RIETI Discussion Paper Series 22-E-092

Similarity and Connectivity of Industrial Networks of Japanese Prefecture Based on Firm-level Data

GOTO, Hiromitsu
Kanazawa Gakuin University

SOUMA, Wataru
Rissho University

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 22-E-092

September 2022

Similarity and Connectivity of Industrial Networks of Japanese Prefecture based on Firm-level Data *

Hiromitsu GOTO
Faculty of Economic Informatics, Kanazawa Gakuin University
Wataru SOUMA
Faculty of Data Science, Rissho University

Abstract

Understanding the industrial structure of inner and inter prefectures is crucial for policymakers to make economic policies according to evidence. To address this issue, using the dataset of financial statements and connections for one million firms in Japan collected by Tokyo Shoko Research Inc., we construct a multiplex supply network with 47 layers equivalent to prefectures. Applying clustering analysis based on the Jensen-Shannon distance between networks and the community detection techniques known as the Infomap method to this multiplex supply network, we clarify industrial structural similarities and differences for each prefecture. Finally, we compare the results for multiplex supply networks and the well-known facts for each prefecture's Input-Output table to evaluate our result's validity and complementarity. Our findings of this study are as follows. First, from 2011 to 2018, the industrial networks of 47 prefectures can be classified into three structural patterns by a degree of urbanization. Second, the hierarchical community structure can be observed using firm-level data. However, this hierarchical community structure cannot be seen in the conventional Input-Output table dataset. Therefore, our findings suggest a new classification approach for prefectures based on similarities in the industrial structure and contribute to a better insight into the geographical characteristics of each region's industrial structure.

Keywords: Production network, Multi-layer network, Structural reducibility, Clustering Analysis, Hierarchical community structure, Input-output analysis

JEL classification: D57, D85, L14, L16

*This study is conducted as a part of the project "Macro-Economy under COVID-19 influence: Data-intensive analysis and the road to recovery" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This research was also supported by grant-in-aid for scientific research (KAKENHI) by JSPS Grant Numbers 20H02391. The author is grateful for helpful comments and suggestions by Hideaki Aoyama (Kyoto Univ.), Yuichi Ikeda (Kyoto Univ.), Yoshi Fujiwara (Univ. Hyogo) and Discussion Paper seminar participants at RIETI.

I. Introduction

Understanding the industrial structure of inner and inter prefectures is crucial for policymakers to make economic policies based on pieces of evidence. Traditionally, the industrial structure has been explored based on Input-Output (IO) tables (Leontief, 1986). Japan has 47 prefectures, and the IO tables for all prefectures have been available since 1990, with revisions every five years. Although the national IO tables have hierarchical sectoral classifications, this classification is not standardized for each prefecture or regional IO table. Furthermore, such classification of firms by industry and geographical region may be too accumulated.

To address this issue, recent studies have begun to concentrate on analyzing firm-level networks based on the comprehensive data of supply chain (Fujiwara and Aoyama, 2010; Atalay et al., 2011; Acemoglu et al., 2012; Luo et al., 2012). The mesoscopic structure of firm-level supply-chain networks can be expected to become an alternative approach to characterize not only industrial structures, including IO tables, but also the nation's economy. Chakraborty et al. (2018) have demonstrated that the firm-level supply-chain networks in Japan have a hierarchical community structure, including overexpressions of industrial and regional components, and suggested the need to replace the conventional industrial classification scheme with a new one based on the real transactions.

In this study, using the dataset of transactions for one million firms in Japan obtained by Tokyo Shoko Research (TSR) Inc., we constructed the Japanese economic system of transactions within and between industries and within and between prefectures as a multiplex supply network with 47 layers equivalent to prefectures. Additionally, we clarified industrial structural similarities and connectivity for prefectures by applying clustering analysis and community detection techniques. Although many studies have used cluster analysis to assess and classify similarities in industrial structure for prefectural and district township-level industry statistics, including IO tables (Irie, 2017; Kondo, 2020), the similarities in inter-industry connections and their mesoscopic structures have been neglected. To overcome this problem regarding network science, we use the methodology to reduce the number of layers to a minimum while maximizing the differentiability between the multi-layer network and the equivalent accumulated one by examining the mesoscale structural similarity of networks, proposed by De Domenico et al. (2015). Moreover, to examine connectivity between prefectures and industrial sectors, we used the map equation method (Rosvall and Bergstrom, 2008), known as Infomap, which is one of the best performing community detection techniques (Lancichinetti and Fortunato, 2009). We compare the findings from both firm-level data and IO table data to evaluate the validity and complementarity of our result.

The rest of this study is organized as follows. First, we explain the details of data in Section II, and the methodology in Section III. Section IV provides some results and discussion regarding similarity and connectivity between the industrial structures of the Japanese prefecture. Finally, the conclusion and research perspective will be presented in Section V.

II. Data

The data for the Japanese supply-chain network are based on a survey conducted by TSR Inc., one of the prominent credit research agencies in Tokyo, and was supplied to us through the Research Institute of Economy, Trade, and Industry (RIETI). Additionally, the data are compiled each September from 2011 to 2018. The survey inquires about firms whom the top 24 suppliers and customers for each, and this form of data collection can expect to avoid including data on one-time trades. Although the replies from large firms with several suppliers and customers become incomplete, these data could be supplemented with data on the other side of the trade. Therefore, we assume in this study that our data provide a good approximation of the true, complete picture of the Japanese economy by combining all submissions from both sides of the trade, which covers approximately one million firms and several million supplier-customer relationships. Furthermore, we can use the information of firms, including firm size measured as sales, profit, and the number of employees, and the classification into industrial sectors and geographical location. For our study, let us limit our investigation to only two characteristics of each firm: the industrial sector and the geographical location of headquarters. The industrial sectors are classified hierarchically into 20 divisions, 99 major groups, 436 minor groups, and more by TSR. To discuss the similarity between the industrial structures of the Prefecture, we concentrated on the industrial sector classified as 99 significant groups. Additionally, the geographical location is converted into a level of 47 prefectures. Because the 99 industry classifications include unclassifiable categories, we extract only the active firms, which can be identified as the geographical location and classified as 98 industry classifications. The numbers of companies and transactions are listed in Table 1.

This study compares the findings from TSR data and the IO table. Although the national IO tables have sectoral classifications of 13, 37, 108, 190, and 397, this classification is not standardized for each prefecture. To address this issue, the RIETI project has created and published the inter-prefecture

IO table on 2011[1], recently. We adopted this inter-prefecture IO table and compared the findings with the TSR2011 data. We should note that the 31 sectoral classifications are available because of the combining process.

Figure 1 depicts the fractions of internal and external transactions to the total number or amounts of transactions for each prefecture. Additionally, external transactions are distinguished between incoming and outgoing flows. As depicted in Figure 1 (left), the ratio of the number of transactions between prefectures in the metropolitan prefectures is comparatively small: particularly, Tokyo, Saitama, Chiba, and Kanagawa in the Kanto region; and Shiga, Kyoto, Osaka, Hyogo, and Nara in the Kansai region. However, the ratio of the number of transactions within prefectures to the total number of transactions in Hokkaido and Okinawa is high. Nevertheless, these are not apparent from the IO table data shown in Figure 1 (right). Note that this distinction may cause the TSR data to fail to reflect the industrial characteristics of transaction values. In this study, we assume that the TSR data can reproduce the actual industrial structure of prefectures through the number of transactions at the firm level and detailed industry classification.

III. Methodology

In this section, we discuss multi-layer network analysis to examine the Japanese prefecture's similarity and connectivity of industrial networks.

A. *Multi-layer Network Representation*

We constructed the multi-layer network reproducing the Japanese economic system of transactions within and between industries and within and between prefectures.

We define a multi-layer network as a pair $\mathcal{M} = (\mathcal{G}, \mathcal{C})$, where $\mathcal{G} = \{G_\alpha; \alpha \in \{1, \cdots, M\}\}$ of the family of graphs $G_\alpha = (V_\alpha, E_\alpha)$, where the set of nodes of layer G_α is denoted as V_α, and M is the number of layers. $\mathcal{C} = \{E_{\alpha\beta} \subseteq V_\alpha \times V_\beta; \alpha, \beta \in \{1, \cdots, M\}, \alpha \neq \beta\}$ is a set of interconnections between the nodes of different layers G_α and G_β with $\alpha \neq \beta$.

Given a layer G_α corresponding to one of the $M = 47$ prefectures in Japan, the N_α nodes corresponding to the industrial sectors are denoted by $V_\alpha = \{s_1^\alpha, \cdots, s_{N_\alpha}^\alpha\}$, and intra-layer neighboring matrix of each layer G_α is denoted by $A^{[\alpha]} = \left(a_{ij}^\alpha\right)$, where the element a_{ij}^α corresponds to the number/amounts of transactions from the industrial sector s_i to s_j in the prefecture α. Alternatively, the cross-layer adjacency matrix $E_{\alpha\beta}$ is the matrix

[1]. https://www.rieti.go.jp/jp/database/r-io2011/index.html (in Japanese)

$A^{[\alpha,\beta]} = \left(a_{ij}^{\alpha\beta}\right)$, which represents transactions between prefectures. The element $a_{ij}^{\alpha\beta}$ corresponds to the number/amounts of transactions from the industrial sector s_i of prefecture α to the industrial sector s_j in the prefecture β.

B. *Similarity measurement and structural reducibility*

Aggregating interactions of a similar nature into single layers can engender different multilayer networks. We use the method to reduce the number of layers to a minimum while maximizing the distinguishability between the multilayer network and the corresponding aggregated one by investigating the mesoscale structural similarity of networks, proposed by De Domenico et al. (2015). At each step, the most similar layers are aggregated, forming a dendrogram describing the further aggregation of layers. Finally, the best aggregation procedure is discovered by cutting the dendrogram at a level equivalent to the best score function, characterizing the best aggregation. Therefore, this approach can provide us with hierarchical clustering of layers and structural reducibility for the multilayer network. We should note that cross-prefecture transactions are neglected to evaluate the similarity between industry transactions within each prefecture.

