A close relationship exists between administrative units and land use pattern. Market areas of different categories of shops affect with each other because of consumers' propensity for one stop multipurpose shopping. It is therefore necessary to analyze the relationship among tessellations to understand not only the relationship itself but also the individual tessellations and their underlying structure.

Comparison and analysis of spatial tessellations depend on the comparability of attributes represented by tessellations. Suppose time series data of land use pattern in the same region. If the data share the same land use categories, the tessellations can be compared in terms of both spatial and attribute properties. Difference in land use category is evaluated as well as geometrical properties of tessellations. Two tessellations are regarded as different even if they are geometrically identical but different in land use pattern. On the other hand, tessellations of different attribute types are comparable only in terms of spatial properties. Difference among census tracts, postal zones, and electoral districts can be evaluated only in their spatial properties. A set of school districts for elementary, secondary, and higher education are not comparable in their attributes even if they are given in the same region. Tessellations are regarded as identical only if they are geometrically identical.

In the former case above, analysis starts with visual comparison of tessellations. Visual analysis is a useful means to obtain research hypothesis about spatial phenomena (Nielson *et al.*, 1997; Kraak, 2003; Slocum *et al.*, 2004; Ware, 2004; Wright, 2006). For time series analysis of tessellations, animated visualization is effective to capture significant or interesting changes in tessellations.

Visual analysis is often followed by quantitative analysis. Statistical measures are available that include χ^2, Kappa index and their extensions (Congalton and Mead, 1983; Rosenfield and Fitzpatrick-Lins, 1986; Monserud and Leemans, 1992; Pontius, 2000, 2002; Fritz and See, 2005). These measures are often used in remote sensing to compare land cover type by location between the actual one and that estimated from satellite images. However, they are applicable to any tessellations defined by the same or comparable variables.

Visual analysis and quantitative measures are also effective for analyzing tessellations of different attribute types. Though the measures mentioned above are not directly applicable, other summary statistics are available such as the average area and perimeter of regions, their variance and standard deviation, spatial mean of their gravity centers, and so forth. Unlike χ^2 and Kappa index, these statistics do not evaluate the difference in tessellations by location. They are calculated for each tessellation separately and compared among tessellations. Consequently, they do not reflect difference in arrangement of regions. Basic geometrical transformations such as translation, rotation and reflection do not affect statistics values.

Difference in spatial arrangement is not negligible in spatial analysis. To resolve this problem, this paper proposes a new method for analyzing spatial tessellations. Spatial tessellations are compared by location, which is followed by summary representation. The method focuses especially on the hierarchical relationship among tessellations which is often found in the real world. A country is divided into states or provinces, each of which is further divided into cities, towns, villages, and counties. A city might be further divided into smaller units defined by streets and avenues. Hierarchical structure is an essential property in the relationship among tessellations.

Section 2 proposes a method for analyzing the relationship among spatial tessellations. Spatial tessellations of different attribute types are discussed. Relationship between a pair of tessellations is evaluated quantitatively in some aspects. Relationship among more than two tessellation is then represented as a diagram and visualized as two trees. Section 3 applies the method to real data to evaluate its advantages and limitations. Two types of spatial tessellations are examined. One is candidate plans for the new administrative system in Japan called Doshusei system. The other is areas covered by branches of private companies. Section 4 summarizes the conclusions with a discussion.

2. Method

This paper analyzes a set of spatial tessellations defined in the same region. Each tessellation can represent the distribution of a different variable such as administrative units, postal zones, and school districts. To compare tessellations, therefore, only their spatial properties should be focused by omitting their attributes.

The method consists of the following steps: 1) classification of the relationship between a pair of tessellations, 2) definition of spatial operations used for evaluating relationship between a pair of tessellations, 3) quantitative evaluation of the similarity between a pair of tessellations, 4) representation of relationship among more than two tessellations, and 5) simplification of the representation proposed. These are explained in the following successively.

2.1 Relationship between a pair of spatial tessellations

Suppose a set of spatial tessellations defined in region S denoted by $\Theta=\{\Omega_1, \Omega_2, \ldots, \Omega_n\}$. Tessellation Ω_i consists of n_i regions, $\{T_{i1}, T_{i2}, \ldots, T_{im_i}\}$. To compare a pair of tessellations, three types of relationships are defined.

1) Identical
2) Globally hierarchical
3) Locally hierarchical

Two tessellations Ω_i and Ω_j are *identical* if any region in Ω_i perfectly fits a region in Ω_j and vice versa. A focus is only on the spatial aspect of tessellations; two tessellations can be identical if one represents administrative units while the other is postal zones.

If two tessellations are not identical, a weaker relationship is considered. Two tessellations Ω_i and Ω_j are *globally hierarchical* if any region in Ω_i is consistently and fully covered by a region in Ω_j or vice versa. Tessellations Ω_1, Ω_2 and Ω_3 in Figure 1 are globally hierarchical with each other. This relationship is often find in administrative and management systems. If any region in Ω_i is fully covered by a region in Ω_j , Ω_i is a *lower level tessellation* of Ω_j. Tessellation Ω_i is generated by further dividing regions in Ω_j into smaller regions. If Ω_i is obtained by aggregating regions in Ω_j, Ω_i is a *higher level tessellation* of Ω_j. In Figure 1, Ω_3 is a lower level tessellation of Ω_2, while Ω_1 is a higher level tessellation of Ω_2.

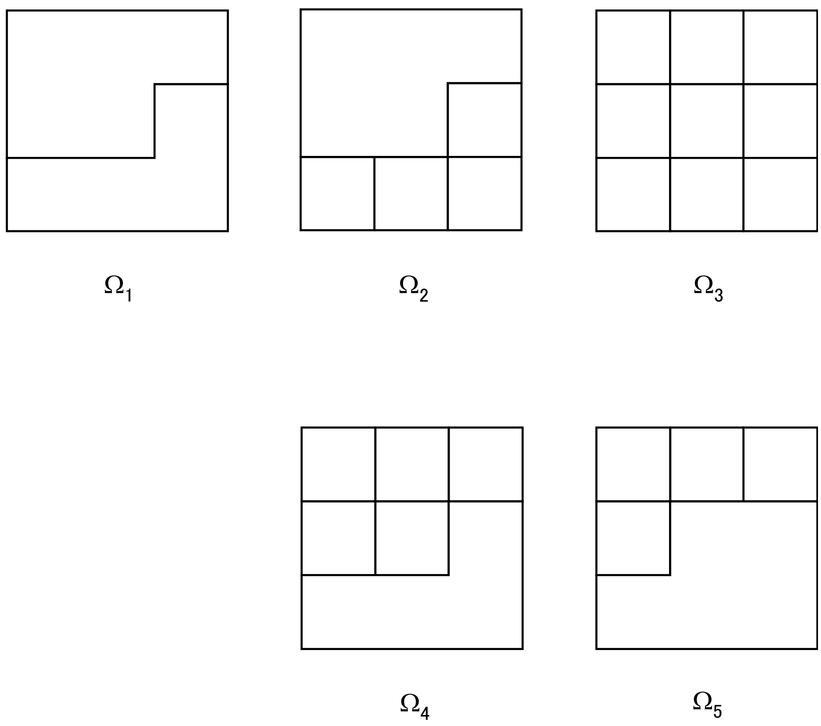

Figure 1 Hierarchical relationship among tessellations. Tessellations Ω_1, Ω_2, and Ω_3 are globally hierarchical with each other. Tessellations Ω_2 and Ω_4 are locally hierarchical.

The above hierarchical relationship applies only when either "cover" or "be covered" relationship consistently exists between tessellations. On the other hand, if the two relationships "cover" and "be covered" exist simultaneously between tessellations, it is called *locally hierarchical*, because hierarchical relationship can be found locally in Ω_i and Ω_j. An example is the relationship between Ω_2 and Ω_4 in Figure 1. A larger polygon in Ω_2 fully covers five squares in Ω_4 while the rest four squares in Ω_2 are fully covered by a polygon in Ω_4. Though Ω_2 and Ω_4 are not globally hierarchical, they are hierarchical at a local scale. This relationship can also be defined as nonexistence of partial overlap of regions in Ω_i and Ω_j.

2.2 Overlay operations for tessellations

This subsection introduces two overlay operations for tessellation. They are used to evaluate and describe relationship among tessellations.

