Journal of Archaeological Science: Reports 51 (2023) 104100

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Sampling variability and centrality score comparisons in archaeological network analysis: A case study of the San Pedro Valley, Arizona

John M. Roberts Jr. [a,*], Emily Dorshorst [a], Yi Yin [b], Matthew A. Peeples [c], Ronald L. Breiger [d], Barbara J. Mills [d]

[a] University of Wisconsin-Milwaukee, United States
[b] Utah Valley University, United States
[c] Arizona State University, United States
[d] University of Arizona, United States

ARTICLE INFO

Keywords:
Archaeological networks
Assemblage data
Bootstrap
Centrality
Sampling variability
San Pedro Valley

ABSTRACT

Archaeological network analysis often focuses on networks in which ties between sites reflect some sort of similarity, such as in artifact assemblages. Site centrality is often of interest, but an apparent difference in two sites' centrality may not be meaningful once sampling variability is considered. We investigate bootstrap assessments of sampling variability in centrality scores of a set of late pre-Hispanic archaeological sites in the San Pedro Valley, U.S. Southwest, for which ceramic assemblage data can be transformed into networks of ceramic similarity. We considered a variety of bootstrap confidence intervals for site eigenvector centrality scores and the implications of these intervals for interpretation of the site's structural importance. In analysis of the San Pedro Valley for CE 1300–1349, small differences among site centrality were not statistically distinguishable, but moderate to large differences were, with conclusions consistent across methods of constructing bootstrap confidence intervals. Similar patterns were evident when examining a broader region in which the Valley is located. It appears that substantive interpretation of site centrality differences often will be justified.

1. Introduction

Network data and analytic methods can inform many archaeological research questions (Brughmans and Peeples, 2017; Mills, 2017; Peeples, 2019). Rather than considering archaeological sites or contexts in isolation, the formal network approach models relations among such contexts as a network in which ties reflect some sort of social connection between contexts or the people who occupied them. Many applications use similarity of sites' artifact assemblages, such as pottery sherds or other type/sourced materials, to construct this network; greater similarity between two assemblages is taken as a stronger tie between the two sites (Mills, et al., 2013a; Mills, et al., 2013b). Researchers then interpret network measures as representing archaeologically significant structural characteristics (Birch and Hart, 2018; Lulewicz, 2019; Peeples and Haas, 2013). Centrality and other characteristics of nodes may drive variation in outcomes for the contexts being studied, and network analysis allows quantification of structural features that otherwise can be discussed only informally.

The nature of material culture and its representation in data introduces inherent uncertainty in archaeological analyses, and archaeological network analysis is no exception. An observed assemblage is one realization of an underlying "true" probability structure that determines the likelihood of finding a particular artifact at a particular site, with this structure ultimately stemming from the use or production of objects by the site's inhabitants. We view the observed assemblage as a sample from that underlying probability structure and expect that different samples would result in different observed assemblages, suggesting that the observed assemblage is subject to sampling variability. That is, the observed assemblage is unlikely to exactly match the underlying probability structure, and the variation in the observed assemblage from different samples is the sampling variability that we are interested in here. This sampling variability in the assemblage implies sampling variability in measures derived from the assemblage, including networks of assemblage similarity and any analysis of those networks. For archaeological interpretations, assessment of this uncertainty due to sampling variability is important in deciding if values such as two sites' network centrality scores are meaningfully different. However, it can be challenging to assess uncertainty in network

* Corresponding author.
E-mail address: jmrob@uwm.edu (J.M. Roberts).

https://doi.org/10.1016/j.jasrep.2023.104100
Received 24 December 2022; Received in revised form 9 May 2023; Accepted 19 June 2023
Available online 25 July 2023

measures; classical descriptive measures typically provide no formula-based standard error. Fortunately, in archaeological networks based on similarity of sites' artifact assemblages, the bootstrap offers a natural method for assessing sampling variability.

In this perspective, small differences in centrality scores may not indicate real differences in the sites' structural importance. Some previous work on assemblage networks has mentioned this (Gjesfjeld, 2015; Mills et al., 2013a; Mills et al., 2013b; Mills et al., 2015; Peeples et al., 2016), but to our knowledge no empirical research has explored this issue in greater depth. In this paper, we investigate bootstrap assessments of sampling variability in network centrality scores of a set of late pre-Hispanic U.S. Southwest sites for which networks can be constructed from ceramic similarity data. We consider confidence intervals for sites' centrality scores and their implications for interpretation of the sites' structural importance. Although focused on a specific time and place, our investigation of sampling variability in centrality scores is pertinent to archaeological network analyses in many other settings.

2. Overview of archaeological networks

Archaeologists have increasingly used network data and social network analysis to understand the structure of relations among a set of actors, using models and methods developed in other fields to study people's interactions with one another, material things, and the natural environment (Brughmans and Peeples, 2017; Collar et al., 2015). In many such studies, archaeologists take a social network approach to investigate ties among archaeological sites. In those analyses, relations among sites, not simple descriptions of site characteristics, are the explicit focus. Interpretation of network analytic measures is enriched by knowledge of the archaeological setting.

Many recent archaeological networks have been constructed from measured similarity between site-level artifact assemblages (Hart and Engelbrecht, 2012; Golitko et al., 2012; Golitko and Feinman, 2015; Habiba et al., 2018; Hart et al., 2017; Mills et al., 2013a, 2013b, 2015, 2018; Östborn and Gerding, 2014; Roberts et al., 2021; Weidele et al., 2016). This similarity is typically calculated from categorical classifications of artifact assemblages, with greater similarity when two sites' assemblages are more alike. For ceramic assemblages, each artifact may be classified into a *ware* or *type* category based on the artifact's physical characteristics and/or design, with raw data giving each site's sherd counts of those categories (Mills et al., 2013b). The measured similarity of categorical distributions at pairs of sites produces a symmetric network of the sites, in which measured similarities are interpreted as network tie weights (see Peeples and Roberts, 2013; Mills et al., 2013a; Peeples and Haas, 2013). Most research to date has calculated tie weights via archaeology's Brainerd-Robinson statistic (Brainerd, 1951; Robinson, 1951) or the equivalent dissimilarity index (Duncan and Duncan, 1955), also equivalent to city block distance between two sites' assemblage profiles. Transforming the continuous weights into traditional binary—present or absent—ties risks loss of information, but many network measures such as node centrality can still rely on the weighted ties (Mills et al., 2013b; Peeples and Roberts, 2013).

As discussed in the Introduction, the fact that such a network is constructed from observed assemblages means that sampling variability in the assemblages introduces uncertainty in the network. This element of uncertainty is our focus in this paper. However, there surely are other important sources of uncertainty in the underlying assemblage data. It is likely that there is some misclassification of artifacts, and different analysts might make different decisions as to which artifacts to use when measuring site similarities, particularly in a classification that includes many fine categories. Also, errors in site occupation dates or, in the ceramic context, ware use or production dates would introduce variability that is not represented in this bootstrap approach. Likewise, there is further uncertainty in any analysis that relies on apportioning objects into different time periods, chooses time periods to highlight, or uses a specific assemblage similarity measurement. These additional sources of uncertainty are not addressed by the bootstrap as used here. While the sampling variability that the bootstrap depicts is likely to be more substantial in practice than these other potential sources of error, the other sources will still be present to some extent in most realistic analyses. These and other concerns, such as differential preservation of artifacts of different kinds or ages, are ubiquitous in archaeological research (see Peeples et al., 2016).

3. The San Pedro Valley, Arizona

Previous network analyses have considered the San Pedro Valley, marked by the "micro-scale" label in Fig. 1, in southeast Arizona, U.S. (Mills et al., 2013a; Mills et al., 2013b; Mills et al., 2015). Larger settlements in the northern portion of the valley have been extensively documented (Clark and Lyons, 2012) and are shown in Fig. 2. The San Pedro Valley is an especially vivid illustration of migration processes, with substantial entry of migrants from northeastern Arizona in the late 13th century, and archaeological accounts have identified the known sites as local or migrant communities (Di Peso, 1958; Gerald, 2019). Settlement origins can be distinguished by architecture (e.g., platform mounds, compounds, pueblos, kivas, and plazas) and decorated ceramics, especially wares such as Maverick Mountain Polychromes which show strong technological and design similarities to ceramics produced immediately earlier in northeastern Arizona (Clark and Lyons, 2012; Woodson, 1999). After CE 1350, there is considerable evidence that hosts and migrants co-occupied a number of sites (Clark and Lyons, 2012). Few settlements persisted past CE 1400, with most large sedentary villages unoccupied by CE 1450, as much of the southern Southwest saw declines in large villages and population coalescence (Hill et al., 2004).

Mills et al. (2013b; 2015) discussed archaeological interpretations of the San Pedro network analyses and highlighted several main points from the analyses. First, the analyses indicate the importance of network centrality early in this period for sites' persistence. Several sites that were highly central prior to the migration remained occupied even after widespread depopulation of the region. Second, after CE 1300, several migrant communities' network centrality was among the highest in the region. This was likely linked to their importance as producers of distinctive and highly valued new ceramic wares. Third, late in the pre-Hispanic period and amidst dramatic depopulation, centrality scores of the remaining sites became more equal, perhaps reflecting decaying cultural distinctions between hosts and migrants (also see Mills et al., 2013a).

