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

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
0305-4403/

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.

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).

	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).

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.