Journal of Archaeological Science 46 (2014) 75–88

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Network analysis of archaeological data: a systematic approach

Per Östborn*, Henrik Gerding

Classical Archaeology and Ancient History, Department of Archaeology and Ancient History, University of Lund, Box 117, S-221 00 Lund, Sweden

ARTICLE INFO

Article history:
Received 27 September 2013
Received in revised form
3 March 2014
Accepted 3 March 2014
Available online 12 March 2014

Keywords:
Similarity networks
Networks
Quantitative methods
Statistical methods

ABSTRACT

Network theory can be employed in two ways in archaeology: it can be used to analyse archaeological data, or it can be used to model a historical process for the purpose of simulating the data. This paper focuses on the first approach. In such analyses, similar archaeological contexts are often connected to form a similarity network. Similarity is treated as a proxy for social or causal relationships. Most often, similarity is defined by the presence of the same kind of find in two contexts. However, to detect relationships effectively, we have to allow any kind of similarity relation to be a criterion for connection, in which different kinds of attributes that characterise the contexts may be mixed. We discuss how such general similarity networks can be used to disclose relational patterns hidden in archaeological data. Statistical tests are necessary to distinguish significant patterns from random patterns. We argue that random permutation tests are well suited for this task, and we introduce appropriate tests of this kind. The methods outlined are compared to other kinds of quantitative data analysis, such as correspondence analysis. We discuss which approach is more suitable for which kind of data. The choice of approach also depends on the questions addressed to the archaeological material.

1. Introduction

In recent years, the use of network theory in archaeological research has become quite popular. Knowing how to apply it in the most meaningful way still remains a problem, however (Brughmans, 2010, 2013a; Knappett, 2013). There are several obstacles to overcome: social networks can only be reconstructed indirectly from the material record and this record is often incomplete and ambiguous.

In this paper, we discuss network theory as a means to make inferences from archaeological data. We call this approach 'network analysis'. We do not discuss the use of network theory to simulate historical processes, an approach which may be called 'network modelling'. A variety of frameworks may be used in the latter approach, such as agent-based models (Graham, 2006b) or gravity models (Rivers et al., 2013).

Starting from the very basics, we develop a map of technical possibilities for network analysis. We review archaeological studies where network analysis has been employed, and position these studies on the map. The points marked on the map reveal that there are unexplored areas.

Most importantly, virtually all network analyses use only a single criterion for the connection of two nodes, for instance the co-presence of a certain find at two sites, or the fact that a given road runs through two sites. Both these conditions reflect a similarity. The basic reason for focussing on similarity is that the chance that two more similar nodes are causally or socially related is larger than the chance that two less similar nodes are related in this way. In a general approach, we have to allow any kind of similarity relation as a criterion for connection, in which different kinds of attributes can be mixed. We try to classify different kinds of attributes, discuss what similarity may mean in each case, and describe how such individual attribute similarities can be combined in an arbitrary logical statement that expresses a similarity condition. Networks created with such a framework in mind may be called *general similarity networks*.

We then discuss how such networks can and should be used in archaeological data analysis. Two complementary approaches are highlighted: exploratory and statistical analysis. Suitable statistical tests are outlined. Limitations of network analysis are also mentioned. The relative merits of network analysis, as compared to correspondence analysis and related methods, are also discussed. We conclude that some questions can be answered by few formal methods other than network analysis.

* Corresponding author. Tel.: +46 738196154.
E-mail addresses: per.ostborn@gmail.com, per.ostborn@klass.lu.se (P. Östborn), henrik.gerding@klass.lu.se (H. Gerding).

http://dx.doi.org/10.1016/j.jas.2014.03.015

The present paper is an attempt to make a systematic survey of possibilities. In a related case study (Östborn and Gerding, 2014), we apply the outlined ideas and methods to analyse the diffusion of fired bricks across Hellenistic Europe. Here, we will briefly mention some results from the case study, to give a hint how the outlined methods can be used in practice.

2. Archaeological networks

A network consists of a set of nodes, some of which are connected by edges. In archaeological applications, the nodes are either contexts, or attributes of contexts.

An archaeological context may be defined as a geographical location where artefacts are found that are interpreted to belong together, in some sense. The information obtained from a context can often be organised as a list of attributes describing the artefacts and their location, where each attribute has a given value.

For example, an attribute may be defined by a given pottery type. Presence of this pottery type may correspond to value 1, and absence to value 0. Alternatively, the value may represent the abundance of the pottery type, as measured by the number of found vessels or the weight of fragments. The position of the context is another attribute, where the value is given by a pair of coordinates. Another attribute where the value can be represented by a pair of numbers is the dating, where the uncertainty may be expressed by a time interval. The sizes of artefacts may also be expressed as numerical intervals. One may also define categorical attributes, such as the function of an excavated building. The possible values might then be *Domestic*, *Public* or *Sacred*. If required, such values can be numerically represented as 1, 2 and 3, respectively.

In any case, whenever a list of attributes is defined, the knowledge about a set of related contexts can be organised as a matrix where each row represents a context, and each column represents an attribute (Fig. 1).

Given such a database, two basic types of networks can be constructed (Fig. 2). In the first type, the contexts are the nodes. In the second type, the attributes are the nodes. Note the symmetry of the two network types. One turns into the other when the rows and columns of the data matrix in Fig. 1 change roles. Since contexts have geographical location, networks of type 1 are spatial. In contrast, the distance between two nodes in networks of type 2 can be defined only topologically, as the minimum number of edges that have to be traversed to reach from one attribute to the other.

If the attribute values are binary, like the presence or absence of a pottery type, all networks of types 1 and 2 can be combined into a single two-mode network (Watts, 2003), which embodies all information contained in the database matrix (Fig. 3). In a two-mode network, there are two classes of nodes, and two nodes can be connected only if they belong to different classes. In the case of archaeological networks, the two classes of nodes are the contexts and the attributes.

	Attribute 1	Attribute 2	Attribute 3	...
Context 1	$Value_{11}$	$Value_{12}$	$Value_{13}$	
Context 2	$Value_{21}$	$Value_{22}$	$Value_{23}$	
Context 3	$Value_{31}$	$Value_{32}$	$Value_{33}$	
⋮				⋱

Fig. 1. An archaeological database organised as a matrix. The attribute values may be either numerical (e.g. abundance of a type of artefact, or numerical measures such as artefact sizes), or categorical (such as the purpose of a building, where a list of possibilities is predefined). Such organisation of archaeological information is required to perform network analysis (Fig. 2).

The two-mode network can be decomposed into single-mode networks of type 1, or single-mode networks of type 2, in many different ways. Most naturally, in the decomposition into a type-1 network, two contexts are connected whenever they are linked to at least m common attributes in the two-mode network (the contexts share the value 1 of at least m attributes). In the decomposition into a type-2 network, two attributes are naturally connected whenever they are linked to at least n common contexts (the attributes have the same value 1 in at least n contexts).

3. Geographical networks and space syntax

Geographical networks where archaeological contexts are connected by known physical routes can also be said to belong to one of the two network types shown in Fig. 2. Each route defines an attribute. If the route runs through a given context, or starts or ends there, the attribute value is 1, otherwise, it is 0. In networks of type 1, two contexts may be connected whenever they both have value 1 of some attribute: that is, they are connected by the same route. In networks of type 2, two attributes may be connected whenever some context has value 1 of both attributes, that is, the routes cross, start or end at the same place. Since the attribute values are binary, the complete structure can be represented as a single two-mode network (Fig. 3).

Graphs constructed in space syntax (Ferguson, 1996; Grahame, 2000; Stöger, 2011; Thaler, 2005) are analogous to geographical networks. In this case the role of the archaeological context is played by a spatial unit like a street (axial analysis)[1] or a room within a building (access analysis), whereas each crossing or doorway that links such units defines one attribute. If the doorway links two rooms, the corresponding attribute value is 1 for these two units, whereas it is 0 for all other units. Different structural measures pertaining to the graph as a whole or to individual nodes can be measured, for example 'integration value' and 'control value'. These correspond to closeness centrality[2] and betweenness centrality[3] in the terminology of network theory (Valente, 1995; de Nooy et al., 2005).

If we do not know the physical routes that mediated contact between contexts, there are methods for constructing a hypothetical route network (Jiménez and Chapman, 2002; Herzog, 2013). The simplest such method is proximal point analysis (PPA). In this approach, each context is connected to its n nearest neighbours, where the integer parameter n may be varied (see sections 5 and 8).

However, PPA does not take into account that if a context C_2 is located approximately along the straight line from context C_1 to context C_3, then a route from C_1 to C_3 often passes C_2. In the corresponding geographical network, C_1 and C_2 should then be connected by an edge, as well as C_2 and C_3, but not C_1 and C_3, even if they are close. In a so-called Gabriel graph, C_1 and C_3 are connected if and only if there is no context C_2 placed within a circle fitted between C_1 and C_3 so that the straight line from C_1 to C_3 becomes the diameter. The notion of a Gabriel graph can be generalized to a

[1] In axial analysis a delimited urban environment is divided into "convex spaces", which are then superimposed by a number of "axial lines", representing the longest and fewest visual lines that are needed to connect all convex spaces. These axes are regarded as "potential movement lines" and often coincide with the streets.

[2] The closeness centrality of a node is the inverse of the average distance from this node to all the other nodes. Distance is measured as the number of edges that have to be traversed to get from one node to another along the shortest possible path.

[3] The betweenness centrality of a node is the number of such shortest paths that pass through it, given the set of all shortest paths between all possible node pairs.