4. Bootstrap

Archaeological networks that are based on artifact counts are a natural fit with the bootstrap resampling framework (Efron and Tibshirani, 1993; Roberts et al., 2021). The resampling uses each site's observed distribution of artifacts into classification categories; at each site, many repeated samples of the same size as the site's original assemblage are taken with replacement from the site's observed data. Because data likely reflect separate data collection efforts at the various sites, when resampling it is appropriate to set the sample size at each site to the site's observed number of artifacts. Then, each resampled assemblage will yield a site-by-site similarity network of the type discussed above. A network measure of interest can be calculated from each resampled network, resulting in an estimated sampling distribution for this measure. The resulting assessment of sampling variability can be used for hypothesis tests, confidence intervals, and other purposes.

This bootstrap has been used in a small number of archaeological network studies (e.g., Gjesfjeld, 2015; Lulewicz, 2019; Mills et al., 2013b; Peeples et al., 2016). While the bootstrap is not guaranteed to "work" in all possible situations (Bickel and Freedman, 1981; Chernick, 2007), recent simulation results suggest that it does provide reasonable estimates of sampling variability in archaeological networks (Roberts

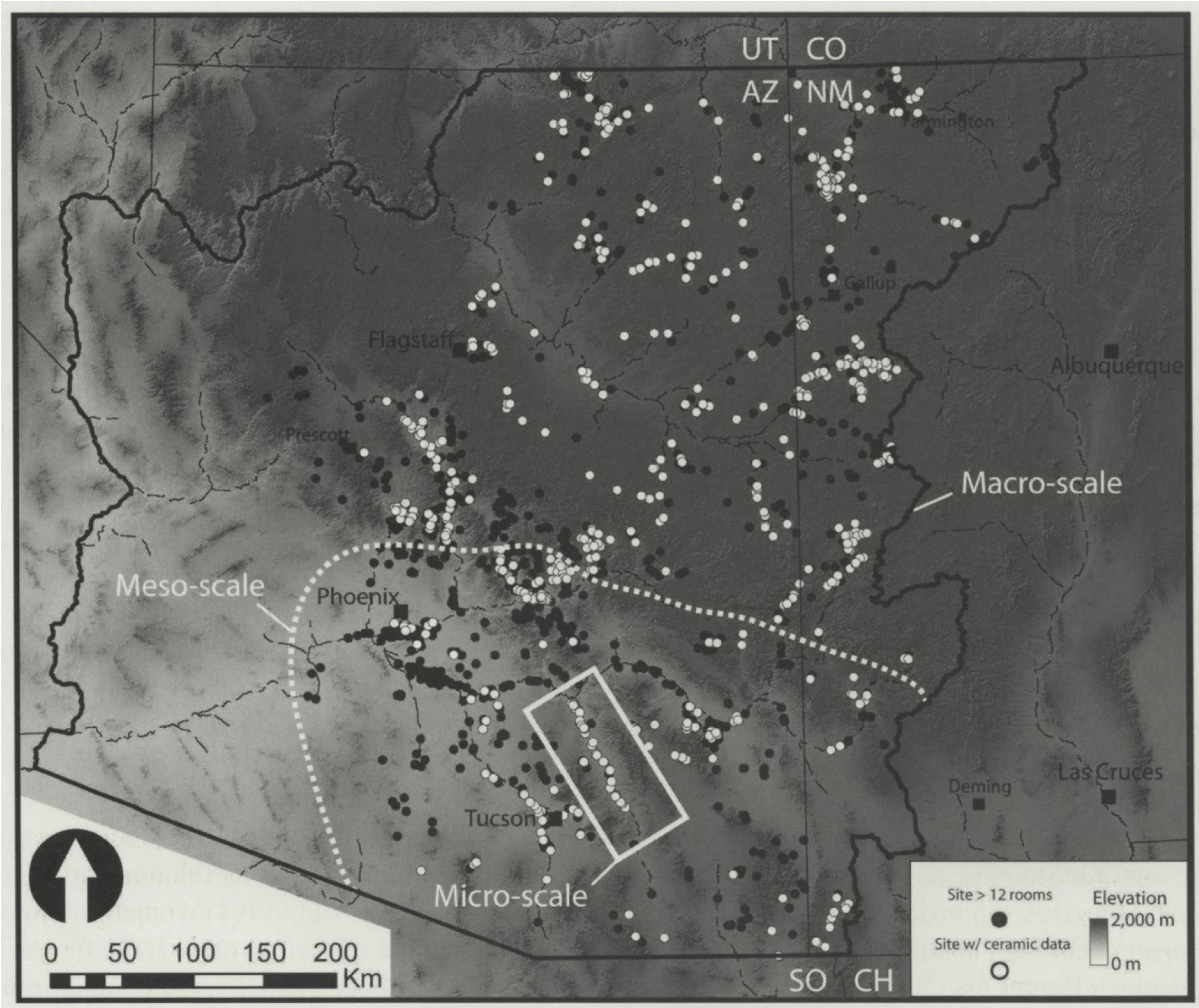

Fig. 1. Archaeological sites in Arizona and western New Mexico, with the San Pedro Valley and neighboring regions indicated.
Source: Mills et al., 2015

et al., 2021). As noted above, the sampling variability indicated by the bootstrap is not the only source of noise in archaeological network data (see Peeples et al., 2016).

There are a variety of approaches to constructing bootstrap confidence intervals for an unknown parameter from the distribution of parameter estimates; here, the unknown parameter would be a site's "true" centrality that would be calculated from underlying (unobserved) ware probabilities. For example, Efron's (1979) percentile method takes the values that define the lowest and highest 2.5 % of the distribution as the endpoints of a 95 % confidence interval for the parameter of interest—in our case, the site's true eigenvector centrality. Other variations on the bootstrap confidence interval are discussed below.

5. Current study

5.1. Data

Our data come from a larger project involving over 1,600 sites that were occupied between CE 1200 and 1500 in Arizona and New Mexico. We focused on relatively large residential sites (at least thirteen rooms) in the San Pedro Valley, which are generally well-known and -documented (Clark and Lyons, 2012), and the broader region in which the Valley is situated. Ceramic data at these sites indicate ware and type classifications of ceramic sherds. Wares are defined by technological attributes, such as appearance and production techniques. Types are finer classifications, here largely based on surface decoration, nested within wares. Network analyses to date have typically used ware-level classification (e.g., Mills et al., 2013a; Mills et al., 2013b; Mills et al., 2015), usually focusing on decorated ceramics with likely ceremonial use and ideological importance (see Mills et al., 2013b; Mills, 2016). In this paper, networks reflect similarity in ware frequencies of decorated ceramics only.

To consider temporal change in networks, long-occupied sites' ceramic assemblages must be apportioned to shorter time intervals. We used Roberts et al.'s (2012) method for apportioning into 25-year intervals and combined periods into 50-year intervals. The method relies on a common trajectory of ceramic types' popularity over their production spans and requires population history estimates for all sites; we used a model-based approximation of Hill et al.'s (2004) approach (see also Bernardini et al., 2021), described in detail in the Supplementary Material. Aggregating type data into wares results in a series of tables giving estimated ware counts at each occupied site in each 50-year period, with "CE 1300" shorthand for "the period CE 1300–1349." The Supplementary Material provides all ceramic data.

Mills et al. (2015) emphasized the importance of networks at different spatial scales. There, the microscale was represented by the northern San Pedro Valley. The mesoscale included that microregion along with much of the Arizona portion of the Basin and Range

