Journal of Archaeological Science 92 (2018) 63–72

Contents lists available at ScienceDirect

Journal of Archaeological Science

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

Social networks and similarity of site assemblages

Habiba [a,*], Jan C. Athenstädt [a], Barbara J. Mills [b], Ulrik Brandes [a]

[a] University of Konstanz, Germany
[b] University of Arizona, USA

ARTICLE INFO

Article history:
Received 24 March 2017
Received in revised form
11 November 2017
Accepted 15 November 2017

ABSTRACT

There have been a number of similarity measures developed in a variety of research domains. Generally, these measures are developed for a specific context and later reused in other contexts and applications, depending on their ease of use and perceived applicability. While there might be statistical reasons to use a particular similarity index, the results of other measures should be taken into account as well, as various similarity measures do not necessarily have similar contextual meaning. Two entities can be similar with respect to a certain similarity criterion but may be distinct in terms of another. Thus, an understanding of the mathematical logic behind a method is crucial to the interpretation of the resulting network of similarities. We review a number of methods from the literature, for constructing similarity networks among disparate entities, regarding their applicability on data from archaeological sites. Formally, given an N X p matrix of N entities with p distinct classes of attributes, how are the entities comparable to each other with respect to the kinds of attributes they share? We distinguish three qualitatively different families of similarity measures for deducing relationships among entities that may meaningfully map onto various distinct social phenomena, such as migration, material acquisition, and movement of goods and skills, among others. Entities can be compared based on: (a) non{uniform weighting of attributes, (b) asymmetric dominance relationships, and (c) rank correlations. We ground the significance and distinction of these classes of measures by giving comparative and contextual examples of selected methods on a case study of archaeological collections pertaining to 1200–1500 CE from the US Southwest region. We attempt to elucidate the differences in outcomes and their meanings when choosing various similarity methods for comparing disparate entities.

1. Introduction

Researchers in the field of archaeology generally have to rely on sparse and fragmented information to understand the social behavior of the populations under study. Given the material discovered at different archaeological sites, one — but certainly not the only — way of estimating the strength of a relationship between them is by evaluating how "similar" they are to each other. Calculating pairwise similarities between site assemblages results in a network that can be seen as a proxy for social interactions and has become one popular basis for analyzing social networks in archaeology (e.g (Hart and Engelbrecht, 2012; Mills et al., 2013b, 2015; Munson, 2013)).

Measuring similarity among entities is one of the most applied techniques in multivariate data analysis. Yet, similarity in and of itself has yet to be concisely defined. A simple — and slightly circular — definition of it, "is a numerical measure of the degree to which two data objects are alike" (Tan et al., 2005). What makes two entities "alike" can vary depending on what the data represents, the type of attributes, and how the attributes are compared. In general, two entities are similar if they share many categorical attributes, or if the values of their numerical attributes are relatively "close". Dissimilarity — the complement of similarity — especially distance measures, have also been frequently used to compare entities.

There have been a number of similarity/dissimilarity measures developed in a variety of domains, such as, natural language processing, information retrieval (Manning et al., 2008, Mihalcea et al., 2006, Santini and Jain, 1999), computational biology (Heringa, 2001, Song et al., 2008), and cluster analysis (Balcan et al., 2008, Strehl et al., 2000, Tan et al., 2005), among others. Most of these measures are grounded in theoretical justifications for distinctive and specialized comparisons that do not necessarily have similar

* Corresponding author.
E-mail address: habibah@gmail.com (Habiba).

https://doi.org/10.1016/j.jas.2017.11.002
0305-4403/

contextual meanings. That is, two entities can be similar with respect to a certain measure but may be distinct in terms of another similarity index. This is one of the downsides of having an abundance of such methods. Many of them seemingly estimate the same general concept yet are operationalized by different procedures and on different bases. As a consequence, the results they generate, may not bear a clear correspondence to the abstract concept of similarity that they are meant to mimic.

Application of network methods in archaeology has increased considerably in the last decade (Brughmans, 2010, Collar et al., 2015). (Knappett, 2013) provides a comprehensive state-of-the-art guide to the main themes and approaches of network analysis for archaeologists. Trends of migrations and movements (Mills, 2011, Mills et al., 2013a, 2013b), exchange of ideas and diffusion of technology (Golitko and Feinman, 2015, Östborn and Gerding, 2014), intra-community social and political dynamics (Munson, 2013, Munson and Macri, 2009, Scholnick et al., 2013, Paris, 2014), and transformation of social landscapes over different social and temporal scales (Mills, 2007) are some of the topics network methods have been used to address in archaeology. In recent years, multiple studies have been published on the reconstruction of networks of similarities, based on the production, consumption, and deposition of ceramic assemblages, most notably in the geographic region of the US Southwest during the late Prehispanic period (Borck et al., 2015, Mills, 2007, 2011, Mills et al., 2013a, 2013b, 2015, Peeples and Roberts, 2013). Using the Brainerd-Robinson (BR) index (Brainerd, 1951, Robinson, 1951), networks are reconstructed that are based on similarities of consumption of ceramics among the settlements at various spatial and temporal scales. This network view of site similarities provides a supplemental approach in systematically exploring the social, political, and economical patterns of interaction among settlements in the region during that period. In other areas of the world, the BR index has also become a common way for comparing assemblages and assessing similarity including (Mesoamerica) Europe (Golitko, 2015, Golitko et al., 2012, Golitko and Feinman, 2015) and Northeast North America (Hart, 2016, Hart and Engelbrecht, 2012).