First, the similarities between the layers are computed using the quantum Jensen-Shannon distance to construct the dendrogram explaining the aggregation procedure. Considering two density matrices $\boldsymbol{\rho}^{[\alpha]}$ and $\boldsymbol{\rho}^{[\beta]}$, we can quantify to what extent $\boldsymbol{\rho}^{[\alpha]}$ is different from $\boldsymbol{\rho}^{[\beta]}$, by the means of the Kullback-Liebler divergence,

$$\mathcal{D}_{\mathrm{KL}}(\boldsymbol{\rho}^{[\alpha]}||\boldsymbol{\rho}^{[\beta]}) = \mathrm{Tr}\left[\boldsymbol{\rho}^{[\alpha]}\left(\log_2\left(\boldsymbol{\rho}^{[\alpha]}\right) - \log_2\left(\boldsymbol{\rho}^{[\beta]}\right)\right)\right] \ . \tag{1}$$

As the symmetrized Kullback-Liebler divergence, the Jensen-Shannon divergence between two density matrices $\boldsymbol{\rho}^{[\alpha]}$ and $\boldsymbol{\rho}^{[\beta]}$ is defined as

$$\begin{aligned}\mathcal{D}_{\mathrm{JS}}(\boldsymbol{\rho}^{[\alpha]}||\boldsymbol{\rho}^{[\beta]}) &= \frac{1}{2}\mathcal{D}_{\mathrm{KL}}(\boldsymbol{\mu}^{[\alpha,\beta]}||\boldsymbol{\rho}^{[\alpha]}) + \frac{1}{2}\mathcal{D}_{\mathrm{KL}}(\boldsymbol{\mu}^{[\alpha,\beta]}||\boldsymbol{\rho}^{[\beta]}) \\ &= h\left(\boldsymbol{\mu}^{[\alpha,\beta]}\right) - \frac{1}{2}\left[h\left(\boldsymbol{\rho}^{[\alpha]}\right) + h\left(\boldsymbol{\rho}^{[\beta]}\right)\right]\end{aligned} \tag{2}$$

where $\boldsymbol{\mu}^{[\alpha,\beta]}$ is the average of two density matrices, $\boldsymbol{\mu}^{[\alpha,\beta]} = \frac{1}{2}(\boldsymbol{\rho}^{[\alpha]} + \boldsymbol{\rho}^{[\beta]})$, and $h(\boldsymbol{\rho}^{[\alpha]})$ is the von Neumann entropy of the density matrix $\boldsymbol{\rho}^{[\alpha]}$. For the set $\boldsymbol{\mathcal{A}} = \{A^{[1]}, A^{[2]}, \cdots, A^{[M]}\}$ of intra-layer adjacency matrices, we can compute the von Neumann entropy $h_{A^{[\alpha]}}$ of layer α, where it can be written in terms of the set $\left\{\lambda_1^{[\alpha]}, \lambda_2^{[\alpha]}, \cdots, \lambda_N^{[\alpha]}\right\}$ of the eigenvalues of rescaled Laplacian matrix associated to the adjacency matrix $A^{[\alpha]}$ of layer α,

$$h_{A^{[\alpha]}} = -\sum_{i=1}^{N} \lambda_i^{[\alpha]} \log_2 \left(\lambda_i^{[\alpha]}\right) . \tag{3}$$

The Laplacian spectrum well includes rich information about the multiscale structure of undirected graphs, e.g., the second smallest eigenvalue of the Laplacian matrix is known as the algebraic connectivity as proposed by Fiedler (1973). Moreover, the relationship between the spectrum and community structure has been empirically validated by Newman (2006), and the definition of the Laplacian matrix for directed graphs and these properties have been discussed by Chung (2005). Hence, we assume that the Jensen-Shannon divergence, built upon the Laplacian spectrum, is suitable for pattern recognition between networks concerning mesoscale structural similarity. After computing the Jensen-Shannon distance matrix between all pairs of layers, we perform hierarchical clustering of layers.

Subsequently, we aggregate some of the original layers $\mathcal{A} = \{A^{[1]}, A^{[2]}, \cdots, A^{[M]}\}$ and obtain a reduced multilayer network $\mathcal{C} = \{C^{[1]}, C^{[2]}, \cdots, C^{[X]}\}$ with $X \leq M$ layers. The von Neumann entropy $H(\mathcal{A})$ of a multilayer network is computed as the sum of von Neumann entropies of its M layers, that is, $H(\mathcal{A}) = \sum_{\alpha=1}^{M} h_{A^{[\alpha]}}$. Furthermore, we quantify the distinguishability between the multilayer network $\mathcal{C}$ and the equivalent aggregated graph $A = A^{[1]} + A^{[2]} + \cdots + A^{[M]}$ through the relative entropy,

$$q(\mathcal{C}) = 1 - \frac{\bar{H}(\mathcal{C})}{h_A} , \tag{4}$$

where $\bar{H}(\mathcal{C})$ is the entropy per layer of the multi-layer network $\mathcal{C}$,

$$\bar{H}(\mathcal{C}) = \frac{H(\mathcal{C})}{X} = \frac{\sum_{\alpha=1}^{X} h_{C^{[\alpha]}}}{X} . \tag{5}$$

The larger relative entropy $q(\mathcal{C})$ conforms to the more distinguishable multilayer network $\mathcal{C}$ from the corresponding aggregated graph A. Therefore, the relative entropy $q(\bullet)$ is used as the quality function for the resulting partition.

Finally, we selected the partition that maximizes the relative entropy, $q_{\max}(\bullet)$. The reducibility of a multilayer network $\mathcal{A}$ of M layers can be characterized as

$$\chi(\mathcal{A}) = \frac{M - M_{\text{opt}}}{M - 1} , \tag{6}$$

where M_{opt} is the number of layers consistent with the maximum value of relative entropy $q_{\max}(\bullet)$.

C. *Hierarchical community structure*

The Infomap method (Rosvall and Bergstrom, 2008), one of the most thriving community detection algorithms on single networks, is a flow-based and information-theoretic approach to identify an efficient code for minimizing the length of the description of the random walk for generating a module partition $\mathcal{M}$ to divide n nodes into m communities. Subsequently, the average single-step description length is defined as

$$L(\mathcal{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p_{i\circlearrowright} H(\mathcal{P}_i) \ . \tag{7}$$

The first term arises from the movements of the random walker across modules, where $q_{\curvearrowright}$ is the probability that the random walker switches communities, and $H(\mathcal{Q})$ depicts the average description length of the community index codewords given by the Shannon entropy. The second term arises from the intra-community movement of the random walker, where the weight $p_{i\circlearrowright}$ represents the fraction of the movements within the community, and $H(\mathcal{P}_i)$ represents the entropy of the intra-community movement. Furthermore, this approach has been expanded to a hierarchical map equation (Rosvall and Bergstrom, 2011) that decomposes a network into communities and subcommunities.

In this study, the same industrial sector belonging to various prefectures is considered to have different nodes, thus creating a single-layer network and applying this methodology. We discovered the hierarchical communities using the multicoding Infomap method, and we use the "Level" index to denote the hierarchy of communities; communities at the 2nd level represent subcommunities at the 1st level. To define the hierarchical communities and to evaluate the connectivity between prefectures, we use the attribution with the prefecture consistent with the layer.

IV. Results and Discussions

We applied the method explained in Section III for the industrial networks constructed by the firm-level (TSR) data from 2011 to 2018 and for the IO table data from 2011 to investigate the validity and complementary of the finding. Notably, the number of nodes N for each layer corresponds to the number of industrial sectors; $N = 98$ for the TSR data and $N = 31$ for the IO table data. Although the findings from the TSR data do not reflect the industrial characteristics of the transaction value, this study anticipates reproducing the actual industrial structure of prefectures through the number of transactions per firm and detailed industry classification.

A. *Similarity between industrial structures of inner-prefecture*

The results of structural reducibility according to the TSR data and IO table data are presented in Table 2. We found that the multilayer networks of prefecture industries based on the TSR data cannot be reduced except for the TSR2018 case. However, the prefecture layers of the industrial network based on the IO table data can be reduced to only two layers, corresponding to Tokyo and the others. This result implies that while firm-level data is suitable for representing the mesoscopic characteristics of industries in each prefecture, data from the IO table is unsuitable owing to over-aggregation.

We demonstrated the dendrograms resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$ and the color-coded map of Japan based on the results of hierarchical clustering and structural reducibility in Figure 2 to Figure 9. The upper and lower parts of Figure 2 depict the result based on TSR2011 data and IO table 2011, respectively. In the IO table 2011 case, because the 1st-level clustering is equivalent to the maximum of distinguishable cutting, the result is reflected in the coloring of the map. Because the prefecture layers on the TSR case, except for 2018, cannot be reduced, we have shown the color-coded map based on the result of the 2nd-level cluster. Figure 9 depicts the result based on the TSR2018 data, and the three reducible pairs of prefectures belonging to the cluster colored in red are presented in different colors: Tochigi and Fukuoka; Gunma and Okayama; and Niigata and Hyogo.

We observed the three structural patterns at the 2nd-level clusters each year from 2011 to 2018: a) prefectures including three major metropolitan areas, including Tokyo, Osaka, and Aichi; b) prefectures surrounding the main metropolitan areas; and c) the other prefectures far from the central cities, including the Hokkaido and Kyushu region. We should note that even though we ignored connections between prefectures, the clusters that reflected geographic adjacencies are observed. These results suggest that the inner prefecture's industrial mesoscopic structure can be classified by a degree of urbanization.

B. *Connectivity between prefectures in community structure*

We detected the hierarchical communities using the multi-coding Infomap method. Table 3 and Table 4 present the statistics for the hierarchical communities found using the multicoding Infomap method for the industrial networks based on the TSR data and IO table data, respectively. "# of com." is the number of all communities, "# of irr.com." is the number of fundamental communities, which are communities that do not have any subcommunities. "# of nodes" represents the number of nodes, including industries, in irreducible communities. We determined the number of nodes "# of nodes" in irreducible communities at each level. We discovered that most of the nodes belong to the 2nd-level communities. Therefore, we limit our discussion of subcommunities' properties to those of the 2nd level.