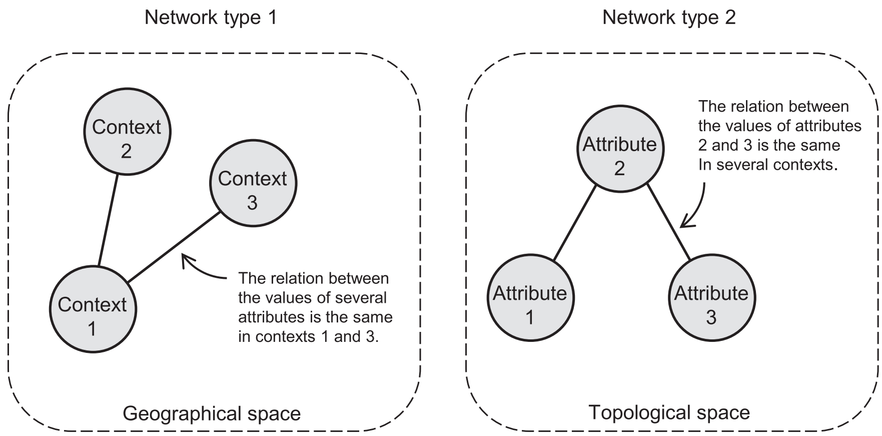

Fig. 2. Two network types can be defined given an archaeological data base expressed as in Fig. 1. In a network of type 1, the contexts are the nodes. A criterion for connecting nodes 1 and 3 with an edge could be that at least three pottery types are co-present in contexts 1 and 3. In a network of type 2, the attributes are the nodes. A criterion for connecting nodes 2 and 3 with an edge could be that pottery types 2 and 3 are co-present in least at three contexts.

Beta skeleton, which is a family of graphs defined for the continuous parameter β. For $\beta = 1$ the Beta skeleton and the Gabriel graph is the same. For $\beta < 1$, the circle used in the definition of the Gabriel graph is squeezed to a thinner lens with C_1 and C_3 at the end points. For $\beta = 0$ this lens is reduced to the straight line from C_1 to C_3. In this limit case, C_1 and C_3 are connected whenever there is no context C_2 placed exactly along this line.

To be realistic, any effort to reconstruct route networks has to take into account travel costs in terms of time and energy along different potential paths (Sindbæk, 2007a; Herzog, 2013; Verhagen, 2013). These costs depend heavily on topography, and travel by sea is much faster than travel by land. Distances between contexts in a cost landscape may thus be very different from distances in the geographical landscape.

4. Flat and hierarchical databases

It is common knowledge that there is no perfect way to organise a database. The architecture may be hierarchical or 'flat', or a combination of both. We say that a database matrix (Fig. 1) has hierarchical elements whenever the value of one attribute limits the possible values of another attribute in the same context. In other words, the values cannot vary independently. To take an example from our case study (Östborn and Gerding, 2014), let the function of a fired brick define one attribute, with possible values *Wall* and *Column*, and let the brick shape define another attribute, with possible values *Rectangular* and *Circular*. We know that whenever the shape is circular, the functional role of the brick is to be part of a column.

The criteria used to create edges in networks (Fig. 2) are often 'democratically' defined: they do not distinguish one attribute from another, or one value from another. Thus a 'flat' architecture is preferable in such network analyses. It is not necessary however, as long as the presence of hierarchical elements is kept in mind in statistical analyses (see Section 13).

5. Review of network analysis in archaeological research

In this section we discuss some studies in archaeology and ancient history that used network analysis. We focus on the technical aspects of the analyses, to be able to get a picture of what improvements are possible in this respect. Tom Brughmans (2010, 2013a) provides broader reviews. Good surveys of the subject are also found in Bentley and Maschner (2003), Malkin et al. (2009), and Knappett (2011, 2013).

Cyprian Broodbank (1993, 2000) visualised possible patterns of interaction in the Early Bronze Age Cyclades with proximal point analysis (PPA). In PPA, networks are constructed where each context is connected to its n closest neighbours. Anna Collar (2007, 2013) also used PPA to study the diffusion of religious innovations in the Roman Empire. In her study, the contexts were not settlements or towns, but epigraphic evidence from different cults. Networks generated in PPA are of type 1, where the relevant attribute is spatial position. The attribute relation that defines the criterion for connection is that of proximity rank.

PPA can help identify the most important routes of communication, and their collective topology. However, it should be kept in mind that long range, direct connections probably occur in most

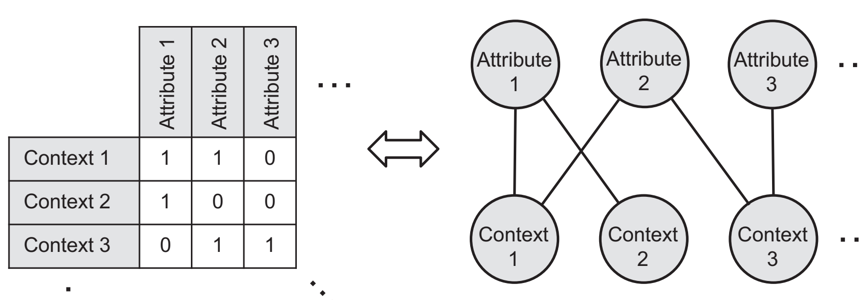

Fig. 3. If the values of all attributes are binary (Fig. 1), the two types of archaeological networks (Fig. 2) can be combined to a single two-mode network. A context is connected to an attribute when the attribute value is 1 in this context.

societies, even though they are less common, and sometimes 'weaker' in the sense that the amount of information that flows through them is smaller. This rarity and weakness may not be essential, however. Even 'weak' ties can change the structure of interaction dramatically, in accordance with the catch phrase 'the strength of weak ties' (Granovetter, 1973).

Leif Isaksen (2006, 2008) and Shawn Graham (2006b) have analysed geographical networks in the form of ancient route networks. Isaksen (2006, 2008) studied the transport network of Roman Baetica, using the Antonine itineraries and the Ravenna Cosmography as sources. Towns were treated as nodes, and their relative importance was judged by the closeness[2] and betweenness[3] centrality measures (Valente, 1995; de Nooy et al., 2005). Centrality measures are sensitive to absent or corrupt edges, and Isaksen is well aware that the source data are limited and probably biased. These empirical problems are common to all archaeological data, and highlight the fact that only network properties that are fairly robust to absent or corrupt information should be interpreted historically.

Graham (2006b) used the Antonine itineraries to study and compare the structure of the Roman route networks in Britain, Gaul, Iberia and Italy. To this end, he calculated quantities such as characteristic path length,[4] diameter[5] and cohesion.[6] Since such structural quantities describe global network properties rather than point to individual nodes, they are less sensitive to limitations in empirical data than are centrality measures. Of course, the data may have structural flaws, in the sense that they are not representative. In that case global structural quantities become unreliable regardless.

Søren Sindbæk (2007a) employed spatial networks of type 1 to reconstruct trade routes in Viking Age Scandinavia. He used the co-presence of any of 31 types of artefacts as a criterion to connect 71 excavated contexts pairwise. Since the co-presence of artefacts presumably generates a large amount of false positives as an indicator of direct contact, he made an additional assumption: contact between two contexts took place via intermediate contexts, where the shortest such route was chosen. The resulting trade network looks very realistic, and gives valuable insights that are not easily obtained without network analysis. The network is qualitatively interpreted to be scale-free,[7] where the trade is governed from a small number of hubs. However, this network structure arises almost automatically from the additional assumption. The degree of clustering[8] becomes very small because in a triangle of three contexts, any direct connection between the two most distant contexts is removed, and the links between the other two contexts are strengthened correspondingly. Thus, by assumption, such a network can never represent a small world[9], whereas the relative importance of settlements with many pottery types located near the geographical centre increases, making them more hub-like. In a related study of the emergence of towns in Viking Age Scandinavia (Sindbæk, 2007b), the author again more or less assumed a scale-free network, and gained compelling insights from this assumption.

The fact that an assumed network structure fits the data gives support to the assumption, but such an analysis does not discriminate between alternative assumptions. In a similar fashion, Irad Malkin (2011) found an impressive array of empirical evidence to support the idea that the classical Greek world was a small world, whereas less effort was put into the exclusion of other possible network structures (Ruffini, 2012; Brughmans, 2013).

In a recent study, Sindbæk (2013) again analysed communication routes in Northern Europe during the Viking Age with network techniques. This time he chose another method to reduce the presumed large number of false positives than to assume that contact occurred via geographically intermediate contexts. Instead, he required that pairs of contexts had to have at least *two* artefact types in common if they were to be connected. The strategy follows the basic idea that with a higher degree of similarity, the probability that the contexts are historically related increases.

Fiona Coward (2010, 2013) studied the presumed change in the social network structure that occurred in the early Neolithic, using the material record from the Near East. Like Sindbæk, she used networks of type 1 based on co-presence of artefacts to infer social relationships between contexts. The total number of co-present artefact types defined the strength of the relation. In contrast to Sindbæk (2007a), she did not make additional network assumptions that affect the outcome of the analysis, and she performed statistical tests to determine the significance of the temporal trends that were found among the structural network quantities. However, the idea that the early Neolithic societies evolved towards a small world[10] was only qualitatively established.

In a similar way, Colby Phillips and Erik Gjesfjeld (2013) reconstructed social relationships in the Kuril Islands through the geochemical identification of common clay sources for ceramics that are distributed in the region. The existence of the same clay source in two contexts was used as a proxy for ceramic exchange relationships. Statistical tests were performed to determine the extent to which the resulting similarity networks deviated from those expected by chance.

Barbara Mills and co-workers (2013) explored the evolution of relations between sites in the US Southwest in the late pre-Hispanic period. They used spatial networks of type 1, where each context was described by the attributes defined by a list of pottery types. The value of attribute j at context i is the fraction f_{ij} of the total number of finds at site i that belong to pottery category j.