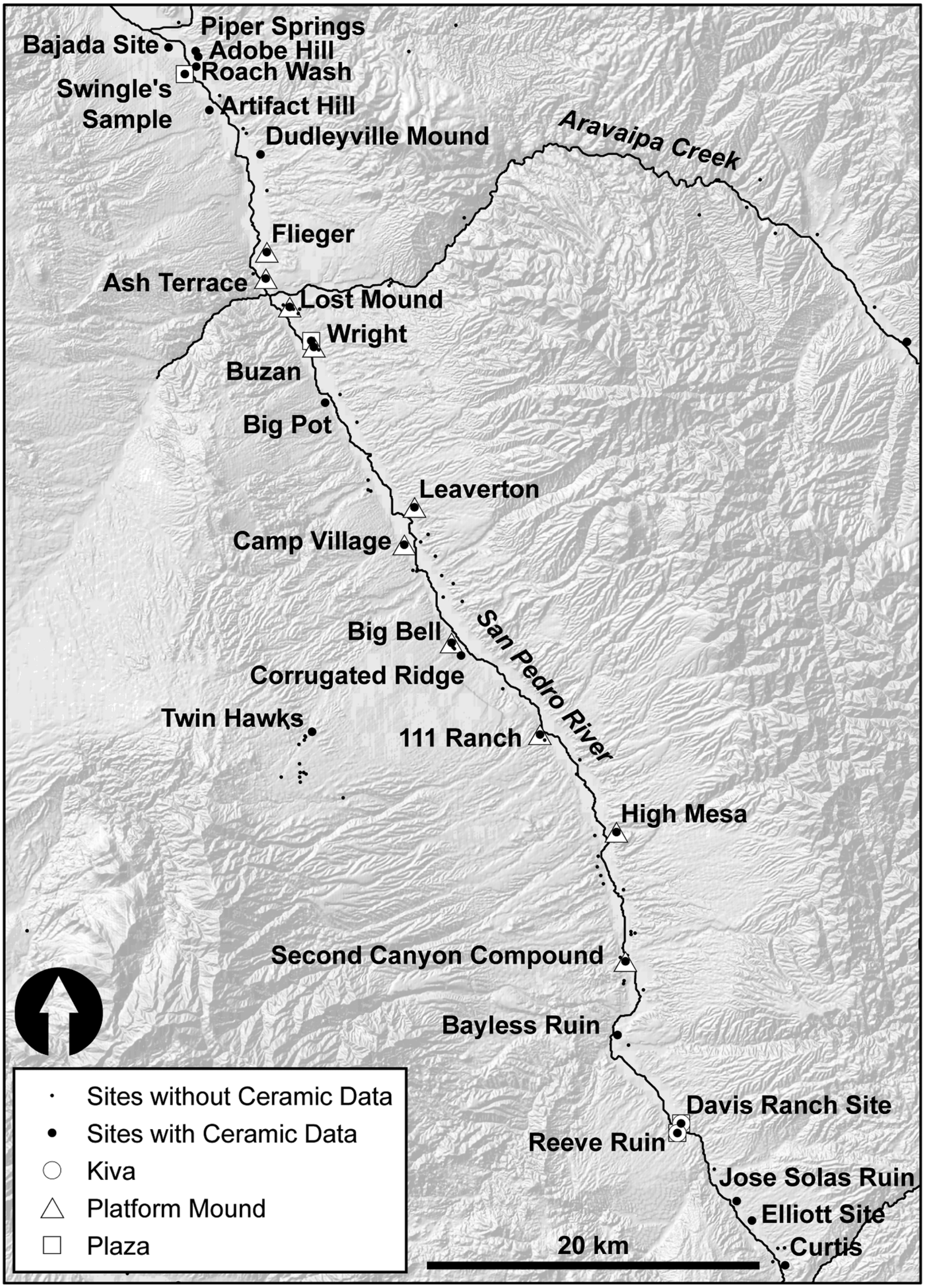

Fig. 2. Archaeological sites in San Pedro Valley, Arizona (Artifact Hill and Roach Wash are not included in the analysis).
Source: Mills et al., 2013b

physiographic province. Culturally, this region mostly aligns with the Hohokam archaeological culture area. The macroscale covered the entire project area in Arizona and New Mexico west of the North American Continental Divide. This work brought archaeology's emphasis on varying geographical scales into the domain of network analysis. Here, we drew on this emphasis by first considering the San Pedro sites as a microregion before analyzing the larger mesoregion.

5.2. Network analysis and bootstrap

From the dissimilarity index D_{ij}, we took $(1 - D_{ij})$ as the weight on tie (i, j), so that greater similarity corresponds to a greater tie weight. While bootstrap resampling involved all of the wares in each site's assemblage, D_{ij} used only the subset of decorated wares. In previous work on ceramic similarity networks, eigenvector centrality was identified as a substantively appropriate measure (Mills et al., 2013a, 2013b, 2015, 2018) because, in Borgatti's (2005) terms, it is reasonable to conceive of ideological and cultural influence as prominent "flows" in this network. An actor's eigenvector centrality is proportional to the sum (weighted by tie strengths) of the centralities of others to whom the actor is tied, with scores given by the eigenvector for the largest eigenvalue of the network's adjacency matrix **A** (Bonacich, 1972). Here we normalized the centrality scores so that 1 represents a typical score within the network.

At both regional scales, we created 10,000 bootstrap replications by sampling with replacement from each site's observed ceramic assemblage (including decorated and undecorated wares). Each replication involved a with-replacement sample of the observed size from each site's assemblage. Resampling was at the level of types and aggregated into ware-level data for analyses; this approach accommodates the possibility of different types within a ware having different production spans. Processing of each replication was as discussed above for the observed data: apportioning into 50-year periods, construction of a network of site similarities, and calculation of site eigenvector centralities. Data for subsequent analyses therefore consisted of 10,000 centrality scores for each site. In principle, a site's assemblage might not overlap with any other site, creating an isolate in the network. Other possibilities could also leave the network disconnected, such as if the bootstrap left a site with no decorated sherds in a particular period. Such issues would affect calculation of eigenvector centrality. However, this was not encountered in any of the 10,000 bootstrap replications for San Pedro, and only extremely rarely for the broader region, so it had no material impact on the results below.

We considered several classic approaches to constructing bootstrap confidence intervals, listed here. For an overview that covers many of these approaches, see Manly (1997), with further details in Efron and Tibshirani (1993) and Chernick (2007) and theoretical justifications in Efron and Tibshirani (1993) or the articles cited in the Supplementary Material. These classic bootstrap confidence intervals are representative of methods used by practicing researchers, but certainly do not exhaust the possibilities available in the vast and ongoing technical literature that has developed around the bootstrap. We also did not take up the often-heated debates in that literature concerning the relative merits of the different approaches. We used a variety of confidence interval methods simply to check if the resulting confidence intervals were roughly similar, not to declare one method or another superior. Our analysis included the following confidence intervals, described further in the Supplementary Material: (i) the standard bootstrap confidence interval, with a bootstrap standard error used to construct a traditional confidence interval; (ii) Efron's (1979) percentile confidence interval, discussed above; (iii) Hall's (1986) percentile confidence interval, based on the distribution of differences between the observed estimate and the bootstrap estimates; (iv) the bootstrap-t percentile confidence interval from the double bootstrap, in which the iterated bootstrap permits calculation of a t-statistic in each bootstrap replication; (v) Booth and Hall's (1994) calibrated percentile confidence interval from the double bootstrap, with the iterated bootstrap providing coverage estimates for adjustment of the confidence interval; and (vi) Efron's BCa confidence interval, in which the confidence interval's endpoints are adjusted for bias and acceleration.

5.3. Summary of analytic plan

We can summarize the steps in our analyses as follows.

1. Data: Counts of classified (by type) ceramic sherds by sites; site occupation spans; use or production spans for types, possibly specific to sites.
2. Apportioning: Use site and type time data to apportion sherd counts to time periods for sites that were occupied for more than one period.
3. Network: Create site-by-site network of decorated assemblage similarity; calculate eigenvector centrality scores for the sites in this network.
4. Bootstrap: Resampling from each site's assemblage; for each site, draw a resample of the same size as the site's observed assemblage; repeat many (here 10,000) times; for each bootstrapped dataset, repeat (2) and (3); collect the eigenvector centrality scores for each site from each bootstrapped dataset.
5. Confidence intervals: For each site, apply one or more confidence interval methods to the collection of bootstrapped eigenvector centrality scores; if desired, use bootstrap eigenvector centrality scores from two sites to make confidence intervals for differences in scores or confidence regions from plots.

The Supplementary Material includes the software code used to carry out our analyses.

6. Results

6.1. Analysis of the San Pedro Valley

We first took the San Pedro Valley sites as the whole network. Because more sites were occupied in CE 1300 than in other periods, we focused on the results for the 21 sites occupied in the CE 1300 interval and explored whether different confidence interval methods produce similar results with these data. The sites' observed eigenvector centrality ranged from 0.554 for Second Canyon Compound to 1.196 for Swingle's Sample. A visualization of the binarized network is given in the Supplementary Material.

Fig. 3 shows the various confidence intervals for Piper Springs' centrality; the dot marks the site's observed centrality. Although the differences among the methods at Piper Springs were large relative to those at other sites, in absolute terms the differences were quite small. The Y-axis scale makes clear that any apparent differences across the methods are minor and unlikely to appreciably affect interpretations; note that the pattern of a longer standard bootstrap interval is consistent across sites. (Figures for Bayless Ruin and Dudleyville Mound, and a table reporting all confidence intervals for all sites, are provided in the Supplementary Material.) Also, the standard deviations of the upper and lower bounds of each site's confidence intervals across the six methods were small: the largest of these standard deviations was roughly 0.035, compared to a standard deviation of the site centrality estimates of 0.187, and many were much smaller. (These standard deviations are displayed in the Supplementary Material.) For these data, then, the choice of confidence interval method does not appear to be too important.