From Table 5 to Table 13, we demonstrate the features of the 1st- and 2nd-level communities, including more than 150 nodes in the TSR data, and over 50 nodes in the case of the IO table data, respectively. Parentheses denote the percentage of prefectures. We have outlined only communities that contain two or more prefectures with a share of 5% or more. Moreover, from Figure 10 to Figure 18, we indicate the results for communities across prefectures by color coding them on the map of Japan. The color-coded maps of Japan are based on the findings of hierarchical communities at the 1st (left) and the 2nd (right) levels. The label of colors correspond to the index corresponding tables, from Table 5 to Table 13. However, when more than 70% of nodes belonging to one community form one prefecture, the prefectures are colored by dark gray. These results frequently show that industrial networks are geographically clustered. However, as is clear from the 2011 comparison in Figure 10 and Figure 11, the same prefectural communities (Shikoku, Kyushu, and others) are observed in the 1st-level communities for both the TSR data and IO table data, whereas the results for the 2nd level are different. Using the IO table data for 2011, communities with various prefectures could not be found except for Kanto (Tokyo, Saitama, and Chiba) and Kagoshima-Okinawa. Although this study cannot conclude whether this difference is caused by the overly coarse-grained industry classifications in the IO tables or by differences in network weights (number of transactions/amounts), the results from firm-level data reveal communities with robust industrial connectivity among prefectures that are not evident in the results from the IO tables.

We discovered that there are pairs of prefectures belonging to the same community at the 2nd level from 2011 to 2018: Tokyo, Ibaraki, Kanagawa, Chiba, Saitama, and Gunma; Ishikawa and Toyama; Gifu and Aichi; Hiroshima and Okayama; and Shimane and Tottori. However, we observed that the Kanto community extended into Tohoku and Hokkaido between 2012 and 2015, as depicted in Figure 12 to Figure 15, and split again into the Kanto and Tohoku communities in 2016, as shown in Figure 16. These results might cause a temporary decrease in the number of firm-level transactions in the Tohoku region due to the Great East Japan Earthquake in 2011 and suggest that firm-level data can help visualize changes in the time-series structure of industries, which is challenging using an IO table.

V. Conclusion

In this study, we analyzed the industrial multiplex network with 47 layers corresponding to prefectures using the firm-level data of transactions for one million firms in Japan. Moreover, we proposed new classification techniques for prefectures based on similarities in industrial structure to give a better understanding of the geographic characteristics of each region's industrial structure. We applied hierarchical clustering analysis based on the quantum Jensen-Shannon distance, which showed the mesoscale structural similarity of industrial structures of inner-prefecture. Applying the approach proposed by De Domenico et al. (2015), we determined the hierarchy in which prefectures belonging to the same cluster have the highest identifiability across mesoscale industrial structures using von Neumann entropy to assess the significance of clustering. Because of the over-aggregation of industrial sectors in the IO table data, we could distinguish only Tokyo and the other prefectures in terms of the mesoscale feature of industrial structures. However, we found the three structural patterns at the 2nd-level clusters by a degree of urbanization: prefectures, including three primary metropolitan areas; prefectures surrounding the main metropolitan areas; and the other prefectures far from the central cities. We also used the hierarchical community detection known as the Infomap method (Rosvall and Bergstrom, 2008) to explore connectivity between prefectures and industrial sectors. Although we could not find communities with various prefectures except for Kanto (Tokyo, Saitama, and Chiba) and Kagoshima-Okinawa based on the IO table data, the analysis of the firm-level data showed pairs of prefectures belonging to the same community at the 2nd level from 2011 to 2018: Tokyo, Ibaraki, Kanagawa, Chiba, Saitama, and Gunma; Ishikawa and Toyama; Gifu and Aichi; Hiroshima and Okayama; and Shimane and Tottori.

In this study, we conducted a survey of industrial structure at the prefectural level to compare it with inter-prefectural IO tables and contribute to policy making at the prefectural level. Concerning policy implications, similar economic policies may be effective among prefectures with high similarity, and when considering economic policy within a prefecture, a group of prefectures with high connectivity should consider more economic spillover effects among prefectures with high connectivity. As studied in spatial economics, these regional cluster structures' origins are expected to be investigated in the future by utilizing microscopic data.

However, there may be some potential limitations in this study. We expected that the firm-level data reproduce the actual industrial structure of prefectures through the number of transactions and detailed industry classification in this study because of data availability. Although the firm-level data collected by TSR include the address of the headquarters of each firm, it should be the address of the place of business as in the IO table data. Future study should consider the weight of the network reflecting the industrial and geographical characteristics of transaction values to assess the validity of the results in more detail. Therefore, we emphasize again that IO table data plays an essential role in investigating the actual industrial structure of Japan prefectures. The firm-level analysis helps us understand the more microscopic features of industrial structure at shorter time intervals.

Acknowledgments

This study is conducted as a part of the Project "Macro-Economy under COVID-19 influence: Data-intensive analysis and the road to recovery" undertaken at the Research Institute of Economy, Trade, and Industry (RIETI). This research was supported by a grant-in-aid for scientific research (KAKENHI) by JSPS Grant Numbers 20H02391. The author is grateful for helpful comments and suggestions by Hideaki Aoyama (Kyoto Univ.), Yuichi Ikeda (Kyoto Univ.), Yoshi Fujiwara (Univ. Hyogo), and Discussion Paper seminar participants at RIETI.

References

Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Atalay, Enghin, Ali Hortacsu, James Roberts, and Chad Syverson, "Network structure of production," *Proceedings of the National Academy of Sciences*, 2011, *108* (13), 5199–5202.

Chakraborty, Abhijit, Yuichi Kichikawa, Takashi Iino, Hiroshi Iyetomi, Hiroyasu Inoue, Yoshi Fujiwara, and Hideaki Aoyama, "Hierarchical communities in the walnut structure of the Japanese production network," *PloS one*, 2018, *13* (8).

Chung, Fan, "Laplacians and the Cheeger inequality for directed graphs," *Annals of Combinatorics*, 2005, *9* (1), 1–19.

Domenico, Manlio De, Vincenzo Nicosia, Alexandre Arenas, and Vito Latora, "Structural reducibility of multilayer networks," *Nature communications*, 2015, *6* (1), 1–9.

Fiedler, Miroslav, "Algebraic connectivity of graphs," *Czechoslovak mathematical journal*, 1973, *23* (2), 298–305.

Fujiwara, Yoshi and Hideaki Aoyama, "Large-scale structure of a nation-wide production network," *The European Physical Journal B*, 2010, *77* (4), 565–580.

Irie, Hiroaki, "Industrial Structure of Kansai Economy Based on the 201Input-Output Tables (in Japanese)," *The Bulletin of The Junior College Division of Kindai University*, 2017, *50* (1), 1–7.

Kondo, Satoshi, "An Analysis on Regional Economies Based on a Classification of 203 Areas in Japan in Terms of Industrial Structure (in Japanese)," *Journal of household economics*, 2020, *52*, 33–48.

Lancichinetti, Andrea and Santo Fortunato, "Community detection algorithms: a comparative analysis," *Physical review E*, 2009, *80* (5), 056117.

Leontief, Wassily, *Input-output economics*, Oxford University Press, 1986.

Luo, Jianxi, Carliss Y Baldwin, Daniel E Whitney, and Christopher L Magee, "The architecture of transaction networks: a comparative analysis of hierarchy in two sectors," *Industrial and Corporate Change*, 2012, *21* (6), 1307–1335.

Newman, Mark EJ, "Finding community structure in networks using the eigenvectors of matrices," *Physical review E*, 2006, *74* (3), 036104.

Rosvall, Martin and Carl T Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, 2008, *105* (4), 1118–1123.

– **and** – , "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems," *PloS one*, 2011, *6* (4), e18209.

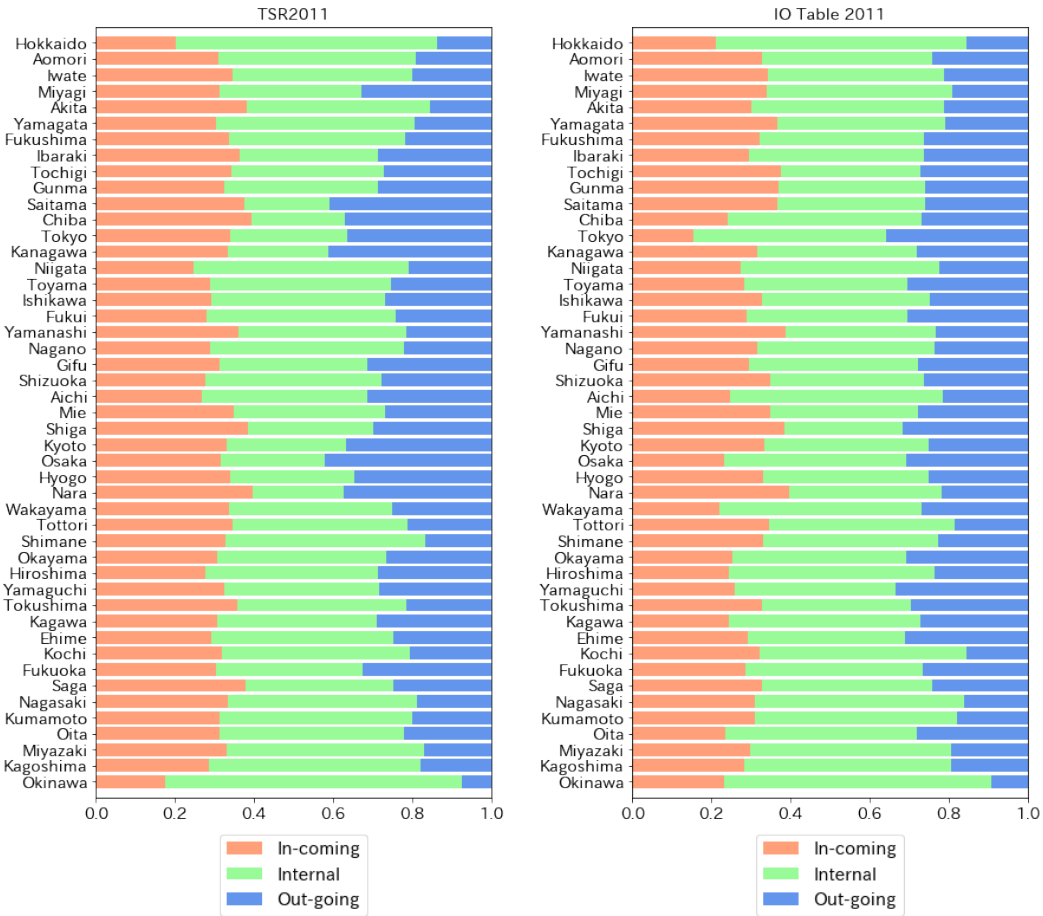

Figure 1. The fractions of internal and external transactions to the total number or amounts of transactions for each prefecture based on the data of TSR2011 (left) and IO Table 2011 (right). The external transactions are distinguished between the incoming and outgoing flows. Notably, ratios in the TSR data do not reflect the transaction values between firms.