The similarity $S_{ii'}$ between two contexts i and i' was defined as $S_{ii'} = 1 - 1/2\sum_{j=1}^{n}\left|f_{i'j} - f_{ij}\right|$, where n is the number of predefined pottery types. If the proportions of found pottery types are identical in the two contexts i and i', then $S_{ii'} = 1$. If no pottery types are co-present, then $S_{ii'} = 0$. This similarity measure is a variant of the so-called Brainerd-Robinson coefficient (Cowgill, 1990). Note that the similarity measure above is more detailed than the one used by Sindbæk (2007a) and Coward (2010, 2013), which just take the number of co-present artefact types into account, ignoring their relative abundance.

However, to create similarity networks, a criterion must be chosen to decide which contexts are similar enough to be connected. Mills et al. chose $S_{ii'} \geq 3/4$. They identified central nodes and calculated a number of structural network quantities from different time periods. Based on these quantities they put forward hypotheses about the evolution of societal organisation and social relations in the region from 1200 CE to 1450 CE. Among other things, they concluded that the spatial range of contacts increased with time. As described above, their method may not lead to robust conclusions, since they depend on the arbitrary similarity criterion $S_{ii'} \geq 3/4$ (Peeples and Roberts Jr., 2013). However, Mills et al. circumvented this problem by using the full matrix S as input in most of their calculations. Formally, this is no longer network analysis,

[4] The average of all the shortest paths between all possible node pairs (see footnote 2).

[5] The longest of all the shortest paths in the network.

[6] The fraction of all possible edges that are realised.

[7] Scale-free networks are characterised by the existence of hubs connected to many more other nodes than would have been expected by chance.

[8] Consider all nodes triplets (C_1, C_2, C_3) such that C_1 is connected to C_2, and C_2 to C_3. The clustering coefficient is the fraction of those triplets in which C_1 is also connected to C_3.

[9] Small-world networks are characterised by a high clustering coefficient (footnote 8) and a small diameter or characteristic path length (footnotes 4 and 5).

since the matrix elements $S_{ii'}$ of S are (almost) continuous numbers in the interval [0, 1] rather than binary (edge or no edge), but all examined quantities were analogous to the ones used in ordinary network analysis.

In each of the studies cited above, all attributes used to characterise a given context were of the same kind. To exemplify, some studies used pottery types as attributes, others used roads, and still others used proximity to other contexts. But no study described a single context by the presence or absence of some pottery type, *and* the fact that some road passed the context or not, *and* the proximity to other contexts. An exception to this limitation is Graham (2006a), who studied the Roman brick industry in the Tiber Valley. He used both the co-appearance of a given brick stamp, and the same clay source as similarity criteria to connect two brick contexts.

Emma Blake (2013) also combined two criteria to construct a similarity network. She studied interactions among regional groups in pre-Roman west-central Italy, using the distribution of imported goods to trace interaction between sites. She connected all pairs of contexts that shared a given type of goods, *and* were no more than 50 km apart.

Treating stamps as brick attributes, Graham (2006a) also constructed networks of type 2 (Fig. 2), connecting stamps that occur in the same contexts. If we identify the stamp with the manufacturer, this creates a social network. As another example of the use of purely topological networks of type 2, let us mention that Brughmans (2010) studied a database of Roman tableware from the eastern Mediterranean, and found relations between pottery forms as defined by their co-presence in the same contexts. He also combined such networks with networks of type 1 to represent the database as a single two-mode network (Fig. 3).

To conclude, network analysis has already given insights into archaeological research that would have been hard to obtain by other means. However, more or less well-motivated network assumptions that bias the analysis are introduced on occasion, and researchers rely on qualitative reasoning at times, making the network perspective more of a metaphor than a method, as discussed by Knappett (2013). The full potential of quantitative network analysis has yet to be realised.

On the other hand, the outcome of the analysis is sometimes over-interpreted; some researchers try to gain more insight from the network perspective than it has potential to offer. A common mistake is to use networks to find support for one hypothesis, without excluding others. This pitfall is hard to avoid due to the incompleteness of archaeological data. The less information you have, the harder it is to falsify your ideas. Nevertheless, to make network analysis a truly scientific tool in archaeology, methods have to be introduced to circumvent this problem. To meet this need, statistical hypothesis testing is becoming more and more common in network analysis of archaeological data.

6. Alternatives to network analysis

The aim of network analysis is to reduce the amount of information in a large database matrix so that patterns and structures that are hidden in the tangle are disclosed. This aim is the same as that in other formal analysis methods employed by archaeologists. The most common such approach is the one that forms the basis for the related methods of principal component analysis, correspondence analysis, and factor analysis (Baxter, 1994, 2003; Bølviken et al., 1981).

In this approach, the data is represented as points in a multidimensional space. The points may be contexts, the coordinates of which are given by the values of their attributes (Fig. 4). Alternatively, the points may be attributes, the coordinates of which are given by the values in each of the contexts. For the sake of illustration, we focus on the first case.

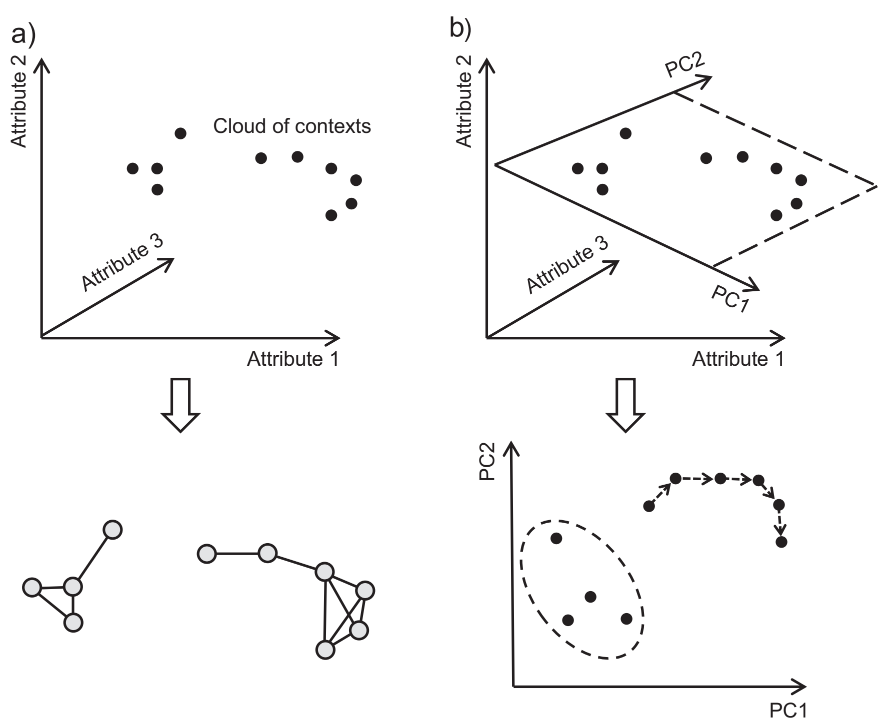

Fig. 4. Two methods to reduce the amount of data in archaeological databases so that patterns are disclosed. a) Network analysis. Similar enough contexts are connected. Only topological relations are preserved from the original space of attributes. b) Projection analysis. The cloud of contexts is projected onto the plane spanned by the first two principal components PC1 and PC2. Encircled contexts are interpreted as a related group in a cluster analysis. The strand of contexts is interpreted as a temporal sequence in a seriation analysis.

Most often, the cloud of points is projected onto the two-dimensional plane that comes as close to it as possible (Fig. 4b). This plane is spanned by what are called the first two principal components. The basic idea is just to look at the cloud of points from the angle that gives the best frontal view of it – just like in the theatre, where you choose a seat right in front of the stage, if you get the chance. The actors and the props are never perfectly aligned; some are in the back and some are in the front, but on average you always get the best view of what is going on if you look straight at the stage. In the theatre, the plane of the curtain defines the two first principal components.

In this two-dimensional projection, it is easier to perform *cluster analysis*, where clusters of close points are interpreted to be related groups of contexts, or to perform *seriation*, in which snake-like strands of points are interpreted to be temporal sequences of contexts (Fig. 4b). The latter is a well-known method for relative dating.

Let us contrast these kinds of 'projection analyses' with network analysis, remembering that the term projection analysis denotes the common features of correspondence analysis, factor analysis and principal component analysis. In projection analysis, the reduction in the amount of data is achieved by describing each context with two 'effective attributes', defined by the two first principal components, disregarding all the original attributes, whatever their number is. In network analysis, all pairs of contexts that fulfil a prescribed relation are connected (Fig. 4a). The data reduction results from the subsequent removal of the information that was used to check whether this relation holds.

In effect, projection analysis makes it possible to see clearly the position of a single context in the collective of all contexts. In contrast, network analysis seeks to identify the individual relations between context pairs, and from these relations reconstruct the structure of the relations of the collective. It keeps only the topology of the similarity relations, whereas projection analysis keeps a measure of distance in similarity space, a measure of dissimilarity. These differences mean that the two methods best suit different kinds of databases, and are effective for answering different kinds of questions.

First and foremost, there is no natural way to apply projection analysis when some attributes are categorical, so that their values are not possible to order. For instance, let the functional role of a fired brick define one attribute, with possible values *Wall*, *Door frame* and *Column*. It is not possible to say that *Wall* is 'closer' to *Door frame* than to *Column*; the three values cannot be ordered along a coordinate axis in any meaningful way. If the database matrix contains such attributes – as our database does – network analysis is the only reasonable option.

Projection analysis cannot, in itself, be used to reconstruct the collective relational structure, to determine quantities such as diameter or clustering coefficient, quantities that may reflect the organisation of the society that gave rise to the material record. This is another advantage of network analysis.

Both network and projection analysis can be used for cluster analysis and seriation. Network components are naturally identified with clusters. If the values of all attributes are possible to order, and the study does not aim at revealing relational societal structure or complex evolution, then projection analysis is probably the best choice. This is because the preserved numerical measure of dissimilarity gives more detailed information to fulfil the task. In particular, if all attribute values are relative abundances of different artefact types, then the database is perfectly suited for correspondence analysis.

