Journal of Archaeological Science 40 (2013) 3001–3010

Contents lists available at SciVerse ScienceDirect

Journal of Archaeological Science

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

To binarize or not to binarize: relational data and the construction of archaeological networks

Matthew A. Peeples [a,*], John M. Roberts Jr. [b]

[a] Archaeology Southwest, 300 N. Ash Alley, Tucson, AZ 85701, USA
[b] Department of Sociology, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

ARTICLE INFO

Article history:
Received 12 February 2013
Received in revised form
10 March 2013
Accepted 12 March 2013

Keywords:
Social network analysis
Network centrality
Sensitivity analysis
U.S. Southwest

ABSTRACT

Over the last several years, network methods and models from the social and physical sciences have gained considerable popularity in archaeology. Many of the most common network methods begin with the creation of binary networks where links among some set of actors are defined as either present or absent. In most archaeological cases, however, the presence or absence of a specific kind of relationship between actors is not straightforward as we must rely on material proxies for assessing connections. A common approach in recent studies has been to define some threshold for the presence of a tie by partitioning continuous relational data among sites (e.g., artifact frequency or similarity data). In this article, using an example from the U.S. Southwest, we present a sensitivity analysis focused on the potential effects of defining binary networks from continuous relational data. We show that many key network properties that are often afforded social interpretations are fundamentally influenced by the assumptions used to define connections. We suggest that, although network graphs provide powerful visualizations of network data, methods for creating and analyzing weighted (non-binarized) networks often provide a better characterization of specific network properties.

1. Introduction

Over the last several decades, social network analysis (SNA) has gained considerable popularity in a number of social science fields, including cultural anthropology and archaeology. Explorations of networks across these domains of research draw on varying theoretical models, analytical tools, and methods of visualization. All of these approaches are linked, however, through a general *relational* perspective in which an understanding of the connections among actors is seen as essential for understanding and predicting the behavior of and outcomes for those actors. SNA has been applied to a wide range of issues such as the link between network position and success (Burt, 1992, 2005; Granovetter, 1973), the development of social groups and identity (Coleman, 1988, 1990), and the influence of social relations on the diffusion of innovation or behavior (Rogers, 2003; Valente et al., 2005), among many others.

Many of the most frequently used formal methods for analyzing network data today originated in the mathematical field of graph theory. Social scientists saw the potential of graph theoretic methods for systematically analyzing and visualizing social relationships by the first half of the twentieth century (Freeman, 2004:69–72), and these graph-based methods have come to dominate contemporary perspectives on SNA (Wasserman and Faust, 1994). By the 1960s, a small number of archaeologists and researchers working in related historical fields had also made forays into graph theory (Doran and Hodson, 1975; Irwin-Williams, 1977; Kendall, 1969; Peregrine, 1991; Pitts, 1965, 1979; Rothman, 1987; Santley, 1991), but these methods never gained the traction in archaeology that they did in the broader social sciences. This is perhaps somewhat surprising given that anthropologists played a major part in the development of SNA (e.g., Barnes, 1954; Bott, 1955; Hage and Harary, 1983; Mitchell, 1969, 1974; Nadel, 1957; Wolfe, 1978). In recent years, however, archaeologists have increasingly begun to analyze network data with formal SNA approaches based on well established models imported from sociology, physics, and complexity science (Bernardini, 2007; Brughmans, 2010; Cochrane and Lipo, 2010; Coward, 2010; Golitko et al., 2012; Graham, 2006; Hart and Engelbrecht, 2012; Isaksen, 2007, 2008; Knappett, 2011; Knappett et al., 2008, 2011; Knappett and Rivers, 2013; Mizoguchi, 2009; Munson and Macri, 2009; Pailes, 2012; Peeples, 2011; Peeples and Haas, 2013; Sindbæk, 2007; Terrell, 2010a, 2010b). As Brughmans (in press) notes, this recent influx of network methods in archaeology has been largely distinct from earlier archaeological applications of graph theory. Thus, the most common methods and models being used in archaeology today

* Corresponding author. Tel.: +1 520 882 6946; fax: +1 520 882 6948.
E-mail address: mpeeples@archaeologysouthwest.org (M.A. Peeples).

0305-4403/$ – see front matter
http://dx.doi.org/10.1016/j.jas.2013.03.014

were developed with little reference to the specific nature of archaeological data (see Knappett, 2011).

In this study, we explore one particular aspect of network data and methods that we argue has, to date, received insufficient attention in the archaeological networks literature: the creation of binary networks in which connections between actors are defined as either present or absent. Binary networks are used quite frequently in contemporary applications of SNA. The most common method for visualizing and analyzing binary network data is to produce network graphs where actors are depicted as nodes and relations between them (also called ties or edges) are depicted as lines. Such binary network graphs are appealing because they provide an easily grasped visual representation of the potentially complex set of relationships among actors (Freeman, 2005). Beyond this, many of the most common network metrics are designed to be used with binary network data (Wasserman and Faust, 1994). Importantly, however, binary networks do not always provide a good fit for the richer, continuous relational data often available to archaeologists (see also Rivers et al., 2013).