Merging overlay of Ω_i and Ω_j, denoted by $O_M(\Omega_i, \Omega_j)$, is to make an overlay of Ω_i and Ω_j with removing the boundary lines that exist only in one tessellation (Figure 2). Regions in Ω_i and Ω_j are merged, and as a result, $O_M(\Omega_i, \Omega_j)$ is a higher level tessellation of Ω_i and Ω_j. *Dividing overlay* $O_D(\Omega_i, \Omega_j)$ is to make an overlay of Ω_i and Ω_j with keeping all the boundary lines in both Ω_i and Ω_j. It divides regions into smaller ones as shown in Figure 2. If Ω_i and Ω_j are globally hierarchical, $O_M(\Omega_i, \Omega_j)$ and $O_D(\Omega_i, \Omega_j)$ and are identical to the higher and lower level tessellation between Ω_i and Ω_j, respectively.

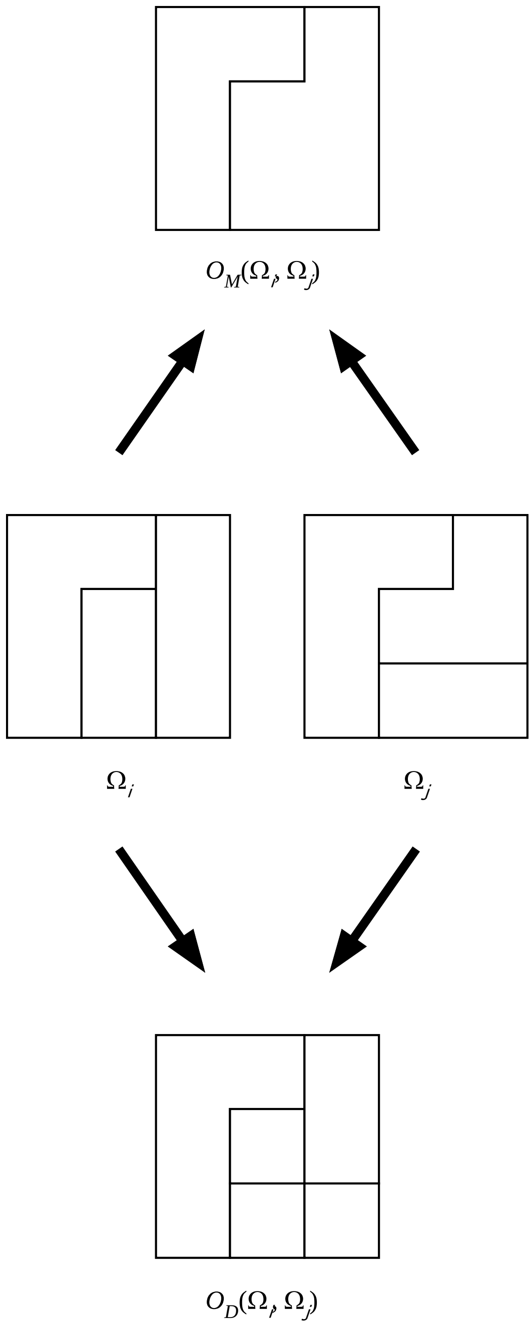

Figure 2 Merging and dividing overlay for (Ω_i, Ω_j), denoted by $O_M(\Omega_i, \Omega_j)$ and $O_D(\Omega_i, \Omega_j)$, respectively.

2.3 Measurement of the similarity between spatial tessellations

This subsection discusses quantitative evaluation of the similarity between tessellations. Mathematical representation is introduced to describe tessellations and their relationship.

Tessellation Ω_i is represented by a binary function:

$$\rho(\mathbf{x};T_{ij})=\begin{cases}1 & \text{if } \mathbf{x}\in T_{ij}\\ 0 & \text{otherwise}\end{cases}. \tag{1}$$

Two points $\mathbf{x}_1$ and $\mathbf{x}_2$ are contained in the same region in Ω_i if

$$\begin{aligned}\sigma(\mathbf{x}_1,\mathbf{x}_2;\Omega_i)&=\begin{cases}1 & \text{if } \sum_j \rho(\mathbf{x}_1;T_{ij})\rho(\mathbf{x}_2;T_{ij})=1\\ 0 & \text{otherwise}\end{cases}.\\ &=\sum_j \rho(\mathbf{x}_1;T_{ij})\rho(\mathbf{x}_2;T_{ij})\end{aligned} \tag{2}$$

The area of region T_{ij} is given by

$$A(T_{ij})=\int_{\mathbf{x}\in S}\rho(\mathbf{x},T_{ij})\,\mathrm{d}\mathbf{x}. \tag{3}$$

Granularity of tessellation Ω_i is the ratio of point pairs contained in the same region in Ω_i to all the pairs in S. Mathematically it is defined as

$$\begin{aligned}G(\Omega_i)&=\frac{\int_{\mathbf{x}_2\in S}\int_{\mathbf{x}_1\in S}\sigma(\mathbf{x}_1,\mathbf{x}_2;\Omega_i)\,\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2}{\{A(S)\}^2}.\\ &=\frac{\sum_j\{A(T_{ij})\}^2}{\{A(S)\}^2}\end{aligned} \tag{4}$$

The domain of granularity is $0<G(\Omega_i)\leq 1$. It is a function of the area of regions; it is large when Ω_i consists of a small number of large regions.

One measure of the similarity between tessellations is the difference of granularity called *granularity distance* denoted by $D_G(\Omega_i, \Omega_j)$:

$$D_G(\Omega_i,\Omega_j)=\left|G(\Omega_i)-G(\Omega_j)\right|. \tag{5}$$

Granularity distance is one of the basic measures of similarity between tessellations. It is calculated from the size of regions of each tessellation, without comparing two tessellations by location. Consequently, if a tessellation is a mirror image of the other, they give the same granularity value. Tessellations Ω_2 and Ω_5 in Figure 1 have the same granularity though they are different in the arrangement of regions.

To treat such cases, this paper examines every pair of points in S in terms of whether they are contained in the same or different regions in two tessellations. The relationship between two points is classified into four types.

R_{11}: Both points are contained in the same region in both Ω_i and Ω_j.
R_{10}: Both points are contained in the same region in Ω_i but in different regions in Ω_j.
R_{01}: Both points are contained in different regions in Ω_i but in the same region in Ω_j.
R_{00}: Both points are contained in different regions in both Ω_i and Ω_j.

The ratio of point pairs of relationship R_{kl} is denoted as $m_{kl}(\Omega_i, \Omega_j)$. Measure $m_{11}(\Omega_i, \Omega_j)$ is given by

$$m_{11}\left(\Omega_i,\Omega_j\right)=\frac{1}{\left\{A(S)\right\}^2}\int_{\mathbf{x}_1\in S}\int_{\mathbf{x}_2\in S}\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_i\right)\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_j\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2 . \tag{6}$$

Similarly,

$$\begin{aligned}m_{10}\left(\Omega_i,\Omega_j\right)&=\frac{1}{\left\{A(S)\right\}^2}\int_{\mathbf{x}_1\in S}\int_{\mathbf{x}_2\in S}\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_i\right)\left(1-\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_j\right)\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2\\&=\frac{1}{\left\{A(S)\right\}^2}\int_{\mathbf{x}_1\in S}\int_{\mathbf{x}_2\in S}\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_i\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2-m_{11}\left(\Omega_i,\Omega_j\right)\\&=G\left(\Omega_i\right)-m_{11}\left(\Omega_i,\Omega_j\right)\end{aligned} , \tag{7}$$

$$\begin{aligned}m_{01}\left(\Omega_i,\Omega_j\right)&=\frac{1}{\left\{A(S)\right\}^2}\int_{\mathbf{x}_1\in S}\int_{\mathbf{x}_2\in S}\left(1-\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_i\right)\right)\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_j\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2\\&=G\left(\Omega_j\right)-m_{11}\left(\Omega_i,\Omega_j\right)\end{aligned} , \tag{8}$$

and

$$\begin{aligned}m_{00}\left(\Omega_i,\Omega_j\right)&=\frac{1}{\left\{A(S)\right\}^2}\int_{\mathbf{x}_1\in S}\int_{\mathbf{x}_2\in S}\left(1-\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_i\right)\right)\left(1-\sigma\left(\mathbf{x}_1,\mathbf{x}_2;\Omega_j\right)\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2\\&=1+m_{11}\left(\Omega_i,\Omega_j\right)-G\left(\Omega_i\right)-G\left(\Omega_j\right)\end{aligned} . \tag{9}$$

The relationship among these measures is summarized as shown in Table 1.

