

Response to PCI Archaeology Reviewers' Comments on Our Manuscript

Dear Editor and Reviewers:

Thank you for taking the time to review our manuscript and providing insightful comments. Our corresponding responses are presented below.

Kind regards,
Sakahira et al.

For editor: Professor James Allison

>In my opinion, the manuscript should be revised to address the issues raised by the reviewers. In particular, it should make clear how the current study relates to the previously published work, and focus more on the novel contributions of the current study. Adding details about exactly how the bootstrap procedure was implemented, providing a better explanation of what was gained by using the bootstrap, and providing more explanation about what the current study adds to already published work would make the manuscript much stronger.

Our response

Thank you for your advice. We have made significant additions and revisions regarding the content on bootstrapping. As a result, we believe that the current paper significantly diverges from our previous study.

Below, we have addressed the comments from each reviewer.

The modified parts are indicated by the green fluorescent marker in our tracked changes file.

For reviewer: Professor Matthew Peeples

>Overall, I think this is a potentially useful approach that would be relevant in a range of settings where archaeological network methods have frequently been applied. I think the text is relatively easy to follow for the most part, with a couple of places where I think some additional technical detail is necessary to avoid confusion. In addition to this, I make a few additional suggestions below regarding wording and very nitty gritty details that could help avoid any confusion. Also, in the current draft I'm not sure if a convincing argument is made for why the bootstrap test helps to evaluate the clusters as defined specifically, but I could be missing something where clarifications are suggested below. Also, in the current draft I'm not sure if a convincing argument is made for why the bootstrap test helps to evaluate the clusters as defined specifically, but I could be missing something where clarifications are suggested below.

Our response

Thank you for your meticulous evaluation and insightful comments on our paper. We have addressed all your comments (R1.1–R1.11) below.

The modified parts are indicated by the green fluorescent marker in our tracked changes file.

Your comment R1.1

>I would suggest that work by Gjesfjeld (2015) is also particularly relevant to the current study and it might be useful to reference this:

Gjesfjeld, Erik 2015 Network Analysis of Archaeological Data from Hunter-Gatherers: Methodological Problems and Potential Solutions. *Journal of Archaeological Method and Theory* 22(1):182–205.

Another potentially relevant study by Roberts et al. (full-disclosure, I'm also an author on this study but do think it's relevant here) is perhaps useful:

John M. Roberts, Yi Yin, Emily Dorshorst, Matthew A. Peeples, Barbara J. Mills. 2021. Assessing the performance of the bootstrap in simulated assemblage networks, *Social Networks*, 65:98-109.

Our response to your comment R1.1

Thank you for providing the useful references. I have included the following brief additions to each of the papers.

Line 123–124

"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)."

Line 130–138

“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.”

Your comment R1.2

>What really needs to be outlined here is how were the data tables randomized for the replicates in the bootstrap (row-wise, column-wise, both). Since the term "bootstrap" is often used for a range of procedures (especially by archaeologists), it would be useful to describe here what was done. For example, when I went to the code, I could see that sampling was done with replacement which is the typical definition of the nonparametric bootstrap, but there is enough variation in the archaeological literature on this that I think the details need to be made explicit in the text here.

Our response to your comment R1.2

We apologize for any confusion caused. We have carried out bootstrapping for each column based on obsidian provenances. To enhance understanding, we have included a description of the bootstrapping process in the text, revised the R script accordingly, and incorporated a comment (refer to “Bootstrap_for_CAA2022.R”).

Line 242–258

“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.”

Your comment R1.3

>Further, in the "kara_df" object created in the simulation sample size by row was held constant which is an important point that needs to be discussed in more detail. I have further comments on the code below but I also think adding additional comments in the R code would help with these missing details as well.

Our response to your comment R1.3

We have modified the R code and added a comment.

Additionally, we have merged the bootstrap R scripts, which were divided into two files due to differences in the target, into one, and added the calculation of the network density (refer to “Bootstrap_for_CAA2022.R”).

Your comment R1.4

>A more informative approach might be to for example show a boxplot/dotplot or similar visual of within cluster, between cluster, and no cluster values from which these means are derived. This would allow for the assessment and discussion of any outliers or other interesting properties of the distributions. As the number of points are fairly small a visual like a dotplot would probably work better than alternatives.

Our answer to your comment R1.4

Based on your advice, we have included Figures 3–8. Please note that only box plots are available for the cosine similarity within each cluster (Figures 4 to 8) because the number of dots is too large for display in the figures.

Furthermore, we have added the following text to the results section concerning these figures.

Line 272–300

“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.”

Your comment R1.5

>As for figures 3 through 5, if color is possible, it would be useful here as well.

Our response to your comment R1.5

Based on your suggestion, we have colored Figures 9–11.

Additionally, we have thoroughly revised the text pertaining to those figures as follows:

Line 349–369

“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).”

