



A different approach to similarity networks in Archaeology - Similarity Network Fusion

Joel Santos based on peer reviews by **Matthew Peeples** and 1 anonymous reviewer

Timo Geitlinger (2024) Similarity Network Fusion: Understanding Patterns and their Spatial Significance in Archaeological Datasets. Zenodo, ver. 3, peer-reviewed and recommended by Peer Community in Archaeology. <https://doi.org/10.5281/zenodo.7998239>

Submitted: 02 June 2023, Recommended: 02 April 2024

Cite this recommendation as:

Santos, J. (2024) A different approach to similarity networks in Archaeology - Similarity Network Fusion. *Peer Community in Archaeology*, 100347. [10.24072/pci.archaeo.100347](https://doi.org/10.24072/pci.archaeo.100347)

Published: 02 April 2024

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

This is a fascinating paper for anyone interested in network analysis or the chronology and cultures of the case study, namely the Late prehistoric burial sites in Dorset, for which the author's approach allowed a new perspective over an already deeply studied area [1]. This paper's implementation of Similarity Network Fusion (SNF) is noteworthy. This method is typically utilized within genetic research but has yet to be employed in Archaeology. SNF has the potential to benefit Archaeology due to its unique capabilities and approach significantly.

The author exhibits a deep and thorough understanding of previous investigations concerning material and similarity networks while emphasizing the innovative nature of this particular study. The SNF approach intends to improve a lack of the most used (in Archaeology) similarity coefficient, the Brainerd-Robinson, in certain situations, mainly in heterogenous and noisy datasets containing a small number of samples but a large number of measurements, scale differences, and collection biases, among other things. The SNF technique, demonstrated in the case study, effectively incorporates various similarity networks derived from different datatypes into one network.

As shown during the Dorset case study, the SNF application has a great application in archaeology, even in already available data, allowing us to go further and bring new visions to the existing interpretations. As stated by the author, SNF shows its potential for other applications and fields in archaeology coping with similar datasets, such as archaeobotany or archaeozoology, and seems to complement different multivariate statistical approaches, such as correspondence or cluster analysis.

This paper has been subject to two excellent revisions, which the author mostly accepted. One of the revisions was more technical, improving the article in the metadata part, data availability and clarification,

etc. Although the second revision was more conceptual and gave some excellent technical inputs, it focused more on complementary aspects that will allow the paper to reach a wider audience. I vividly recommend its publication.

References:

[1] Geitlinger, T. (2024). Similarity Network Fusion: Understanding Patterns and their Spatial Significance in Archaeological Datasets. Zenodo, 7998239, ver. 3 peer-reviewed and recommended by Peer Community in Archaeology. <https://doi.org/10.5281/zenodo.7998239>

Reviews

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.5281/zenodo.7998240>

Version of the preprint: 1

Authors' reply, 28 March 2024

To whom it may concern

Many thanks to both reviewers for their thorough reviews. As young researcher I could majorly benefit from their/your feedback and have learned a lot trying to implement their/your suggestions. I almost implemented all of the comments. Without wanting to point out every single spelling mistake I corrected, here are the major changes I did to my first manuscript:

Reviewer I:

- Metadata: I enriched my code with a Markdown passage, expanding on the installation of SNFpy, the data (including descriptions of the variables), and the functionality of the code. I also added the Markdown as README file.
- I added a file with the sites (including the used coordinates) and further attributes used in the discussion section. Furthermore, I added files with the resulting matrices of the SNF analysis.
- I added the requested citation (Habiba 2015) and tried to further expand on some methodological details (spectral clustering).
- I added details about strength of certain connections and also expanded on the importance single individual burials might have had on the formation of the clusters (cf. also the general section further below).

Reviewer II:

- I added more information about the case study area, some general remarks about the burial ritual during the discussed time period and improved the explanation of the aim of the case study. Now the link between landscape archaeology and my approach should be clearer.
- I tried to make my analytical decisions more explicit. I also addressed the findings from my analysis more extensively in my conclusion section and connected them with the aim/question of my study.
- I corrected all of the smaller issues pointed out in the last part of the review (mainly typos).