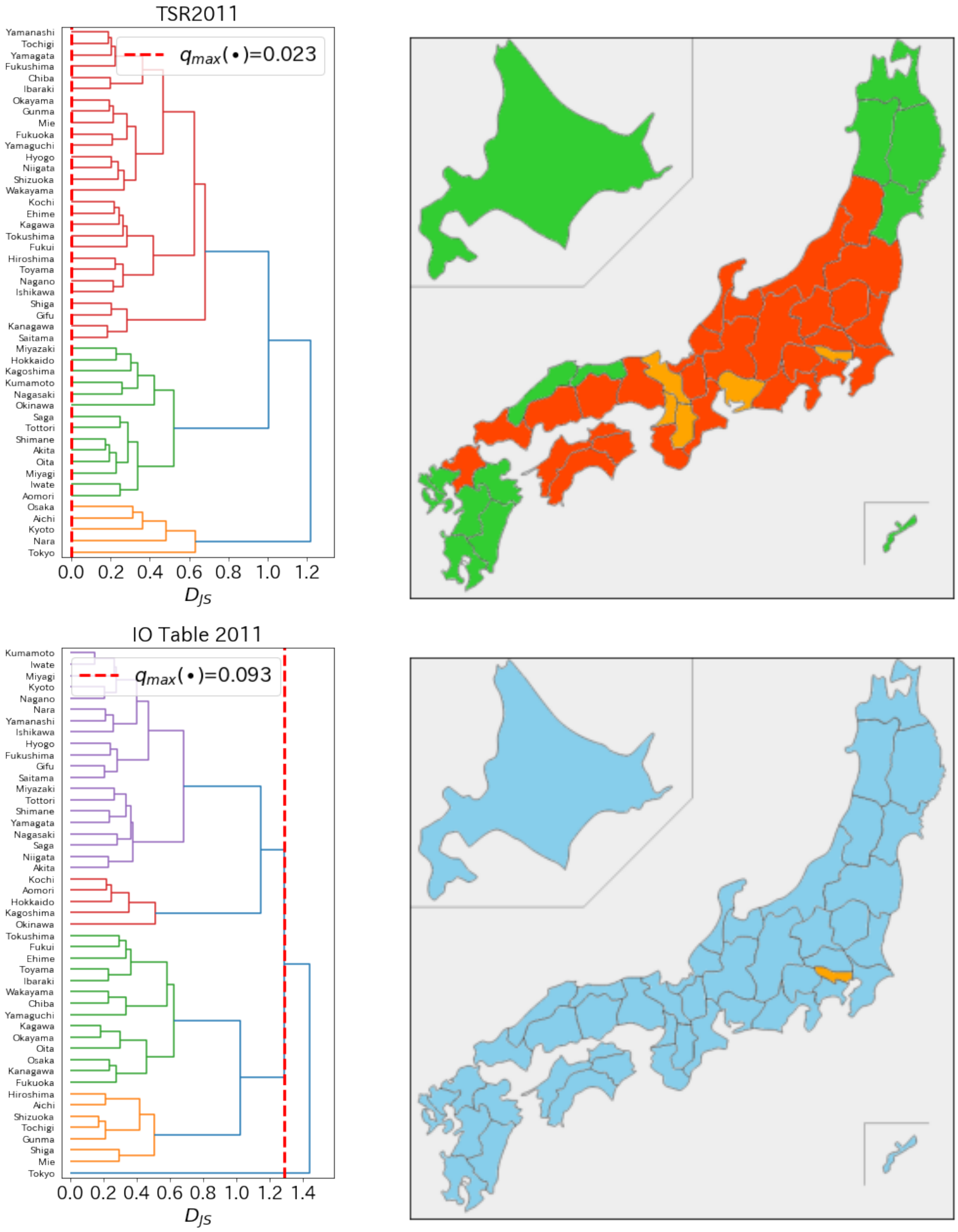

Figure 2. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. The upper and lower figures show the result based on the TSR2011 data and IO Table 2011, respectively. Because the prefecture layers in the TSR2011 case cannot be reduced, we show the color-coded map based on the finding of 2nd-level clusters. In the IO table 2011 case, because the 1st-level clustering corresponds to the maximum of distinguishable cutting, the result is reflected in the coloring of the map.

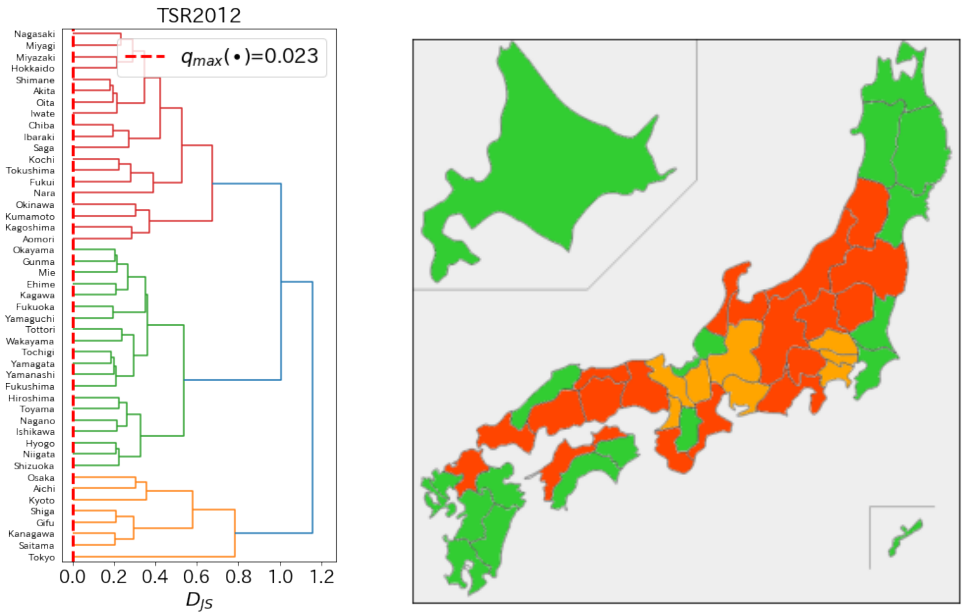

Figure 3. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. Because the prefecture layers in the TSR2012 case cannot be reduced, we demonstrate the color-coded map based on the result of 2nd-level clusters.

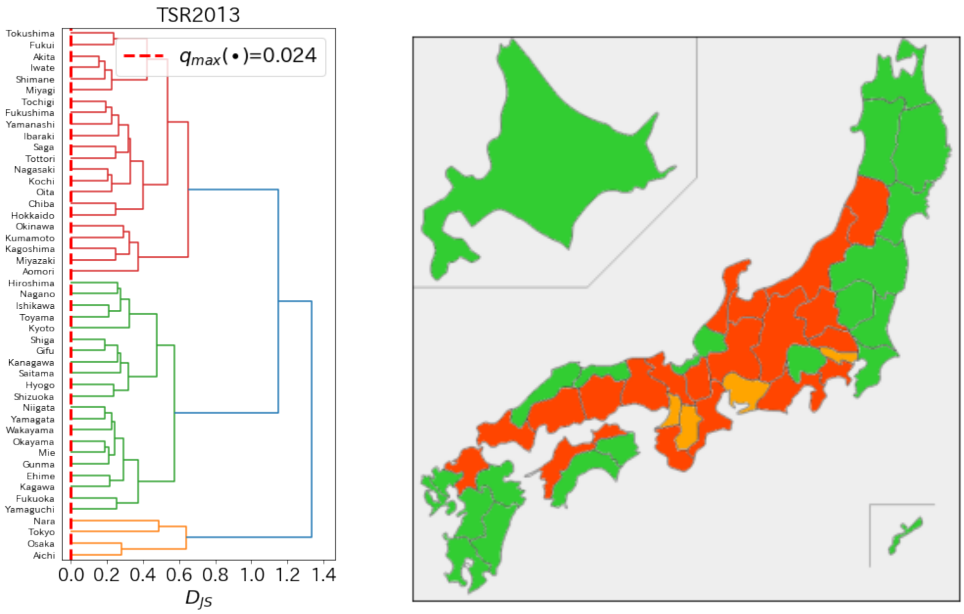

Figure 4. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. Because the prefecture layers in the TSR2013 case cannot be reduced, we show the color-coded map based on the result of 2nd-level clusters.

