




# Archeofrag: an R package for refitting and spatial analysis in archaeology

Sébastien Plutniak<sup>1</sup>

<sup>1</sup> TRACES Laboratory, Toulouse, France

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Martin Fleischmann](#) 

## Reviewers:

- [@joeroe](#)
- [@zoometh](#)

Submitted: 17 February 2022

Published: unpublished

## License

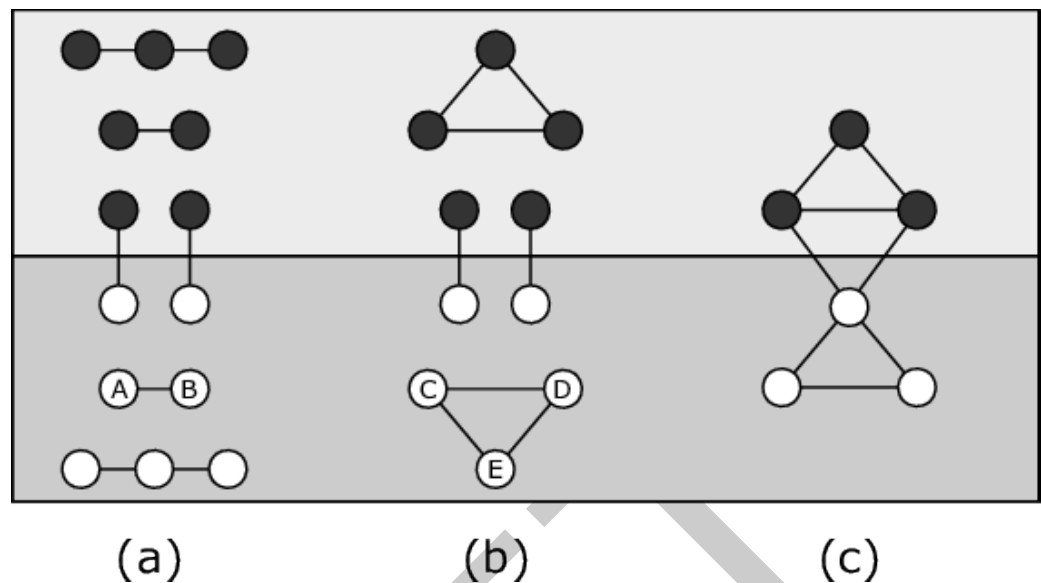
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Distinguishing between spatial entities is fundamental in archaeology since archaeologists deal with spatial phenomena at multiple scales of analysis. During an excavation, objects are discovered within various types of spatial units e.g., stratigraphic layers, pits, hearths and houses. Spatial units are far from being raw data, and the identification and determination of their boundaries is the result of conjoint lines and methods of investigation, to name only a few: field observations, geoarchaeology, sedimentology, and the study of archaeological “refits”. Refitting fragments belonged to the same object at some moment in the past. More precisely, archaeologists deduce former connection relationships from the symmetry and the possibility of contact of significantly large surface areas from two fragments, which can be physically adjusted (the fragments “refit”). Here, “connection” is used as a shorthand to refer to the connection relationship that existed in the past between two areas of an object before they were broken into fragments. Archaeological refitting analysis has several aims:

1. to reconstruct objects,
2. to determine technical sequences (e.g., stone tool manufacture), and
3. to determine the reliability of spatial units and their possible admixture due to pre- and post-depositional processes.

This analysis has long been used for the latter aim, benefiting from multiple methodological improvements (for an overview, see Czesla et al. (1990), Schurmans & De Bie (2007)). These methods have relied on the comparison between the number of refits between different spatial units and within these units. However, it has been demonstrated that considering the number of refits without considering their topology can lead to misleading interpretations. A method, coined TSAR “Topological Study of Archaeological Refitting”, was developed to overcome this issue using graph theory to model the topology of the relations between fragments (Plutniak (2021b), Plutniak (2022b)). This renewed approach distinguishes between ambiguous cases (Figure 1), and is much more robust and less sensitive to the lack of information than count-based methods, thus resulting in a more accurate evaluation of the reliability of the boundaries between spatial units. Archeofrag is an R package (R Core Team, 2021) implementing the TSAR method (Plutniak, 2022a).



**Figure 1: Three examples (a-c) of two layers with internal refitting ( $n=6$ ) and inter-layer refitting ( $n=2$ ).** Although the numbers of relationships are equal in all examples, their archaeological interpretation are very different: relevant distinction between the two layers in (a); relevant distinction with higher confidence about the fragmented objects' initial location in (b); doubtful distinction between layers in (c).

## Statement of need

The use of R in archaeology has increased slowly albeit constantly, during the last two decades. However, the development of R packages for specific archaeological needs is an even more recent phenomenon. Only a few packages are available for spatial analysis in archaeology, notably for stratigraphic analysis:

- [stratigraphr](#): package in its early development phase to visualise and analyse stratigraphies as directed graphs (Harris matrices).
- [tabula](#): generic package to visualise remain counts, which can also be used to compare layers ([Frerebeau, 2019](#)).
- [recexcavAAR](#): package for 3D reconstruction and analysis of excavations ([Schmid & Serbe, 2017](#)).
- [archaeoPhases](#): package for Bayesian analysis of archaeological chronology to define stratigraphic phases ([Philippe & Vibet, 2020](#)).

