Social Network Analysis of Ancient Japanese Obsidian Artifacts Reflecting Sampling Bias Reduction

Fumihiro Sakahira [1]*, Hiro'omi Tsumura[2]

[1] Faculty of Information Science and Technology, Osaka Institute of Technology, Hirakata City, Osaka, Japan, ORCID: 0000-0002-1228-4069
[2] Faculty of Culture and Information Science, Doshisha University, Kyotanabe, Kyoto, Japan

*Corresponding author
Correspondence: fumihiro.sakahira@oit.ac.jp

Abstract

This study aims to investigate the dynamics of obsidian trade networks during the Jomon period (approximately 15,000 to 2,400 years ago), the hunting and gathering era in Japan. To improve regional representation and reduce the distortions caused by small sample sizes, we performed clustering based on a large-scale dataset and conducted social network analysis. The research results revealed that the trade networks during the Jomon period were not constant; they expanded throughout the southern Kanto region during the Middle Jomon period (5,500–4,500 years cal BP) and ceased to function during the Late Jomon period (4,500–3,200 years cal BP). Furthermore, to enhance the readability and interpretability of the dataset, we implemented clustering using the density-based spatial clustering of applications with noise (DBSCAN) method. The results showed that in every time division of the Jomon period, the mean intra-cluster cosine similarity of each cluster was higher than the similarity between sites outside the clusters, confirming the reasonableness of an analysis considering regional representation. In addition, to verify the robustness of the network in the social network analysis after clustering, we also performed a bootstrap simulation analysis. The results showed high network robustness and demonstrated that the sampling after clustering had minimal impact on this study's findings.

Keywords: social network analysis, obsidian artifact, DBSCAN, clustering, ancient Japan, Jomon period

Introduction

This study aims to reveal the changes in obsidian trade networks during the Jomon period (15,000 to 2,400 years ago), the hunting and gathering era in Japan. We conducted clustering using a large-scale dataset to improve regional representation and reduce the distortion caused by small sample sizes, and then performed a social network analysis. Obsidian is a type of volcanic glass that was used for making sharp stone tools and processing food and wood materials (Ono, 2011). In archaeology, the similarities and differences in artifacts are used as indicators of contact and relationships between groups (Freund, 2013). As obsidian provenances are limited, identifying them is essential for understanding trade networks and resource procurement (Freund, 2013). Shells and jade ornaments from the Jomon period have been found in regions of Japan far from their production sites, suggesting the existence of extensive trade (Hashiguchi, 1999). However, the Jomon period spans approximately 13,000 years, during which cultural transitions can be observed; therefore, it is hypothesized that the trade range was not constant and instead expanded and contracted over time. To investigate the expansion and contraction of the Jomon period trade networks, we conducted a social network analysis of obsidian artifacts. This approach allowed us to clarify how trade networks changed over time.

The Kanto region is located in the eastern part of the Japanese mainland, and its obsidian provenance analysis is considered to be of the highest quality and quantity in the world (Tsumura & Tateishi, 2013). In this study, we focus on obsidian from the Jomon period in the Kanto region. According to a survey conducted in 2011, approximately 21,000 obsidian artifacts had been found at over 270 sites (Nihon-kokogaku-kyokai 2011 nendo tochigi-taikai-jikkoiinkai, 2011). However, when dealing with large-scale data, social network analysis graphs can become overly complex, making it difficult to derive useful interpretations.

In archaeology, it is important to consider that archaeological sites, artifacts, and features represent only a portion of what originally existed. In particular, with chemical analysis methods such as obsidian provenance analysis, it is difficult to target all excavated items due to constraints associated with excavation periods and budgets. The dataset used in this study also includes sites where only a few artifacts or, in extreme cases, just one artifact per site have been analyzed (Tsumura & Tateishi, 2013). When the sample size of obsidian at each site is small, the regional composition ratio may be distorted, potentially affecting the results (Golitko & Feinman, 2015). To address this issue, this study conducts clustering by region to improve the readability and interpretability of the dataset and then applies social network analysis. This approach can help reduce the distortion caused by small sample sizes.

Related Work

Obsidian Analysis of Japan's Kanto Region

Regarding the analysis of obsidian provenances in the Kanto region, Suzuki (1973, 1974) investigated trends in provenances and timing, and Warashina and Higashimura (1988) collected and organized information on obsidian and sanukaito provenances. Since the late 1980s, the proliferation of X-ray fluorescence analysis equipment has led to an increase in obsidian provenance analyses, and various studies focusing on archaeological issues across the Kanto region have been conducted (Kanayama, 1994; Kojo, 1996; Daikuhara, 2008; Ikeya, 2009). Furthermore, Sugihara and Kobayashi (2008) and Tsutsumi (2018) investigated resource development and supply from specific provenances from the Paleolithic to the middle Yayoi period (–2,000 years cBP).

Subsequently, the Japanese Archaeological Association compiled a collection of obsidian provenance analyses in the Kanto region in 2011 (Nihon-kokogaku-kyokai 2011 nendo tochigi-taikai-jikkoiinkai, 2011). Tsumura and Tateishi (2013) used these materials and statistical analysis methods to verify the patterns of provenances and consumption sites in the Kanto region during the Jomon period. As a result, the authors suggested that the obsidian trade network changed over time. They also quantitatively analyzed the relationship between provenances and consumption sites; however, the dynamics of the trade network among consumption sites have not been sufficiently investigated, and there remain many unexplained details. It is difficult to visualize and interpret large amounts of data using conventional methods, and social network analysis has only recently been established as a tool in archaeology.

Social Network Analysis of Obsidian Artifacts

Regarding research using social network analysis to study obsidian trade networks, there have been several such studies in areas like Mesoamerica and New Zealand. For example, Golitko et al. (2012) assumed that the inland land trading network in Mesoamerica collapsed, and the coastal maritime trading network developed at the end of the Classical period. In addition, Golitko and Feinman (2015) suggested that the hierarchy and scale of the network decreased over time, indicating that the economy of Mesoamerica was not centralized. Furthermore, through a social network analysis of obsidian provenances, Ladefoged et al. (2019) observed that the selection of provenances in Maori society in 15th-century New Zealand was influenced by the community to which they belonged.

These studies used social network analysis of obsidian provenances to represent archaeological sites and provenances of obsidian as "nodes." Nodes are supplemented with attribute information such as geographic location, estimated age, and the amount or percentage of obsidian at the provenance. Links are established based on the similarity between nodes (i.e., similarity in the proportion of obsidian), reflecting the relationship between them. Social network analysis focuses on these nodes and their relationships, adopting an approach that considers the system a combination of the two (Ladefoged et al., 2019).

Impact of Sampling

In the social network analysis of the obsidian trade, the data size typically ranges from several hundred to several thousand obsidian artifacts. For example, Ladefoged et al. (2019) analyzed 2,404 obsidian artifacts from 15 sites, Meissner (2017) analyzed 2,630 obsidian artifacts from 796 sites, and Mills et al. (2013) analyzed 4,805 obsidian artifacts. Golitko et al. (2012) and Golitko and Feinman (2015) used data from 121 and 242 sites, respectively, although they did not specify the exact number of obsidian artifacts used in their social network analyses. In contrast, the present study used a large dataset of approximately 21,000 obsidian artifacts from over 270 sites (Nihon-kokogaku-kyokai 2011 nendo tochigi-taikai-jikkoiinkai, 2011). However, a drawback of such a large dataset is that the resulting social network graph may be too complex to yield useful interpretations.

Archaeological data such as sites, artifacts, and structures are often only a partial representation of what actually existed. In particular, the chemical analysis techniques used in obsidian provenance studies do not typically analyze all excavated artifacts due to constraints related to excavation durations and budgets. The dataset used in the present study includes sites where only a few or even only one artifact was analyzed for obsidian (Tsumura & Tateishi, 2013). In such cases, there is a risk of bias in regional composition and therefore of biased results (Golitko & Feinman, 2015). Consequently, Golitko and Freiman (2015) excluded obsidian samples of less than 10 per site from their study. They also mentioned combining sets of sites from specific time periods to create a pooled set of frequencies for the entire region but did not provide suggestions for specific methods.

Owing to the aforementioned situation in archaeology, it is natural to consider sampling variability in network analysis based on the similarity of artifact assemblages (Roberts et al., 2021). In social network analysis, studies that consider sampling effects have shown that node-level indicators such as degree centrality are susceptible to sampling effects, while network indicators such as distance, centrality, and diameter are robust to node removal (Wey et al., 2008). Regarding the assessment of sampling variability, Mills et al. (2013) used bootstrap simulation analysis to verify a dataset from the American Southwest and found that while individual node scores may vary due to sampling, summary statistics at the network level, such as centrality, are relatively stable. Gjesfjeld (2015) conducted a social network analysis on hunter-gatherers in Northeast Asia during the time period of 2,500–500 years cal BP. The analysis was based on compositional data from ceramic artifacts found in the Kuril Islands. Bootstrap simulation and sensitivity analysis were used to evaluate network indicators and determine the stability of these network structures. The results indicated that even with incomplete archaeological data, the variation in the indicators of network analysis was minimal and did not significantly impact the overall interpretation of the network. Roberts et al. (2021) proposed a method that employs bootstraps to assess sampling variability in network analysis, specifically focusing on the similarity of artifact assemblages. Their results demonstrated that bootstrap simulation is effective for assessing sampling variability in network analyses.