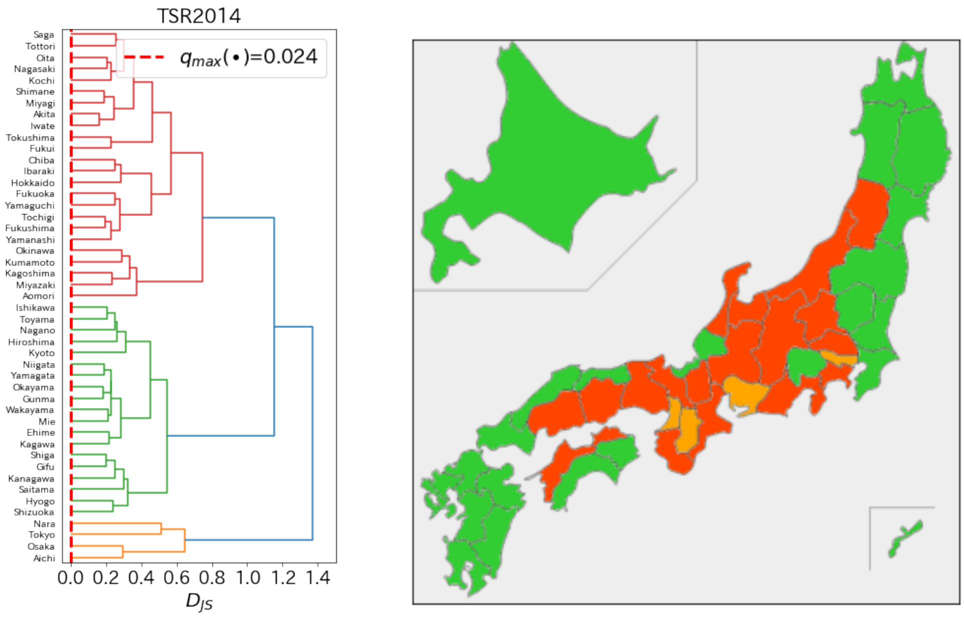

Figure 5. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. Because the prefecture layers in the TSR2014 case cannot be reduced, we show the color-coded map based on the result of 2nd-level clusters.

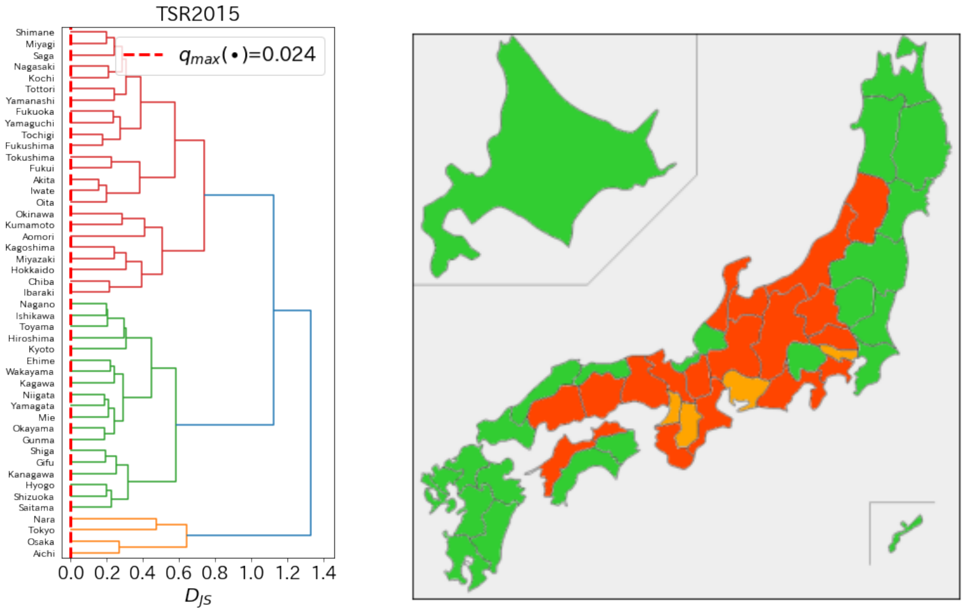

Figure 6. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to thehierarchical clustering analysis and structural reducibility. Because the prefecture layers in the TSR2015 case cannot be reduced, we show the color-coded map based on the result of 2nd-level clusters.

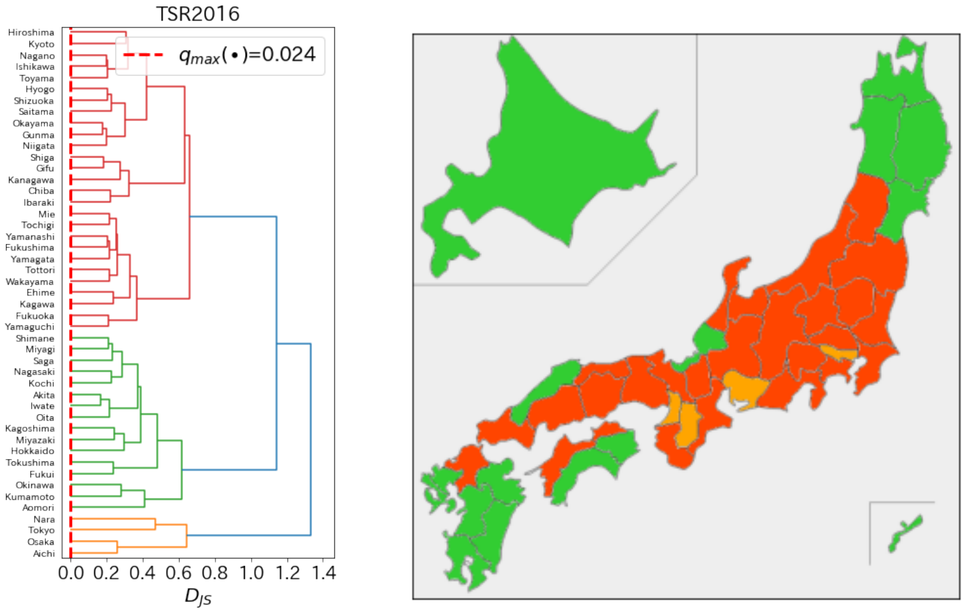

Figure 7. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. Because the prefecture layers on the TSR2016 case cannot be reduced, we show the color-coded map based on the result of 2nd-level clusters.

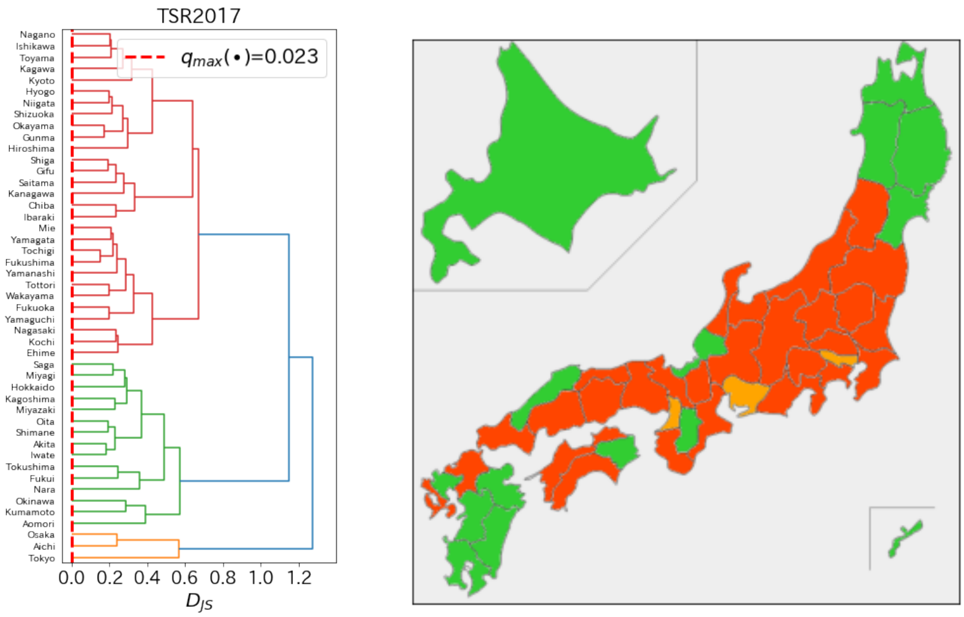

Figure 8. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. Because the prefecture layers on the TSR2017 case cannot be reduced, we show the color-coded map based on the result of 2nd-level clusters.

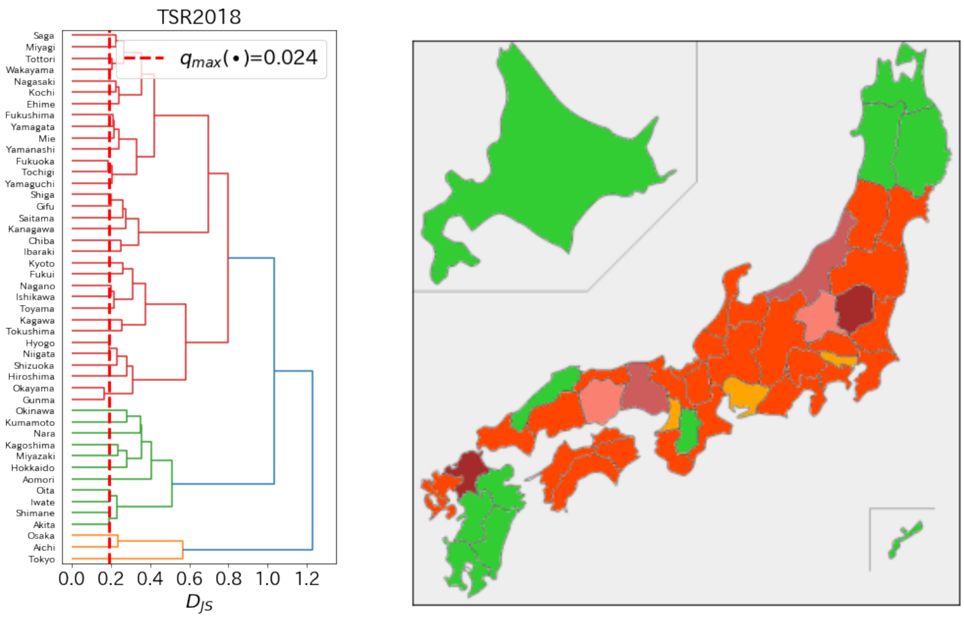

Figure 9. Left: the dendrogram resulting from hierarchical clustering with the dashed red lines identifying the maximum of the quality function $q_{\max}(\bullet)$. Right: the color-coded map of Japan according to the hierarchical clustering analysis and structural reducibility. In the TSR2018 case, we indicate the color-coded map based on the result of 2nd-level clusters. The reducible three pairs of prefectures belonging to the cluster colored in red are shown in different colors: Tochigi and Fukuoka, Gunma and Okayama, and Niigata and Hyogo.