The mixing of different kinds of attributes causes problems in most quantitative analysis. However, such mixing is sometimes necessary to encode all relevant information about a context. Quantitative analysis of mixed attributes has been attempted several times, for instance by Philip and Ottaway (1983). They performed a cluster analysis of Cypriot hooked-tang weapons, in which they mixed numerical attributes describing weapon size with categorical attributes describing their shape qualitatively. Their cluster analysis had nothing to do with network analysis or projection analysis. Network analysis is not the only approach to quantitative analysis in the presence of categorical attributes.

7. Seriation versus complex evolution, projections versus networks

We noted above that correspondence analysis and other forms of projection analysis are often more suitable for seriation than network analysis. On the other hand, network analysis is needed to reconstruct spatio-temporal processes that evolve along several fronts. We may call this 'complex evolution' (Fig. 5). The reason why network analysis is necessary for this task is that we need to keep track of pair-wise relations between contexts in order to resolve different fronts or branches of evolution.

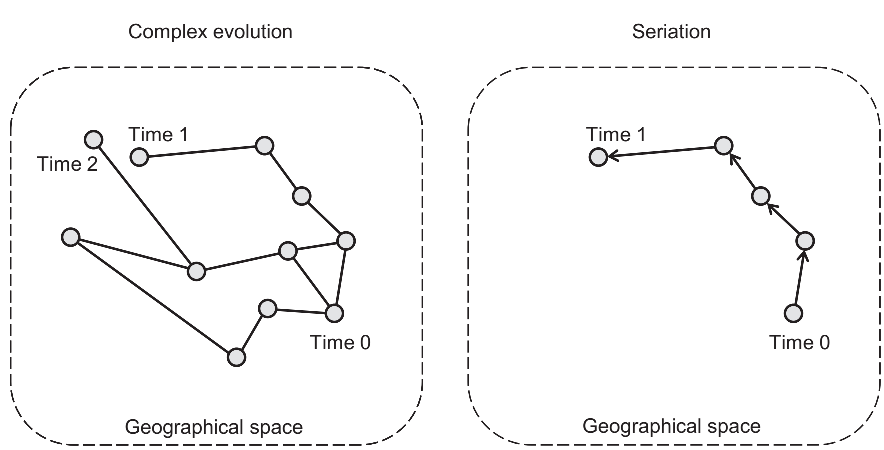

Fig. 5. Spatio-temporal processes that evolve along several fronts may be called complex evolution. Network analysis can be used to reconstruct such processes. Correspondence analysis is better suited for seriation. Contexts are positioned at their geographical location to stress the spatio-temporal nature of the processes. Similar enough contexts are connected. In the seriation, similar contexts are judged to be temporally close. In this example they also happen to be spatially close, indicating a gradual evolution of the cultural expression across the landscape.

Complex evolution in the historical or social sciences may be compared to the branching that occurs in biological evolution. The difference is that in biology, one species can give rise to two new species, but two species never combine to yield a third. The network forms a tree. In contrast, cultural traits can both divide and recombine. A social or societal network that maps complex evolution may contain both branches and loops.

In the diffusion of Hellenistic fired bricks, there is one clear example of complex evolution and the confluence of different trends (Gerding and Östborn, forthcoming). A particular mode of construction, one consisting of external brickwork covering a core of carelessly placed and heavily mortared bricks, evolved over several centuries in a number of distinct stages. By the mid-second century BCE it reached full maturity in Northern Italy, from whence it spread southwards. Parallel to this development, the tradition of re-using roof tiles as ersatz bricks diffused from Southern Italy northwards, steadily becoming more and more common and regularised. In Rome the two techniques merged, an event which very quickly led to the substitution of large Hellenistic bricks for thin triangular bricks, resembling the ersatz bricks. This resulted in the typical Roman construction method known as *structura testacea*. These two trends of development are reflected in similarity network analyses, where two network branches first crystallize in Northern and Southern Italy, respectively, and then merge in the area around Rome in the last decades BCE (Östborn and Gerding, 2014).

8. General similarity networks

In a general similarity network, any kind of similarity relation can be used as a criterion to connect two nodes. To be able to describe how such arbitrary similarity relations can be defined, we first classify different types of attributes in Table 1. The geographical location of a context can always be treated as one of the attributes of a context.

Attributes of types 1), 2) and 3) may be continuous or discrete numbers. Attributes of type 4) are always discrete. The number of finds is always an integer. Dividing with the total number of finds gives a relative abundance as a rational number. Attributes of type 5) are binary, and can be represented as 1 or 0. Attributes of type 6 can be represented as an integer in some predefined set, like 1, 2, 3 or 4. The hierarchy makes the ordering meaningful. Attributes of type 7) can also be represented as an integer in a predefined set. It has to be remembered that in this case the ordering is arbitrary.

To be able to mix these kinds of attributes in a similarity criterion, first we have to define what similarity might mean for each individual attribute type.

A) For attribute types 1), 2) and 4), a distance in attribute space between the values of two contexts *i* and *i'* can be defined. In case the attribute is given as a single number, the Euclidean distance $|v_{i'} - v_i|$ is the obvious measure of dissimilarity. If the value is given as a vector $(v_1, v_2, \ldots, v_n)$, the Euclidean distance $\sqrt{\sum_{j=1}^{n}(v_{i'j} - v_{ij})^2}$ is again the natural choice. For a vector of relative abundances, the Brainerd-Robinson coefficient $1/2\sum_{j=1}^{n}|v_{i'j} - v_{ij}|$ is frequently used (Cowgill, 1990; Mills et al., 2013), where $0 \le v_{ij} \le 1$ and $\sum_{j=1}^{n} v_{ij} = 1$.

B) For attributes of type 3), it is natural to say that two intervals are similar if and only if they overlap. Two settlements whose dating overlaps *could* be from the same time period and thus associated. We get a binary measure of similarity. Other options are possible, for instance the Euclidean distance between the upper end of the lower interval, and the lower end of the upper interval.

C) For attributes of types 5) and 7), the only reasonable choice is to say that contexts which have the same attribute value are similar, others are not. We again get a binary measure.

D) For attributes of type 6), the obvious dissimilarity measure is the number of steps between the two contexts in the hierarchy. This can be expressed as the Euclidean distance $|v_{i'} - v_i|$, where the values are integers.

E) We can also define relative similarity: one context is similar to another if the attribute value of the second context is closer to that of the first context than are the values of most other contexts. Such similarity ranking is the basis of proximal point analysis (see Sections 3 and 5). It is most naturally defined for attributes of types 1), 2), 4) and 6).

Table 1
Attribute types with examples.

	Attribute type	Examples
1	Numerical value	The weight of a find; the concentration of a trace element in the material
2	Numerical vector	The three dimensions of a block (length, width, thickness); the location of a context (latitude, longitude)
3	Intervals	The dating of a context (not before year *X* and not after year *Y*); the three dimensions of a block, given as the smallest and largest occurring values in a group of similar blocks
4	Abundances	The number of vessels of pottery type *X* found in the context
5	Incidences	Does pottery type *X* occur in the context or not?
6	Hierarchical categories	The importance of a settlement (village, small town, regional centre, or capital)
7	Categories	Types of Greek columns (Corinthian, Doric and Ionian)

We condense four kinds of similarity criteria to connect context *i* to context *i'* that address an individual attribute (Table 2).

Criterion iv means that there are fewer than *n* contexts *i''* for which the difference between the values of *A* in *i* and *i''* is smaller than their difference in *i* and *i'*. If $n = 1$, only the most similar contexts are connected.

In principle, all possible combinations of such criteria that address several attributes at the same time, using the logical operators AND, OR and NOT, should be allowed as a similarity criterion. Let us use attributes from our case study of fired bricks as examples (Östborn and Gerding, 2014). Consider, for instance, the following criterion (Fig. 6):

Connect all pairs of brick contexts that are no more than 300 km apart, whose intervals of dating overlap, and in which the function or the shape of the bricks is the same.

An important combined criterion is the following: connect any two contexts for which at least *m* attributes *A* fulfil an individual similarity criterion of type i), ii) or iii). Simply put:

Connect all pairs of contexts that have at least m attributes in common, no matter which.

de Nooy et al. (2005) call this kind of construction an *m-slice*. The importance of this criterion is that *m* defines a general level of similarity. The basic idea is that the higher level of similarity, the higher probability that two contexts are causally or socially related. Such relations are, of course, what we are looking for. The ideal is to be able to represent all relevant relations between contexts as edges, and to avoid all edges that do not represent any relevant relations. Of course, this goal is impossible to reach, but adjusting *m* is a useful tool to come as close to it as possible.

Note that the logical operator OR is necessary to get non-trivial networks if we combine criteria of type iii only. If AND is the only

Table 2
Four types of similarity criteria addressing a given attribute *A*.

i	The difference between the values of *A* is smaller than a given number.
ii	The intervals of possible values of *A* overlap.
iii	The values of *A* are the same.
iv	Relative similarity: the similarity rank with respect to *A* is high.

operator that occurs, then if context C_1 is connected to C_2, and C_2 to C_3, then C_1 is also connected to C_3. The result is a network in which all nodes are connected to all the others within a given network component.[10] The edges then lose their meaning, and can be removed if we keep in mind which nodes belong to which component. In effect, we have made a simple cluster analysis, identifying clusters for which some set of attributes is the same. In other words, we have made distribution maps of different forms. Even if these networks are trivial, they show the generality of the similarity network approach: it can be used to produce all possible kinds of distribution maps, apart from more interesting networks.

Further flexibility in network construction is obtained if we make it possible to choose subsets of contexts, and define a similarity network exclusively on such a subset. This is achieved if we allow the types of criteria for connection of two contexts that are shown in Table 3).