Problem Formulation

This study conducted a social network analysis of obsidian artifacts to investigate the expansion and contraction of the trade network in the Jomon period. To improve the readability and interpretability of the large dataset we used and reduce the distortion caused by small sample sizes, we clustered the obsidian samples at each site by region and performed a social network analysis.

In addition to the findings from Sakahira and Tsumura (2023), this study:

- Evaluated the distribution of cosine similarities among clusters based on obsidian composition by provenance, after clustering using the DBSCAN algorithm. Additionally, the distribution of cosine similarities between sites within a cluster and sites outside of a cluster was assessed.
- To enhance interpretation, the composition of obsidian by provenance was incorporated into each cluster during the network analysis.
- To evaluate the effectiveness of this method in reducing distortion and ensuring network robustness, bootstrap simulation analyses were performed in the clustered social network analysis.

Materials and Methods

Dataset of Obsidian Assemblages

This study focused on obsidian artifacts excavated from Jomon period sites in the Kanto region. The Kanto region is located in the eastern part of Honshu and is surrounded by Tokyo Bay, Sagami Bay, the Pacific Ocean, and mountainous areas to the north and northwest (Figure 1). The obsidian artifacts brought to southern Kanto have been found to have originated from islands further south in Tokyo Bay and the surrounding mountainous areas. These obsidian artifacts were transported by sea from the island areas and brought to the consuming areas via a route that diverted to the north from the mountainous area to the northwest (Sugihara & Kobayashi, 2008; Tateishi, 2010).

The dataset for this study was based on the results of previous obsidian provenance analyses conducted on Jomon period sites in the Kanto region and compiled by the Japan Archaeological Association at the Tochigi meeting in 2011 (Nihon-kokogaku-kyokai 2011 nendo tochigi-taikai-jikkoiinkai, 2011). Although this dataset was compiled in 2011, it is still valuable because of the vast amount of data it comprises and because it includes obsidian provenances that have been reported in the years since. The present study's analysis focused on eight main production areas: 1) Takahara-yama, 2) Wada-toge, 3) Omegura, 4) Suwa, 5) Tateshina, 6) Kozu-shima, 7) Hakone, and 8) Amagi. For convenience, Wada-toge, Omegura, Suwa, and Tateshina are collectively referred to as the "Shinshu group" and are considered to belong to the mountainous area known as the "Central Highlands." Several other production areas were excluded from the analysis due to the small number of obsidian artifacts that have been found there.

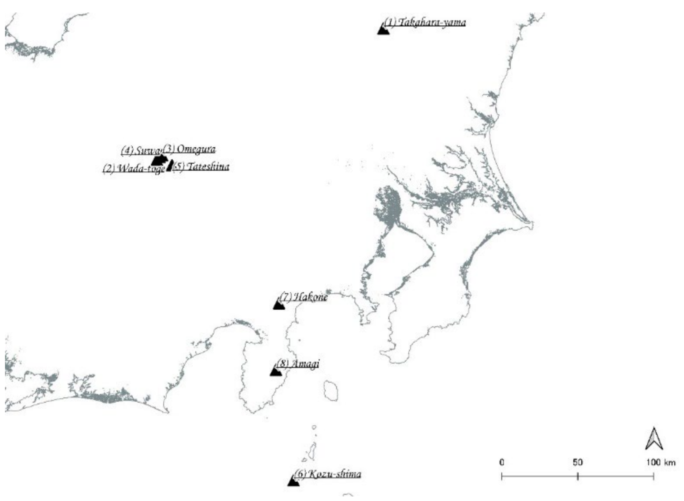

Figure 1 - Location of major obsidian provenance areas.

Clustering

As mentioned earlier, to improve the readability and interpretability of the data and reduce the distortion caused by a small sample size, we performed clustering by region and summarized the results as aggregate values for each region. Assuming that adjacent sites have interactions and share information, we applied the density-based spatial clustering of applications with noise (DBSCAN) algorithm (a density-based algorithm for discovering clusters in large spatial databases with noise) (Ester et al. 1996) to group the geographical locations of the sites. Many other clustering methods do not consider noise and assign all sites to clusters, which can result in sites being clustered even if they cannot access each other. However, the DBSCAN algorithm defines regions as clusters based on the number of points (density) within a radius (ε value) (minPts). If the density within the region exceeds a certain threshold, the cluster expands, but if there are no nearby points within the radius, it is considered noise (Figure 2). The ε value is determined based on the factor at issue (such as physical distance), and the minPts is the optimal size of the minimum cluster. In this study, we set the ε value to 10 km, which is commonly accepted as the activity range of the ancient Jomon people (Akazawa, 1982; Koizumi, 2016). The minPts was set to a minimum requirement of three, which is essential for cluster growth in the DBSCAN algorithm. The DBSCAN algorithm was used for each of the five divisions of the Jomon period.

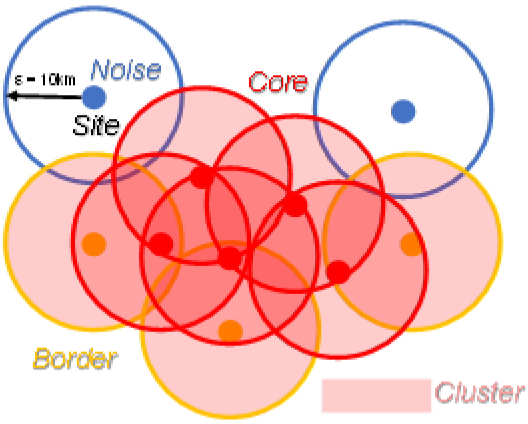

Figure 2 - Image of clustering using the DBSCAN method.

We treated these clusters as a single region, summed up the obsidian provenances in each region, and calculated the proportion of obsidian provenances in each cluster.

The composition ratio (R) was defined by the following equation:

$$R_{i,j} = \frac{N_{i,j}}{T_i},$$

where $R_{i,j}$ indicates the composition ratio of provenance *j* in cluster (or single site) *i*, T_i indicates the total number of analyzed obsidian artifacts in *i*, and $N_{i,j}$ indicates the number of obsidian artifacts of provenance *j* in cluster *i*.

As mentioned above, a small number of obsidian samples may distort the regional composition ratio and potentially affect the results (Golitko & Feinman, 2015). Therefore, we excluded clusters with fewer than 30 obsidian artifacts from the analysis. On the other hand, sites without geographical relationships forming clusters but with more than 30 obsidian artifacts were used as single sites for the analysis by calculating the obsidian provenance composition ratio in the same way as for the clusters.

Similarity

We calculated similarity and performed social network analysis for each period division. Following Ladefoged et al. (2019), we measured the similarity of the obsidian provenance compositions between clusters, between each cluster and individual sites, and within each cluster by calculating cosine similarity. We calculated the provenance composition ratio for each cluster and individual site from the total number of obsidian artifacts and treated them as vectors. Specifically, since this study included eight provenances, they were represented as eight-dimensional vectors.

The cosine similarity (Sim) was expressed by the following formula:

$$Sim_{A,B} = \cos\theta = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|},$$

where $Sim_{A,B}$ represents the similarity between *A* and *B* (where *A* and *B* are clusters or individual sites, and $a\rightarrow$ and $b\rightarrow$ are vectors corresponding to *A* and *B*, and */ /* indicates the magnitude of the vector). If the provenance compositions of *A* and *B* are similar, the direction of vectors $a\rightarrow$ and $b\rightarrow$ becomes close, and the value of $\cos\vartheta$ approaches 1. Conversely, if they are dissimilar, the value approaches 0.

Network Analysis

We created an undirected network based on the cosine similarity of obsidian provenance composition ratio between clusters and single sites. This network revealed the relationships between consumption sites for each period. Each cluster or single site was represented as a node, and a link was generated between nodes when the cosine similarity between them exceeded 0.9. The value of 0.9 was chosen for convenience, to improve the readability of the figures. Changing the value to a lower one would not have affected the overall trend of the results. We also calculated the network density for these networks for each period.

The network density (D) was defined as the ratio of the number of actual links in the network to the total number of possible links in the network. Density was expressed by the following equation:

$$D = \frac{2m}{n(n-1)},$$

where *n* represents the number of nodes in the network and *m* represents the number of links. The density value varies within the range of 0 to 1, such that the closer the value is to 1, the higher the network density, indicating a close relationship. Conversely, values close to 0 indicate that there are few relationships in the network.

When the threshold is not set, the network density is equivalent to the mean cosine similarity between each node pair. In this case, the network density does not need to satisfy the condition that the cosine similarity is greater than 0.9.