Table 1 Relationship among the measures of similarity between tessellations

			Sum
	$m_{11}(\Omega_i, \Omega_j)$	$m_{10}(\Omega_i, \Omega_j)$	$G(\Omega_i)$
	$m_{01}(\Omega_i, \Omega_j)$	$m_{00}(\Omega_i, \Omega_j)$	$1-G(\Omega_i)$
Sum	$G(\Omega_j)$	$1-G(\Omega_j)$	1

Using these measures, we can identify the relationship between Ω_i and Ω_j. Tessellations Ω_i and Ω_j are identical if $m_{10}(\Omega_i, \Omega_j)=m_{01}(\Omega_i, \Omega_j)=0$. If $m_{01}(\Omega_i, \Omega_j)=0$ or $m_{10}(\Omega_i, \Omega_j)=0$, Ω_i and Ω_j are globally hierarchical. In the former case Ω_i is a higher level tessellation of Ω_j, while Ω_i is a lower level tessellation of Ω_j in the latter case.

They also permit us to evaluate the closeness of the relationship between Ω_i and Ω_j. *Global structural distance* between Ω_i and Ω_j is defined as

$$\begin{aligned} D_{GS}\left(\Omega_i,\Omega_j\right) &= \min\left(m_{10}\left(\Omega_i,\Omega_j\right), m_{01}\left(\Omega_i,\Omega_j\right)\right) \\ &= \min\left(G\left(\Omega_i\right)-m_{11}\left(\Omega_i,\Omega_j\right), G\left(\Omega_j\right)-m_{11}\left(\Omega_i,\Omega_j\right)\right). \\ &= \min\left(G\left(\Omega_i\right), G\left(\Omega_j\right)\right)-m_{11}\left(\Omega_i,\Omega_j\right) \end{aligned} \tag{10}$$

It is zero if Ω_i and Ω_j are globally hierarchical, and increases as the hierarchical relationship is weakened.

Local hierarchy, on the other hand, is evaluated by using overlay operations defined earlier as well as these measures. Tessellations Ω_i and Ω_j are locally hierarchical if any region in Ω_i does not partially overlap with a region in Ω_j. In other words, local hierarchical relationship can be evaluated by examining if partial overlap exists between tessellations. In Figure 3, region T_{ik} (hatched area) in Ω_i (solid lines) and T_{jl} (gray shaded area) in Ω_j (broken lines) partially overlap with each other. In such a case, pairs of points exist such as those indicated by $\mathbf{x}_1$ and $\mathbf{x}_2$. Points $\mathbf{x}_1$ and $\mathbf{x}_2$ are contained in different regions in both Ω_i and Ω_j, but the region containing $\mathbf{x}_1$ in Ω_i and that of $\mathbf{x}_2$ in Ω_j overlap with each other.

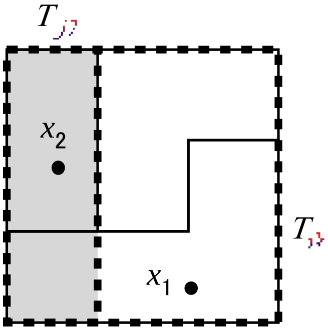

Figure 3 Partial overlap of regions in different tessellations Ω_i and Ω_j.

Local hierarchy is evaluated by detecting point pairs such as $\mathbf{x}_1$ and $\mathbf{x}_2$. It is easily done by applying merging overlay to Ω_i and Ω_j. This operation merges regions T_{ik} and T_{il} into one region denoted by $T_{ik} \cup T_{il}$. Relationship between pairs of points in $T_{ik} \cup T_{il}$ is classified into one of the four types R_{11}, R_{10}, R_{01}, or R_{00}. If Ω_i and Ω_j are locally hierarchical, any pair of points in $T_{ik} \cup T_{il}$ is classified into R_{11}, R_{10}, or R_{01}. Therefore, the ratio of point pairs in R_{00} indicates whether Ω_i and Ω_j are locally hierarchical.

Local structural distance between Ω_i and Ω_j is defined as

$$\begin{aligned} D_{LS}\left(\Omega_i,\Omega_j\right) &= G\left(O_M\left(\Omega_i,\Omega_j\right)\right) - m_{11}\left(\Omega_i,\Omega_j\right) - m_{10}\left(\Omega_i,\Omega_j\right) - m_{01}\left(\Omega_i,\Omega_j\right) \\ &= G\left(O_M\left(\Omega_i,\Omega_j\right)\right) - m_{11}\left(\Omega_i,\Omega_j\right) - G\left(\Omega_i\right) + m_{11}\left(\Omega_i,\Omega_j\right) - G\left(\Omega_j\right) + m_{11}\left(\Omega_i,\Omega_j\right) \\ &= G\left(O_M\left(\Omega_i,\Omega_j\right)\right) + m_{11}\left(\Omega_i,\Omega_j\right) - G\left(\Omega_i\right) - G\left(\Omega_j\right) \end{aligned} \tag{11}$$

Using

$$\begin{aligned} D_{GS}\left(\Omega_i,\Omega_j\right) &= \min\left(m_{10}\left(\Omega_i,\Omega_j\right), m_{01}\left(\Omega_i,\Omega_j\right)\right) \\ &= \min\left(G\left(\Omega_i\right) - m_{11}\left(\Omega_i,\Omega_j\right), G\left(\Omega_j\right) - m_{11}\left(\Omega_i,\Omega_j\right)\right), \\ &= \min\left(G\left(\Omega_i\right), G\left(\Omega_j\right)\right) - m_{11}\left(\Omega_i,\Omega_j\right) \end{aligned} \tag{12}$$

we have

$$D_{LS}\left(\Omega_i,\Omega_j\right) = G\left(O_M\left(\Omega_i,\Omega_j\right)\right) + \min\left(G\left(\Omega_i\right), G\left(\Omega_j\right)\right) - D_{GS}\left(\Omega_i,\Omega_j\right) - G\left(\Omega_i\right) - G\left(\Omega_j\right) \tag{13}$$

It is zero if Ω_i and Ω_j are locally hierarchical, and increases as the hierarchical relationship is weakened.

As seen in Equation (12), local structural distance is measurable by performing merging and dividing overlay and calculate the granularity of resulting tessellations. Local structural distance is zero if Ω_i and Ω_j are locally hierarchical. It increases with the collapse of local hierarchy.

2.4 Representation of relationship among spatial tessellations

So far we have discussed the relationship between a pair of spatial tessellations. To treat more than two tessellations simultaneously, we propose a method for representing the relationship among spatial tessellations as a diagram.

Applying merging overlay to tessellations Ω_1 and Ω_2, we obtain $O_M(\Omega_1, \Omega_2)$. Tessellation $O_M(\Omega_1, \Omega_2)$ is an identical or a higher level tessellation of Ω_1 and Ω_2, and consequently, the relationship of global hierarchy is a partial order. Tessellation $O_M(\Omega_1, \Omega_2)$ is an immediate successor of Ω_1 and Ω_2. Similarly, since dividing overlay also gives global hierarchy, tessellation $O_D(\Omega_1, \Omega_2)$ is an immediate predecessor of Ω_1 and Ω_2.

From tessellations Ω_1, Ω_2 and Ω_3, merging overlay generates $O_M(\Omega_1, \Omega_2)$, $O_M(\Omega_1, \Omega_3)$, and $O_M(\Omega_2, \Omega_3)$ The new tessellations yields another tessellation, which is as a result a single tessellation because any pair of tessellations in $O_M(\Omega_1, \Omega_2)$, $O_M(\Omega_1, \Omega_3)$, and $O_M(\Omega_2, \Omega_3)$ yields the same result. Application of dividing overlay also generates a set of tessellations.

The set of all the original tessellations and those generated by merging and dividing overlay operations is a finite partially ordered set. To represent the set and the partial order, this paper propose a diagram called *overlay diagram*. This diagram represents tessellations as nodes and overlay operations as links. It contains a Hasse diagram as a subset (Anderson, 2002; Davey and Priestley, 2002; Pemmaraju and Skiena, 2003), which is often used as a visual representation of a finite partially ordered set and its partial order. Figure 4 is an example of overlay diagram for the tessellations obtained from $\{\Omega_1, \Omega_2, \Omega_3\}$ by overlay operations, where the vertical axis represents the granularity of tessellations.