An example of the use of a criterion of type I is to create a similarity network exclusively with contexts where fired bricks are used as columns. An example of the use of a criterion of type II is to create a network where all contexts *possibly* belong to the period 200–100 BCE.

9. Incomplete data

There is one problem with general similarity networks that will almost always surface: that some attribute value is unknown for some contexts. For instance, if a rectangular fired brick is found in the ground at some distance from the nearest building, it is impossible to know whether its function was to be part of a wall, a door frame, or something else. A reasonable solution is to assign the value -1 to such unknown attributes in the database matrix, with the additional condition that this value can never be the basis for a connection. Lack of information about two contexts cannot be treated as a similarity.

10. Can attributes be nodes in general similarity networks?

Unfortunately, the versatile commands to create general similarity networks that were described above cannot be used straightforwardly in networks of type 2 (Fig. 2), where the attributes are the nodes. (Recall the classification of attribute types in Table 1.)

Even though contexts and attributes play symmetrical roles in database matrices in the sense that rows and columns may be interchanged (Fig. 1), there is also a fundamental asymmetry between them. All contexts are equivalent in the sense that they are described by the same array of attributes. Any two contexts have the potential to be identical, if all the attribute values match. In contrast, if we allow different types of attributes according to Table 1, then, of course, these are never equivalent, even though they all apply to the same array of contexts.

[10] A network component is a group of nodes such that it is possible to reach from any node to any other within the group by traversing some number of edges, and such that there are no edges that link a node in the group to any node outside the group.

In practice, this means that we cannot always formulate a single similarity criterion for the connection of two attributes that applies to all possible pairs of attributes. Consider, for example, the following three attributes: 1) the number of found vessels of a certain type (abundance), 2) the importance of the settlement (hierarchical category), and 3) the location of the settlement (measured value in vector format).

We may define a similarity criterion for the connection of attributes 1 and 2 as follows: divide the vessel abundances into the same number of bins as there are predefined levels of importance of a settlement. Connect the attributes when the numerical correlation between this discretised vessel abundance and the settlement importance is large enough. This similarity condition cannot be used, however, to investigate the correlation between vessel abundance and location: that is, the possible geographical clustering of the vessel type. Some other similarity measure has to be introduced in this case.

In such a situation, the edges between different attribute pairs become qualitatively different, since they are the result of different similarity conditions. It is not very meaningful, then, to describe the resulting mesh of different kinds of correlations between attributes as a single network.

In conclusion, if we want to study similarity networks of type 2, we should choose a set of attributes of the same kind from Table 1, and choose a single similarity criterion that applies to all attribute pairs in this set. Brughmans (2010) conducted a study of this sort, in which 16 pottery forms from the Roman East defined 16 incidence attributes. These were connected if the corresponding pottery forms were co-present in different contexts.

11. Similarities and differences

Until now, all the criteria for the connection of two nodes that we have discussed have been based on similarity. It does not make sense to take the opposite road and use only difference as criterion, since all meaningful interpretation of empirical data involves causal relationships between facts, and differences alone cannot be proxies for such relationships. However, similarities in combination with differences can.

We may, for instance, be interested in the occurrence of cultural innovations in a culturally homogeneous milieu. In the network picture, such a milieu is defined by a group of very similar contexts. We may then create a network where we connect any pair of contexts that are very similar, but also show some differences that correspond to a local innovation in one of the contexts.

Differences are needed to create directed networks: it is not possible to attach an arrow to an edge that is defined by similarity only. Directed networks may be useful to determine the direction of the diffusion of an innovation. Say that we want to study the direction of cultural influence between contexts C_1 and C_2. The necessary distinction between the contexts may be temporal, for example. If some finds in C_1 predate similar finds in C_2 by some amount of time that is larger than zero, but not too large for direct transfer, we may connect the sites with a directed edge from C_1 to C_2.

Directed networks need not relate to directions in space or time. Consider, for instance, the way cultural influences may trickle down from the top layer of society to lower hierarchical levels. In an attempt to capture this flow in a network, we may classify contexts with a hierarchical attribute A with possible values 1, 2, 3, and so on, where 1 is the highest societal level. We may then draw an arrow from any context where the value of A is m, to any other context that is similar enough, and where the value of A is $m + 1$.

We see that a truly general framework for creating a network in archaeological data analysis should allow both similarities and

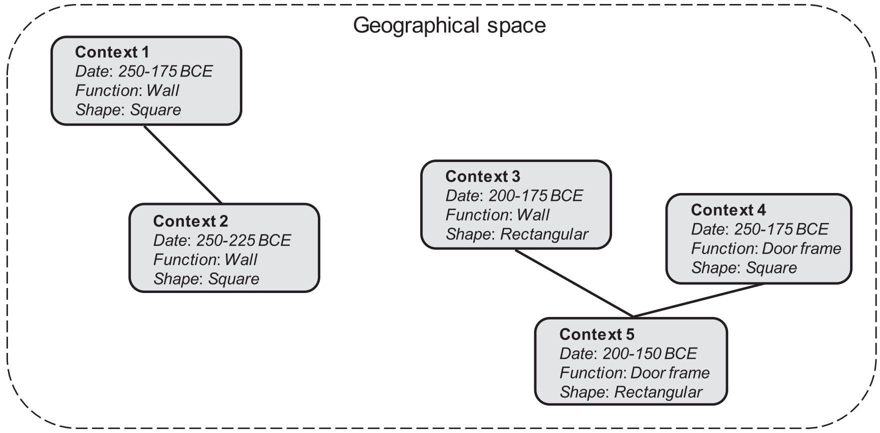

Fig. 6. An example of a similarity network defined by the criterion given in the main text. Each brick context is described by four attributes: *Location*, *Date*, *Function*, and *Shape*. The reason why contexts 1 and 3 are not connected is that they are assumed to be more than 300 km apart.

differences in the formulation of connection criteria. With a bit of semantic sophistry we stick to the name 'general similarity networks', making the concept of similarity so general that even differences are incorporated.

A simpler method to take differences into account is to divide networks defined by similarity alone into smaller pieces and look for differences between the pieces. This can be done if connection criteria like those in Table 3 are used, criteria that choose a subset of the contexts to include in the network. In this way trends in the data can be detected. The most common approach is to place the contexts into temporal bins according to their dating, and to construct one network per bin. Such networks are often called 'time slices'. In this way, Coward (2010, 2013) studied the evolution of social contacts between Neolithic settlements, and Mills et al. (2013) explored the evolution of relations between sites in the US Southwest.

12. The philosophy of general similarity networks

What can similarity networks be used for, and how should they be used? Most importantly, there is no single similarity network that in some sense represents the 'true network'. Some edges will always be false positives in terms of social or causal relations, and the absence of some other edges will always be false negatives. This leads us to the following conclusions:

1) A single edge should never be interpreted as a social or causal relationship between a pair of contexts.
2) Structural properties of similarity networks can be interpreted as a reflection of the structure of the underlying relationships, but properties that are interpreted in this way must be robust when the similarity criterion is varied (within reasonable limits).

Keeping this in mind, we see two ways to use similarity networks, one more formal, and one less so.

Table 3
Similarity criteria that choose subsets of contexts, like subsets shown in distribution maps.

I	The attributes have the same given value, or belong to the same set of values of interest
II	Values given as intervals both overlap a common limited interval of interest

A) The more formal strategy is to look for network structure that, apart from being robust, also differs significantly from the structure of networks in which geographical location or other attributes are assigned randomly. This means that statistical tests are carried out where we look for significant deviations from the null hypothesis of randomness.
B) The less formal strategy is to play around with similarity networks, to ask the database a lot of network questions, and see what patterns it returns. Intriguing answers can be followed up by careful thought, by more formal network analysis, or by further studies in the literature or in the field.

The latter strategy requires an interactive computer program that quickly returns network images that can be visually inspected, and network quantities that can be analysed and compared (see below). Such a program can be seen as an archaeologist's extended brain. After all, trying to see patterns in the material, playing around with similarities, is what the archaeologist does all the time in her or his mind.

There are two advantages with playing computer games. First, the capacity of the computer is much higher than that of the human brain; it can handle and find patterns in larger amounts of data. Second, the computer has no preferences. Humans fall in love with their hypotheses. You see one pattern but you are blind to others, be it a hypothetical trade route or something else. The computer shows all patterns of a given type hidden in the data, so that you have to consider all the possibilities.

On the other hand, the computer has a serious drawback. It needs a predefined database matrix, and it is impossible to represent all knowledge about a material and its context in matrix format. The archaeologist has intuitive knowledge of the relative importance of different factors. Also, there is always peripheral knowledge, not directly related to the material, which happens to be important in any given circumstance.

The output of network analysis should therefore not be overinterpreted. No quantitative or statistical exactness can be guaranteed. There is always some arbitrariness in the way we define our database matrix. Even more importantly, we can never be sure that our material record is a representative sample. At best, similarity network analysis is a versatile, yet systematic tool to formulate qualitative hypotheses.

The choice of reasonable similarity criteria relies on the qualitative judgement of the archaeologist, as does the choice of the range of reasonable variations of these criteria to check for the robustness of observed patterns. Some shared attributes are much more important markers of historical relationships between contexts than others. For instance, the occurrence of identical stamps on fired bricks from two contexts is a stronger indication of social relationship than are identical sizes. For this reason also, it is partially misguided to regard general similarity network analysis as a quantitative method. It is, rather, a qualitative method that produces quantitative output.

Finally, let us metaphorically compare similarity network analysis with projection analyses such as correspondence analysis (see Section 6). The latter methods aim to find the angle from which you get the optimal view of the data. In contrast, studying general similarity networks the way we suggest is regarding the bushy shrubbery of data from many different angles, taking a pause when you get a glimpse of something interesting. In particular, if you see the same intriguing pattern from many different angles, you become confident that there is something significant hiding in the bushes, that it is not just the leaves and branches playing tricks on your eyes.