Bootstrap Simulation

In this study, a clustering method and the DBSCAN method were used to reduce the distortion of obsidian provenance composition ratios caused by sampling effects on small samples. To test the effectiveness of this approach in reducing distortion and the robustness of the network in the clustered social network analysis, we conducted a simulation using the non-parametric bootstrap method on the data clustered with the DBSCAN method. In this study, we assessed whether the cosine similarity and network density derived from the current archaeological sample fall within the expected range of population cosine similarity and network density as estimated from bootstrap simulation. Furthermore, we examined whether clustering enhances the reduction of distortion and the robustness of the network by comparing the outcomes of bootstrap simulations for each cluster after clustering and for each site without clustering.

In this study, obsidian was randomly selected within each cluster, with the number of selections based on the total number of obsidian from each provenance in that cluster. Duplication was allowed, and the selection probability was based on the composition of obsidian stones from each provenance within each cluster. We then calculated the simulated cosine similarity and network density for the social network analysis. This simulation was repeated 100 times, and the mean and standard deviation of the cosine similarities and network densities from the 100 simulations were calculated and compared with the actual data.

Results and Discussion

Clustering

Based on the results of clustering using DBSCAN, some clusters were excluded from the analysis, as they contained less than 30 obsidian artifacts. For details of the number of clusters and single sites for each period, as well as the total number and composition ratios of obsidian artifacts by provenance, please refer to Sakahira and Tsumura (2023).

Table 1 shows the cosine similarity between clusters and between single sites and clusters for each period, which verified whether the clustering by DBSCAN ensured regional representativeness. The results showed that for each division of the Jomon period, the mean cosine similarity within each cluster was higher than the similarity between sites not belonging to the cluster. For example, in period 1, the mean cosine similarity of sites not belonging to a cluster (no cluster) was 0.280, which was lower than the values for B1, B2, B4, and B5.

The distribution of the cosine similarity of pairs between clusters in each period category is shown in dot and box plots in Figure 3. In the Beginning and Earlier Jomon periods, the cosine similarity between clusters is biased toward high and low pairs, while in the Early Jomon, there are more pairs with lower cosine similarity. However, the Middle Jomon has more pairs with high cosine similarity. In the Late and Last Jomon, pairs are evenly distributed between high and low pairs.

Figures 4–8 show the box plots of the actual cosine similarities between sites within each cluster at each period category, respectively. However, unlike Figure 3, dot plots are not shown because there are too many dots to display. In the Beginning and Earlier Jomon periods (Figure 4), the cosine similarity of site pairs within each cluster is distributed at higher values in the median and first and third quartiles compared to site pairs that do not belong to a cluster (no cluster). In the Early Jomon period (Figure 5), the cosine similarity of site pairs within each cluster is also distributed at higher values overall than in the Beginning and Earlier periods (Figure 4), and higher than site pairs that do not belong to a cluster (no cluster). In the Middle Jomon period (Figure 6), the cosine similarity of site pairs within each cluster is distributed at even higher values than in the previous periods (Figures 4 and 5), with outliers in clusters M2 and M4, but still higher than site pairs that do not belong to a cluster (no cluster), except in cluster M1. In the Late Jomon period (Figure 7), the cosine similarity of site pairs within each cluster is distributed at slightly lower values overall than in the Middle Jomon period (Figure 6), but higher in clusters except cluster L6 than in site pairs that do not belong to a cluster (no cluster). In the Last Jomon period (Figure 8), the cosine similarities of the site pairs within each cluster are all distributed to very high values and are higher than site pairs that do not belong to a cluster (no cluster).

Clusters M1 and L6 in the Middle (Figure 6) and Late Jomon periods (Figure 7), respectively, which have lower distributions than the cosine similarity of pairs of sites not belonging to a cluster, are both located in the midpoint of each obsidian provenance area. Therefore, it is considered that the existence of differences in obsidian source composition ratios at each site within these same clusters, owing to slight differences in location, may have resulted in combinations of sites with lower cosine similarity. Clusters M2 and M4 in the Middle Jomon period, which show many outliers in cosine similarity within the same cluster (Figure 6), were generated as clusters covering a wide geographical area for the DBSCAN algorithm (Figure 10), which may have resulted in pairs of sites that are further apart within a similar cluster having a lower cosine similarity and becoming outliers.

From these results, it can be inferred that nearby archaeological sites hold information on obsidian and the flow of obsidian between each site. It was thus reasonable to aggregate values between adjacent sites by region and analyze them from the perspective of regional representativeness.

Table 1 - Network density and cosine similarity within each cluster and between sites not belonging to a cluster in each period category.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon		Period 3 Middle Jomon		Period 4 Late Jomon		Period 5 Last Jomon	
Network density	0.444		0.200		0.405		0.143		0.256
Mean of actual cosine similarities between clusters	0.623		0.411		0.716		0.535		0.508
Mean of cosine similarities between sites not belonging to a cluster (no Cluster)	0.280		0.402		0.538		0.447		0.644
Mean of cosine similarities within a cluster	0.500		0.692		0.760		0.641		0.987
B1	0.760	E1	0.670	M1	0.421	L1	0.872	T1	0.983
B2	0.717	E2	0.752	M2	0.737	L2	0.800	T2	0.987
B4	0.552	E3	0.672	M3	0.892	L3	0.503	T3	0.984
B5	0.472	E5	0.576	M4	0.835	L4	0.495		
		E6	0.714	M5	0.644	L5	0.682		
		E7	0.767	M6	0.904	L6	0.483		
				M7	0.884	L7	0.650		

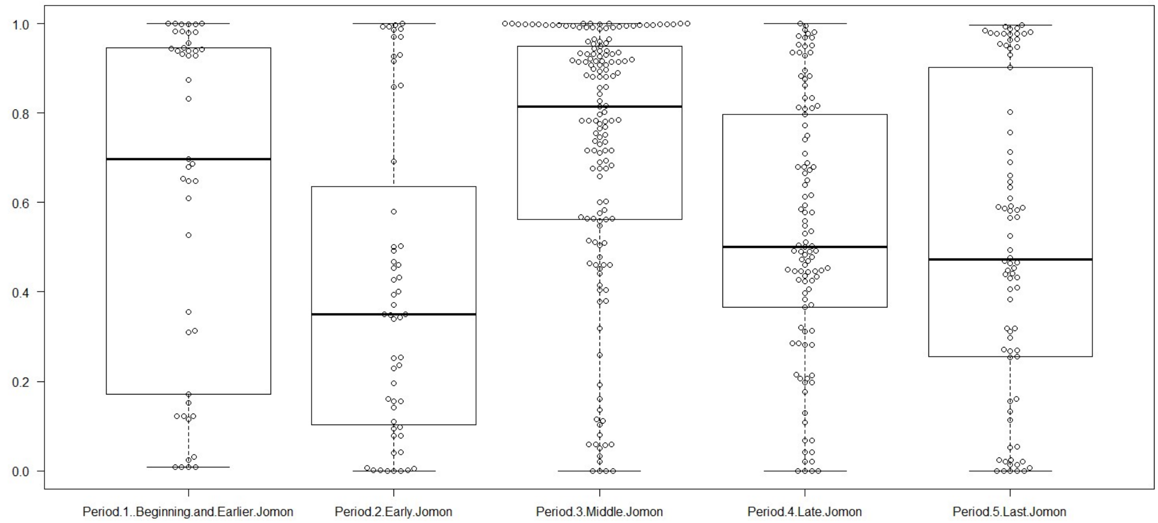

Figure 3 – Dot and Box plots of actual cosine similarities between clusters in each period category. Each dot represents a respective simulation value. The thick line in the middle of the box indicates the median, and the top and bottom of the box indicate the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

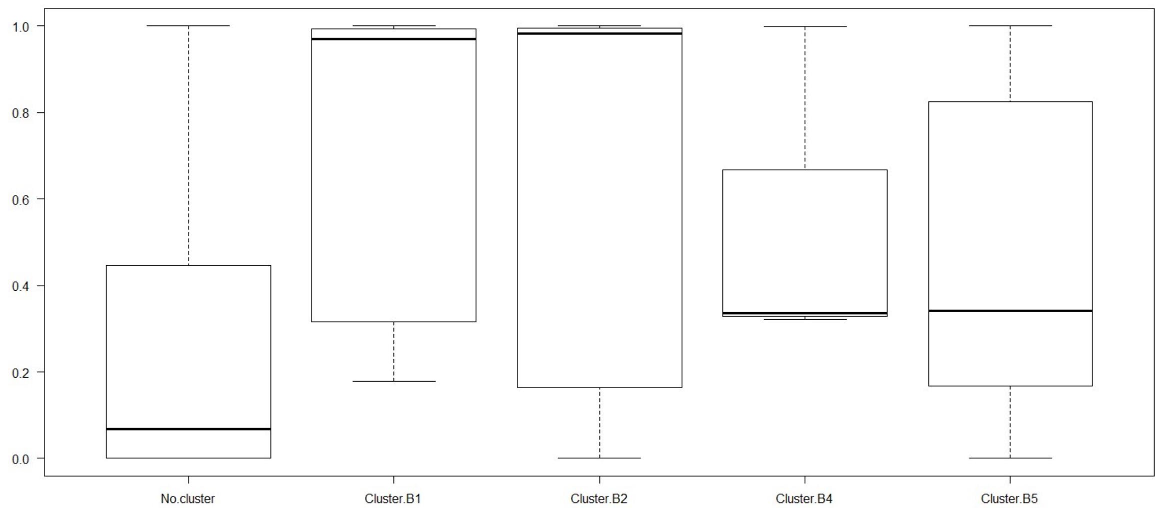

Figure 4 – Box plots of actual cosine similarities within sites not belonging to a cluster and clusters in Period 1 Beginning and Earlier Jomon. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