The purpose of this article is to more explicitly explore network binarization for the analysis of archaeological networks. Using a dataset from the late prehispanic U.S. Southwest, we present a sensitivity analysis focused on documenting the effects of network binarization on the calculation of several common graph- and node-level network indices. As these analyses demonstrate, many common network properties, which are often lent specific social interpretations, can be extremely sensitive to the assumptions and thresholds used to create ties. From this, we suggest that while binary networks often provide powerful and useful visuals and can provide insights into a number of different kinds of network structures and relationships, binarization is often unnecessary for comparisons of network properties when continuous relational data are available. In many cases, the use of weighted network data may also allow for more nuanced interpretations of network characteristics, and would be substantively preferable unless binarization had a strong archaeological justification in the particular research context. Further, we note that there is a large and rapidly growing body of network literature focused on weighted relational data (e.g., Barrat et al., 2004; Brandes, 2001; Freeman et al., 1991; Newman, 2001, 2004; Opsahl et al., 2010), that has yet to be sufficiently leveraged for archaeological analyses.

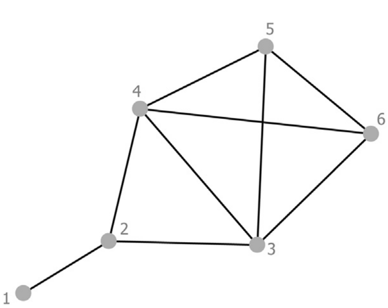
2. Creating and weighting networks

In the simplest terms, a binary network describes a set of actors in which all possible ties among the actors are characterized as either present or absent. Ties in such networks can be directed, where a relation between two nodes is not necessarily reciprocated, or undirected, where relations are assumed to go in both directions. In either case, binary networks can be formally represented as a two-dimensional matrix with a row and a column for each actor, and with 1s and 0s denoting the presence or absence of a tie between each pair of actors in each direction (Fig. 1).

In many contexts where SNA has commonly been applied, the creation of binary networks seems very natural. For example, if the network relation of interest is that of "published in the same journal", a 1 in entry (i, j) of the matrix could indicate that i and j published in the same journal at least once, while a 0 could indicate that they did not. Similarly, a network indicating which members of an organization interact with each other could show a 1 or 0 when a pair did or did not interact over the course of some interval of observation. There are also archaeological cases in which binary networks are wholly appropriate and relatively straightforward to define. Explorations of transportation networks based on well established paths (e.g., roads, rivers, trails, etc.) are an example, as connections between nodes can often be relatively easily defined as either present or absent (e.g., Jenkins, 2001; Pailes, 2012; Peregrine, 1991; Rothman, 1987). For many other classes of archaeological data, however, the substantive problem is far more complicated, as our data do not always relate to the interactions we are interested in measuring in any simple way. For example, if we want to define a network of exchange among settlements, we might decide to create ties between sites that share a certain class of object (e.g., Mizoguchi, 2009). The meaning of such ties is complex, as different potential social processes may drive the distributions of objects, in addition to problems associated with differential sampling and preservation. Nonetheless, again in this case the binary representation may be appropriate.

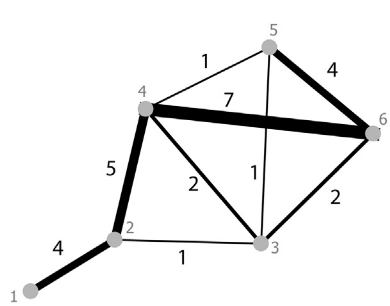
Yet even in cases like those described above, relations may in fact be better represented in terms of their weights or strengths rather than simply their presence or absence. Perhaps a count of the number of journals in which authors co-occurred is preferable to the binarized data above, and the frequency or duration of members' interaction may tell us more about the structure of the organization than the binary measure (Fig. 2). In a transportation network, we might conceptualize a stronger relationship between two points connected by a busy highway than two points connected by a dirt path, and wish to weight ties to reflect such differences. Weights in network data can reflect a simple external categorical classification (i.e., 1-dirt path, 2-paved road, 3-highway), or some measure of the volume of flows across those ties. In general, as these brief examples suggest, weighted networks can capture details about interaction that are subsumed in binary networks.

Many kinds of data frequently used to create archaeological networks lend themselves particularly well to the definition of weighted ties. For example, Brughmans (2010) presents a co-occurrence network of Roman table wares where ties among sites are weighted by the *number* of wares that co-occur (see also Coward, 2010). Another common form of archaeological data that has been used in recent archaeological network analyses is frequency or proportional data such as the relative frequencies of sourced obsidian objects (Golitko et al., 2012) or counts of ceramic types or styles (Cochrane and Lipo, 2010; Hart and



	1	2	3	4	5	6
1	--	1	0	0	0	0
2	1	--	1	1	0	0
3	0	1	--	1	1	1
4	0	1	1	--	1	1
5	0	0	1	1	--	1
6	0	0	1	1	1	--

Fig. 1. An example of a binary network and the associated matrix.



	1	2	3	4	5	6
1	--	4	0	0	0	0
2	4	--	1	5	0	0
3	0	1	--	2	1	2
4	0	5	2	--	1	7
5	0	0	1	1	--	4
6	0	0	2	7	4	--

Fig. 2. An example of a weighted network and the associated matrix.