13. Statistical analysis

Network analysis can only be seen as a scientific method in archaeology if it can be used to reject hypotheses as well as finding support for them. To this end, statistical tests are necessary. If some pattern in the data lacks statistical significance, it should be rejected in the sense that there is no support for it in the material record. Of course, new data may alter the situation.

Just as we should check whether an observed network pattern is robust in exploratory analyses when the similarity criterion for connection is varied, we should check whether a feature that appears to be statistically significant stays significant when the criterion is varied. That is, to be trusted, statistical significance should be robust within a set of similarity networks that we judge to be reasonable.

Also, as mentioned above, we must keep in mind the obvious fact that we can never be sure that the database contains a representative sample. Any hypothesis put forward that is based on statistical significance should not be presented as a statistically established fact, but as an idea strengthened by statistical investigation.

In this section we discuss what kind of statistical tests are suitable for deciding the questions that we think can be addressed by using general similarity networks. We will not give a general overview of statistical methods suitable for network analysis. To this end, Hanneman and Riddle (2005) or Butts (2008a) may be consulted.

We concluded in Section 10 that general similarity networks where the contexts are the nodes are easiest to work with. Such networks are intrinsically spatial (Fig. 2). Therefore, statistical similarity network analysis is well suited to detect significant spatio-temporal processes, such as migration, cultural change or diffusion of innovations.

Basically, the detection of such processes is done by interpreting the edges in the networks as potential rays of causal influences between geographical locations. Such an interpretation may provide indication of complex evolution, as described in Fig. 5.

13.1. Significant similarities

It is hard to test statistically the significance of a specific network pattern as a reflection of causal relationships, but we can at least exclude the null hypothesis that there is no causal process going on at all. In plain language, we should avoid trying to find meaning in meaningless patterns of random similarities.

Under the null hypothesis there is no correlation between the distance between two contexts and the degree of similarity of their attributes. Thus, if we demonstrate a significant correlation, we have shown that there were most probably one or more processes going on. Such a correlation may mean that similar contexts tend to be either closer than dissimilar contexts, or farther apart. In the first case we have shown that the influences that carried the spatial process were mostly short range – neighbour to neighbour. The second case might occur, for instance, if there were two distant trade centres with close contact by sea. Tests of this type were performed by Coward (2013), as she analysed similarity networks in the Neolithic Near East.

We need a method to produce random networks to compare with the true ones. Similarity networks are all about the relations between attribute values. The values themselves do not matter. Therefore, instead of generating new attribute values randomly, we suggest reshuffling the existing values in the database matrix (Fig. 1). In so doing, random relations between the attributes of different contexts are produced, to be compared with the true relations.

Statistical tests that reshuffle the existing data randomly are called *random permutation tests* (Good, 1994). They boil down to a random re-labelling of data points. In our case this means re-labelling the elements in the database matrix (Fig. 1) so that each label *ij* is used for exactly one element both before and after the relabelling. Apart from the fact that permutation tests target the relational essence of similarity networks, they have the advantage of allowing one to not bother about the underlying statistical distribution of attribute values; permutation tests are *non-parametric*.

To exemplify, we are interested in the question of whether more similar contexts on average are placed geographically closer than suggested by chance, but not in the question of whether there is a geographical cluster of contexts in some region that cannot be explained by chance. The location themselves do not matter to us.

This means that we are not suggesting *spatial analysis* (Baxter, 2003; Crema et al., 2010; Hodder and Orton, 1976), which focuses on the spatial patterns of finds. Traditional statistical methods aimed at testing for spatial processes do study such patterns. The basic idea is just that in a spatial process – say the diffusion of an innovation – it is more probable that pairs of adopters are neighbours than in the case of independent, random adoptions (Geary, 1954; Moran, 1948, 1950). From the archaeological perspective, this would give rise to spatial clustering of similar contexts. Naïve application of such methods can produce false positive results, however. Neighbours tend to have things in common that produce non-homogeneous spatial patterns of adoption even in the absence of causal relations between individual adoptions. Taking our case study of fired bricks as an example (Östborn and Gerding, 2014): there may be a good supply of clay in some region that increases the probability that bricks are produced in this area. Jukka Nyblom et al. (2003) developed methods to circumvent these so-called *confounding covariates*.

However, as already stated, we propose a different method. To test for the existence of a spatial process, we suggest reshuffling the locations of the contexts randomly, keeping all the other attributes unchanged (Fig. 7). The result of this method is that the set of possible spatial edges are the same in the true database as in the randomised database. The distributions of the edge lengths can therefore be straightforwardly compared.

We may, for example, compare the median edge length in a true similarity network with median edge lengths in randomised networks defined by the same similarity criterion. If fewer than 5% of the random median edge lengths are shorter than the true edge length (or fewer than 5% are longer than the true median edge

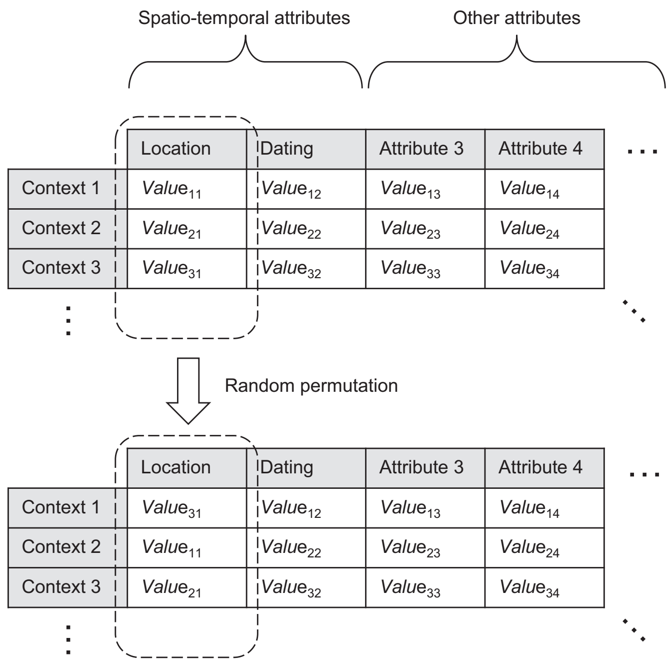

Fig. 7. A random permutation of the database matrix where the locations of the contexts are reshuffled, but the other attribute values are left untouched. Such randomised matrices can be used to decide whether a spatial process is suggested by the data, and reveal some of its characteristics. In an analogous way, a random permutation of the datings can be used to detect a temporal process. The dating may be described by a temporal interval, a phase (a temporal bin), or both. Compare Fig. 8.

length), then we may say that there is a spatial process going on at significance level $p < 0{,}05$.

Having performed such a basic test, we may look for more subtle differences between the true and random edge lengths distribution. Such an investigation may reveal interesting characteristics of the spatial process.

We employed these methods in the study of Hellenistic fired bricks (Östborn and Gerding, 2014), and concluded that there was indeed a diffusion process going on, at a very high level of significance. The median edge length was much closer than expected by chance, suggesting that the diffusion was mostly short-range. Very long edges were also present, but the distribution of edge lengths fell off more quickly for distances greater than 250 km. This finding suggests that 250 km was a typical radius of action for craftsmen during this period.

If the contexts are dated, we may test for a *temporal* process in an analogous manner to that of testing for a *spatial* process. Namely, if earlier contexts are causally related to later contexts, similar contexts should be placed closer along the temporal axis than expected by chance. We may then reshuffle the datings (Fig. 7), and investigate whether similar contexts are significantly closer in time in the true network than in these randomised networks defined by the same similarity condition.[11] If we obtain statistical evidence for a spatial as well as for a temporal process, we may say that the data suggest that a spatio-temporal process is going on.

The next step is to characterise this process (or these processes). To do so, structural network quantities such as the clustering coefficient or the characteristic path length may be calculated

[11] Note that, to obtain relative dating by seriation, we *assume* that there is a temporal process going on: i.e., that more similar contexts are closer in time. Here we take the opposite road and assume that the contexts are already dated, and investigate whether there is a temporal process going on.

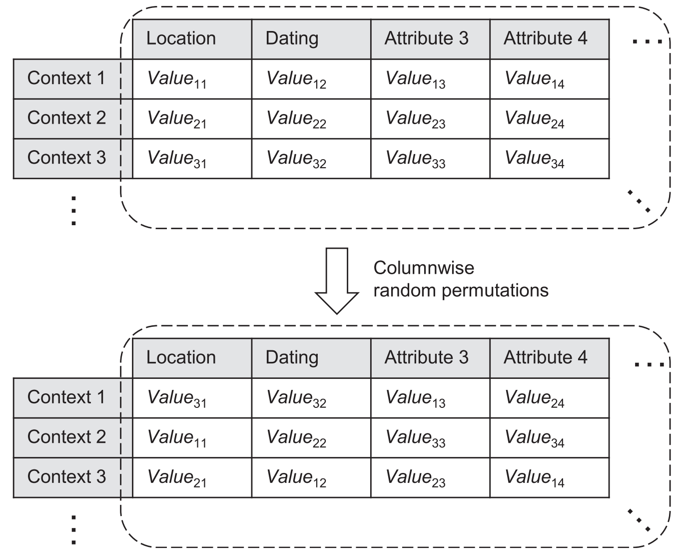

Fig. 8. A random permutation of the database matrix where the values of all attributes are reshuffled: that is, the values in each column are moved to random new rows in the same column. Such randomised matrices can be used to detect significant network structure hidden in the true database. Compare Fig. 7.

(Valente, 1995; de Nooy et al., 2005). The values of these quantities are then compared with those of random networks created by reshuffling the values of *all* attributes independently, not just the location or the dating (Fig. 8).