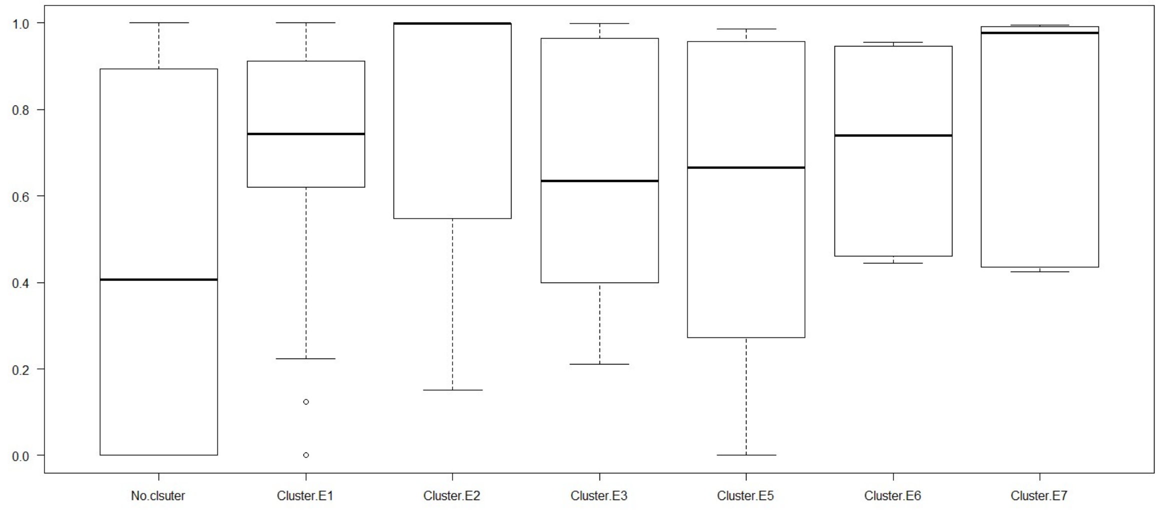

Figure 5 – Box plots of actual cosine similarities within sites not belonging to a cluster and clusters in Period 2 Early Jomon. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

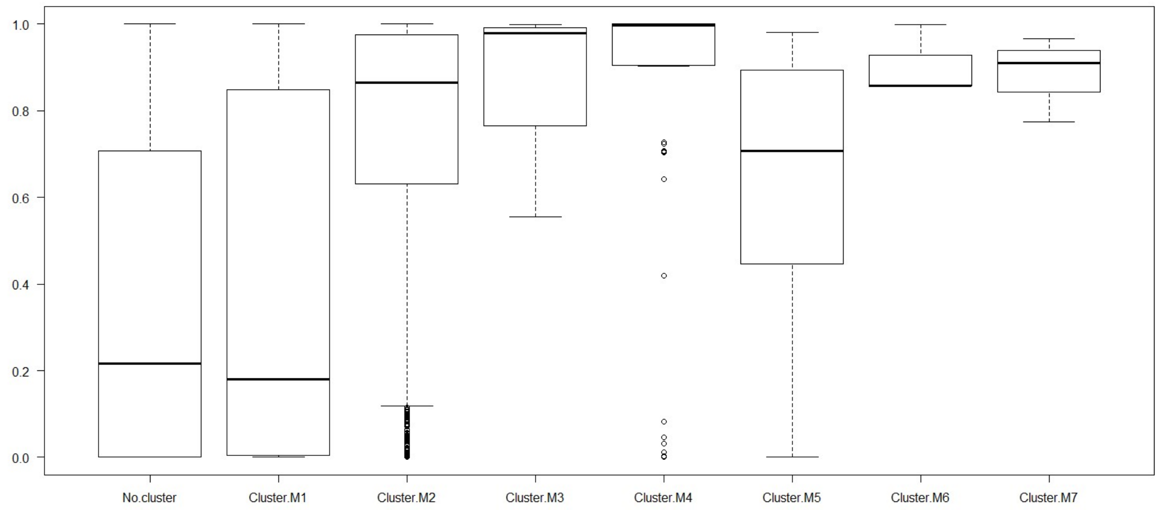

Figure 6 – Box plots of actual cosine similarities within sites not belonging to a cluster and clusters in Period 3 Middle Jomon. The thick line in the middle of the box indicates the median, the top and bottom of the box the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile), respectively.

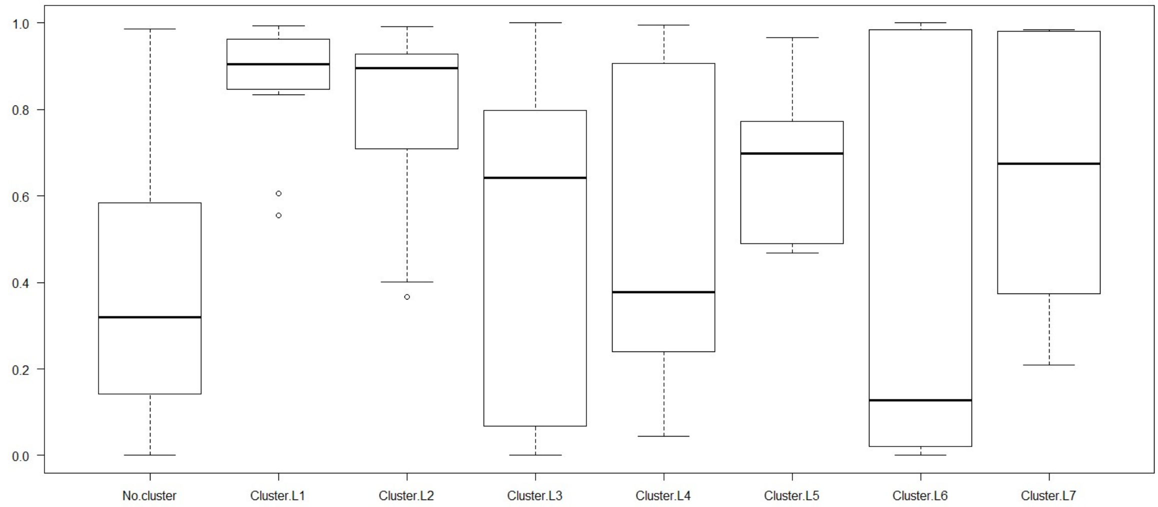

Figure 7 – Box plots of actual cosine similarities within sites not belonging to a cluster and clusters in Period 4 Late Jomon. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

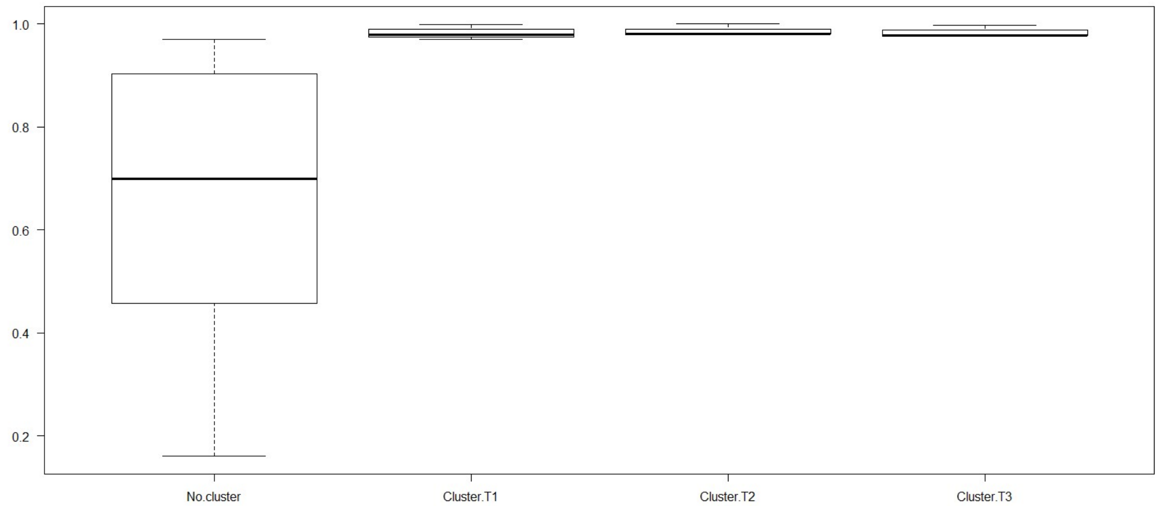

Figure 8 – Box plots of actual cosine similarities within sites not belonging to a cluster and clusters in Period 5 Last Jomon. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

Social Network Analysis

The results of the social network analysis were described in our previous study (Sakahira & Tsumura, 2023). However, in this study, we have created a new pie chart to show the composition of obsidian from different provenances in each cluster, and we have added it to the network analysis. Therefore, this study focuses on the compositional ratios of obsidian from each provenance, mainly presented as pie charts.