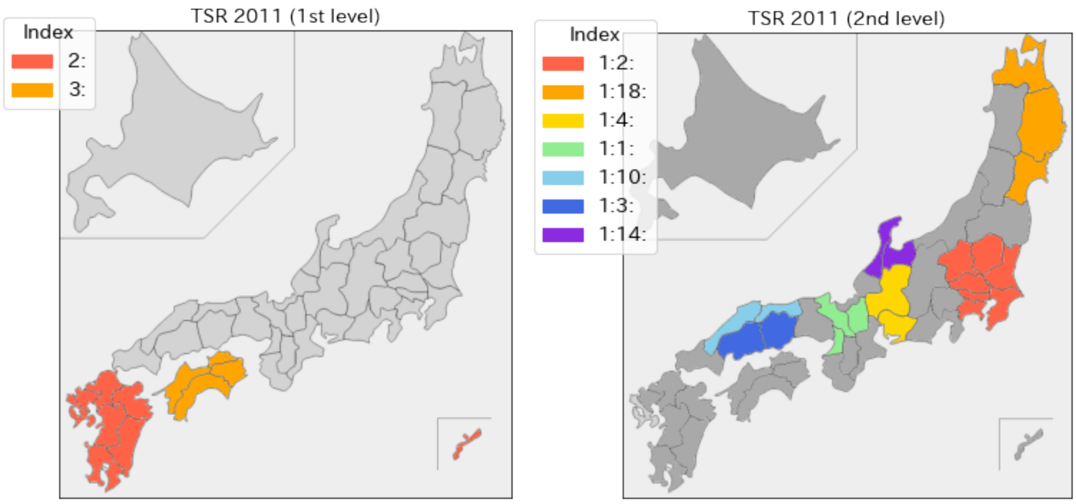

Figure 10. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2011 data. The label of colors corresponds to the index in Table 5. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

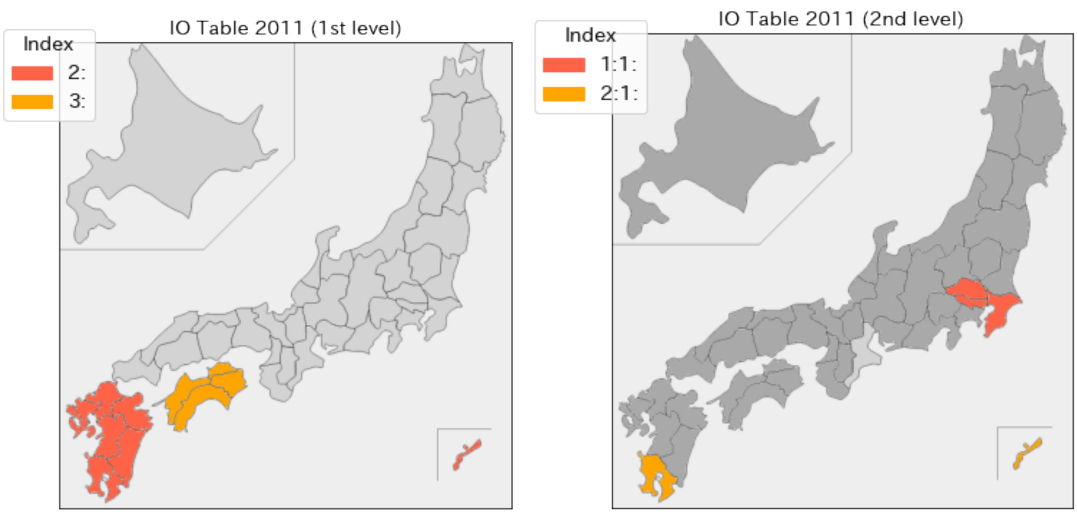

Figure 11. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the IO table 2011 data. The label of colors corresponds to the index in Table 6. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

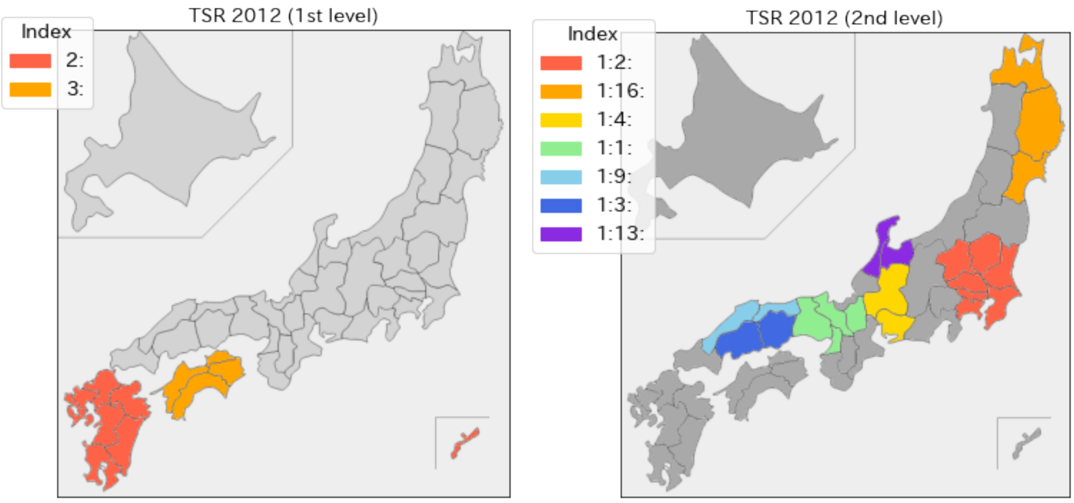

Figure 12. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2012 data. The label of colors corresponds to the index in Table 7. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

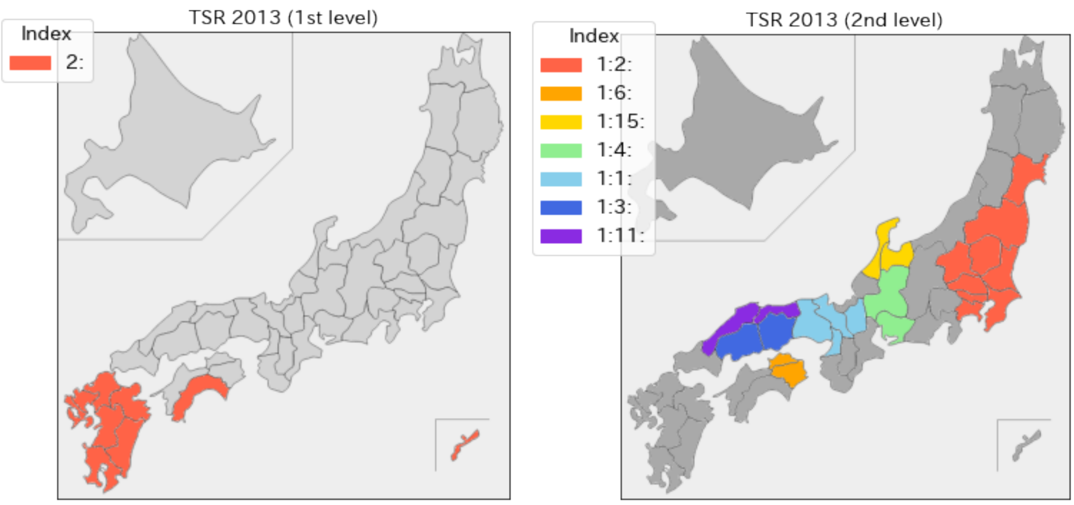

Figure 13. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2013 data. The label of colors corresponds to the index in Table 8. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

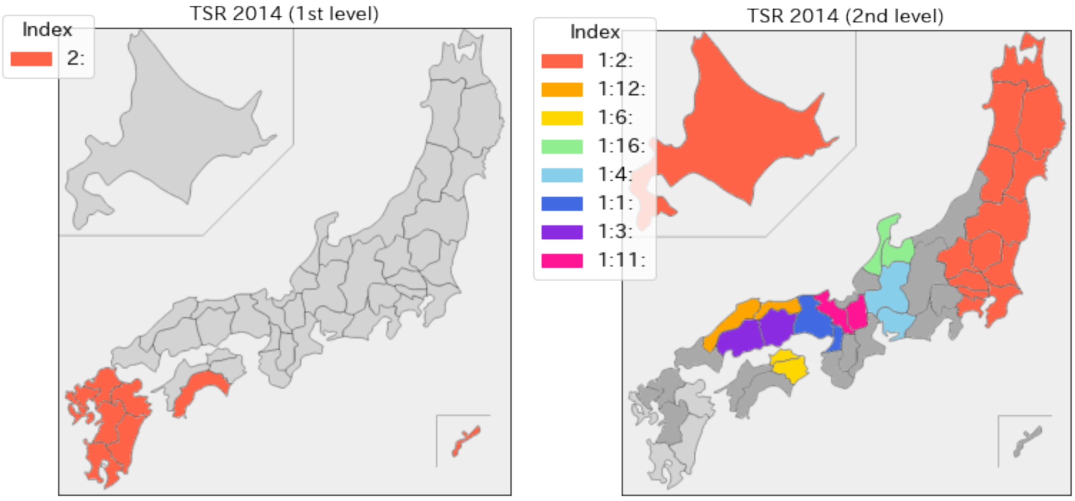

Figure 14. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2014 data. The label of colors corresponds to the index in Table 9. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