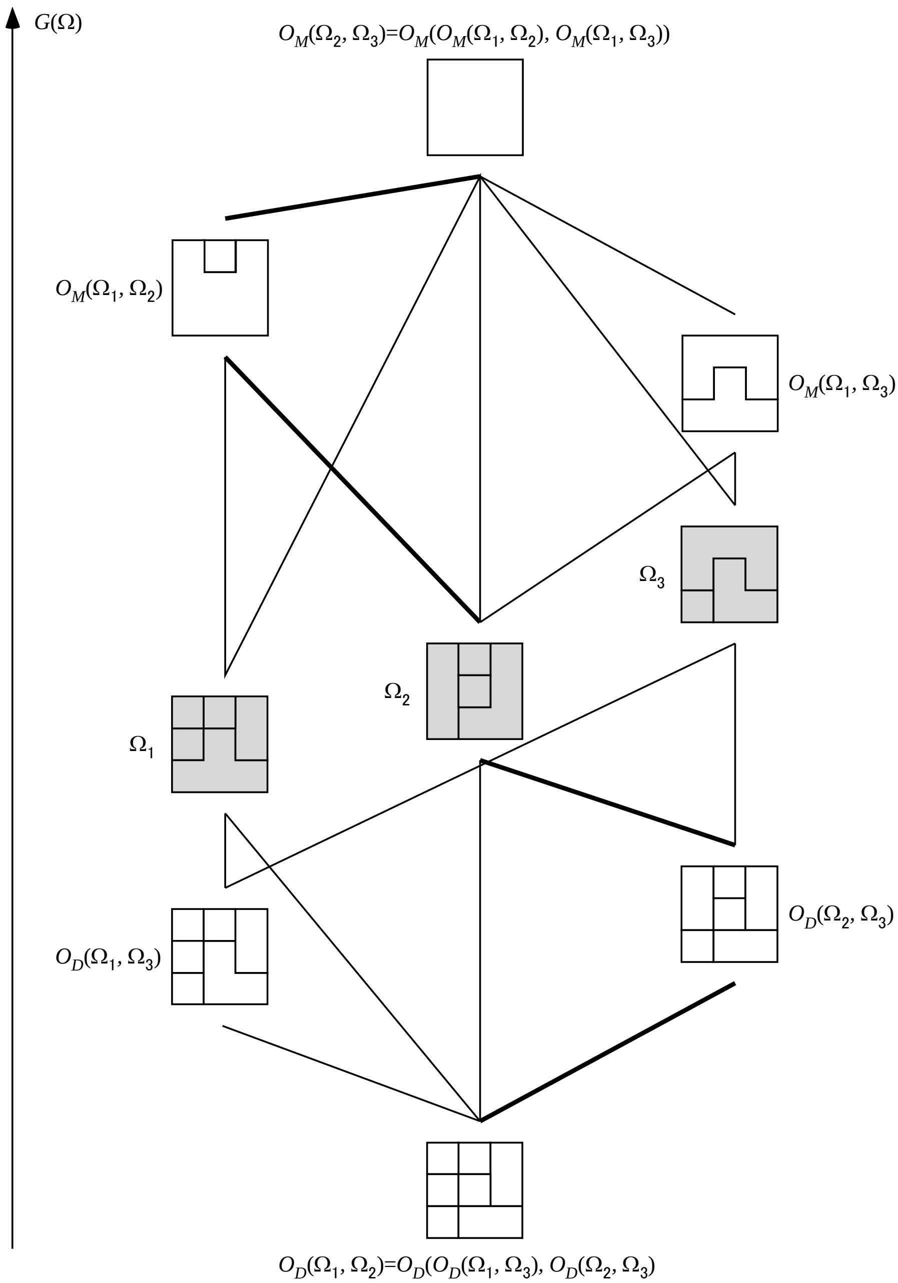

Figure 4 Overlay diagram for tessellations obtained from $\{\Omega_1, \Omega_2, \Omega_3\}$ by overlay operations. Links indicate the overlay operations. The vertical axis represents the granularity of tessellations.

Links directly connecting tessellations indicate overlay operations, and consequently, global hierarchical relationship. Since transitivity holds for the partial order, any pair of tessellations connected by a set of links in the same direction are globally hierarchical. Tessellations $O_M(\Omega_2, \Omega_3)$, $O_M(\Omega_1, \Omega_2)$, Ω_2, $O_D(\Omega_2, \Omega_3)$, and $O_D(\Omega_1,$

Ω_2) are all globally hierarchical as indicated by bold lines in Figure 4. A tessellation at a lower level is generated by simply dividing the regions of a tessellation at an upper level. To the contrary, merging regions of a tessellation, we obtain tessellations at upper levels connected by a set of links of the same direction.

In the overlay diagram shown in Figure 4, the measures of similarity between tessellations are represented as the vertical distance. As seen in Figure 5, granularity distance between Ω_i, Ω_j is given by the distance between Ω_i and Ω_j in the vertical direction. Global structural distance $D_{GS}(\Omega_i, \Omega_j)$, which is defined by Equation (10), is the vertical distance between either $G(\Omega_i)$ or $G(\Omega_j)$ and $G(O_D(\Omega_1, \Omega_2))$. Local structural distance $D_{LS}(\Omega_i, \Omega_j)$ (Equation (11)) is given by subtracting the distance between either $G(\Omega_i)$ or $G(\Omega_j)$ and $G(O_D(\Omega_1, \Omega_2))$ from that between $G(O_M(\Omega_1, \Omega_2))$ and either $G(\Omega_i)$ or $G(\Omega_j)$.

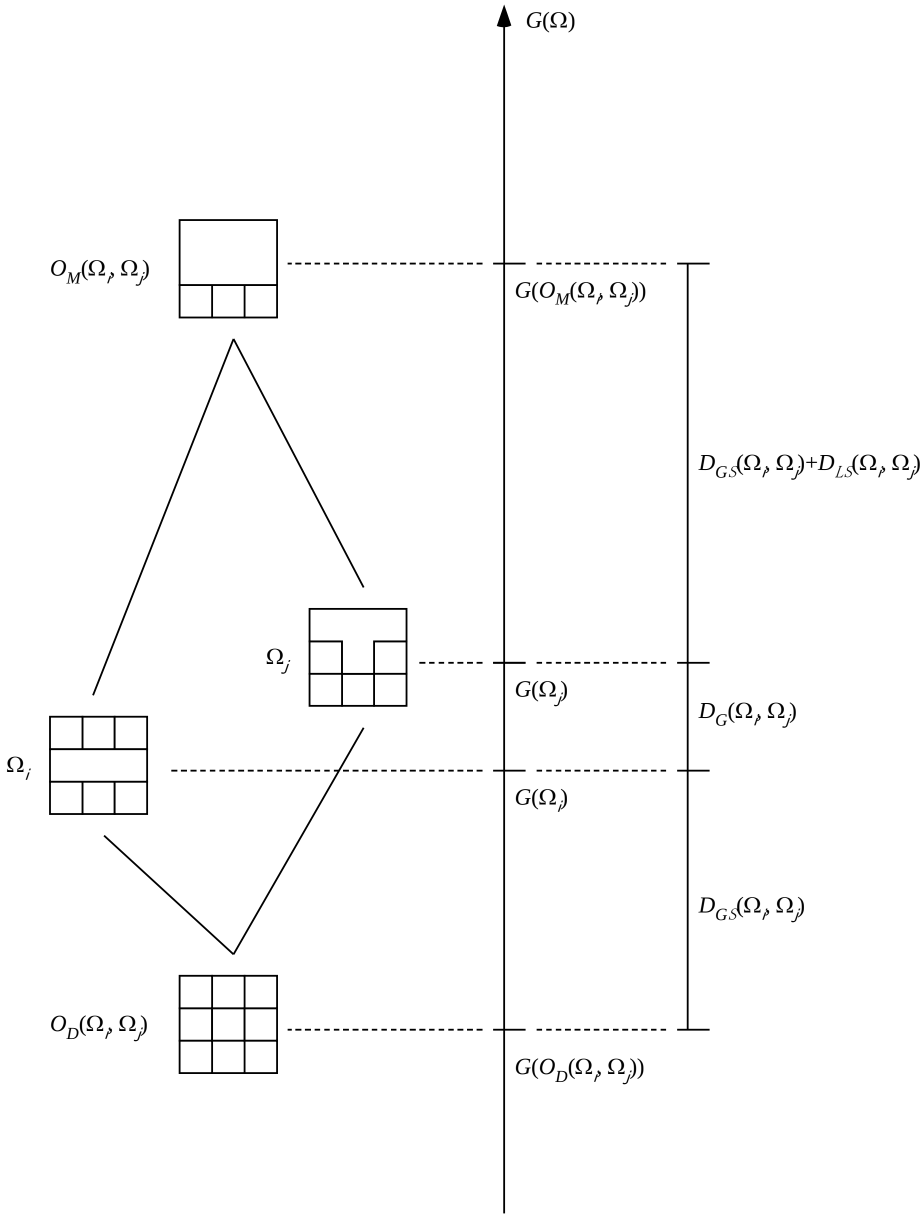

Figure 5 Measures of similarity between tessellations Ω_i and Ω_j.