Fig. 4 presents Efron's percentile 95 % confidence intervals for all sites. We first consider whether incorporating sampling variability via the confidence interval changes any interpretations implied by the original centrality score estimates, initially looking for overlapping confidence intervals as a crude indication of statistically indistinguishable scores. Among the four sites with the highest estimated centrality, confidence intervals overlap a great deal. In the group of ten sites with the next highest estimated centrality, there is also considerable overlap among the confidence intervals, and between these two groups, the degree of overlap depends on which pair is being examined. The overall impression is still of the second group having lower centrality, but not all pairs appear statistically distinguishable. The six sites in the third group are clearly distinguishable from the other groups. Although within the group there is substantial overlap, the lowest centrality site, Second Canyon Compound, is obviously much less central than any other site. Note that in general the confidence intervals were shorter for sites with larger observed decorated assemblages, but there are exceptions to this pattern, and the correlation between observed number of decorated sherds and length of Efron's percentile confidence interval was only –0.39 (–0.49 when using sites' ranks on these variables). Fig. 5 gives 95 % confidence intervals under Booth and Hall's first method using the double bootstrap, with the overall impression quite like that from Fig. 4. (The Supplementary Material gives equivalent figures for the other

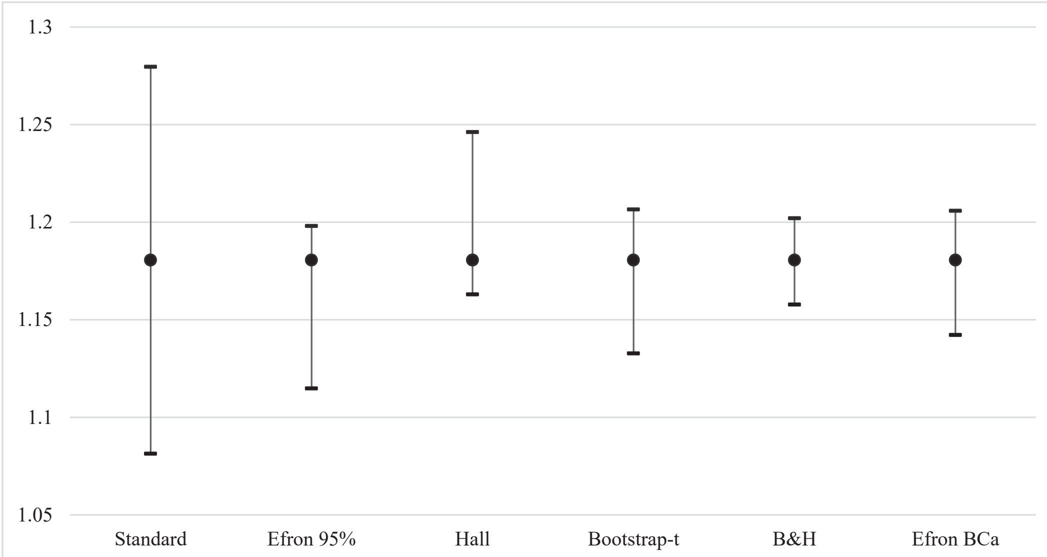

Fig. 3. Comparison of bootstrap confidence intervals for Piper Springs, CE 1300.

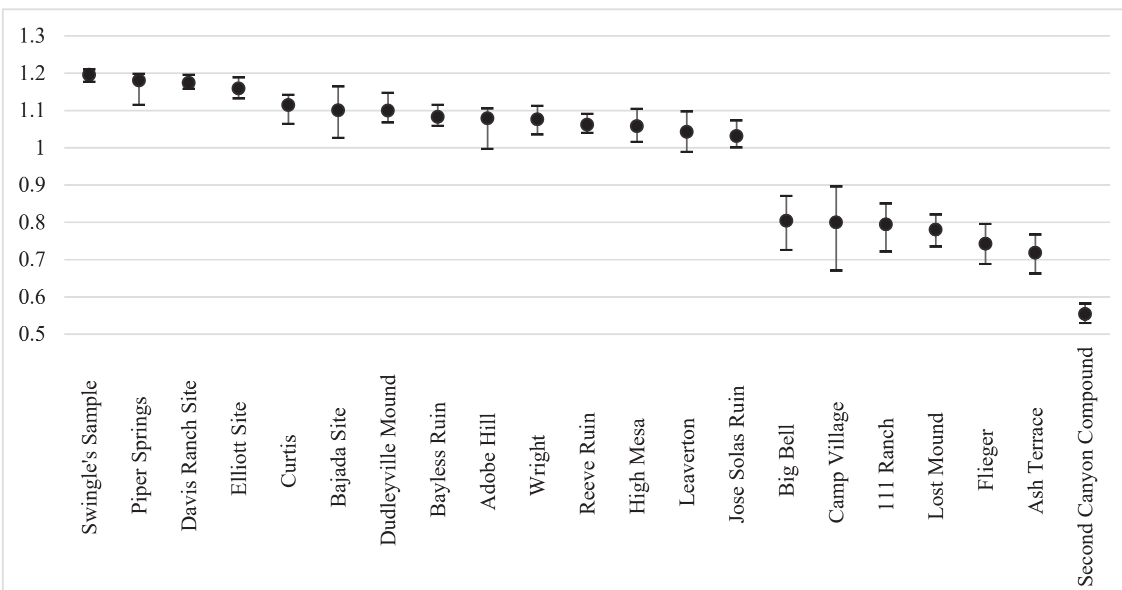

Fig. 4. Efron's percentile 95 % confidence intervals, San Pedro, CE 1300.

confidence intervals.).

We also explicitly examined centrality differences in particular pairs of sites, expressed as $c_i - c_j$ for additive comparisons and as c_i/c_j for multiplicative comparisons. By calculating these quantities in each bootstrap replication, confidence intervals for the difference between two sites can be constructed in the same way as a confidence interval for a single centrality score, with non-independence between the sites' scores reflected in the analysis. We illustrate this via Efron's percentile 95 % confidence intervals for the additive and multiplicative differences in the pairs {Wright, Lost Mound}, {Camp Village, Ash Terrace}, {Swingle's Sample, José Solas Ruin}, and {Bayless Ruin, Reeve Ruin}. These were chosen to highlight two pairs with quite similar observed eigenvector centrality scores and two with quite different scores.

Fig. 6 shows confidence intervals for the additive differences between these sites' centrality scores $c_i - c_j$; the Supplementary Material contains a corresponding figure for the confidence intervals for multiplicative differences c_i/c_j. No difference between scores is indicated by a value of 0 in the additive case and a value of 1 in the multiplicative case. When we informally considered overlap between the site confidence intervals in these pairs via Figs. 4 and 5, Wright appeared distinct from Lost Mound, and Swingle's Sample likewise appeared distinct from José Solas Ruin. However, the confidence intervals for Bayless Ruin and Reeve Ruin overlap, and the intervals for Camp Village and Ash Terrace substantially overlap. When we consider the differences in these pairs more formally with confidence intervals for the differences in Fig. 6, the conclusions change a bit. The confidence intervals in Fig. 6 for additive differences between Wright and Lost Mound and between Swingle's Sample and José Solas Ruin exclude 0, and the confidence intervals for the multiplicative difference between Wright and Lost Mound (1.304, 1.468) and Swingle's Sample and José Solas Ruin (1.114, 1.193) likewise exclude 1. These conclusions agree with the informal comparison of the sites' separate confidence intervals. However, for Bayless Ruin vs. Reeve Ruin, the confidence intervals for the difference exclude 0 for the additive comparison (0.001, 0.041) and 1 for the multiplicative

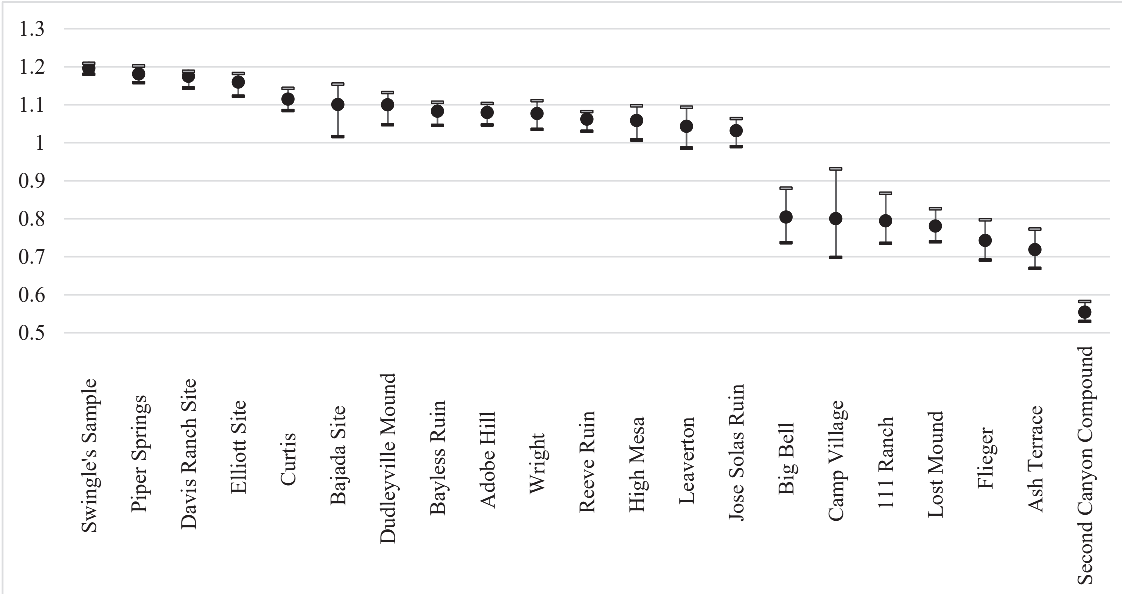

Fig. 5. Booth and Hall 95 % confidence intervals, San Pedro, CE 1300.