In the Early Jomon Period, each cluster contained obsidian from nearby provenances. For example, clusters E5 and E7 and site e8 in the coastal area were dominated by obsidian from Kozu-shima, an island product, while cluster E2 and site e10 were dominated by obsidian from nearby Omegura, and clusters E1 and E3 were dominated by obsidian from nearby Suwa (Figure 9).

In the Middle Jomon period, obsidian from island provenances spread throughout the southern Kanto region. Except for cluster M3 and site m15, the majority of clusters and sites had over one-third of their obsidian coming from Kozu-shima (Figure 10).

In the Late Jomon period and beyond, the distribution of obsidian from island provenances became limited, and obsidian from inland provenances began to appear. Clusters L3, L5, and L6, and some surrounding sites were dominated by obsidian from nearby Suwa, while clusters L1 and L7 and site l11 were dominated by obsidian from nearby Takahara-yama (Figure 11).

Additionally, we discovered that the network density between clusters and the cosine similarity between sites within clusters during the Middle Jomon Period (Table 1) were higher than those before the Early Jomon Period and after the Late Jomon Period. These results suggest that the obsidian trading network developed throughout the southern Kanto region during the Middle Jomon Period and ceased to function during the later period. For more details of these analyses, please refer to Sakahira and Tsumura (2023).

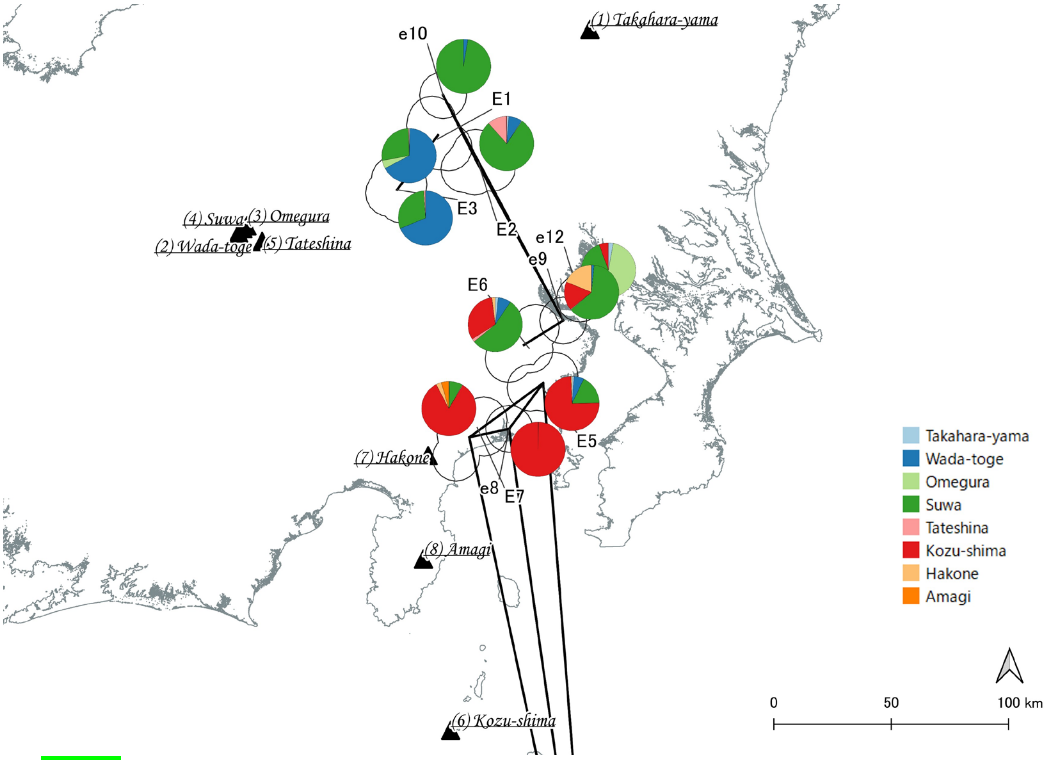

Figure 9 - Network among the consumption areas in Period 2, the early Jomon period (7,000–5,500 years cal BP). Clusters are represented by uppercase characters and single sites by lowercase characters. Pairs with a cosine similarity greater than 0.9 in the composition ratio of each provenance area are linked. White circles indicate clustered areas. Pie charts show the composition ratio of each cluster by provenance.

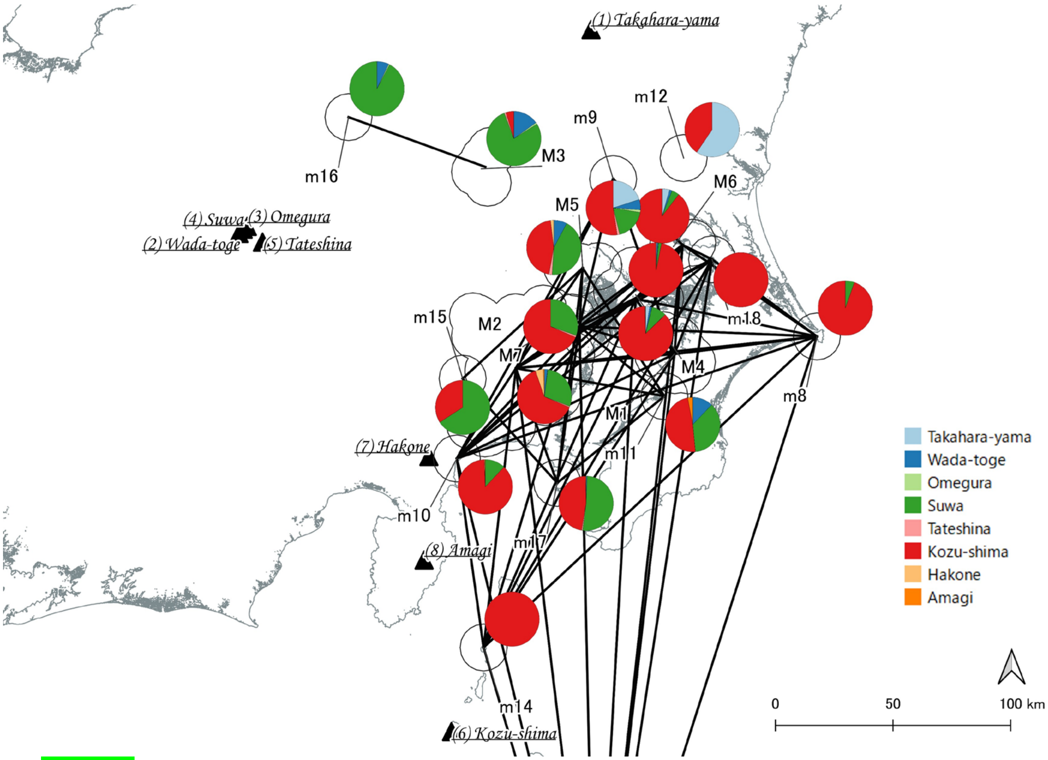

Figure 10 - Network among the consumption areas in Period 3, the middle Jomon period (5,500–4,500 years cal BP). Clusters are represented by uppercase characters and single sites by lowercase characters. Pairs with a cosine similarity greater than 0.9 in the composition ratio of each provenance area are linked. White circles indicate clustered areas. Pie charts show the composition ratio of each cluster by provenance.

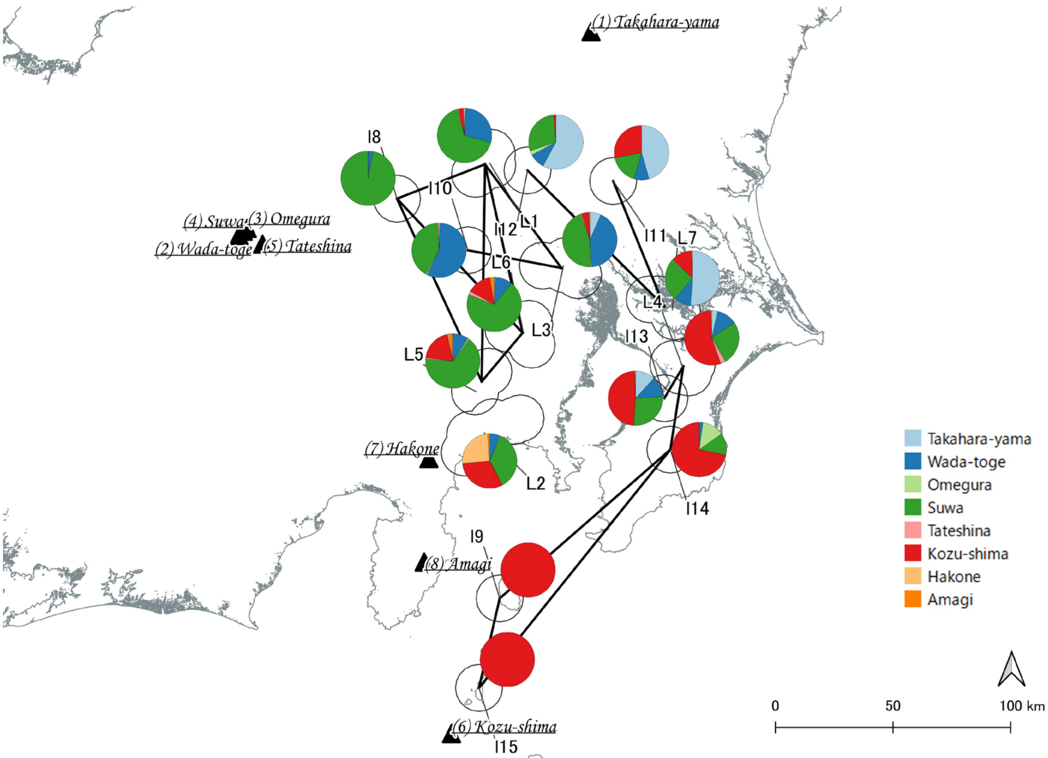

Figure 11 - Network among the consumption areas in Period 4, the late Jomon period (4,500–3,200 years cal BP). Clusters are represented by uppercase characters and single sites by lowercase characters. Pairs with a cosine similarity greater than 0.9 in the composition ratio of each provenance area are linked. White circles indicate clustered areas. Pie charts show the composition ratio of each cluster by provenance.