As seen above, overlay diagram provides a framework for analyzing the relationship among spatial tessellations. Two tessellations connected by a set of links are globally hierarchical. If two tessellations are locally hierarchical, $D_{LS}(\Omega_i, \Omega_j)=0$ so that the tessellations and their immediate successor and predecessor form a parallelogram (recall Figure 5). If two tessellations are almost globally hierarchical, the shorter link between the tessellations and their predecessor is almost negligible so that the tessellations look almost directly connected by a link. Using overlay diagram, we can also grasp the overall distribution of granularity of tessellations. Consequently, the relationship among tessellations can be analyzed globally in aspects of granularity and hierarchy.

2.5 Simplification of visual representation

Given a set of tessellations $\Theta=\{\Omega_1, \Omega_2, ..., \Omega_n\}$, theoretically, we can always construct a overlay diagram through the procedure mentioned earlier. A few number of tessellations yields a simple diagram effective for visual analysis. However, overlay diagram can be very dense and complicated when it is created from numerous tessellations. Such a diagram is not suitable for visual analysis because it contains too much information to visualize as one figure.

One solution to this problem is to visualize only the essential structure of the relationship among spatial tessellations. In fact, since overlay diagram is not uniquely determined for a given partially ordered set, there are many possibilities for a given set of tessellations.

For a simple but effective visualization, we propose a tree representation called *overlay tree*. Given a set of tessellations Θ, we consider all the possible pairs of tessellations. We evaluate their merging overlay operation on a certain criteria to choose the one that gives the best result. If Ω_i and Ω_j are chosen, they are overlaid to generate their successor $O_M(\Omega_i, \Omega_j)$. From the set $\Theta \cup O_M(\Omega_i, \Omega_j)-\{\Omega_i, \Omega_j\}$, a pair of tessellations are chosen on the same criteria. This process gradually decreases the number of tessellations, and it halts when only one tessellation is left. The result of this process is called *merging overlay tree*. Using dividing overlay instead of merging overlay, we also obtain a tree representation, which is called dividing overlay tree.

Overlay trees should omit unnecessary details to present only the essential structure of the relationship among tessellations. Shorter and fewer links are desirable to keep the tree as simple as possible. We thus propose four methods to create a tessellation tree. They are different in the criteria used for choosing tessellations to be overlaid in tree construction.

The first method uses $D_{GS}(\Omega_i, \Omega_j)+D_{LS}(\Omega_i, \Omega_j)$ as the criterion for merging overlay tree. As seen in Figure 5, it is the length of the link between the tessellation of larger granularity and its successor. Tessellations of smaller $D_{GS}(\Omega_i, \Omega_j)+D_{LS}(\Omega_i, \Omega_j)$ are overlaid earlier, independently of their granularity. This implies that similarity in structure is more emphasized than that in granularity. From pairs of original tessellations, the one that gives the smallest $D_{GS}(\Omega_i, \Omega_j)+D_{LS}(\Omega_i, \Omega_j)$ are extracted to generate their successor. The new set of tessellations are similarly evaluated to generate another new tessellation. Merging overlay is performed repeatedly until only one tessellation is left. Dividing overlay tree is similarly constructed by using the distance $D_{GS}(\Omega_i, \Omega_j)$ instead as the criterion.

The second method uses $D_{GS}(\Omega_i, \Omega_j)+D_{LS}(\Omega_i, \Omega_j)+D_G(\Omega_i, \Omega_j)$ and $D_{GS}(\Omega_i, \Omega_j)+D_G(\Omega_i, \Omega_j)$ as criteria for merging and dividing overlay trees, respectively. Distances $D_{GS}(\Omega_i, \Omega_j)$ and $D_{LS}(\Omega_i, \Omega_j)$ evaluate structural similarity between tessellations while $D_G(\Omega_i, \Omega_j)$ focuses on the similarity of granularity. Consequently, the second method put less emphasis on structural similarity than method 1.

In the above methods, tessellations to be overlaid can be chosen from anywhere in overlay diagram. Tessellations of very large granularity may be chosen first in merging overlay tree. This inevitably makes links longer because tessellations of small granularity are overlaid later on those of larger ones. To obtain shorter links, method 3 adopts $G(O_M(\Omega_i, \Omega_j))$ and $G(O_D(\Omega_i, \Omega_j))$ for merging and dividing overlay trees, respectively. Tessellations of smaller $G(O_M(\Omega_1, \Omega_2))$ are chosen earlier so that the overlay tree grows upward gradually from original tessellations. As a result, method 3 put more emphasis on granularity similarity than methods 1 and 2. The location of tessellations combined is considered first, while structural similarity is not explicitly evaluated.

Method 4 takes into account simultaneously the similarity in structure and granularity and the location of tessellations in overlay diagram. It uses $G(O_M(\Omega_i, \Omega_j))+D_{GS}(\Omega_i, \Omega_j)+D_{LS}(\Omega_i, \Omega_j)$ and $G(O_D(\Omega_i, \Omega_j))-D_{GS}(\Omega_i, \Omega_j)$ for merging and dividing trees, respectively. Granularity $G(O_M(\Omega_i, \Omega_j))$ represents the location of tessellations, while $D_{GS}(\Omega_i, \Omega_j)+D_{LS}(\Omega_i, \Omega_j)$ measures the structural difference. Consequently, in merging overlay tree, tessellations of similar structure with small granularity are overlaid earlier.

Using these measures, we obtain two overlay trees from the original tessellations. They often look similar to dendrograms used in cluster analysis. However, overlay trees are different from dendrograms in the following aspects. First, original tessellations are not always located at the end of the tree. If two tessellations are globally hierarchical, they are often connected directly by a link and the higher level tessellation is located in the middle of the tree. Second, total similarity between tessellations cannot be evaluated by link length. In dendrograms, longer links indicate less similarity between elements. In overlay trees, on the other hand, one overlay operation is usually represented by two links that indicate different kinds of similarity. The shorter link represents the structural difference between tessellations while the longer one is the summation of both structural and granularity differences.

Consequently, only longer links should be considered to evaluate total similarity between tessellations.

3. Application

Using the method proposed above, this section analyzes two sets of tessellations both of which covers the whole area of Japan. One is a set of candidate plans for new administrative system called Doshusei system, a set of administrative units now under consideration. The other is a set of areas covered by branches of private companies. They consist of 33 and 34 tessellations, respectively. All the tessellations are determined based on prefecture units and cover the whole area of Japan. Every region of tessellations consists of one or more prefectures, and every prefecture is covered by a single region. Figure 6 shows an example of Doshusei system.

Figure 6 A Doshusei system. Thin and thick lines indicate the boundary of prefectures and new administrative units, respectively.

A focus is put on the comparison of four different methods for constructing overlay trees. Merging and dividing overlay trees were constructed for both sets of tessellations by using four different methods. As a result sixteen trees were obtained. Their properties are summarized as the total and average lengths of links in Table 2.

Table 2 Total and average length of links in overlay trees.

Tessellations	Tree type	Methods	Total length	Average length
Doshusei system	Merging overlay	1	6.4962	0.2030
		2	6.0598	0.1836
		3	5.9765	0.1828
		4	6.0210	0.1830
	Dividing overlay	1	2.0914	0.0654
		2	1.3889	0.0421
		3	1.1227	0.0351
		4	1.1942	0.0351
Company branches	Merging overlay	1	5.5908	0.1694
		2	4.9588	0.1503
		3	4.8103	0.1458
		4	4.8231	0.1462
	Dividing overlay	1	1.5165	0.0460
		2	1.4830	0.0449
		3	1.0177	0.0308
		4	1.0638	0.0322

From Table 2 we first notice that merging overlay trees are longer than dividing overlay trees. This is because merging overlay trees put more emphasis on structural than granularity similarity; merging overlay operation detects even a slight difference in location of boundary lines. Consequently, as seen in Figure 7, merging overlay sometimes generates a tessellation far more different from the original ones than dividing overlay.