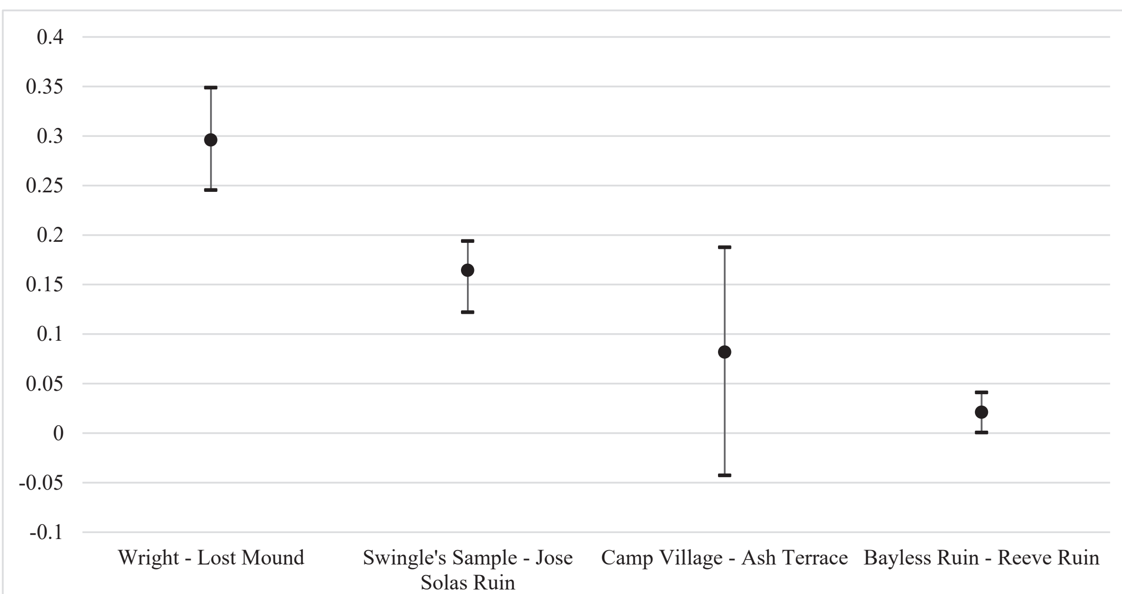

Fig. 6. Efron's percentile 95 % confidence intervals for additive comparisons of site centrality, CE 1300.

comparison (1.001, 1.039), a different conclusion than in the informal comparison. The formal Camp Village vs. Ash Terrace comparison agrees with the informal comparison.

A different way of considering these comparisons between sites is to construct two-dimensional confidence regions for pairs of scores. To our knowledge, methods for bootstrap confidence regions for such pairs (e.g., Yeh and Singh, 1997) have not been standardized, so we used a simple approach that shares the spirit of Efron's percentile method for a one-dimensional confidence interval and is akin to bootstrap confidence regions for points in correspondence analysis (Greenacre, 1984; Ringrose, 1992). In Fig. 7, each point represents the two sites' centrality scores in a single bootstrap replication, with the graph showing the 9,500 (95 %) such points closest in Euclidean distance to the point of means of the scores across the 10,000 replications. The superimposed circular or oval shape is the convex hull, and the 45° line represents equality of the two scores. Treating the convex hull as a 95 % confidence region, if the line does not pass through the hull, then the two centralities seem statistically distinct. When, on the other hand, the line passes through the convex hull, the two scores are not statistically distinguishable. Note that beyond the way in which eigenvector centrality scores are conceptually related to each other, normalization of the scores introduces some inherent dependence of one site's reported score on the others. But unless the number of sites is very small, this probably has little impact.

For several pairs, the conclusion from this approach is like that from the confidence intervals for additive and multiplicative differences. Wright and Lost Mound (and, in the Supplementary Material, Swingle's Sample and Jose Solas Ruin), appear to have distinct centrality scores because the 45° line does not pass through the convex hull, while Camp Village and Ash Terrace do not appear distinct. These interpretations agree with those from the confidence intervals for additive and multiplicative differences above. On the other hand, for Bayless Ruin and

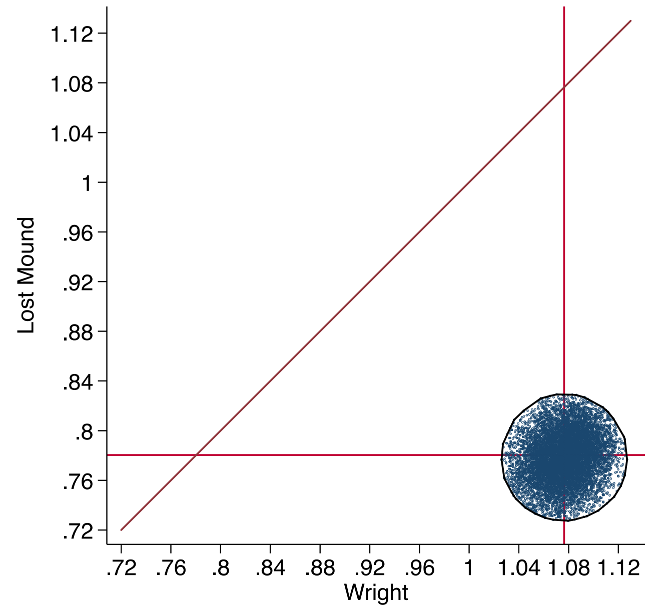

Fig. 7A. Convex hull of points representing centrality of Lost Mound and Wright, CE 1300, for the 95 % of bootstrap replications closest to the observed centrality for these sites, indicated by horizontal and vertical lines. The 45° line indicates equality of the sites' centrality.

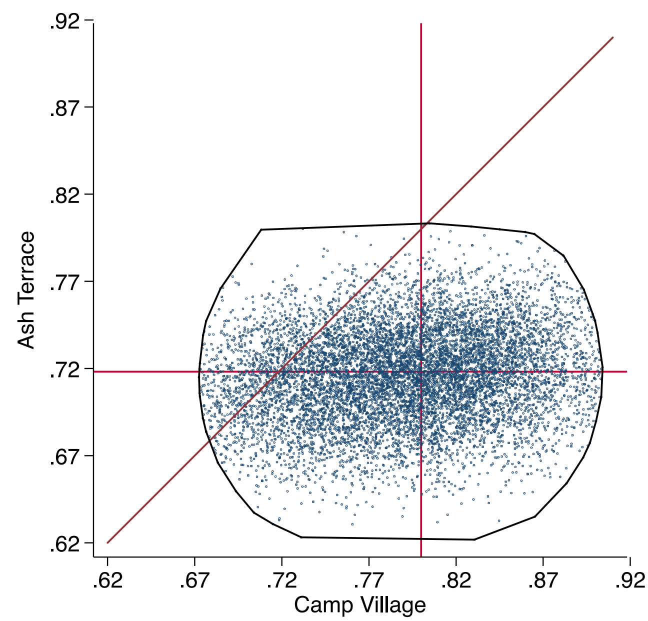

Fig. 7B. Convex hull of points representing centrality of Ash Terrace and Camp Village, CE 1300 for the 95 % of bootstrap replications closest to the observed centrality for these sites, indicated by horizontal and vertical lines. The 45° line indicates equality of the sites' centrality.

Reeve Ruin, the 45° line passes through the convex hull, while the confidence interval for the difference between these two sites in Fig. 6 sits just above the equality threshold (0 for the additive comparison, and 1 for the multiplicative comparison shown in the Supplementary Material). Note, however, that these confidence intervals are very close to the threshold, so that a different confidence interval method, a slightly different confidence level, or a different (due to chance) set of bootstrap replications might indicate no difference, and the 95 % confidence region indicated by the convex hull is also by its nature somewhat informal. The comparison of these two sites is thus less clear cut and

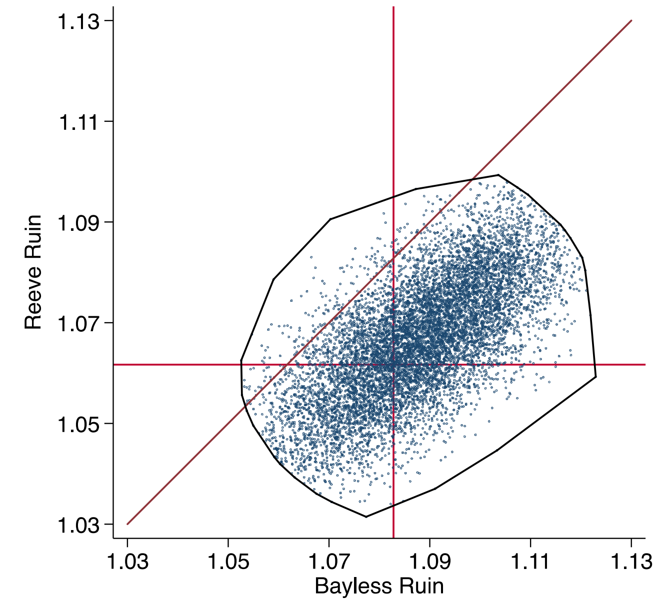

Fig. 7C. Convex hull of points representing centrality of Reeve Ruin and Bayless Ruin, CE 1300 for the 95 % of bootstrap replications closest to the observed centrality for these sites, indicated by horizontal and vertical lines. The 45° line indicates equality of the sites' centrality.

would bear further investigation if there were substantive interest in the specific comparison of Bayless Ruin to Reeve Ruin.