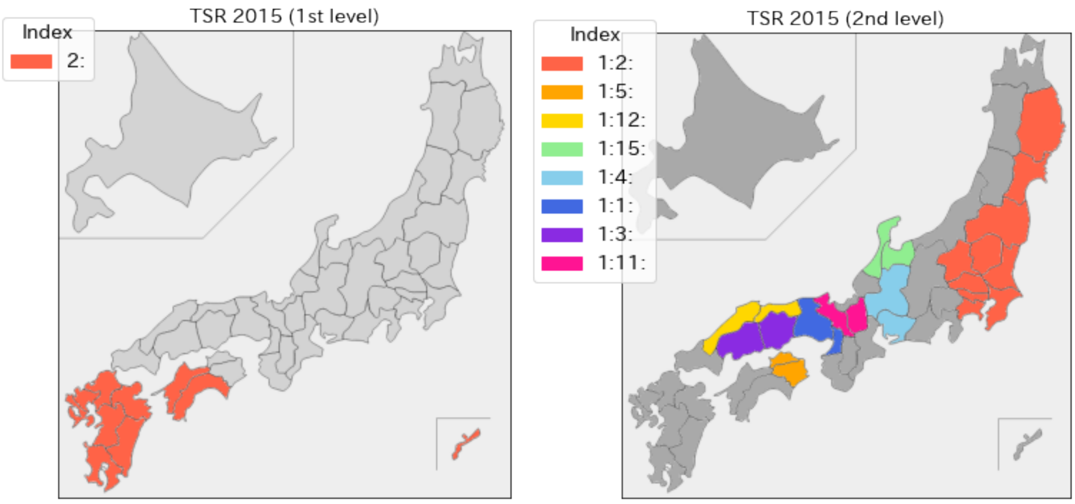

Figure 15. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2015 data. The label of colors corresponds to the index in Table 10. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

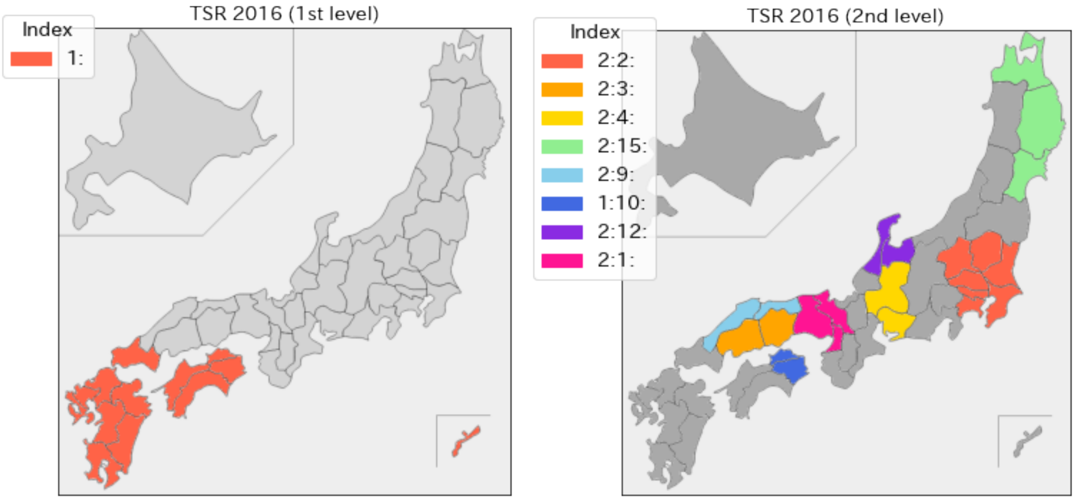

Figure 16. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2016 data. The label of colors corresponds to the index in Table 11. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

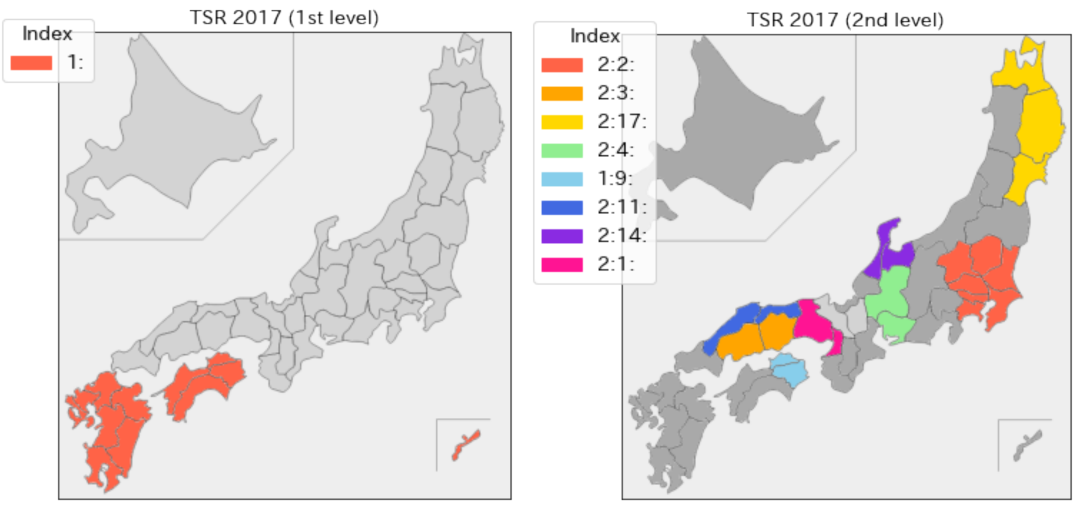

Figure 17. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on the TSR2017 data. The label of colors corresponds to the index in Table 12. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

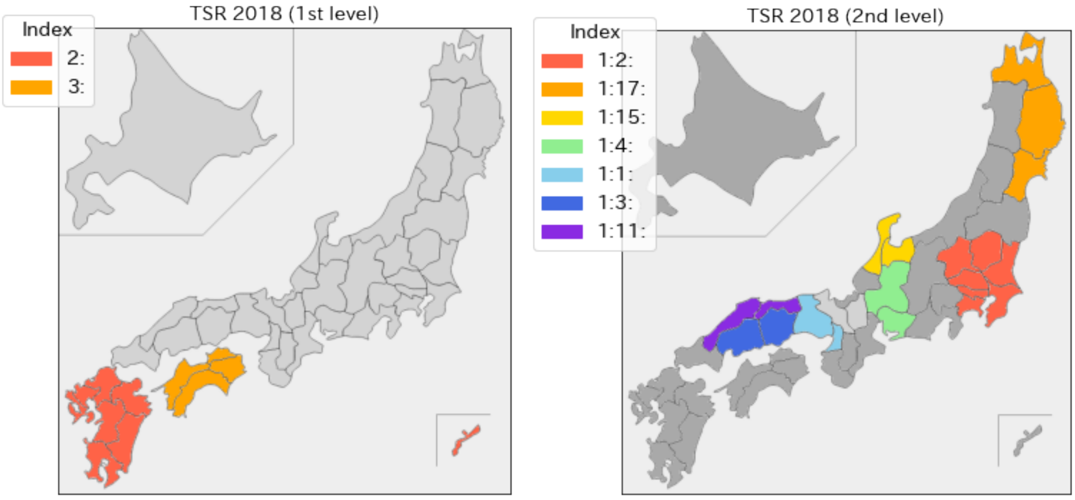

Figure 18. The color-coded maps of Japan based on the results of hierarchical communities at the 1st (left) and the 2nd (right) level based on TSR2017 data. The label of colors corresponds to the index in Table 13. When over 70% of nodes belonging to one prefectural layer form one community, these prefectures are colored with dark gray.

Table 1. The number of firms and transactions extracted from the TSR data from 2011 to 2018. The extracted firms could be active, identify the geographical location, and classify as 98 industry classifications.

Year	# of firms	# of transactions
2011	1,003,304	4,498,690
2012	1,026,152	4,585,893
2013	1,065,382	4,840,612
2014	1,071,622	4,893,952
2015	1,073,218	4,940,266
2016	1,074,365	4,991,876
2017	1,069,554	5,016,434
2018	1,074,802	5,084,967

Table 2. Comparison of reducibility results based on the TSR and IO table data. N is the number of nodes for each layer corresponding to the number of industrial sectors. M_{opt} is the number of layers corresponding to the maximal value of the quality function ($q_{\max}(\bullet)$) obtained through the greedy hierarchical clustering procedure and the value of the reducibility χ.

TSR	$N = 98$		
Year	M_{opt}	$q_{\max}(\bullet)$	χ
2011	47	0.023	0.000
2012	47	0.023	0.000
2013	47	0.024	0.000
2014	47	0.024	0.000
2015	47	0.025	0.000
2016	47	0.024	0.000
2017	47	0.023	0.000
2018	44	0.023	0.065

IO Table	$N = 32$		
Year	M_{opt}	$q_{\max}(\bullet)$	χ
2011	2	0.093	0.978

Table 3. Statistics for the hierarchical communities detected using the multicoding Infomap method for the industrial networks based on TSR data. "# of com." is the number of all communities, and "# of irr.com." is the number of irreducible communities, which are communities that do not have any subcommunities. "# of nodes" denotes the number of nodes, including industries, in irreducible communities.

Year	Level	1	2	3	4
2011	# of com.	377	47	9	-
	# of irr. com.	374	45	9	-
	# of nodes	374	3,885	344	-
2012	# of com.	353	40	8	-
	# of irr. com.	350	38	8	-
	# of nodes	350	3,909	344	-
2013	# of com.	297	37	4	-
	# of irr. com.	295	36	4	-
	# of nodes	295	4,219	92	-
2014	# of com.	286	31	18	3
	# of irr. com.	284	29	17	3
	# of nodes	284	2,945	1,332	45
2015	# of com.	277	37	4	-
	# of irr. com.	275	36	4	-
	# of nodes	275	4,239	92	-
2016	# of com.	269	40	7	-
	# of irr. com.	267	38	7	-
	# of nodes	267	3,989	350	-
2017	# of com.	268	39	10	-
	# of irr. com.	266	37	10	-
	# of nodes	266	3,990	350	-
2018	# of com.	261	45	10	2
	# of irr. com.	258	43	9	2
	# of nodes	258	3,998	335	15

Table 4. Statistics for the hierarchical communities found using the multicoding Infomap method for the industrial network based on IO table data. "# of com." is the number of all communities, and "# of irr.com." is the number of irreducible communities, which are communities that do not have any subcommunities. "# of nodes" represents the number of nodes, including industries, in irreducible communities.

Year	Level	1	2	3
2011	# of com.	3	44	3
	# of irr. com.	0	43	3
	# of nodes	0	1,395	62

Table 5. Characteristics of hierarchical communities based on the TSR2011 data at the 1st and 2nd level, including over 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are outlined.