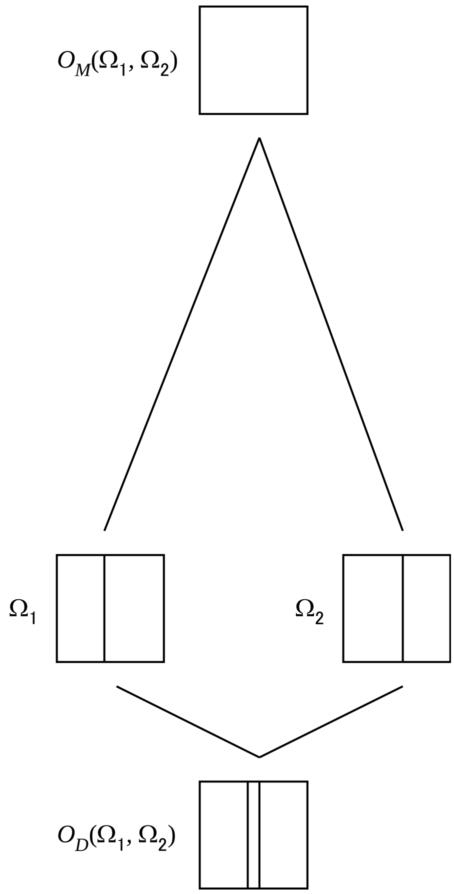

Figure 7 Merging and dividing overlay of two tessellations. A slight difference in location of a boundary line greatly affects the merging overlay while it is not critical in dividing overlay.

Table 2 also shows that methods 3 and 4 yields trees of shorter links than methods 1 and 2. It is reasonable because methods 3 and 4 put more emphasis on the simplicity of tree structure. Among the four methods, method 1 gives the longest tree in all the cases. This is consistent with the discussion on merging and dividing overlay trees that emphasis on hierarchical similarity yields longer links.

Excessive sensitivity to hierarchical inconsistency found in merging overlay trees is not a desirable property of visual representation. We thus focus on dividing overlay tree in discussing the result of analysis in detail hereafter.

Figure 8 and Figure 9 show the dividing overlay trees for Doshusei systems and company branches, respectively. Each dividing overlay operation is represented by two tessellations, their predecessor and their connecting links. A shorter link indicates the structural difference while the longer link is the summation of structural and granularity differences.

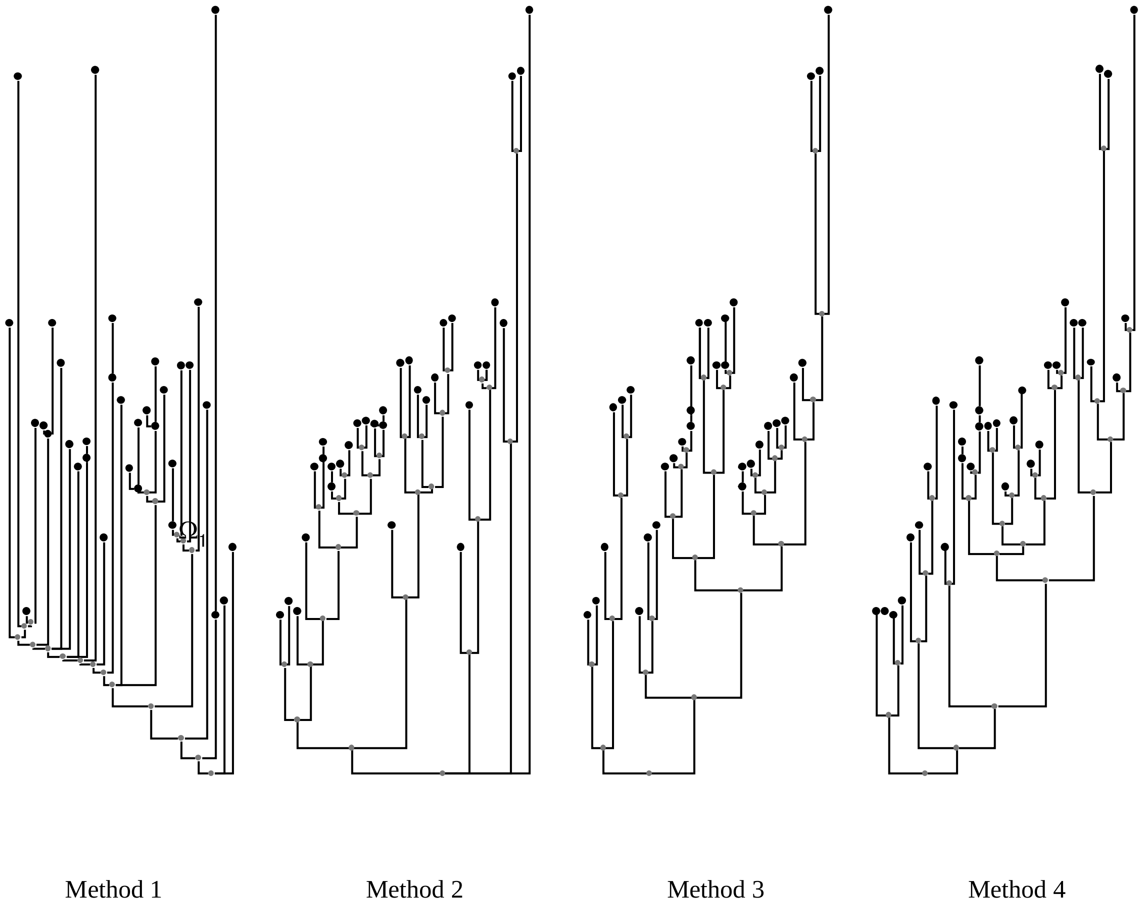

Figure 8 Dividing overlay trees for Doshusei systems. Black circles indicate Doshusei systems while small gray circles indicate predecessors generated by dividing overlay operation.