In this way the significance of the overall topological features of the networks are tested rather than just the spatial or temporal ones. Of course, such statistical tests of network structure can be performed even if we are not interested in any spatio-temporal process. We may only want to characterise social or societal relational structure as reflected in the material record. For example, if the true clustering coefficient is larger than 95% of the clustering coefficients in the randomised networks (defined by the same similarity condition), and the true characteristic path length is not significantly longer than the randomised path lengths, then the data support a small-world hypothesis (Watts and Strogatz, 1998).

If the database has hierarchical elements, things become a bit more complicated. In the randomisation we have to make sure that each context is assigned attribute values that are mutually compatible. Recall the example concerning fired bricks in Section 4. In a randomised network, we have to make sure that the attribute value *Circular* of the attribute *Shape* never occurs in the same context as the value *Wall* of the attribute *Function*.

In the analysis of the database of Hellenistic fired bricks (Östborn and Gerding, 2014), we found that the clustering coefficient was much higher than expected by chance, whereas the diameter was quite small. The data thus indicated that the social network responsible for the dissemination of fired bricks was a small world during the Hellenistic period. On the other hand, there were no hubs of such importance that the similarity networks became scale-free. The data thus indicate that the brick diffusion took place in a ‘democratic’ fashion between small or medium-sized towns, instead of being controlled by a few important centres of influence.

The outlined random permutation tests can be seen as a subclass of so-called conditional uniform graph tests (Butts, 2008b). In such tests, the structure of the actual network is compared with the structure of random networks that fulfil certain conditions. They may, for example, be conditioned to have the same number of nodes and edges as the actual network. The distribution of the

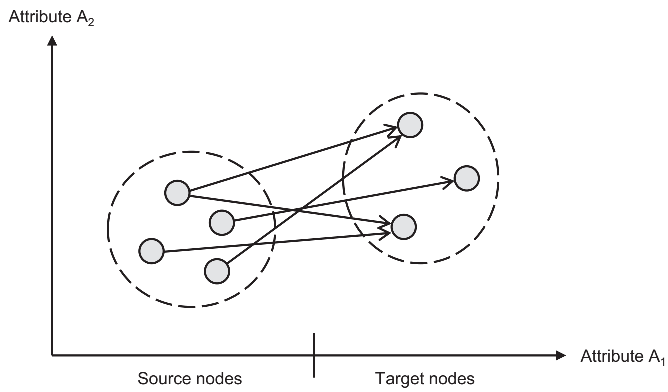

Fig. 9. Illustration of a test of the significance of the difference between the values of attribute A_2 among the source nodes and the target nodes in a directed network. In this example, the values of A_2 tend to be slightly higher among the target nodes. In a more general directed network, the source nodes and target nodes need not be two separate groups. A target node for one directed edge can be the source node for another such edge.

random value of the network measure of interest is then estimated by creating random networks repeatedly. These are taken from a uniform distribution – all random networks that fulfil the conditions are regarded as equally likely. Phillips and Gjesfeld (2013) performed tests of this kind to conclude that some reconstructed social networks in the Kuril Islands deviated significantly from those expected by chance.

13.2. Significant differences

The aim of the tests discussed above is to see whether the structure of a similarity network differs significantly from that expected by chance. Let us now discuss tests to detect significant differences between two groups of nodes in the same similarity network. To do so, we assume that we have created a directed network. The question we ask is: are there any significant differences between the source nodes and the target nodes of the directed edges?

In the simplest case, the source and target nodes are two separate groups defined by the condition that they have two different values of some attribute A_1, or belong to different sets of values of this attribute (compare Table 3). This situation is illustrated in Fig. 9. We may have a similarity condition that links some of the nodes in the first group to some of the nodes in the second group. Then we choose some other attribute A_2, and see whether the values of A_2 differ significantly between the sources and targets of these links. In this form, the question we ask is similar to those asked in regression analysis: if we change variable A_1, is there a significant response in variable A_2? If so, how large is it, and which direction does it take?

In a more general setting, the source and target nodes need not belong to separate groups. For instance, we may want to link similar contexts that differ in the dating by a small amount, where the earlier context is the source, and the later one is the target. Then, of course, the target context of one link may be the source of another.

To perform a test of this kind, the actual directed network may be compared to a random directed network where the values of attribute A_2 are randomly permuted among the contexts. The permutation shown in Fig. 7 is a special case of such a randomisation, where the location corresponds to attribute A_2. As usual, we use the same criterion to link two nodes in the true and in the randomised networks.

Consider the set of all N directed edges in a network, and let k be an index that labels the individual edges. Further, let d_k be the difference in the value of A_2 between the target node and source node of edge k. If the absolute value of the sum $\sum_{k=1}^{N} d_k$ in the true network differs from zero more than do 95% of the corresponding sums in the randomised networks, then the difference in the value of A_2 is significant at level $p < 0{,}05$.

Note that a test of this kind can be performed regardless of the nature of attribute A_2 (Table 1). However, we must modify the definition of d_k depending on which type of attribute we are working with. For example, if A_2 is the location of a context, d_k is naturally defined as the spatial vector that stretches from the source context to the target context. The sum $\sum_{k=1}^{N} d_k$ becomes a vectorial sum of directions. If A_2 is a categorical attribute (such as the function of a fired brick in a construction), it does not make sense to talk about the size or direction of the difference between the source and the target. We may set $d_k = 0$ whenever the value is the same in the source and the target, and $d_k = 1$ otherwise.

As a simple example of the above strategy, we may want to investigate whether there is a significant geographical movement of some type of finds from one time period to another. In this case attribute A_1 is the dating and A_2 is the location. This kind of question is traditionally addressed by the creation of two time slices, so that the centres of gravity of the two distribution maps may be compared.

Here, we link each context from the early period to all contexts from the late period that are similar enough to the source context. The directed links define spatial vectors d_k, which are summed vectorially to yield $D = \sum_{k=1}^{N} d_k$. Before the summing, it is reasonable to downplay the importance of long vectors between distant contexts by normalising all vectors d_k to unit length, so that they become only indicators of direction. The resultant D indicates the size and direction of the trend of geographical movement. If its size is not significant, there is no support in the material for the hypothesis that a geographical movement of the cultural expression of interest took place.

Tests of this kind were performed to see whether there were any significant directions in the diffusion of Hellenistic fired bricks (Östborn and Gerding, 2014). We found a significant trend that the diffusion progressed towards the south-west from the north Aegean in the fourth and third centuries BCE. During the second half of the Hellenistic period, there was an even clearer trend that the diffusion progressed towards the north-west, mainly along the Italian peninsula.

These trends are visible to the naked eye in the database of Hellenistic fired bricks. The statistical analysis merely demonstrated their significance. It is easy to get fooled. Two small groups of brick contexts from different time periods sometimes look geographically well separated, but the movement is not significant due to the small sample size. On the other hand, we may have two large groups of contexts that seem to overlap almost perfectly, but where we can detect a significant movement due to the large sample size.

14. Network analysis software

To use general similarity networks the way we propose, we would like to have software in which it is possible to do the following: 1) create general similarity networks with shorthand commands that can express arbitrary similarity criteria, 2) display spatial networks of type 1 on a geographical map, 3) display networks as you do in social network analysis – with similar nodes close to each other – to visualise the relational structure 4) get basic network quantities in return automatically when a network is created, so that the network structure that results from different similarity criteria can be quickly compared with each other, 5) explore the created network visually and numerically, highlighting components, central nodes, shortest paths, and so on, 6) obtain database information about each node by clicking on it, and 7) perform basic statistical analysis of spatial, temporal and structural patterns.

The wish list is long. Naturally, there is no existing software that is fully satisfactory. Geographical information systems (GIS) may fulfil some of the needs (Batty, 2005; Eiteljorg II and Limp, 2008), and social network analysis programs such as Pajek may fulfil others (de Nooy et al., 2005). However, in preparing our case study (Östborn and Gerding, 2014), we determined that we could not rely on any existing software, but had to write our own. This program is custom-made for our database of fired brick contexts, and its functionality is limited to those questions we were interested in. We see the need for new network analysis software that can handle all kinds of archaeological databases, given that they are expressed in a predefined form, and answer most network questions that we would like to ask.

15. Conclusions

In this paper, we have reviewed how network analysis has been previously used in archaeological research. We concluded that there is unfulfilled potential in the approach. To take a step forward, we have introduced the notion of 'general similarity networks', a flexible framework in which all kinds of similarity relations (including well defined differences) can be used as proxies for causal or social relationships and define links between archaeological contexts.

It is important to use general similarity network analysis not just to support historical hypotheses, but also to reject them (in the sense that there is no support for them in the data at hand). To this end, we have proposed specific types of random permutation tests. They capture the relational nature of networks, and the random reshuffling of existing entries in the database means that no additional assumptions have to be made; all information needed for the tests is contained in the database that is analysed.

Based on the discussion in the preceding sections, we suggest that general similarity network analysis is the preferred formal archaeological analysis method in the following circumstances:

1. when the aim is to infer spatio-temporal processes from a spatially distributed material, in particular if we hope to be able to detect 'complex evolution', where cultural traits evolve along several fronts
2. when the aim is to infer the relational structure of a society from a material record
3. when the aim is to perform cluster analysis in a situation where the archaeological contexts are best described by a mixture of qualitatively different attributes, some of which are categorical

There are most probably other circumstances in which general similarity network analysis may be a valuable tool.