General:

- I revised my writing and tried to sharpen my arguments. I tried to stick to more direct sentences.
- I revised the discussion section in general, to make my argument as clear as possible and remove inconsistencies.
- I also majorly revised the section about the importance of burial goods per graves. I realised that I made a slight mistake with calculating the underlying numbers, changing the results of the correlation analysis. Thus, I adjusted the discussion section accordingly.
- I didn't change the general design of the images; the visibility problem pointed out by the first reviewer are due to the poor resolution in the word file and shouldn't occur on the original image. Increasing the diameter of the edges also made the picture appear very fuzzy. I thus decided to stick to the original design.

Many thanks and kind regards,
Timo Geitlinger

Decision by [Joel Santos](#), posted 27 August 2023, validated 27 August 2023

I recommend the paper for future publication. However, it should be reviewed by the author first.

I recommend the author look at the revisions suggested by the two reviewers. One is more technical, allowing us to go deeper into the technical part of the article, including the calculations used. The second one is more conceptual and, although giving some good technical inputs, focuses more on complementary aspects that will allow the paper to reach a wider audience.

Reviewed by [Matthew Peeples](#), 16 June 2023

I'm reviewing this article as a specialist in archaeological network methods. I have little knowledge of the regional/temporal context that is the focus of the case study.

Overall, I found this to be an interesting application of a new (to archaeology) method of dealing with multiplex archaeological similarity data through the Similarity Network Fusion approach. I had only a general familiarity with the SNF approach from the Wang et al. 2014 paper before reading this article but, after digging into the details, I agree with the authors that there is considerable potential for archaeological applications. This is a nice introduction to the approach for an archaeological audience and an interesting set of data to serve as a test case. I have suggestions for a few additions and clarifications but overall, I think this piece will be useful to those interested in archaeological network science.

Replication

In the process of the review, I re-ran all of the provided Python code to replicate those results. I ran the code in Python 3.8.3 in a Conda environment in RStudio. Once I installed the relevant snfpy package and edited the output locations in the code (as indicated in the header comments), it worked flawlessly on the first attempt. The only comment I have on the code is that it would be useful to add a few additional comments to specify details on certain steps just so it is easy to find a specific function at a glance. Also, why not include the installation instructions for snfpy right in the intro comments. As for the data, good metadata descriptions of variables and what they mean would also be useful in the final version. For example, I was unclear what the column headers in the "Individuals" file referred to.

As far as I am able to tell, the results I got in my replication seem to be a match for those in the figures, but I could not fully evaluate that as I did not have access to site location information for sites. I understand

that this is sometimes restricted, but if it could be included in the final paper (even jittered at some scale) that would allow for replication of the full results here. At the very least, I would suggest including the output files with clusters and unscaled similarities for each time period and dataset with the final article as an alternative check. There are a few details in the ArcGIS Pro analysis I replicated in Python as I don't have access to ArcGIS right now (rescaling, defining 25 highest similarity connections, and plotting) but the text was clear enough that could do that.

Comments on text and figures

I think the text generally reads well and provides a good introduction to similarity networks in archaeology including some of the common ways that people have defined edges and edge weights. It might be worth adding a citation to work by Habiba and colleagues (2018) which provides an overview of the advantages and disadvantages of some of the potential measures for archaeological data including the cosine similarity measure you use here. I would also suggest that you add a sentence or two to the details of your cluster detection and spectral algorithm discussion on page 5 to briefly explain what these methods do in simple terms.