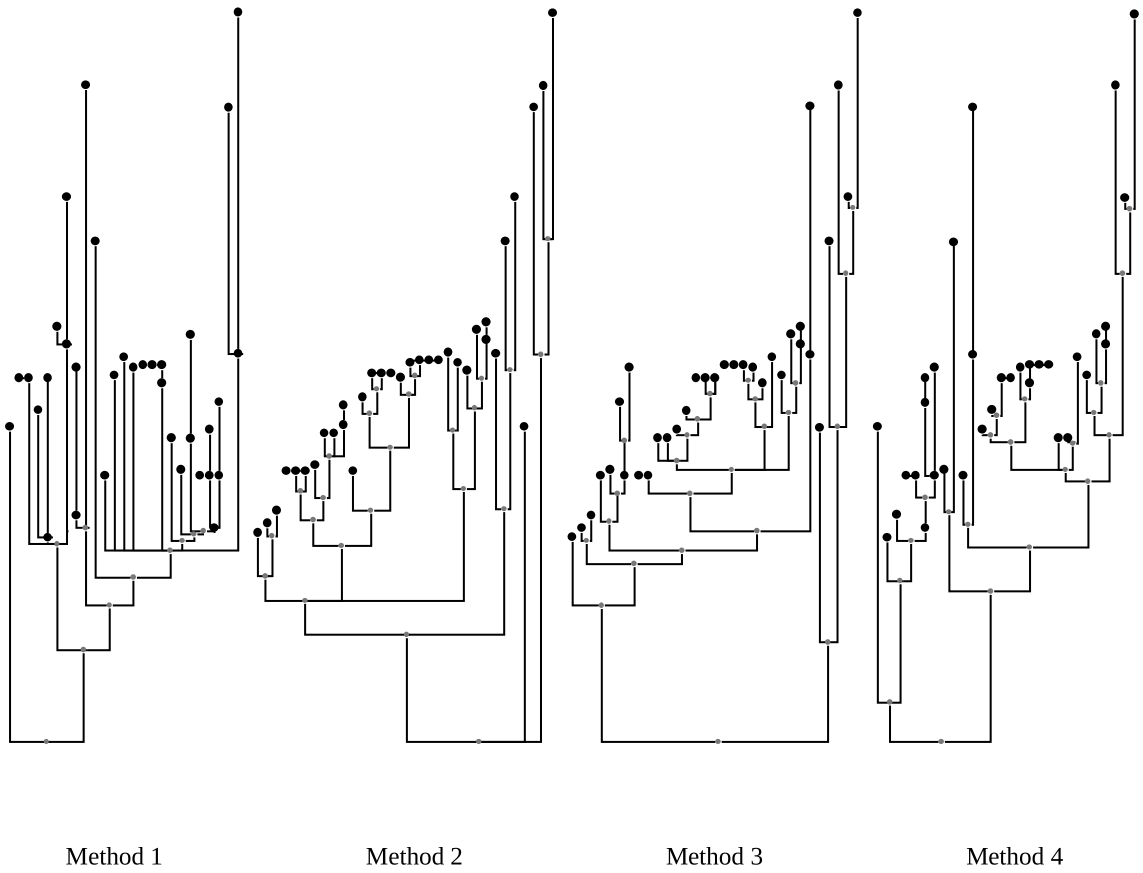

Figure 9 Dividing overlay trees for company branches. Black circles indicate Doshusei systems while small gray circles indicate predecessors generated by dividing overlay operation.

As mentioned earlier, tessellations connected with are considered similar at least in structure. Tessellations connected with only one vertical link are globally hierarchical. A set of more than two similar tessellations form a cluster connected by a tree.

The figures look consistent with the summary statistics in Table 2. Method 1 clearly generates longer links than the others. Tessellations are overlaid at low granularity level so that the links are inevitably long. On the other hand, trees generated by methods 3 and 4 have shorter links. Method 3 tends more to combine tessellations of similar granularity, while method 4 often combines tessellations of different granularity. The figures also shows that the tessellation tree heavily depends on the method by which it is created. Trees look different within the same set of tessellations while trees generated by the same method look similar even if they are based on different set of tessellations.

Among the four methods, method 1 does not seem helpful to understand the relationship among tessellations. Since it puts too much emphasis on structural similarity between tessellations, generated trees are comb-like revealing only that all the tessellations are different at a similar degree. Method 2, on the other hand, shows quite different trees, which is not expected from Table 2. Links connecting original tessellations are very short, which implies that method 2 tends to combine only tessellations of similar granularity. The average length of links is longer than that of methods 3 and 4 because the links connecting tessellations generated by combination operation are very long in method 2. Though the trees generated by methods 3 and 4 look similar to that by method 2, they consist of shorter links as seen in Table 2. This is because these methods aim to keep the tree as compact and simple as possible. Both methods evaluate the granularity of tessellations, and method 4 also considers the similarity in structure and granularity between tessellations. Difference between methods 3 and 4 can be confirmed in Figure 9 where method 4 connects tessellations of different granularity more than method 3.

Among the four methods, method 4 seems relatively effective for understanding the relationship among tessellations. It generates compact and simple trees useful for visual analysis, taking into account various aspects of similarity in tessellations. We thus examine the trees generated by method 4 in more detail.

Figure 10 shows the tessellation trees of Doshusei systems and company branches with identification numbers. The figure also shows "natural" clusters of tessellations, that is, tessellations similar in structure, granularity, or both of them. As seen in Figure 10, each cluster consists tessellations of two or more different granularity. This indicates that method 4 considers similarity among tessellations in both structure and granularity.

Let us focus on the cluster on the right end of tessellation tree of Doshusei system. Configuration of tessellations in this cluster is shown in Figure 11. As seen the figures, this cluster consists of tessellations of two different granularity levels. Tessellations {21, 22, 26} consist of 4 to 6 regions while {1, 8, 12, 17, 19} has 9-12 regions. This cluster can be further classified into three subgroups: {1, 8, 22} and {12, 21, 26} and {17, 19}. Each group shares a similar structure though the granularity is different within the group. For instance, tessellations 1 and 8 are lower level tessellations of 22, and they form a subgroup of tessellations that shares a similar structure. Tessellation 22 is characterized by small regions in the center of Japan. This property is inherited by tessellations 1 and 8, though they consist of smaller regions over the whole area. In subgroup {12, 21, 26}, tessellations 21 and 26 look quite similar, and tessellation 12 can be obtained by further dividing either 21 or 26. Tessellations 17 and 19 are similar to 1, 8, and 12 in granularity, but different in structure especially in the center of Japan.

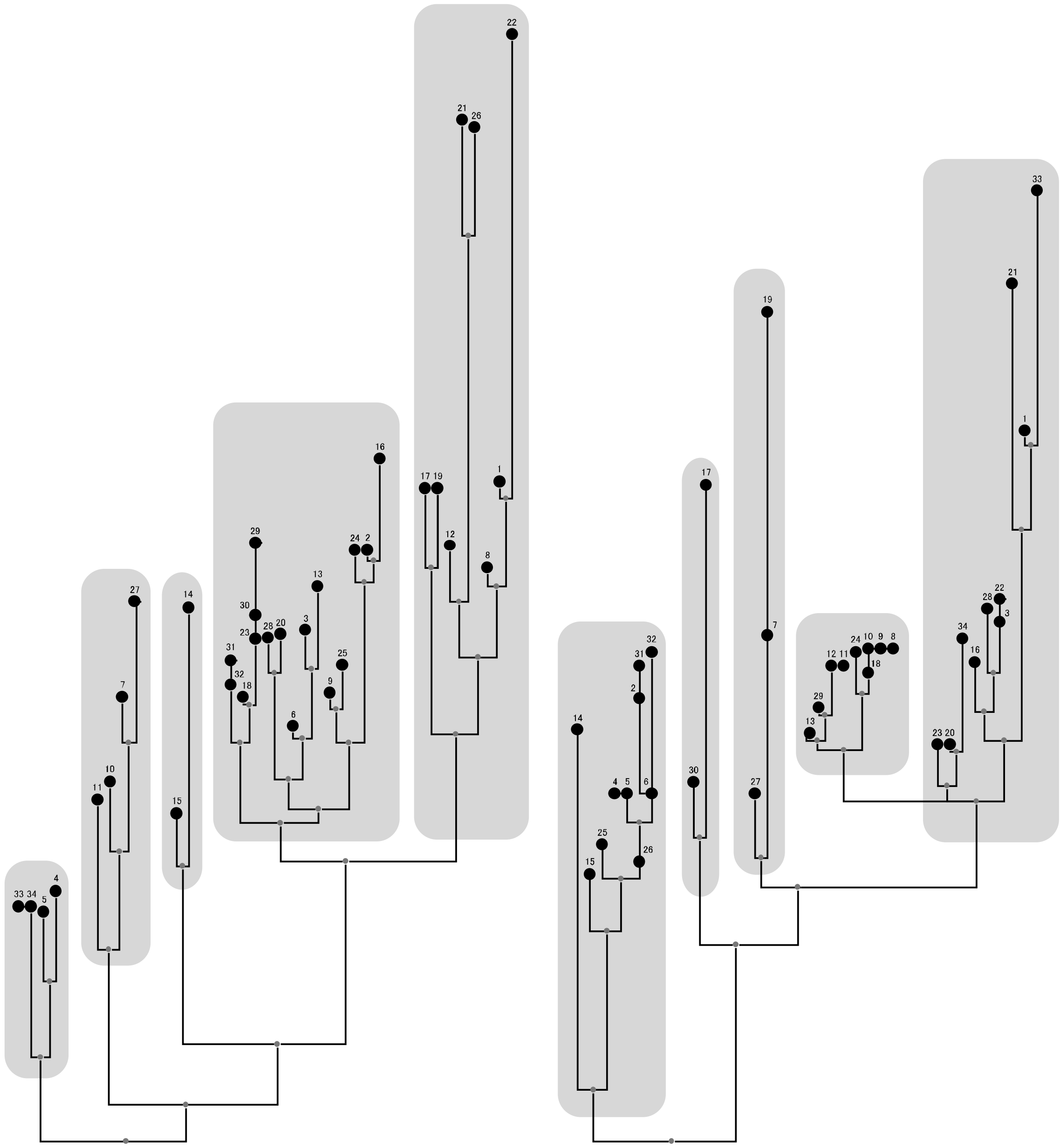

Figure 10 Tessellation trees of Doshusei systems (left) and company branches (right). Black circles with identification numbers indicate original tessellations while small gray circles indicate tessellations generated by dividing overlay operation. Gray shaded areas are clusters of tessellations considered similar in some aspects.