Bootstrap Simulation

One hundred simulations were performed using the bootstrap method, both from cluster-by-cluster aggregation results after clustering using the DBSCAN algorithm and from site-by-site aggregation results without clustering. The results of each simulation were used to calculate the cosine similarity for each period category. The distribution of the cosine similarity of pairs between clusters after clustering by the DBSCAN algorithm and the cosine similarity of pairs between sites without clustering are shown in dot and box plots in Figures 12 and 13, respectively. Comparing these, except for the Early Jomon period, the cosine similarity values differ significantly without and after clustering.

The width of the distribution of cosine similarity in the simulation appears to be narrower without clustering than after clustering, except for the Last Jomon period. This is also confirmed by the standard deviations in Tables 2 and 3. Specifically, in the Beginning and Earlier Jomon periods, the standard deviation of the cosine similarity in the simulations after clustering was 0.016, compared to 0.015 without clustering. In the Early Jomon period, both values were equal to 0.012; nevertheless, in the Middle Jomon period, the latter (0.010) was smaller than the former (0.016). In the Late Jomon period, the latter (0.010) was smaller than the former (0.014). However, in the Last Jomon period, the latter (0.022) was greater than the former (0.014). The standard deviation of the cosine similarity was smaller without clustering than after clustering, except in the Last Jomon Period. This is because after clustering, the sites were grouped, and the number of pairs of sites for which the cosine similarity was measured was smaller than that without clustering. For example, in the Middle Jomon Period, which has the largest differences, 149 sites existed, and the number of cosine similarity pairs was 11,026 without clustering. However, with clustering, seven clusters and 11 sites existed, and the number of cosine similarity pairs was 153. Therefore, the larger the number of pairs, the more stable the value of the standard deviation. In the Last Jomon Period, 26 sites existed, and the number of cosine similarity pairs was 325 without clustering. However, with clustering, three clusters and ten sites existed, and the number of cosine similarity pairs was 78. Thus, in the Last Jomon period, the standard deviation was not higher after clustering than without clustering, probably because the number of pairs decreased less after clustering.

Additionally, as mentioned earlier, there is a difference in the mean values of the cosine similarity after and without clustering; thus, the coefficient of variation was calculated to assess their variability relative to each other (Tables 2 and 3). The coefficient of variation is the value of the standard deviation divided by the mean value. Even after calculating the coefficient of variation, the after-clustering values remained equal to the without clustering values, or the latter was smaller than the former, except in the Last Jomon Period.

Although the effect of clustering is difficult to observe when only examining the variation in the simulation values, the effect of clustering becomes evident when comparing actual and simulated values. For example, in the Beginning and Earlier Jomon Periods, without clustering, the actual cosine similarity between sites was 0.493, and the mean of the simulation was 0.477 (Tables 2 and 3), with a difference of 0.016 (Table 4). However, after clustering, the actual cosine similarity between clusters was 0.623, and the mean of the simulation was 0.614 (Tables 2 and 3), with a difference of 0.009 (Table 4). Thus, in all period categories, the difference between actual and simulated values was better after clustering than without clustering (Table 4). Moreover, in the bootstrap simulations that after clustering, the actual cosine similarity values were within one standard deviation of the mean of the cosine similarity values from 100 simulations across all category periods (Table 2). Conversely, in the bootstrap simulations without clustering, the actual cosine similarity did not fall within one standard deviation of the mean of the cosine similarity values from 100 simulations in any of the periods, with the exception of the final Jomon period (Table 3). These suggest that clustering by region can reduce the distortion of obsidian provenance composition ratios due to sampling effects on small samples.

Network densities based on cosine similarity calculated by 100 bootstrap methods for both the per-cluster composition ratio after clustering by the DBSCAN algorithm and the per-site composition ratio without clustering were calculated. The distribution of network density among clusters after clustering and among sites without clustering is shown in dot and box plots in Figures 14 and 15, respectively. Comparing these, except for the Early and Late Jomon Periods, the cosine similarity values differ significantly without and after clustering.

The width of the distribution of network density in the simulation also appears to be narrower without clustering than after clustering, except for the Last Jomon Period. This is also confirmed by the standard deviations in Tables 5 and 6. Specifically, in the Beginning and Earlier Jomon periods, the standard deviation of the network density of the simulations after clustering was 0.056, compared to 0.032 without clustering. In the Early Jomon period, the latter (0.014) was smaller than the former (0.020). In the Middle Jomon period, the latter (0.012) was smaller than the former (0.034). In the Late Jomon period, the latter (0.010) was smaller than the former (0.016). However, in the Last Jomon period, the latter (0.037) was greater than the former (0.010). Except for the Late Jomon Period, the standard deviation of the network density without clustering was smaller than the value after clustering, for reasons similar to those of the cosine similarity described above. Additionally, the coefficient of variation was smaller without clustering than after clustering, except for the Late Jomon Period.

The difference between actual network density and simulated values after and without clustering confirms the effect of clustering (Table 7). For example, in the Beginning and Earlier Jomon Periods, the difference without clustering was 0.034. By contrast, the difference after clustering was 0.031. For all period categories, the difference between actual and simulated values was equal to or better after clustering than without clustering (Table 7). However, the clustering effect was smaller in network density (Table 7) than in cosine similarity (Table 4). The reasons for this could not be elucidated in detail in this paper; nonetheless, as mentioned earlier, it may be related to the fact that network indicators are robust to the removal of nodes, as mentioned by Wey et al. (2008).

In the bootstrap simulations after clustering, the actual network density was within one standard deviation of the mean network density from the 100 simulations for all periods, with the exception of the Last Jomon Period (Table 5). Conversely, in the bootstrap simulations without clustering, the actual network density was not within one standard deviation of the mean network density from the 100 simulations for either the Beginning and Earlier Jomon Periods or the Last Jomon Period (Table 6).

These results showed that the social network analysis of the network after clustering using the DBSCAN algorithm had high robustness. The results also confirmed that this study's sampling had little effect on its results. Therefore, it is suggested that the DBSCAN clustering method used in this study is applicable to other archaeological themes where missing data and sampling effects are issues.

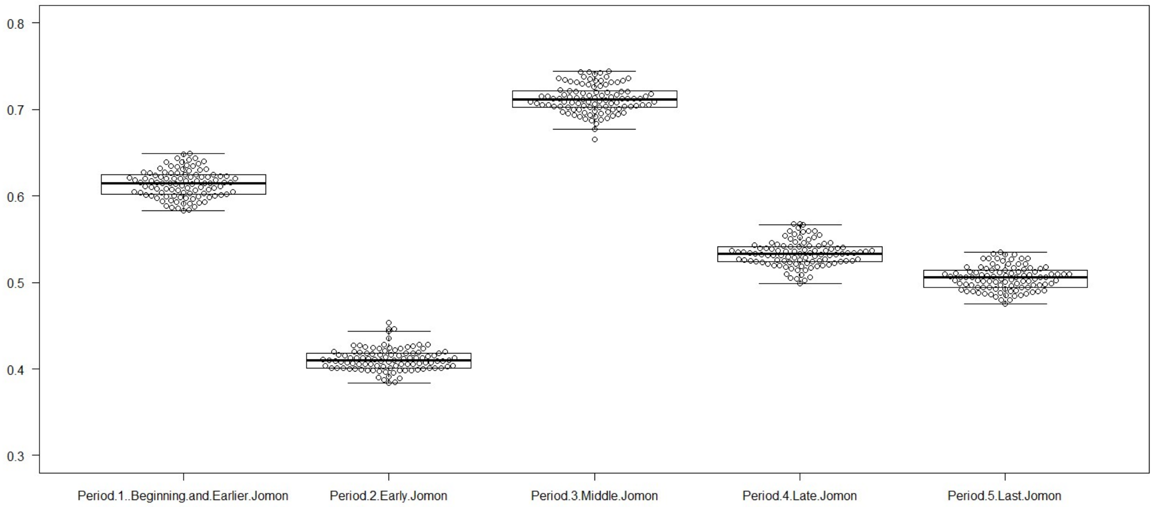

Figure 12 – Dot and Box plots of simulated cosine similarities between clusters after clustering in each period category. Each dot represents a respective simulation value. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