In general, I could follow all of the methods you used in the text pretty easily and I think the analytical decisions you made on choosing clustering methods and defining cutoffs on what to display were generally reasonable. In the discussion of the results, you focus primarily on the topological clusters and their spatial distribution and don't talk about variation in network structures or positions much directly. This is okay but there may be additional interesting factors worth at least mentioning there. For example, I noted in the EBA time period when I recreated your network that cluster 1 represents a dense cluster of nodes all connected with edges in the top 10% of weight whereas cluster 0 represents nodes that are much more weakly connected overall. Things like this are lost a bit when you only visualize the top 25 nodes, but additional non-spatial network graphs with lower thresholds (like the 0.9 quantile or something) might be informative of additional patterns in these data.

I was happy to see discussion of the objects, attributes, etc. associated with specific clusters as that is where my thoughts went as I read but I found a few sections of this discussion a bit unclear. For example, on page 15 you state "the networks covering a great variety of aspects of the burial data seem to cluster sites with a ratio of more than one good per site in different clusters than the sites with fewer goods." Do you mean sites with more than 1 good per individual here? Also in the discussion of the EBA network you talk about one cluster having more than one burial good per grave and the other having consistently less than 3. Those are mutually exclusive sets so this is a bit confusing. I would suggest going back through this entire discussion section and just checking to make sure that the discussion of groups and attributes is consistent with the data and clearly stated. When messing with the data you provided, I also noticed that in general there were clusters that formed for dataset 1 including all of the sites with single individual burials and those with multiple burials. Since you are discussing ratios and numbers of grave goods per grave in your discussion, it may be useful to discuss the potential impact of grave numbers on the formation of topological clusters. In addition to this, since the data are spread across lots of files and are quite diverse, it would actually be helpful to have some summary tabular information on grave goods, monument type, and all of the other things you include by cluster available. This may be too big a table for the text but would be a useful document for the supplement. I produced a file like this for my review and it was quite useful in addressing questions I had about the performance of the SNF and clustering procedures.

The figures were generally good though I had to zoom in to > 100% to see the weakest edges displayed on the maps. If it doesn't make things look too busy, you might want to increase the weight of all the edges a bit.

Minor issues

One small typo I caught on page 5 near the bottom of the first full paragraph on the page. “restricted the representation of the nodes to the 25 highest similarities” I believe should be “restricted the representation of the edges to the 25 highest similarities”.

Suggested reference:

Habiba, Athenstädt, Jan C., Mills, B. J., & Brandes, U. (2018). Social networks and similarity of site assemblages. *Journal of Archaeological Science*, 92, 63–72. <https://doi.org/10.1016/j.jas.2017.11.002>

Reviewed by anonymous reviewer 1, 18 July 2023

Review:

Similarity Network Fusion: Understanding Patterns and their Spatial Significance in Archaeological Datasets

As a disclaimer, I have no personal experience with Similarity Network Fusion (SNF) and have not checked the submitted Python script. General impressions

Overall, this is a very interesting and thorough study, aiming to introduce archaeologists to methods of Similarity Network Fusion (SNF). Despite some errors and inconsistencies (see below), the manuscript is also well-written and logically structured.

The author demonstrates that he is aware of previous work on material and similarity networks while highlighting how this study extends or differs from them, and is able to explain the basics of relevant approaches in a clear manner. He presents how the data was selected, pre-processed, analysed and visualised, while also specifying which software was used in the process and arguing for choices made at various stages of the research. Although he does not seem to mention them in the manuscript (?), transparency is further supported by sharing the project's .csv files - containing the first and second datasets - as well as the Python script.

One thing that is lacking is, however, some more information on the case study and (cultural and ritual) contexts of the studied graves. With the *longue durée* perspective, it is of course difficult to find common traits to describe the area or phenomenon of interest. With reference to the research objectives (stipulated on pp. 2 and 3), knowing a bit more about the area, people and ritual customs they practiced, might prove valuable for guiding the research as well as interpreting findings in their cultural and ritual contexts. Abstract and introduction

The paper's title is fitting but, especially if the paper is not to be accompanied by keywords, it could perhaps be more precise, for example specifying what types of patterns are to be understood and/or mentioning the period and area under study.