We mentioned earlier that analysts may prefer to interpret centrality rank instead of actual centrality scores. Rather than confidence intervals, Fig. 8 is a "heatmap" indicating the proportion of the bootstrap replications in which a site had a given centrality rank. The proportions are shown in the figure and the continuous shading is darker as the proportion is larger. Outlined diagonal entries refer to replications in which a site's centrality rank was the same as its rank in analysis of the observed data. First, the extremely dark cells in the corners show that Swingle's Sample and Second Canyon Compound were virtually always the highest and lowest centrality sites across the bootstrap replications. For some other sites, the diagonal entry is darkest because the most frequent rank for that site was its observed rank, even if that largest proportion might fall below 0.5. In other cases, a non-diagonal entry is darkest, when the most frequent rank in the bootstrap replications was not the observed rank. However, the disparity was never great in those cases, and the most frequent ranks usually were adjacent to the observed. The largest variability in rank was for sites in the middle range; for these sites, the precise rank is rather uncertain. Although differing in its focus on ranks, this impression is consistent with the confidence intervals above, and site ranks appear to be reasonably consistent across the bootstrap replications.

We can relate these results to previous discussions of the San Pedro Valley. For instance, Mills et al. (2013b) commented as follows on the CE 1300 period:

> In the following period (AD 1300–1350), as migration into the valley continued, the greatest number of sites was occupied. The two most well-known migrant enclaves, Reeve Ruin and the Davis Ranch site, as well as sites in the vicinity with probable migrant components (i.e. Curtis, Elliott, Bayless Ranch Ruin, Jose Solas Ruin) exceed or rival the centrality scores of first-comer villages such as Flieger and Ash Terrace, with all in the group of sites with above average centrality. We think that this is related to the fact that the migrants became producers of highly valued decorated ceramics called Salado polychromes, and, through exchange of these vessels, their villages established connections with earlier 'well-connected' local sites. The migrant position in the San Pedro social network was enhanced by

	Centrality Rank																				
Site	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Swingle's Sample	.931	.051	.014	.004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Piper Springs	.050	.425	.227	.231	.033	.018	.007	.004	.002	.002	.001	.000	.001	0	0	0	0	0	0	0	0
Davis Ranch Site	.017	.484	.471	.029	.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Elliott Site	.002	.039	.283	.641	.034	.001	.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Curtis	0	0	.000	.019	.323	.322	.170	.071	.040	.023	.016	.010	.006	.001	0	0	0	0	0	0	0
Bajada Site	.000	.001	.005	.045	.279	.161	.141	.086	.072	.055	.045	.042	.036	.034	0	0	0	0	0	0	0
Dudleyville Mound	0	0	.001	.030	.273	.313	.201	.096	.049	.020	.011	.005	.001	0	0	0	0	0	0	0	0
Bayless Ruin	0	0	0	.000	.006	.053	.222	.322	.219	.117	.047	.011	.002	.000	0	0	0	0	0	0	0
Adobe Hill	0	0	0	0	.001	.011	.061	.166	.238	.157	.112	.097	.081	.077	.000	0	0	0	0	0	0
Wright	0	0	0	.001	.033	.075	.117	.128	.135	.153	.150	.115	.070	.024	0	0	0	0	0	0	0
Reeve Ruin	0	0	0	0	.000	.000	.002	.020	.098	.266	.311	.228	.074	.001	0	0	0	0	0	0	0
High Mesa	0	0	0	.000	.011	.030	.052	.072	.093	.123	.154	.192	.172	.100	0	0	0	0	0	0	0
Leaverton	0	0	0	.000	.008	.016	.027	.036	.049	.068	.093	.140	.226	.337	0	0	0	0	0	0	0
Jose Solas Ruin	0	0	0	0	0	0	.000	.000	.005	.017	.061	.160	.331	.426	0	0	0	0	0	0	0
Big Bell	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.372	.294	.172	.115	.039	.008	0
Camp Village	0	0	0	0	0	0	0	0	0	0	0	0	0	.000	.359	.159	.114	.150	.114	.105	0
111 Ranch	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.210	.301	.249	.179	.052	.009	0
Lost Mound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.058	.231	.406	.285	.021	.000	0
Flieger	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.002	.015	.056	.232	.552	.144	0
Ash Terrace	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.001	.004	.039	.223	.734	0
Second Canyon Cmpnd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Fig. 8. Proportions of San Pedro sites' centrality rank across bootstrap replications. Darker shading indicates greater proportions; unshaded cells indicate ranks that did not appear in any bootstrap replications. Entries of ".000" indicate non-zero proportions that round to less than 0.001 (representing fewer than 5 out of 10,000 replications).

their central role in the production and distribution of these decorated ceramics.

Our current analyses seem consistent with these remarks. The confidence intervals for Flieger and Ash Terrace indicate that even when sampling variability is considered, these sites have meaningfully lower centrality than the named migrant sites. The confidence intervals show that the centrality of the migrant sites is difficult to distinguish, but clearly exceeds that of these two first-comer sites. In this case, then, the substantive conclusion does not appear to be affected by sampling variability.

6.2. Analysis of the broader region

We also considered the broader region consisting of the Chihuahuan Lowlands (CL), Papagueria (P), Phoenix Basin (PB), Safford (S), San Pedro (SP), Santa Cruz (SC), Tonto Basin (TB), and Upper Gila (UG) areas, corresponding to the "*meso*-scale" boundary in Fig. 1. Although San Pedro is part of this broader region, its analysis in this wider context is not directly comparable to its analysis as a region unto itself. Partly that is just a consequence of what was previously the whole network now being part of a larger one, so that the nature of the ties between San Pedro sites and those in other subregions could affect the relative centrality of the San Pedro locales. However, a more fundamental difference is that the set of wares designated as decorated is not the same in the two analyses. Previous researchers have made this choice because some of the wares that are relatively uncommon, and therefore distinctive, when considering the San Pedro Valley in isolation are much more common in other parts of the broader region (see Mills et al., 2015).

For instance, red-slipped wares lacking other decoration can be reasonably treated as decorated when considering only San Pedro sites, but these are much more prevalent in data from sites in the Tonto Basin. As wares that are extremely common in sites' assemblages probably have a utilitarian rather than expressive nature, this shift in the definition of decorated seems appropriate for the larger region. One consequence, though, is that changes in relative centrality within San Pedro reflect more than just the impact of considering ties in the rest of the network. The Supplementary Material reports the list of wares considered decorated for the analysis of the broader region. We also restricted our analysis to sites for which the apportioned count of decorated sherds in the site's observed assemblage for that period was at least five; for CE 1300, this left 96 sites. While this low cutoff permitted the inclusion of many sites and a large network, it also left some sites with very wide confidence intervals.

Fig. 9 shows Efron's percentile confidence intervals for sites active in CE 1300 across the broader region, roughly in thirds and ordered by their observed centrality scores. Because of the wide confidence intervals in panel B, the vertical scale differs across the panels. The figure suggests that, as in the isolated analysis of San Pedro, many sites of differing observed centrality genuinely are statistically distinguishable, but small differences in observed centrality are likely not. However, in the broader region there are also sites whose centrality is essentially indistinguishable from that of all other sites, so some modesty is required when interpreting analyses of this larger network.

Fig. 9 includes each site's regional designation. When considered in this broader context, the San Pedro sites appear to have had generally high centrality in this period; note the difference for Second Canyon Compound, from lowest centrality in the San Pedro-only analysis to roughly average in the analysis of the broader region. This high centrality of San Pedro sites is consistent with the Valley's recognized importance for social transformation and migration in the region in this period and its location in the larger region's center. As before, interest may lie in assessing centrality ranks, and the Supplementary Material includes a heat map similar to Fig. 8 displaying variability in sites' centrality rank across the bootstrapped replications.