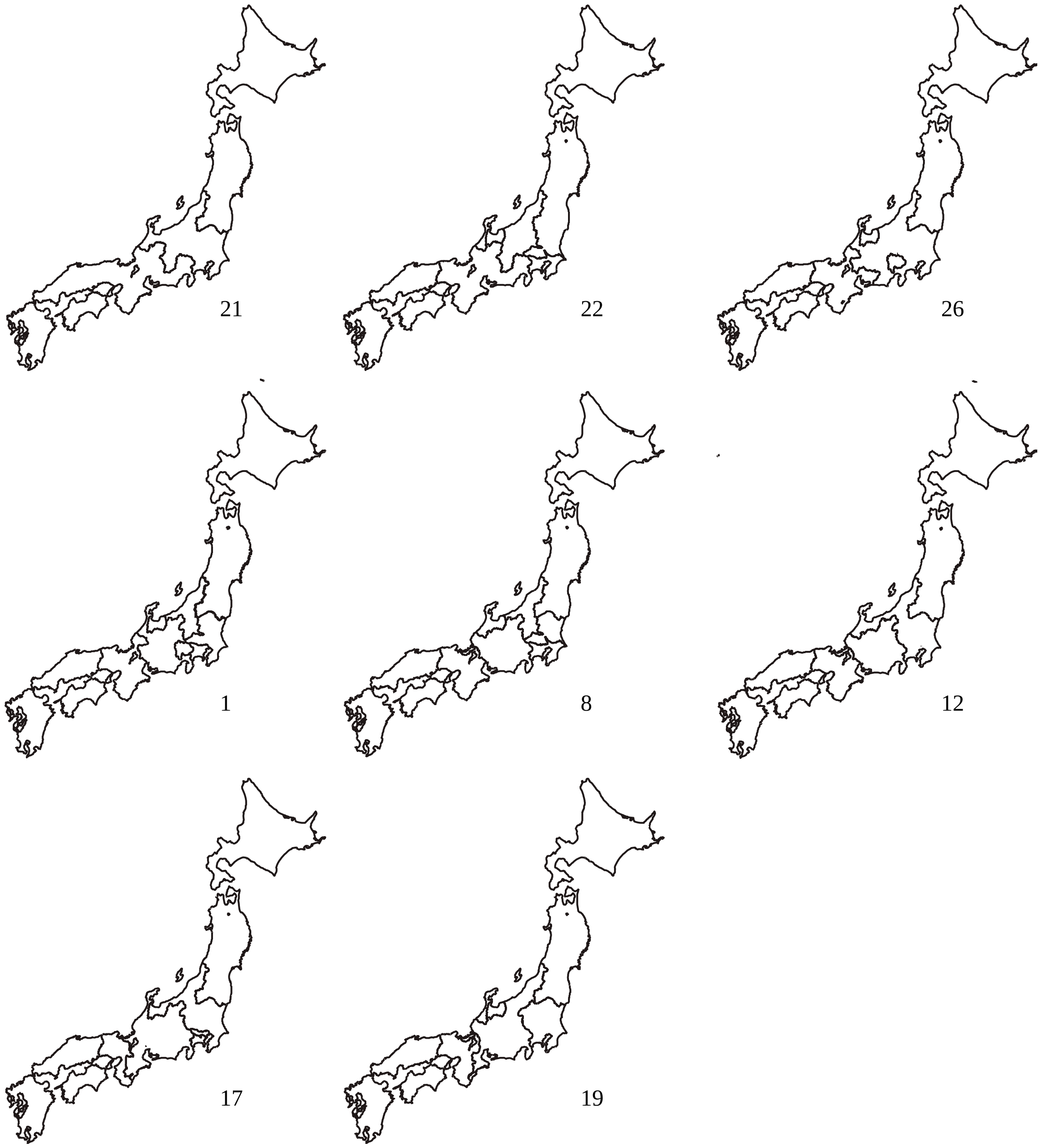

Figure 11 Examples of Doshusei systems.

4. Conclusion

This paper has proposed a new method for analyzing spatial tessellations. It evaluates the similarity between tessellations in three aspects: global hierarchy, local hierarchy, and granularity. They are represented by three measures, whose calculation procedure is also provided. The relationship among tessellations is represented by a overlay diagram, and visualized by overlay trees. The method was applied to the analysis of Doshusei systems and branch areas of private companies. This application reveals the properties of the method and its measures as well as empirical findings.

Focusing on global and local hierarchies and granularity, the method successfully represents the relationship among spatial tessellations as a overlay diagram. The three properties are represented as $D_{GS}(\Omega_i, \Omega_j)$, $D_{LS}(\Omega_i, \Omega_j)$, and $D_G(\Omega_i, \Omega_j)$ and they are fully independent as shown in Figure 5. It enables us to specify an aspect in which we evaluate the similarity among tessellations. It also helps us to interpret overlay trees because these measures are represented as the vertical distances between tessellations.

We finally discuss some limitations of the paper for future research. First, as mentioned above, this paper focuses on three properties of spatial tessellations in evaluation of similarity between tessellations. This makes it easy to understand the meaning of quantitative measures and to interpret overlay trees. However, it also limits our scope of analysis to only the basic properties of tessellations. For instance, granularity is evaluated only by its average degree over the whole region. It is desirable to consider its spatial variation and higher moments in comparing tessellations. Other measures such as perimeter and shape index are also useful to describe geometrical properties of tessellations. Second, this paper compares tessellations at every pair of points equally over the whole region. One extension is to put more weight on closer pairs of points to consider the correlation between spatial and attribute similarities explicitly. It is also possible to define a weight matrix as a general function of point pairs; it permits us to take spatial heterogeneity into account in evaluating similarity among tessellations. Third, the method was applied to only two sets of spatial tessellations, both of which are defined based on the same administrative units in the same region. It is clearly necessary to apply the method to a wider variety of tessellations. Examples include tessellations representing other types of spatial phenomena, tessellations in other places, and tessellations of the same phenomena at different times. It would be useful to apply the method to tessellations representing natural phenomena because, unlike the tessellations discussed in this paper, they do not usually share common boundaries.

Acknowledgement

The authors would like to thank A Okabe, Y. Asami, and T. Ishikawa for their valuable comments.

References

Anderson, I. (2002) *Combinatorics of Finite Sets.* Dover.

Congalton, R. G. and Mead, R. A. (1983). A Quantitative Method to Test for Consistency and Correctness in Photointerpretation, *Photogrammetric Engineering and Remote Sensing* 49, 69-74.

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

Fritz, S. and See, L. (2005). Comparison of Land Cover Maps Using Fuzzy Agreement, *International Journal of Geographical Information Science*, 19(7), 787-807.

Kraak, M.-J. (2003) *Cartography: Visualization of Geospatial Data*. Pearson Education.

Nielson, G. M., Hagen, H., Muller, H. (1997) *Scientific Visualization: Overviews, Methodologies, and Techniques*. IEEE.

Monserud, R. A. and Leemans, R. (1992) Comparing global vegetatic maps with Kappa statistic, *Ecological Modelling*, **62**, 275-293.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S.-N. (2000): *Spatial Tessellations - Concepts and Applications of Voronoi Diagrams* . John Wiley & Sons.

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

Pontius Jr., R. G. (2000). Quantification Error versus Location Error in Comparison of Categorical Maps, *Photogrammetric Engineering & Remote Sensing*, 66, 1011-1016.

Pontius Jr, R. G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions, *Photogrammetric Engineering & Remote Sensing*, 68(10), 1041-1049.

Rosenfield, G. H. and Fitzpatrick-Lins, K. (1986). A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2), 223-227.

Sadahiro, Y. (2002): An Exploratory Method for Analyzing a Spatial Tessellation in Relation to a Set of Other Spatial Tessellations, *Environment and Planning A*, 34(6), 1037-1058.

Slocum, T. A., McMaster, R. B., Kessler, F. C. and Howard, H. H. (2004) *Thematic Cartography and Geographic Visualization*. Prentice-Hall.

Ware (2004) *Information Visualization: Perception for Design*. Morgan Kaufmann.

Wright, H. (2006) *Introduction to Scientific Visualization*. Springer.