The abstract is clear and well written. In addition to introducing the topic and highlighting that SNF applications are rare in archaeological network analysis, it summarises the case study, main findings and potential benefits of applying methods of SNF to archaeologically informed similarity networks.

The author has also written a solid introduction that focuses on relevant methods and earlier applications of similarity coefficients for archaeological case studies, while highlighting what is novel about his approach. The main aims of the paper are also specified (end of p. 2). This short section briefly mentions the case study, but whereas network analysis, similarity networks and SNF are treated in more detail, we only return to Dorset and the studied burials in the next part of the paper. It might be useful to already provide a sentence or two on the relevance of the case study in the introduction. There is, however, a logical build up and smooth transition to this (and other) section(s). Data and Methods

Whereas the introduction explains that the study's main aim – to further explore the potential of material network analysis, similarity networks and SNF – is approached with a case study, the next section reveals that

the case study's objective is to gain insight into the cultural and ritual structuration of the studied landscape and burial groups, but also the ritual relationship between graves.

The author clearly explains that the case study is based on data collected by the Grave Goods project, how his approach differs from their analyses and why he chose to focus on the given region and time periods. Whereas the selection of the region was largely dependent on the availability of suitable data, the decision to study two datasets was more directly informed by the researcher's aim to study cultural and ritual structuration and ritual relationship between graves. Here would be a suitable place for adding a bit of contextualisation. What sort of landscape are we in? What do we already know about the region, people living there, how they buried their dead, etc.? And which holes in our knowledge might the applied methodology help us fill?

In addition to describe the content of the two datasets and rationale behind them, the author also gives a rather detailed account of the research steps. The process appears sufficiently transparent, but as mentioned, I have no personal experience with SNF and am not able to judge the relevance of applied tools or appropriateness of the statistical analyses.

That said, the practicalities of the spatial correlation analysis between the topological clusters and five attributes do not appear to have been made explicit. On p. 11, for example, the author rather vaguely states that "there does not seem to be an obvious correlation between the burial clusters and the different sex categories", but it is not clear how this observation was reached. From visual examination and comparison of the maps, or another method?

Discussion and findings
In the discussion of the Burial Goods per Grave we get some answer to such questions: at least where potential correlation was spotted, the author conducted statistical analysis of variance. While this becomes clearer, this section could benefit from some revisions and polishing, since the argument is not always easy to follow. For example, it is specified (on p. 15) that "A similar significance [to the EBA data] could not be detected for the MBA ($\chi^2(1,44)=3.23, p<0.008$) and LIA ($\chi^2(1,21)=0.109, p=0.75$) clusters", yet argued that "the long-running continuity of this correlation is utterly surprising". This seems contradictory.

In general, the author should, however, be complimented for taking on a cautious and critical approach while discussing the findings. For example pointing out that the regional classifications of National Character Area do not mirror the past, although they seem more accurate than modern political borders (p. 5), or arguing that most of the attributes checked for potential correlation did not result in clearly discernible patterns.

Ultimately, his analyses lead him to conclude that two potential spatial clusters and two statistically significant correlations (objects per grave and chronology) "shed light on inter-site connections of ritual and cultural behaviour and past landscape structuration not yet recognised" (p. 18), but I do not think he specifies how. Following the discussion of the results of the correlation analyses, I would therefore like to invite the author to synthesize the findings and address the research aim stipulated on pp. 2 and 3 more directly. What exactly has this study and approach taught us about cultural and ritual structuration of this landscape and burial groups, or the ritual relationships between the graves? And how can this case study, in turn, contribute to discussions on the potential of material network analysis and the employment of SNF for similarity networks by archaeologists?

Tables and figures
Tables and figures are clearly presented with appropriate legends and captions. The table and figure captions are comprehensive and can be understood on their own term, although it might be worth also mentioning the topological clusters in the captions of Tables 3 and 4. Also, there is some inconsistency in the use of and in listing the colours (Figs. 3, 10), a comma missing between white and grey for Fig. 8, and I would add the before 25 (in Figs. 2-16).