7. Conclusion

The results presented here are encouraging with respect to the value of network analysis of archaeological assemblage data and the incorporation of sampling variability into such analysis. First, it did not appear that confidence intervals for eigenvector centrality scores were much affected by the choice of bootstrap confidence interval method. Different methods' confidence limits were quite similar, with little

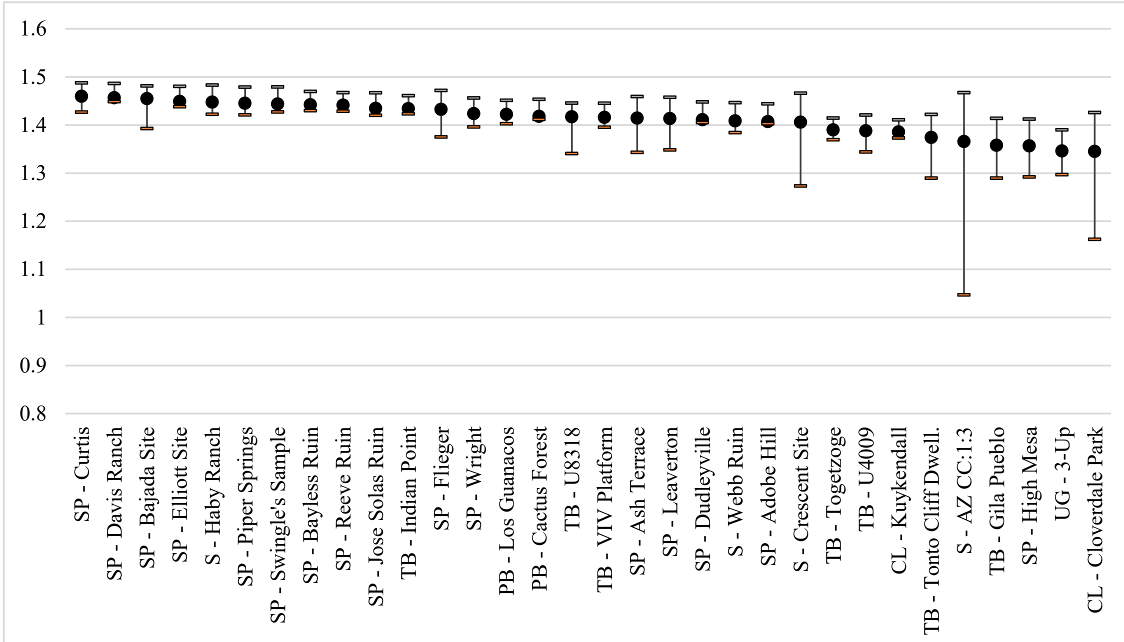

Fig. 9A. Efron's percentile 95 % confidence intervals, broader region, CE 1300; high centrality sites.

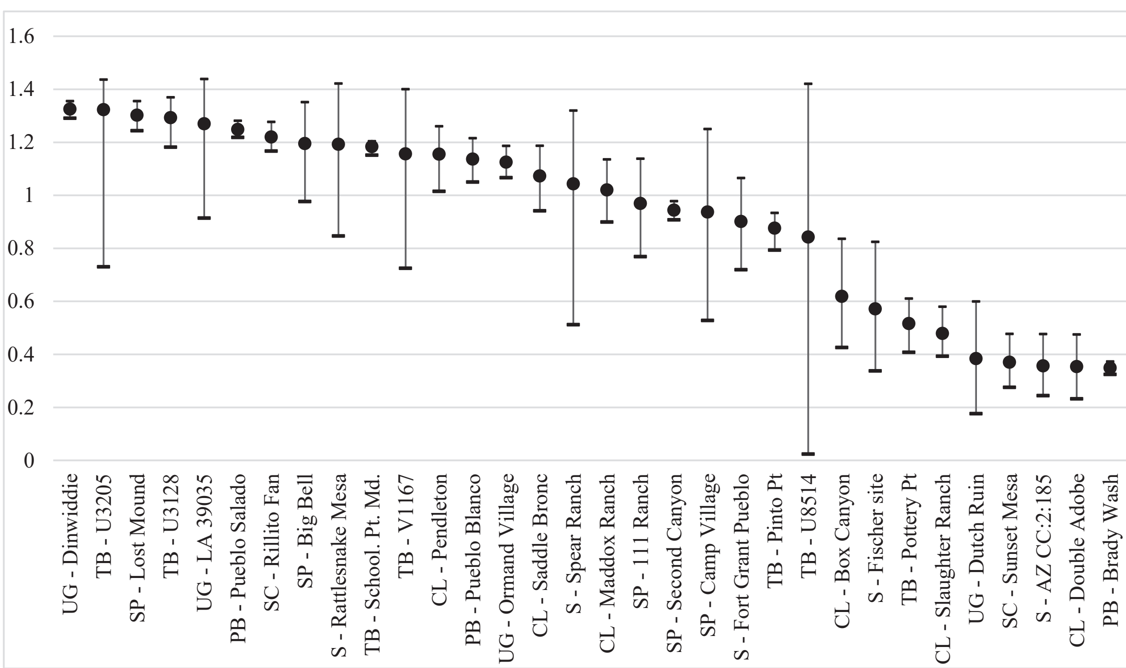

Fig. 9B. Efron's percentile 95 % confidence intervals, broader region, CE 1300; moderate centrality sites.

variation across methods relative to the variation in centrality across sites. Second, while the analysis showed that not all San Pedro sites were statistically distinguishable from each other in centrality, clear groups of similarly central sites could be distinguished. Third, when focusing on centrality ranks rather than centrality scores, we likewise saw that the analysis could reasonably identify groups of similarly central sites. Fourth, this conclusion also seemed to apply in analysis of data from the broader region, though with some sites having very wide centrality confidence intervals due to the small size of their assemblage for some apportioned time periods.

Although focused on a particular time and place, these results give some reassurance that substantive conclusions from analysis of archaeological assemblage networks reflect more than statistical noise. Consideration of sampling variability adds richness to archaeological network analysis and should be a standard part of substantive research in this area. The bootstrap methods used here require that the site similarities used in the network analysis were derived from assemblage counts. If similarities were derived from, say, binary presence-absence data that did not include counts, different strategies would be necessary. But when counts are available, this approach is a straightforward tool for incorporating sampling variability in the form of confidence intervals for centrality scores. Similar confidence intervals could be constructed for sites' other network properties.

Regarding the other sources of uncertainty that are not addressed by

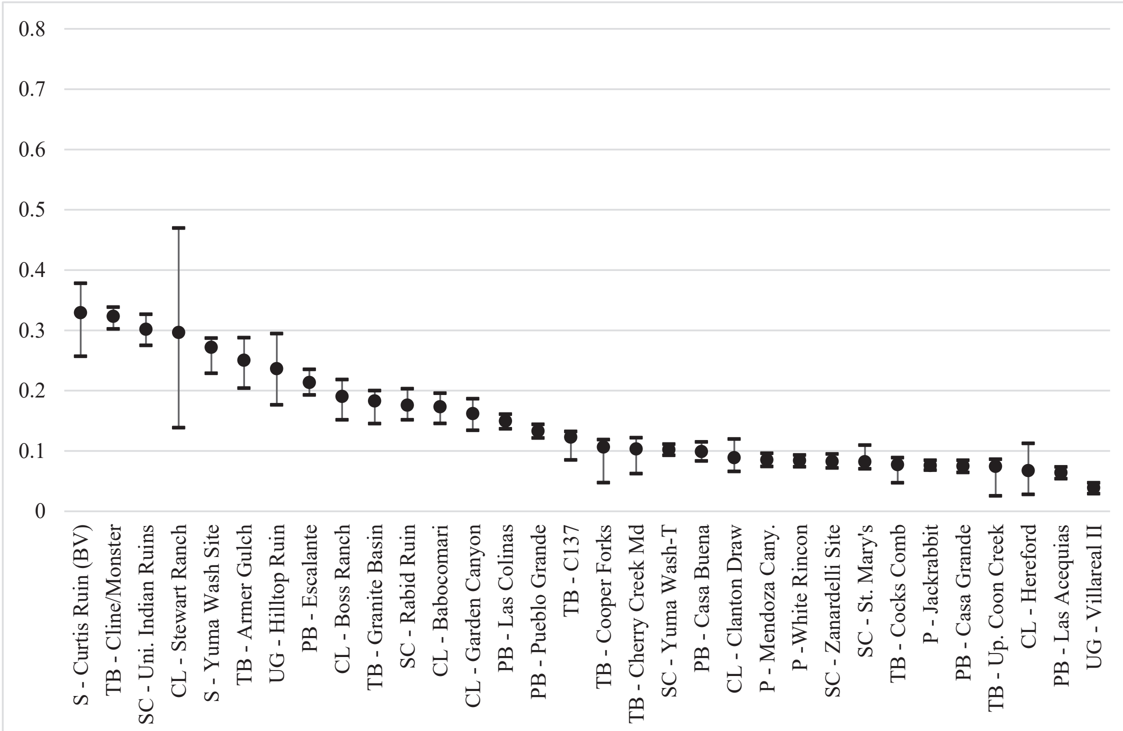

Fig. 9C. Efron's percentile 95 % confidence intervals, broader region, CE 1300; low centrality sites.

the bootstrap approach here, many could be effectively investigated as part of a more comprehensive analysis. For instance, Ladefoged et al. (2019) represented uncertainty in classification by adding artifacts at random to the observed classification. Alternatively, one could randomly reclassify some proportion of the observed distribution. For dates, random perturbations of the start and end dates of site occupation or ceramic wares could account for this source of uncertainty. Archaeological knowledge would inform a researcher's choices of the expected size of such perturbations, with different archaeological contexts marked by varying degrees of uncertainty in dates. For uncertainty stemming from choices of apportioning methods and similarity measurements, analysts can assess the robustness of major findings under different methods and measures. Finally, although we use ceramic ware-based categories to form networks here, similar issues would pertain to a variety of analyses of classified artifacts. Whatever the particular data setting and challenges, the bootstrap methods used here will be a helpful addition to any archaeological network analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data in supplementary files.