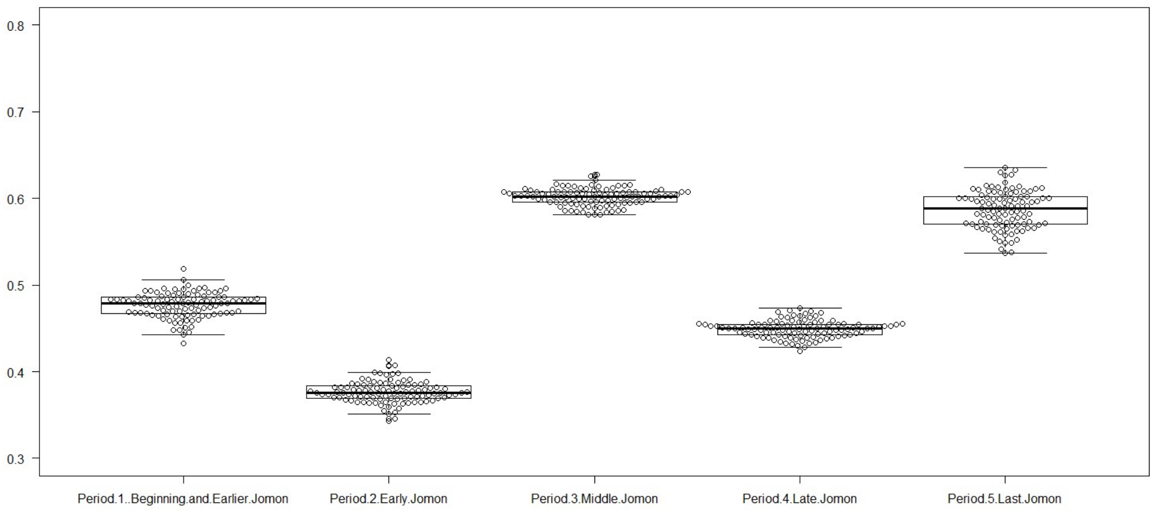

Figure 13 – Dot and Box plots of simulated cosine similarities between sites without clustering in each period category. Each dot represents a respective simulation value. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

Table 2 - Comparison of actual and bootstrap simulation values for cosine similarity between clusters after clustering.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon	Period 3 Middle Jomon	Period 4 Late Jomon	Period 5 Last Jomon
Mean of actual cosine similarities between clusters	0.623	0.411	0.716	0.535	0.508
Mean of simulated cosine similarities between clusters	0.614	0.411	0.712	0.533	0.505
Simulated standard deviation	0.016	0.012	0.016	0.014	0.014
Coefficient of variation	0.026	0.029	0.023	0.026	0.028

Table 3 - Comparison of actual and bootstrap simulation values for cosine similarity between sites without clustering.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon	Period 3 Middle Jomon	Period 4 Late Jomon	Period 5 Last Jomon
Mean of actual cosine similarities between sites	0.493	0.396	0.620	0.466	0.608
Mean of simulated cosine similarities between sites	0.477	0.377	0.602	0.449	0.587
Simulated standard deviation	0.015	0.012	0.010	0.010	0.022
Coefficient of variation	0.031	0.029	0.013	0.022	0.032

Table 4 - Difference between the actual cosine similarity and the mean of the simulation.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon	Period 3 Middle Jomon	Period 4 Late Jomon	Period 5 Last Jomon
After clustering: Difference between the actual cosine similarities between clusters and the mean of the simulation	0.009	0.000	0.004	0.002	0.003
Before clustering: Difference between the actual cosine similarities between sites and the mean of the simulation	0.016	0.019	0.018	0.017	0.021

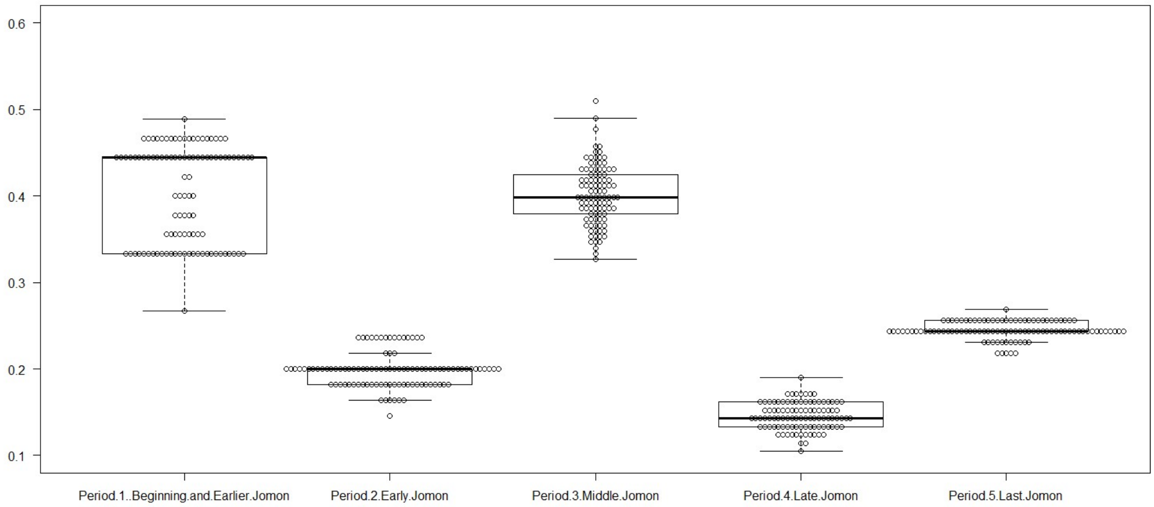

Figure 14 – Dot and Box plots of simulated network density among clusters after clustering in each period category. Each dot represents a respective simulation value. The thick line in the middle of the box indicates the median, and the top and bottom of the box represent the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

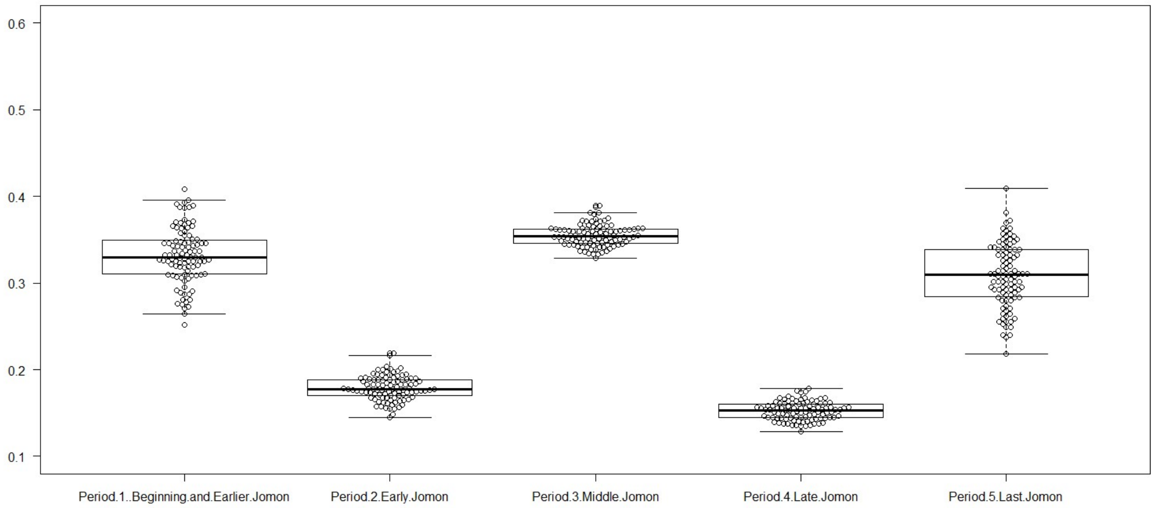

Figure 15 – Dot and Box plots of simulated network density among all sites without clustering in each period category. Each dot represents a respective simulation value. The thick line in the middle of the box indicates the median, and the top and bottom of the box the first and third quartiles, respectively. The bar above the box indicates the range of the first quartile - 1.5* (third quartile - first quartile) and the bar under the box indicates the range of the third quartile + 1.5* (third quartile - first quartile).

Table 5 - Comparison of actual and bootstrap simulation values for network density among all clusters after clustering.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon	Period 3 Middle Jomon	Period 4 Late Jomon	Period 5 Last Jomon
Actual network density	0.444	0.200	0.405	0.143	0.256
Mean of simulated network density	0.403	0.198	0.402	0.146	0.245
Simulated standard deviation	0.056	0.020	0.034	0.016	0.010
Coefficient of variation	0.138	0.101	0.086	0.107	0.042

Table 6 - Comparison of actual and bootstrap simulation values for network density among all sites without clustering.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon	Period 3 Middle Jomon	Period 4 Late Jomon	Period 5 Last Jomon
Actual network density	0.366	0.170	0.358	0.159	0.351
Mean of simulated network density	0.332	0.179	0.355	0.152	0.309
Simulated standard deviation	0.032	0.014	0.012	0.010	0.037
Coefficient of variation	0.096	0.078	0.035	0.068	0.120

Table 7 - Difference between the actual network density and the mean of the simulation.