In practical work, we found that general similarity network analysis was useful in the study of the diffusion of fired bricks in Hellenistic Europe (Östborn and Gerding, 2014). It helped us to support hypotheses more solidly, and in some cases it revealed facts about the material that would otherwise have been invisible to us. In Sections 7 and 13 we briefly mention how it helped us answer questions of type 1 and 2. As for questions of type 3, general similarity network analysis made it possible to separate an early cluster of quite homogeneous brick usage in the fourth and third century BCE from a late cluster in the second and first century BCE. The separation of the contexts into these clusters was not possible if individual brick attributes were analysed, but only appeared when they were combined in general similarity criteria. Therefore traditional distribution maps would have been of no use.

These experiences make us confident that general similarity network analysis may be useful in other archaeological studies as well. However, we see the need for new computer software to be able to perform general similarity network analysis without too much trouble.

Finally, we stress that network analysis is just one of many useful tools that may be placed in the toolbox of the archaeologist. In a discipline where we need to make the most of limited evidence, one's mind should be open to all available tools, quantitative as well as qualitative. Sometimes several tools must be combined to complete a given task, just as the carpenter needs many tools to make a chair. Network analysis is at best a useful tool during the early stages of data analysis, but is never the only one you need.

References

Batty, M., 2005. Network geography: relations, interactions, scaling and spatial processes in GIS. In: Unwin, D., Fisher, P. (Eds.), Re-presenting GIS. Wiley, Chichester.

Baxter, M.J., 1994. Exploratory Multivariate Analysis in Archaeology. Edinburgh University Press, Edinburgh.

Baxter, M.J., 2003. Statistics in Archaeology. Wiley, London.

Bentley, R.A., Maschner, H.D.G. (Eds.), 2003. Complex Systems and Archaeology. University of Utah Press, Salt Lake City.

Blake, E., 2013. Social networks, path dependence, and the rise of ethnic groups in pre-Roman Italy. In: Knappett, C. (Ed.), Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford.

Bølviken, E., Schweder, T., Solheim, L., 1981. Preliminary Lecture Notes for the Nordic Research Course in Multivariate Analysis in Archaeology. University of Tromsø, Tromsø.

Broodbank, C., 1993. Ulysses without sails: trade, distance, knowledge and power in the early Cyclades. World Archaeol. 24, 315–330.

Broodbank, C., 2000. An Island Archaeology of the Early Cyclades. Cambridge University Press, Cambridge.

Brughmans, T., 2010. Connecting the dots: towards archaeological network analysis. Oxf. J. Archaeol. 29, 277–303.

Brughmans, T., 2013a. Thinking through networks: a review of formal network methods in archaeology. J. Archaeol. Method Theory 20, 623–662.

Review of Malkin, I., 2011 Brughmans, T., 2013b. A small greek world. Networks in the ancient mediterranean. Class. Rev. 63, 146–148.

Butts, C.T., 2008a. Social network analysis with sna. J. Stat. Softw. 24, 1–51.

Butts, C.T., 2008b. Social network analysis: a methodological introduction. Asian J. Soc. Psychol. 11, 13–41.

Collar, A., 2007. Network theory and religious innovation. Mediterr. Hist. Rev. 22, 149–162.

Collar, A., 2013. Religious Networks in the Roman Empire: the Spread of New Ideas. Cambridge University Press, Cambridge.

Coward, F., 2010. Small worlds, material culture and ancient Near Eastern social networks. Proc. Br. Acad. 158, 453–484.

Coward, F., 2013. Grounding the net: social networks, material culture and geography in the Epipalaeolithic and Early Neolithic of the Near East (~21,000–6,000 cal BCE). In: Knappett, C. (Ed.), Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford.

Cowgill, G.L., 1990. Why Pearson's r is not a good similarity coefficient for comparing collections. Am. Antiq. 55, 512–521.

Crema, E., Bevan, A., Lake, M., 2010. A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record. J. Archaeol. Sci. 37, 1118–1130.

de Nooy, W., Mrvar, A., Batagelj, V., 2005. Exploratory Social Network Analysis with Pajek. Cambridge University Press, Cambridge.

Eiteljorg II, H., Limp, W.F., 2008. Archaeological Computing. Center for the Study of Architecture, Bryn Mawr.

Ferguson, T.J., 1996. Historic Zuni Architecture and Society: an Archaeological Application of Space Syntax. University of Arizona Press, Arizona.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. Incorp. Stat. 5, 15–145.

Gerding, H., Östborn, P., 2014. The Diffusion of Fired Bricks in Hellenistic Europe forthcoming book.

Good, P., 1994. Permutations Tests for Testing Hypotheses. Springer-Verlag, New York.

Graham, S., 2006a. EX FIGLINIS, the Network Dynamics of the Tiber Valley Brick Industry in the Hinterland of Rome. In: BAR Int. Ser, vol. 1486. Archaeopress, Oxford.

Graham, S., 2006b. Networks, agent-based models and the Antonine Itineraries: implications for Roman archaeology. J. Mediterr. Archaeol. 19, 45–64.
Grahame, M., 2000. Reading Space: Social Interaction and Identity in the Houses of Roman Pompeii: a Syntactical Approach to the Analysis and Interpretation of Built Space. In: BAR Int. Ser, vol. 886. Archaeopress, Oxford.
Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78, 1360–1380.
Hanneman, R.A., Riddle, M., 2005. Introduction to Social Network Methods. University of California, Riverside, CA.
Herzog, I., 2013. Least-cost networks. In: Earl, G., Sly, T., Chrysanthi, A., Murrieta-Flores, P., Papadopoulos, C., Romanowska, I., Wheatley, D. (Eds.), CAA 2012. Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology Southampton, March 2012. Pallas Publications, Amsterdam.
Hodder, I., Orton, C., 1976. Spatial Analysis in Archaeology. Cambridge University Press, Cambridge.
Isaksen, I., 2006. Network analysis of transport vectors in Roman Baetica. In: Clark, J.T., Hagemeister, E.M. (Eds.), Digital Discovery. Exploring New Frontiers in Human Heritage. Archaeolingua, Budapest.
Isaksen, L., 2008. The Application of Network Analysis to Ancient Transport Geography: a Case Study of Roman Baetica. In: Digital Medievalist, vol. 4. http://www.digitalmedievalist.org/journal/4/isaksen/.
Jiménez, D., Chapman, D., 2002. An application of proximity graphs in archaeological spatial analysis. In: Wheatley, D., Earl, G., Poppy, S. (Eds.), Contemporary Themes in Archaeological Computing. Oxbow Books, Oxford.
Knappett, C., 2011. An Archaeology of Interaction: Network Perspectives on Material Culture and Society. Oxford University Press, Oxford.
Knappett, C. (Ed.), 2013. Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford.
Malkin, I., Constantakopoulou, C., Panagopoulou, K. (Eds.), 2009. Greek and Roman Networks in the Mediterranean. Routledge, London.
Malkin, I., 2011. A Small Greek World. Networks in the Ancient Mediterranean. Oxford University Press, Oxford.
Mills, B.J., Clark, J.J., Peeples, M.A., Haas, W.R., Roberts Jr., J.M., Hill, J.B., Huntely, D.L., Borck, L., Breiger, R.L., Clauset, A., Shackley, M.S., 2013. Transformation of social networks in the late pre-Hispanic US Southwest. Proc. Natl. Acad. Sci. U. S. A. 110, 5785–5790.
Moran, P.A.P., 1948. The interpretation of statistical maps. J. Royal Stat. Soc. B 10, 243–251.
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.
Nyblom, J., Borgatti, S., Roslakka, J., Salo, M.A., 2003. Statistical analysis of network data—an application to diffusion of innovation. Soc. Networks 25, 175–195.
Östborn, P., Gerding, H., 2014. The diffusion of fired bricks in Hellenistic Europe: similarity networks disclose spatial processes in archaeological data. J. Archaeol. Method Theory (invited and submitted to a special network issue).
Peeples, M.A., Roberts Jr., J.M., 2013. To binarize or not to binarize: relational data and the construction of archaeological networks. J. Archaeol. Sci. 40, 3001–3010.
Philip, G., Ottaway, B.S., 1983. Mixed data cluster analysis: an illustration using Cypriot hooked-tang weapons. Archaeometry 25, 119–133.
Phillips, S.C., Gjesfeld, E., 2013. Evaluating adaptive network strategies with geochemical sourcing data: a case study from the Kuril Islands. In: Knappett, C. (Ed.), Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford.
Rivers, R., Knappett, C., Evans, T., 2013. What makes a site important? Centrality, gateways and gravity. In: Knappett, C. (Ed.), Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford.
Review of Malkin, I., 2011 Ruffini, G.R., 2012. A small greek world: networks in the ancient mediterranean. Am. Hist. Rev. 117, 1643–1644.
Sindbæk, S.M., 2007a. The small world of the Vikings: networks in early medieval communication and exchange. Nor. Archaeol. Rev. 40, 59–74.
Sindbæk, S.M., 2007b. Networks and nodal points: the emergence of towns in early Viking Age Scandinavia. Antiquity 81, 119–132.
Sindbæk, S.M., 2013. Broken links and black boxes: material affiliations and contextual network synthesis in the Viking world. In: Knappett, C. (Ed.), Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford.
Stöger, J., 2011. Rethinking Ostia: a Spatial Enquiry into the Urban Society of Rome's Imperial Port-town. Leiden University Press, Leiden.
Thaler, U., 2005. Narrative and syntax: new perspectives on the Late Bronze Age palace of Pylos, Greece. In: van Nes, A. (Ed.), 5th International Space Syntax Symposium Proceedings, vol. II. Techne Press, Delft.
Valente, T.W., 1995. Network Models of the Diffusion of Innovations. Hampton Press, Creskill, NJ.
Verhagen, P., 2013. On the road to nowhere? Least cost paths, accessibility and the predictive modelling perspective. In: Contreras, F., Farjas, M., Melero, J.F. (Eds.), Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010. Archaeopress, Oxford.
Watts, D.J., 2003. Six Degrees: the Science of a Connected Age. Vintage, London.
Watts, D.J., Strogatz, S., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.