Acknowledgements

This work was supported by United States National Science Foundation grants 1758606 and 1758690 to the University of Wisconsin-Milwaukee and Arizona State University. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jasrep.2023.104100.

References

Bernardini, W., Peeples, M.A., Kuwanwisiwma, L., Schachner, G., Throgmorto, K., 2021. Reconstructing population in the context of migration. In: Bernardini, W., Koyiyumptewa, S., Schachner, G., Kuwanwisiwma, L. (Eds.), Becoming Hopi: A History. University of Arizona Press, Tucson, AZ, pp. 133–160.
Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Ann. Stat. 9 (6), 1196–1217.
Birch, J., Hart, J.P., 2018. Social networks and northern Iroquoian confederacy dynamics. Am. Antiq. 83 (1), 13–33.
Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. 2 (1), 113–120.
Booth, J.G., Hall, P., 1994. Monte Carlo approximation and the iterated bootstrap. Biometrika 81 (2), 331–340.
Borgatti, S.P., 2005. Centrality and network flow. Soc. Networks 27 (1), 55–71.
Brainerd, G.W., 1951. The place of chronological ordering in archaeological analysis. Am. Antiq. 16 (4), 301–313.
Brughmans, T., Peeples, M.A., 2017. Trends in archaeological network research: A bibliometric analysis. J. Hist. Network Res. 1, 1–24.
Clark, J.J., Lyons, P.D. (Eds.), 2012. Migrants and Mounds: Classic Period Archaeology of the Lower San Pedro Valley, Anthropological Papers 45. Archaeology Southwest, Tucson, AZ.
Chernick, M.R., 2007. Bootstrap Methods: A Guide for Practitioners and Researchers. Wiley Interscience, Hoboken, NJ.
Collar, A., Coward, F., Brughmans, T., Mills, B.J., 2015. Networks in archaeology: Phenomena, abstraction, representation. J. Archaeol. Method Theory 22 (1), 1–32.
Di Peso, C.C., 1958. The Reeve Ruin of Southeastern Arizona. The Amerind Foundation Inc, Dragoon, AZ.
Duncan, O.D., Duncan, B., 1955. Residential distribution and occupational stratification. Am. J. Sociol. 60 (5), 493–503.
Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Ann. Stat. 7 (1), 1–26.
Efron, B., Tibshirani, R.J. (Eds.), 1993. An Introduction to the Bootstrap. Springer US, Boston, MA.
Gerald, R.E., 2019. The Davis Ranch Site: A Kayenta Immigrant Enclave in Southeastern Arizona. In: Lyons, P. (Ed.), D. University of Arizona Press, Tucson.
Gjesfjeld, E., 2015. Network analysis of archaeological data from hunter-gatherers: Methodological problems and potential solutions. J. Archaeol. Method Theory 22 (1), 182–205.

Golitko, M., Feinman, G.M., 2015. Procurement and distribution of pre-Hispanic Mesoamerican obsidian 900 BC–AD 1520: A social network analysis. J. Archaeol. Method Theory 22 (1), 206–247.
Golitko, M., Meierhoff, J., Feinman, G.M., Williams, P.R., 2012. Complexities of collapse: The evidence of Maya obsidian as revealed by social network graphical analysis. Antiquity 86 (332), 507–523.
Greenacre, M., 1984. Theory and Applications of Correspondence Analysis. Academic Press, London.
Habiba, A., Jan, C., Mills, B.J., Brandes, U., 2018. Social networks and similarity of site assemblages. J. Archaeol. Sci. 92, 63–72.
Hall, P., 1986. On the bootstrap and confidence intervals. Ann. Stat. 14 (4), 1431–1452.
Hart, J.P., Birch, J., St-Pierre, C.G., 2017. Effects of population dispersal on regional signaling networks: An example from northern Iroquoia. Sci. Adv. 3, e1700497.
Hart, J.P., Engelbrecht, W., 2012. Northern Iroquoian ethnic evolution: A social network analysis. J. Archaeol. Method Theory 19 (2), 322–349.
Hill, J.B., Clark, J.J., Doelle, W.H., Lyons, P.D., 2004. Prehistoric demography in the Southwest: Migration, coalescence, and Hohokam population decline. Am. Antiq. 69 (4), 689–716.
Ladefoged, T.N., Gemmel, C., McCoy, M., Jorgensen, A., Glover, H., Stevenson, C., O'Neale, D., 2019. Social network analysis of obsidian artefacts and Māori interaction in northern Aotearoa New Zealand. PLoS One 14, e0216420.
Lulewicz, J., 2019. The social networks and structural variation of Mississippian sociopolitics in the southeastern United States. Proc. Natl. Acad. Sci. 116 (14), 6707–6712.
Manly, B.F., 1997. Randomization. Bootstrap and Monte Carlo Methods in Biology, Chapman and Hall, New York, NY.
Mills, B.J., 2017. Social network analysis in archaeology. Ann. Rev. Anthropol. 46, 379–397.
Mills, B.J., Clark, J.J., Peeples, M.A., Haas, W.R., Roberts, J.M., Hill, J.B., Huntley, D.L., Borck, L., Breiger, R.L., Clauset, A., Shackley, M.S., 2013a. Transformation of social networks in the late pre-Hispanic US Southwest. Proc. Natl. Acad. Sci. 110 (15), 5785–5790.
Mills, B.J., Roberts Jr., J.M., Clark, J.J., Haas Jr., W.R., Huntley, D., Peeples, M.A., Borck, L., Ryan, S.C., Trowbridge, M., Breiger, R.L., 2013b. The dynamics of social networks in the prehispanic US Southwest. In: Knappett, C. (Ed.), Network Analysis in Archaeology: New Approaches to Regional Interaction. Oxford University Press, Oxford, pp. 181–202.
Mills, B.J., 2016. Communities of consumption: Cuisines as constellated networks of situated practice. In: Roddick, A.P., Stahl, A.B. (Eds.), Knowledge in Motion: Constellations of Learning Across Time and Place. University of Arizona Press, Tucson, AZ, pp. 247–270.
Mills, B.J., Peeples, M.A., Haas Jr., W.R., Borck, L., Clark, J.J., Roberts Jr., J.M., 2015. Multiscalar perspectives on social networks in the late prehispanic Southwest. Am. Antiq. 80 (1), 3–24.
Mills, B.J., Peeples, M.A., Aragon, L.D., Bellorado, B.A., Clark, J.J., Giomi, E., Windes, T. C., 2018. Evaluating Chaco migration scenarios using dynamic social network analysis. Antiquity 92 (364), 922–939.
Östborn, P., Gerding, H., 2014. Network analysis of archaeological data: A systematic approach. J. Archaeol. Sci. 46, 75–88.
Peeples, M.A., 2019. Finding a place for networks in archaeology. J. Archaeol. Res. 27 (4), 451–499.
Peeples, M.A., Haas Jr., W.R., 2013. Brokerage and social capital in the prehispanic U.S. Southwest. Am. Anthropol. 115 (2), 232–247.
Peeples, M.A., Mills, B.J., Haas Jr., W.R., Clark, J.J., Roberts Jr., J.M., 2016. Analytical challenges for the application of social network analysis in archaeology. In: Brughmans, T., Collar, A., Coward, F. (Eds.), The Connected Past: Challenges to Network Studies in Archaeology and History. Oxford University Press, Oxford, pp. 59–84.
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.
Ringrose, T.J., 1992. Bootstrapping and correspondence analysis in archaeology. J. Archaeol. Sci. 19 (6), 615–629.
Roberts, J.M., Mills, B.J., Clark, J.J., Haas, W.R., Huntley, D.L., Trowbridge, M.A., 2012. A method for chronological apportioning of ceramic assemblages. J. Archaeol. Sci. 39 (5), 1513–1520.
Roberts Jr., J.M., Yin, Y., Dorshorst, E., Peeples, M.A., Mills, B.J., 2021. Assessing the performance of the bootstrap in simulated assemblage networks. Soc. Networks 65, 98–109.
Robinson, W.S., 1951. A method for chronologically ordering archaeological deposits. Am. Antiq. 16 (4), 293–301.
Weidele, D., van Garderen, M., Golitko, M., Feinman, G.M., Brandes, U., 2016. On graphical representations of similarity in geo-temporal frequency data. J. Archaeol. Sci. 72, 105–111.
Woodson, M.K., 1999. Migrations in late Anasazi prehistory: The evidence from the Goat Hill site. Kiva 65 (1), 63–84.
Yeh, A.B., Singh, K., 1997. Balanced confidence regions based on Tukey's depth and the bootstrap. J. Roy. Stat. Soc.: Series B 59 (3), 639–652.