1st-level

Index	Size	Prefecture
2	701	Oita (12.8), Kagoshima (12.7), Okinawa (12.7) Nagasaki (12.6), Kumamoto (12.4), Saga (12.1) Fukuoka (12.1), Miyazaki (11.8)
3	339	Ehime (25.7), Kagawa (25.7) Kochi (24.8), Tokushima (23.3)

2nd-level

Index	Size	Prefecture
1:1	242	Osaka (31.8), Kyoto (31.0), Shiga (28.9)
1:2	641	Tokyo (14.0), Ibaraki (13.9), Kanagawa (13.9) Chiba (13.7), Saitama (13.1), Gunma (12.9) Tochigi (12.8)
1:3	176	Hiroshima (49.4), Okayama (47.2)
1:4	178	Aichi (48.9), Gifu (47.8)
1:10	163	Shimane (50.3), Tottori (49.7)
1:14	164	Toyama (51.2), Ishikawa (48.8)
1:18	256	Aomori (33.1), Miyagi (33.1), Iwate (32.7)

Table 6. Characteristics of hierarchical communities based on the IO Table 2011 data at the 1st and 2nd level, including over 50 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are presented.

1st-level

Index	Size	Prefecture
2	248	Oita (12.5), Kagoshima (12.5), Kumamoto (12.5) Nagasaki (12.5), Okinawa (12.5), Fukuoka (12.5) Miyazaki (12.5), Saga (12.5)
3	134	Kagawa (25.0), Ehime (25.0) Tokushima (25.0), Kochi (25.0)

2nd-level

Index	Size	Prefecture
1:1	94	Chiba (33.0), Tokyo (33.0), Saitama (33.0)
2:1	62	Kagoshima (50.0), Okinawa (50.0)

Table 7. Characteristics of hierarchical communities based on the TSR2012 data at the 1st and 2nd level, including over 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are presented.

1st-level

Index	Size	Prefecture
2	700	Nagasaki (12.9), Oita (12.9), Kagoshima (12.7) Okinawa (12.7), Fukuoka (12.3), Saga (12.1) Kumamoto (12.0), Miyazaki (12.0)
3	346	Ehime (25.4), Kagawa (25.4) Tokushima (24.0), Kochi (24.0)

2nd-level

Index	Size	Prefecture
1:1	360	Hyogo (25.3), Osaka (24.2) Kyoto (22.8), Shiga (21.7)
1:2	652	Tokyo (14.0), Ibaraki (13.8), Kanagawa (13.7) Chiba (13.5), Gunma (12.9), Saitama (12.7) Tochigi (12.6)
1:3	176	Hiroshima (49.4), Okayama (48.3)
1:4	180	Aichi (48.9), Gifu (47.2)
1:9	162	Shimane (50.6), Tottori (49.4)
1:13	167	Toyama (52.1), Ishikawa (47.9)
1:16	254	Miyagi (33.9), Aomori (33.1), Iwate (32.7)

Table 8. Characteristics of hierarchical communities based on the TSR2013 data at the 1st and 2nd level, including more than 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are outlined.

1st-level

Index	Size	Prefecture
2	808	Nagasaki (11.5), Oita (11.4), Kumamoto (11.3) Okinawa (11.1), Kagoshima (11.1), Miyazaki (10.9) Fukuoka (10.9), Saga (10.5), Kochi (10.1)

2nd-level

Index	Size	Prefecture
1:1	325	Hyogo (28.0), Osaka (27.1) Kyoto (26.2), Shiga (12.6)
1:2	835	Ibaraki (10.9), Tokyo (10.9), Chiba (10.8) Kanagawa (10.7), Fukushima (10.7), Miyagi (10.7) Saitama (10.2), Tochigi (10.2), Gunma (10.2)
1:3	180	Hiroshima (48.9), Okayama (48.3)
1:4	181	Aichi (49.2), Gifu (47.5)
1:6	175	Kagawa (51.4), Tokushima (46.3)
1:11	169	Shimane (50.3), Tottori (48.5)
1:15	170	Toyama (51.8), Ishikawa (47.1)

Table 9. Characteristics of hierarchical communities based on the TSR2014 data at the 1st and 2nd level, including more than 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are presented.

1st-level

Index	Size	Prefecture
2	801	Oita (11.5), Nagasaki (11.4), Okinawa (11.4) Kagoshima (11.2), Miyazaki (11.1), Fukuoka (11.0) Kumamoto (10.9), Saga (10.9), Kochi (10.2)

2nd-level

Index	Size	Prefecture
1:1	192	Hyogo (46.4), Osaka (41.7), Nara (5.7)
1:2	1,283	Ibaraki (7.2), Hokkaido (7.2), Tokyo (7.2) Chiba (7.2), Miyagi (7.2), Kanagawa (7.1) Fukushima (7.0), Tochigi (7.0), Gunma (7.0) Yamagata (6.9), Iwate (6.9), Saitama (6.9) Aomori (6.9), Akita (6.8)
1:3	178	Okayama (49.4), Hiroshima (48.3)
1:4	180	Aichi (48.9), Gifu (47.8)
1:6	173	Kagawa (51.4), Tokushima (47.4)
1:11	151	Kyoto (51.0), Shiga (48.3)
1:12	171	Shimane (50.9), Tottori (48.0)
1:16	172	Toyama (50.6), Ishikawa (47.7)

Table 10. Characteristics of hierarchical communities based on the TSR2015 data at the 1st and 2nd level, including over 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are listed.

1st-level

Index	Size	Prefecture
2	898	Nagasaki (10.5), Kumamoto (10.2), Oita (10.2) Ehime (10.0), Kagoshima (10.0), Okinawa (10.0) Fukuoka (9.8), Miyazaki (9.7), Saga (9.6) Kochi (9.5)

2nd-level

Index	Size	Prefecture
1:1	180	Hyogo (48.9), Osaka (44.4)
1:2	916	Tokyo (10.0), Kanagawa (9.8), Ibaraki (9.8) Miyagi (9.8), Chiba (9.7), Fukushima (9.6) Iwate (9.5), Tochigi (9.4), Gunma (9.4) Saitama (9.2)
1:3	178	Hiroshima (49.4), Okayama (47.8)
1:4	177	Aichi (50.3), Gifu (48.6)
1:5	175	Kagawa (51.4), Tokushima (47.4)
1:11	151	Kyoto (51.0), Shiga (47.7)
1:12	170	Shimane (50.6), Tottori (48.2)
1:15	174	Toyama (50.0), Ishikawa (48.3)

Table 11. Characteristics of hierarchical communities based on the TSR2016 data at the 1st and 2nd level, including over 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are listed.

1st-level

Index	Size	Prefecture
1	1,157	Nagasaki (8.1), Oita (8.0), Kumamoto (7.9) Kagoshima (7.8), Ehime (7.8), Okinawa (7.8) Miyazaki (7.7), Kochi (7.6), Fukuoka (7.6) Saga (7.5), Kagawa (7.5), Yamaguchi (7.4) Tokushima (7.3)

2nd-level

Index	Size	Prefecture
1:10	172	Kagawa (50.6), Tokushima (48.3)
2:1	263	Hyogo (33.1), Osaka (30.4), Kyoto (27.4) Shiga (6.1)
2:2	653	Ibaraki (14.1), Tokyo (14.1), Kanagawa (13.8) Chiba (13.6), Saitama (13.2), Tochigi (13.0) Gunma (13.0)
2:3	175	Hiroshima (50.3), Okayama (46.3)
2:4	178	Aichi (50.0) Gifu (48.3)
2:9	171	Shimane (49.7), Tottori (48.5)
2:12	174	Ishikawa (49.4), Toyama (48.9)
2:15	258	Iwate (33.3), Miyagi (33.3), Aomori (32.6)

Table 12. Characteristics of hierarchical communities based on the TSR2017 data at the 1st and 2nd level, including over 150 nodes. Parentheses show the percentage of prefectures, and only communities that include two or more prefectures with a share of 5% or more are listed.

1st-level

Index	Size	Prefecture
1	1,073	Nagasaki (8.8), Kumamoto (8.6), Oita (8.6) Okinawa (8.5), Kagoshima (8.4), Miyazaki (8.4) Ehime (8.3), Kagawa (8.3), Kochi (8.1) Saga (8.1), Fukuoka (8.1), Tokushima (7.7)

2nd-level

Index	Size	Prefecture
1:9	173	Kagawa (51.4), Tokushima (47.4)
2:1	183	Hyogo (47.0), Osaka (42.6)
2:2	652	Tokyo (14.1), Kanagawa (13.8), Ibaraki (13.8) Chiba (13.7), Saitama (13.5), Tochigi (13.0) Gunma (13.0)
2:3	179	Hiroshima (49.7), Okayama (48.0)
2:4	179	Aichi (49.2), Gifu (48.0)
2:11	171	Tottori (49.7), Shimane (49.1)
2:14	174	Ishikawa (49.4), Toyama (48.9)
2:17	257	Iwate (33.5), Miyagi (33.5), Aomori (32.3)

Table 13. Characteristics of hierarchical communities based on TSR2018 data at the 1st and 2nd level, including over 150 nodes. Parentheses indicate the percentage of prefectures, and only communities that contain two or more prefectures with a share of 5% or more are presented.

1st-level

Index	Size	Prefecture
2	714	Kagoshima (12.7), Oita (12.7), Nagasaki (12.6) Miyazaki (12.6), Okinawa (12.6), Kumamoto (12.5) Saga (12.2), Fukuoka (11.6)
3	354	Kagawa (25.1), Ehime (24.6) Kochi (24.3), Tokushima (23.7)

2nd-level

Index	Size	Prefecture
1:1	176	Hyogo (48.9), Osaka (44.3)
1:2	654	Ibaraki (14.1), Tokyo (13.9), Kanagawa (13.8) Chiba (13.8), Saitama (13.3), Gunma (13.0) Tochigi (13.0)
1:3	175	Hiroshima (52.0), Okayama (46.3)
1:4	172	Aichi (49.7), Gifu (48.0)
1:11	172	Shimane (50.6), Tottori (48.8)
1:15	172	Toyama (50.6), Ishikawa (49.4)
1:17	259	Miyagi (34.0), Iwate (33.2), Aomori (32.0)