Archeofrag complements this series of packages with a specific focus on refitting.

## Package overview

Archeofrag mainly uses the [igraph](#) library for graph analysis ([Csárdi & Nepusz, 2006](#)) and also relies on some functions from the [RBGL](#) package ([Carey et al., 2020](#)). It comes with an example data set ([Plutniak, 2021a](#)) containing refitting data on the pottery found during excavations at Liang Abu rock shelter, in Borneo ([Plutniak et al., 2016](#)).

Archeofrag has six main sets of functions:

1. **Data management:** create, check, and transform fragmentation graphs (including edge weighting based on the topological properties of the vertices and optionally, on the size of the fragments and the distance between the locations where they were discovered during excavation)

2. **Visualisation:** represent fragmentation graphs as node-and-edge diagrams
3. **Boundary-related statistics:** count the relationships within and between two spatial units, measure their cohesion and the admixture values
4. **Spatial unit-related statistics:** characterise the topology of a specific set of refitting relationships (e.g., a layer) with several measurements
5. **Simulation:** generate simulated fragmentation graphs, simulate their alteration (missing data); compare an empirical graph with similar simulated graphs and return results in a convenient way
6. **Similarity analysis:** in addition to the topological analysis of refits, functions are available to analyse similarity relationships, which are determined between fragments considered as sharing enough common features (motif, clay, inclusions, etc.) to state they are (and were) parts of the same initial object.

## Data management

Archeofrag is intended to be used with two sources of data, namely the user's empirical data and artificially generated data using its simulation function. User's data must be split into different tables:

- **Fragment table:** each line contains the unique identifier of a fragment, its spatial unit, and optionally additional information.
- **Connection table:** an edge list with the identifiers of two connected fragments by line (e.g., [Table 1](#)).
- **Similarity table** (optionally): each line includes the unique identifier of a fragment and the identifier of a set of similar fragments it belongs to.

**Table 1:** Recording of connection relationships between fragments (illustrated by the examples given in [Figure 1](#)).

id	fragment 1	fragment 2
1	A	B
2	C	D
3	D	E
4	E	C

The package includes functions to generate summary statistics about the fragmentation graph and extract specific sub-graphs (by layer, by component size, etc.).

## Visualisation of fragmentation graphs

The fragmentation graphs are visualised as node-and-edge diagrams. For graphs with only two spatial units and connection relationships, the location of the nodes in the upper and the lower part of the plot are based on their spatial unit.

## Boundary-related statistics: measuring the cohesion and admixture of two spatial units

Evaluating the consistency of spatial units and their division from refits is the first aim of the Archeofrag package. Evaluation follows a three-step procedure, implemented in three related functions:

1. **Weighting the edges.** Three parameters can be combined: the topology of the connection relationships (mandatory), the size of the two connected fragments (optional), and the spatial distance between the location where they were found (optional). The optional parameters implement the hypothesis that two large fragments found far from each other

- 95 suggest more disturbance in the site than two small fragments found very close together  
96 (see an application and details in [Caro et al., 2022](#)).  
97 2. Measuring the internal cohesion of each spatial unit (intuitively, how they are “self-  
98 adherent” to themselves)  
99 3. Measuring the admixture of two spatial units (a summary statistic based on two cohesion  
100 values)

101 Cohesion and admixture values are computed by pairs of spatial units. Results for cohesion  
102 measurements range between  $[0;1]$ , with values towards 0 for low cohesion and towards 1  
103 for high cohesion, with their sum never being superior to 1 for a pair of layers. Results for  
104 admixture measurements range between  $[0;1]$ .

## 105 **Spatial unit-related statistics: fragmentation patterns, technology, hu-** 106 **man behaviour**

107 The second aim of the TSAR method implemented in Archeofrag is to characterise spatial  
108 units based on the topological properties of the connection relationships between the fragments  
109 they contain. Several functions are provided for this purpose, selected for their relevance  
110 in the archaeological context, namely cycle count, path length, and component diameter (a  
111 component is related to an initial object).

112 The archaeological interpretation of the numerical values depends on the type of object (lithic,  
113 pottery, etc.) and their completeness or incompleteness. These values can suggest specific  
114 behaviours related to the production or use of the objects (intentional breaking), and post-  
115 depositional processes (natural breaking, scattering). This aspect of the TSAR method will be  
116 further developed in the future.