Given the length of the paper, four tables seem reasonable, but sixteen colour images might be a bit much. All network visualisations are, however, relevant: Figs. 3-8 present the networks with sex classification (based on the first [Figs. 3-5] and second datatypes [Figs. 6-8]), whereas Figs. 9-14 present the networks with average grave goods per grave per site (based on the first [9-11] and second [12-14] datatypes). Finally, Figs. 15-16 present the similarity networks of Late Iron Age (or Early Bronze Age? See below) sites with sub-period classification (based on the first [15] and second [16] datatypes respectively). Other comments and corrections

Finally, a list of suggestions and things my eyes fell upon:

- In footnote 1, it might be useful to also guide the reader towards a handbook or introduction to network analysis.

- On p. 1, should it be "Similarly, **edges** can represent a multitude of different relationships.." (rather than nodes)?

- Where you refer to Fig. 1 (on p. 2), one could also mention edge lists and spatial networks in the text, so all four examples are mentioned. Also, are the values in these figures made up or based on 'real' data from the case study, as suggested by the spatial network visualisation? If so, this might be worth stressing, to make it clear that the figure was adapted to this study.

- On p. 2, you say that "applications of material networks remained astonishingly restricted". Perhaps you could add a reference with examples of such studies?

- *aspatial* -> *a spatial* (p. 2).

- On p. 3, you state that "SNF provides a set of additional network techniques, which are especially suitable for detecting subtypes and clusters in topological networks". Would it make sense to give a few concrete examples of such techniques?

- On p. 4, "I choose Dorset as the area.."

- At the end of the Material and Methods section, have nodes and edges been confused? I suspect it should read "restricted the representation of the **edges** to the 25 highest similarities" (not nodes) and "cluster classification of sites was joined with the **nodes**" (not edges)?

- On p. 5, "(i. e. fig. 4, 5, 7, 8)" could be re-organised according to the pairs, i.e. linking figs. 4 and 7, and figs. 5 and 8.

- On p. 6, add cluster colours in "apart from the geographically very restricted clusters visible in the LIA object network (fig. 8)".

- Currently, especially the Sex Affiliation, but also the Burial Goods per Grave and Chronology titles get a bit lost. Would it make sense to also start these sections with text rather than figures?

- On p. 11, does "the white cluster in the MBA network derived from a diverse set of datatypes only comprises female burials" rather concern fig. 4 (than fig 7, which does not seem to have a white cluster)?

- If the EBA network referred to on p. 15 is fig. 9, should it be "and the **black** ($\square=0.76, \square=0.53$) cluster" (rather than white)?

- Also, delete one of in "**of of** the small sample size", and *context-dependenct* -> *context-dependent* (p. 15)?

- Note also that the captions of Figs. 15 and 16 say "Late Iron Age" while the text stress that "The refinement of the chronology was solely undertaken for the EBA period" (p. 17).

Please also check the reference lists. For example,

- Brughmans 2010 is referred to (p. 3), but not included in the bibliography.

- There are some typos in the paper title of Blake 2013.

- First names are sometimes written out in capital letters (e.g. in Brughmans & Peeples 2020).

- Do the two references to COOPER, A., GARROW, D., GIBSON, C., GILES, M. & WILKIN, N. (2020 and 2022) refer to one or two sources? They are referenced separately in the text, but the dois are identical and the URL takes the reader to the same place.

- There is occasionally a space before :

- Should it be **Donnellan 2020** (as in the reference list) or **Donnellan 2021** (p. 2)?

- Should it be Morgan & **Whitelaw** (as in the reference list) or **Whitlaw** (p. 3)?

Finally, the paper is well-written, but in need of some language revisions and polishing to sharpen the argument, especially in the Burial Goods per Grave, Mover or Local and Chronology sections.