Period 1 Beginning and Earlier Jomon		Period 2 Early Jomon	Period 3 Middle Jomon	Period 4 Late Jomon	Period 5 Last Jomon
After clustering: Difference between the actual network density among all clusters and the mean of the simulation	0.031	0.002	0.003	0.003	0.011
Before clustering: Difference between the actual network density among all sites and the mean of the simulation	0.034	0.009	0.003	0.007	0.042

Conclusion and Future Work

This study's social network analysis of obsidian artifacts revealed that the trade networks during the Jomon period were not constant, but rather developed throughout the southern Kanto region during the middle Jomon period and ceased to function in the late Jomon period. The use of DBSCAN clustering improved the readability and interpretability of the large dataset and reduced the bias caused by the small sample sizes of each site, thus confirming the validity of analyzing regional representation. Finally, a bootstrap simulation analysis demonstrated the high robustness of the network in the social network analysis after clustering. The impact of sampling on the results of this study was found to be minimal.

In the future, ancient digital elevation data in GIS should be used to consider the ε value of DBSCAN and the geographical distance between production and consumption areas more accurately, as well as to extract regional clusters and calculate the shortest transportation costs between production and consumption areas. This will enable us to determine the shortest distance or route, taking into consideration geographical features such as elevation differences, slopes, and seas (Ladefoged et al., 2019; Tobler, 1993). We plan to address these points as future research tasks.

Acknowledgments

We would like to thank Prof. T. Terano (Chiba University of Commerce), and Dr. M. Kunigami (Tokyo Institute of Technology) for the useful discussions. We would like to thank Editage (www.editage.com) for English language editing.

Funding

This work was supported by JSPS KAKENHI (Grant nos. 21K21323, 22K18156, and 22H00021), Japan. The funding source was not involved in preparing the manuscript or in the collection, analysis, or interpretation of the data.

Conflict of Interest

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, Scripts, Code, and Supplementary Information

Our study used the dataset from Nihon-kokogaku-kyokai 2011 nendo tochigi-taikai-jikkoiinkai (2011). The dataset is part of their research and therefore cannot be included in our proceedings. Instead, we have provided sample data that can be used to validate the R scripts.

References

Akazawa T (1982) Maritime Adaptation of Prehistoric Hunter-Gatherers and Their Transition to Agriculture in Japan. *Senri Ethnological Studies*, **9**, 213–258.

Daikuhara Y (2008) *Jomon-sekki-kenkyu-josetsu*. Tokyo: Rokuichi Shobo. (In Japanese).

Ester M, Kriegel H, Sander J, Xu X. (1996, August 2–4) A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon.

Freund KP (2013) An Assessment of the Current Applications and Future Directions of Obsidian Sourcing Studies in Archaeological Research. *Archaeometry*, **55**, 779–793. https://doi.org/10.1111/j.1475-4754.2012.00708.x

Gjesfjeld, E (2015) Network Analysis of Archaeological Data from Hunter-Gatherers: Methodological Problems and Potential Solutions. *Journal of Archaeological Method and Theory,* **22**, 182–205. https://doi.org/10.1007/s10816-014-9232-9

Golitko M, Feinman GM (2015) Procurement and Distribution of Pre-Hispanic Mesoamerican Obsidian 900 BC–AD 1520: A Social Network Analysis. *Journal of Archaeological Method and Theory,* **22**, 206–2547. https://doi.org/10.1007/s10816-014-9211-1

Golitko M, Meierhoff J, Feinman GM, Williams PR (2012) Complexities of Collapse: The Evidence of Maya Obsidian as Revealed by Social Network Graphical Analysis. *Antiquity*, **86**, 507–23. https://doi.org/10.1017/S0003598X00062906

Hashiguchi N (1999) *Umi o watatta Jomon jin*. Tokyo: Shogakukan. (In Japanese).

Ikeya N (2009) *Kokuyoseki-kokogaku*. Tokyo: Shinsensha. (In Japanese).

Kanayama Y (1994) Jomon jidaizenki ni okeru kokuyoseki-koeki no shutsugen. *Hosei-kokogaku*, **17**, 61–65. (In Japanese).

Koizumi M (2016) Chugoku Shikoku chiho niokeru Jomon kaizuka no tayosei ni kansuru kisoteki-kosatsu. *The Bulletin of the Faculty of Law and Letters: Humanities*, **40**, 75–118. (In Japanese).

Kojo Y (1996) Jomon chuki niokeru Shinshu-san kokuyoseki no Minami Kanto heno hannyuro. *J Archaeological Soc Nippon*, **81**, 340–350. (In Japanese).

Ladefoged TN, Gemmell C, McCoy M, Jorgensen A, Glover H, Stevenson C, O'Neale D (2019) Social Network Analysis of Obsidian Artefacts and Maori Interaction in Northern Aotearoa New Zealand. *PLoS One*, **14**, e0212941. https://doi.org/10.1371/journal.pone.0212941

Meissner, N (2017) A Social Network Analysis of the Postclassic Lowland Maya Obsidian Projectile Industry. *Ancient Mesoamerica*, **28**, 137–56. https://doi.org/10.1017/S0956536116000390

Mills BJ, Clark JJ, Peeples MA, Haas WR, Roberts JM, Hill JB, Huntley DL, Borck L, Brieger RL, Clauset A, Shackley MS (2013) Transformation of Social Networks in the Late Pre-Hispanic US Southwest. *Proceedings of the National Academy of Sciences*, **110**, 5785–5790. https://doi.org/10.1073/pnas.1219966110

Nihon-kokogaku-kyokai 2011 nendo tochigi-taikai-jikkoiinkai (2011) Sekki-jidai niokeru sekizai-riyo no chiiki-so. In: *2011 nendo tochigi-taikai-kenkyuhappyo-shiryoshu*, 7–306. (In Japanese).

Ono A (2011) Obsidian in the Natural Resource Environment: A Methodological Perspective. *Natural Resource Environment and Humans*, **1**, 1–8. (In Japanese with English abstract).

Roberts JM, Yin Y, Dorshorst E, Peeples MA, Mills BJ (2021) Assessing the Performance of the Bootstrap in Simulated Assemblage Networks. *Social Networks,* **65**, 98–109. https://doi.org/10.1016/j.socnet.2020.11.005

Sakahira F, Tsumura H (2023) Tipping Points of Ancient Japanese Jomon Trade Networks from Social Network Analyses of Obsidian Artifacts. *Frontiers in Physics*, **10**, 1015870. https://doi.org/10.3389/fphy.2022.1015870

Sugihara S, Kobayashi S (2008) Scientific Analysis of an Obsidian Source and its Distribution, with Special Reference to Obsidian Quarried in the Kozu Island, off the Pacific Coast of Japan. *Memoirs of the Institute of Humanities, Meiji University*, **62**, 97–229. (In Japanese with English abstract).

Suzuki M (1973) Chronology of Prehistoric Human Activity in Kanoto, Japan. Part I. *Journal of the Faculty of Science, University of Tokyo*, V, 4(3), 241–318.

Suzuki M (1974) Chronology of Prehistoric Human Activity in Kanoto, Japan. Part II. *Journal of the Faculty of Science, University of Tokyo*, V, 4(4), 395–469.

Tateishi T (2010) *Study on Material and Information Exchange in the Jomon Period: Analysis of Jomon Pottery and Obsidian Lithics Using Tcientific techniques* (Dissertation/Ph.D. thesis). Sakura (Chiba): The Graduate University for Advanced Studies SOKENDAI. (In Japanese with English Abstract).

Tobler W (1993) *Three Presentations on Geographical Analysis and Modeling (National Center for Geographic Information and Analysis Technical Report 93-1)*. Santa Barbara, CA: University of California, Santa Barbara.

Tsumura H, Tateishi T (2013) Transition of the Network of the Obsidian Distribution in Kanto Region, the Jomon Period. *Zooarchaeology*, **30**, 377–393. (In Japanese with English abstract).

Tsutsumi T (2018) Shinshukokuyosekigensanchi no shigenkaihatsu to kyokyu o megutte. *Shimanekenkodaibunka Cent kenkyuronshu*, **19**, 153–168. (In Japanese).

Warashina T, Higashimura T (1988) Sekki-genzai no sanchi-bunseki. In: *Kokogaku to KanrenkagakuKamaki yoshimasa sensei koki-kinen-ronshu-kankokai*, p. 447–491. (In Japanese).

Wey T, Blumstein DT, Shen W, Jordán F (2008) Social Network Analysis of Animal Behaviour: A Promising Tool for the Study of Sociality. *Animal Behaviour*, **75**, 333–44. https://doi.org/10.1016/j.anbehav.2007.06.020