## 117 **Simulation of fragmentation graphs**

118 The simulation function generates connected fragments scattered within one or two spatial  
119 units (see Plutniak ([2021b](#)) for details). It can be set with multiple parameters:

- 120 ■ Number of initial objects or fragments.
- 121 ■ Final number of fragments.
- 122 ■ Final number of connection relationships.
- 123 ■ Proportion of fragments in each spatial unit, before post-depositional processes.
- 124 ■ Proportion of components in each spatial unit.
- 125 ■ Proportion of fragments likely to move from a spatial unit to the other one.
- 126 ■ Whether to only disturb the fragments from spatial unit 1 or spatial unit 2.
- 127 ■ When applying fragmentation, increase the likelihood for objects with more fragments  
128 being selected.
- 129 ■ Number of initial hypothetical spatial unit.
- 130 ■ Whether generating only planar graphs.

131 Generating only planar graphs is interesting since this corresponds to the fragmentation observed  
132 in some specific archaeological contexts (e.g., pottery with simple shapes, small sets of refits).  
133 However, the run time of this function is doubled when this constraint is used.

## 134 **Resources and examples**

135 Archeofrag is available on [CRAN](#) and the code of the development version is available on  
136 [Github](#). A [Vignette](#) and a [Shiny application](#) demonstrate the package.

137 At the time of this submission, Archeofrag has been applied to two archaeological sites:

- 138 ■ Liang Abu, Indonesia (Plutniak ([2021b](#)), Plutniak et al. ([forthcoming](#)))
- 139 ■ Taï cave, France (Caro et al. ([2022](#)))

## Acknowledgements

I thank Luce Prignano, Claire Manen, Joséphine Caro, and Oliver Nakoinz for their valuable comments during the development of this package, which was finalised at the *Institut für Ur- und Frühgeschichte of Kiel*, with the support of a “Short-term research grant” from the *Deutscher Akademischer Austausch Dienst (DAAD)*.

## References

- Carey, V., Long, L., & Gentleman, R. (2020). *RBGL: An interface to the BOOST graph library*. <https://doi.org/10.18129/B9.bioc.RBGL>
- Caro, J., Manen, C., Baux, A., & Plutniak, S. (2022). Les productions céramiques du Néolithique ancien et moyen: approches céramo-stratigraphique, technologique et morpho-stylistique. In C. Manen (Ed.), *Le Taï (Remoulins – Gard). Premières sociétés agropastorales du Languedoc méditerranéen (6<sup>e</sup>–3<sup>e</sup> millénaire avant notre ère* (pp. 616–687). Archives d'Écologie Préhistorique.
- Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, 1695(5), 1–9. <http://igraph.org>
- Cziesla, E., Eickhoff, S., Arts, N., & Winter, D. (Eds.). (1990). *The big puzzle: International symposium on refitting stone artefacts*. Holos.
- Frerebeau, N. (2019). Tabula: An r package for analysis, seriation, and visualization of archaeological count data. *Journal of Open Source Software*, 44(4), 1821. <https://doi.org/10.21105/joss.01821>
- Philippe, A., & Vibet, M.-A. (2020). Analysis of archaeological phases using the r package ArchaeoPhases. *Journal of Statistical Software*, 93. <https://doi.org/10.18637/jss.v093.c01>
- Plutniak, S. (2021a). *Refitting pottery fragments from the liang abu rockshelter, borneo*. Zenodo. <https://doi.org/10.5281/zenodo.4719577>
- Plutniak, S. (2021b). The strength of parthood ties. Modelling spatial units and fragmented objects with the TSAR method – topological study of archaeological refitting. *Journal of Archaeological Science*, 136, 105501. <https://doi.org/10.1016/j.jas.2021.105501>
- Plutniak, S. (2022a). *Archeofrag: An r package for refitting and spatial analysis in archaeology*. <https://doi.org/10.5281/zenodo.6338443>
- Plutniak, S. (2022b). L'analyse topologique des remontages archéologiques: la méthode TSAR et le package R archeofrag. *Bulletin de la Société préhistorique française*, 119(1), 110–113.
- Plutniak, S., Araujo, A., Agus Oktaviana, A., Sugiyanto, B., Chazine, J.-M., & Ricaut, F.-X. (forthcoming). Mainland-coastal interactions in east borneo: Inter-site comparison and bayesian chronological models of two late pleistocene-holocene sequences (liang abu and kimanis rock shelters). *Journal of Island and Coastal Archaeology*.
- Plutniak, S., Araujo, A., Puaud, S., Ferrié, J.-G., Oktaviana, A. A., Sugiyanto, B., Chazine, J.-M., & Ricaut, F.-X. (2016). Borneo as a half empty pot: Pottery assemblage from liang abu, east kalimantan. *Quaternary International*, 416, 228–242. <https://doi.org/10.1016/j.quaint.2015.11.080>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schmid, C., & Serbe, B. (2017). *recexcavAAR: 3D reconstruction of archaeological excavations*. <https://CRAN.R-project.org/package=recexcavAAR>

<sup>184</sup> Schurmans, U., & De Bie, M. (Eds.). (2007). *Fitting rocks: Lithic refitting examined*.  
<sup>185</sup> Archaeopress. ISBN: 978-1-4073-0012-2

DRAFT