Your comment R1.6

>Similarly, in the discussion of the bootstrap simulation, it would be useful to see distributions of values rather than just means and standard deviations. Overall, I had the most trouble understanding what had actually been done with the bootstrap test and how this relates to the "stability" of cluster

solutions. This section/topic needs considerable more description for clarity. Specifically, if you resample data with replacement a bunch of times for a given cluster solution and get similar results, what does that tell you about the validity of that cluster solution and why? I think I'm missing something here.

Our response to your comment R1.6

Based on your comment, we have added figures (Figures 12–15) depicting the distribution of values related to the results of the bootstrap simulation. Furthermore, we have included references to these figures in the text as follows. As a result, we have made substantial additions to this section.

Line 391–465

“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).’

Furthermore, the following statement about the purpose of bootstrapping was also added to Section “Materials and Methods”.

Line 246–251

“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.”

Your comment R1.7

>Specifically, if you resample data with replacement a bunch of times for a given cluster solution and get similar results, what does that tell you about the validity of that cluster solution and why? I think I’m missing something here.

Our response to your comment R1.7

We apologize for the lack of clarity. To address your question, we have added the following sentence.

Line 242–251

“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.”

Your comment R1.8

>I would recommend that the authors provide the data used in the final version. If there is some reason that this cannot be done, that reason should be explained in the text and sample data of a similar format should be provided instead so that the code is at least testable.

Our response to your comment R1.8

Original data cannot be provided, and the reasons for this are described in the Data, Scripts, Code, and Supplementary Information section, as quoted below, along with sample data that can be used for code verification.

Line 544–545

“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.”

Your comment R1.9

>I could be wrong here, but since this only outputs the mean and not the full distribution the line "sd(simil_vec)" then is the standard deviation of the means across the sampling distribution or the standard error of the mean rather than the SD of values. I could be missing something here but that detail and wording should be checked.

Our response to your comment R1.9

We apologize for any confusion. We have calculated the mean and standard deviation from 100 simulations of the mean, whether within or between clusters, for cosine similarity.

Additionally, we have included comments in the R script (refer to “Bootstrap_for_CAA2022.R”).

Your comment R1.10

>I think if details of the bootstrap and its puprose are described in greater detail in the text that would be helpful.

Our response to your comment R1.10

Based on your comment, we have added the following sentences to provide more information on bootstrapping.

Line 242–258

“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.”

Your comment R1.11

>Minor things:

A few minor things I noticed...

Line 168 - "minPts is the optimal size of the minimum cluster". Isn't it just the minimum size? I wasn't sure what optimal meant here.

Line 198 - The equation text here is a bit confusing as it isn't the standard vector notation most are probably used to. Since cosine similarity is the dot product of two vectors divided by their norms, I would suggest denoting vectors and magnitudes in typical vector notation. For example, in LaTeX it would be:
$$S_c(a,b) = \frac{a \cdot b}{\|a\| \|b\|}$$

where the dot product of vectors a and b are the numerator and the denominator is the product (not dot product as indicated here) of the magnitudes of a and b.

Line 208, Could you provide a little text on why a particular threshold was selected for links?

Our response to your comment R1.10

We have corrected each of these. The corresponding lines are as follows.

Line 187–188

We have added the following sentence.

"The minPts was set to a minimum requirement of three, which is essential for cluster growth in the DBSCAN algorithm."

Line 197, 216, and 232

The description of the respective formulae has been corrected.

Line 226–228

We have added the following sentence.

"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."

For anonymous reviewer

>As far as I can tell the only original contribution of the current manuscript is the section on the bootstrap simulation. That is 9 lines of the manuscript.

I will leave it to the editors to decide whether this level of repetition with a previously published article is too much and prevents the publication of the manuscript in the conference proceedings (for me, it seems like it is), but at the very least the authors should be much more explicit about the similarities between this manuscript and their previously published work, and the novel contribution of this manuscript.

Our response

Thank you for your comment, and we appreciate your concerns. To address both your concerns and the comments received from Professor Matthew Peeples, we have made significant additions. These additions distinguish this work from our previous paper, and we believe we have effectively addressed your concerns.

The modified parts are indicated by the green fluorescent marker in our tracked changes file.

- Line 241–258: Subsection “Bootstrap Simulation” in Section “Materials and Methods.”
- Line 272–300: Subsection “Clustering” in Section “Results and Discussion.”
- Line 348–369: Subsection “Social Network Analysis” in Section “Results and Discussion.”
- Line 390–465: Subsection “Bootstrap Simulation” in Section “Results and Discussion.”
- Eleven new figures and four new tables were created, and two figures were revised.
 - Newly created: Figures 3–8, Figures 12–15, and Tables 3, 4, 6, and 7.
 - Revised: Figures 9–11

We then specified the differences from our previous study as follows.

Line 144–152

“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."*