In this work we selectively review some of the more frequently used similarity measures from the literature in relation to specific concepts in archaeology. Such an approach has been outlined by Östborn and Gerding (Östborn and Gerding, 2014). We compare these similarity measures to the BR index, which is currently most widely used in archaeological research. We argue that it is crucial to choose a method that corresponds to the specific research question and show that it is important to use and compare multiple methods. This can lead to a more nuanced picture of the historical and social contexts being explained by the type of proxy data used to represent social interactions of different kinds. Lastly, we apply some of the proposed methods to the dataset from the US Southwest that was used in (Mills et al., 2013a) and compare the resulting networks.

In Table 1 we list a set of measures that we use as a base for the methods proposed in this paper. A comprehensive survey on similarity/dissimilarity measures can be found in (Choi, 2008, Choi et al., 2010, Everitt and Rabe-Hesketh, 1997).

2. Proposed methods

This work focuses on the following aspects of constructing similarity networks. Section 2.1 gives the formal definitions of similarity measures that transforms an N X p multivariate matrix of N entities represented by p attributes into an N X N similarity matrix of the entities. In Section 2.2 we provide simulated examples to underscore certain characteristics of the Brainerd-Robinson technique, that is frequently used in archaeology for establishing similarities among entities. The objective of this exercise is to emphasize the relevance of application of a measure of choice in a given realistic social process. Furthermore, in Table 1 we list the high level characteristic of the proposed methods to distinguish them from each other. In Section 2.3, we make the case for non-uniform and unequal significance of attributes representing an entity and propose a transformation that assigns variable weights to attributes based on their assumed significance before the application of a similarity measure. Finally, in Section 2.4 we outline the approach to reconstruct cross-temporal networks of similarities that naturally translate to a number of real social processes, such as, movement, migration, and trading among others. Table 2 lists the notations used in the following sections.

2.1. Similarity measures

In the following we give detailed descriptions for the selected similarity measures used for reconstructing networks of interactions among entities. We describe weighted, unweighted, symmetric, asymmetric similarities, and rank correlations among uniformly weighted attributes.

Dominance relationship: An entity x dominates an entity y, if and only if, $S_y \subset S_x$, that is S_y is a proper subset of S_x.

Relationships among groups of people in a geographically proximal setting are not necessarily symmetric. For example, there are power, status, resources, and economic disparities that result in asymmetric dynamics among participating entities. In many such cases, the relationship can be more logically contextualized as supplier-consumer, source-sink, or politically dominant-subordinate relation. The dominance relationship proposed here captures the most basic form of such imbalanced types of relationship among entities. This method establishes dominance relationship between a pair of entities as a subset or containment relationship. Two entities represented by a common set of attributes are related to each other through complete containment of the attribute set of one in the other. The dominant site is larger and encompasses all the attributes of the dominated site. Mathematically, it encodes the partial order relation among a set of entities. This dominance relationship is built of subset relations, however, there can be many other ways dominance relation can be constructed that emulates different forms of social dominance. This method can be further refined into binarized and non-binarized dominance.

Binarized Dominance: An entity x dominates entity y if it contains all attributes of y.

$$\text{Dominance}_1(x, y) = \begin{cases} 1 & \text{if } \; S_y \subseteq S_x \\ 0 & \text{otherwise} \end{cases}$$

Non-binarized Dominance: An entity x dominates entity y if each attribute of x is quantitatively greater than the corresponding attribute of y. In the case that the attributes are indicating quantity of objects found of a certain type, this relationship indicates that larger sites dominate smaller ones. Therefore this measure is particularly sensitive to a consistent sampling strategy.

$$\text{Dominance}_2(x, y) = \begin{cases} 1 & \text{if } \; Q_{x,i} > Q_{y,i} \, \forall i \in [1, p] \\ 0 & \text{otherwise} \end{cases}$$

Brainerd-Robinson (BR) index: The Brainerd-Robinson index compares the similarity in the proportions of values of attributes.

$$BR(x, y) = 1 - \sum_{i=1}^{p} \left| \frac{x_i}{\sum_{i=1}^{p} x_i} - \frac{y_i}{\sum_{i=1}^{p} y_i} \right| \Big/ 2$$

This measure is specifically developed in archaeology for