Journal of Archaeological Method and Theory (2025) 32:19
https://doi.org/10.1007/s10816-024-09688-z

RESEARCH

Filling the Gaps—Computational Approaches to Incomplete Archaeological Networks

Deborah Priß[1] · John Wainwright[1] · Dan Lawrence[2] · Laura Turnbull[1] · Christina Prell[3] · Christodoulos Karittevlis[4] · Andreas A. Ioannides[4]

Accepted: 4 December 2024

Abstract

Networks are increasingly used to describe and analyse complex archaeological data in terms of nodes (archaeological sites or places) and edges (representing relationships or connections between each pair of nodes). Network analysis can then be applied to express local and global properties of the system, including structure (*e.g.* modularity) or connectivity. However, the usually high amount of missing data in archaeology and the uncertainty they cause make it difficult to obtain meaningful and robust results from the statistical methods utilised in the field of network analysis. Hence, we present in this paper manual and computational methods to (1) fill gaps in the settlement record and (2) reconstruct an ancient route system to retrieve a network that is as complete as possible. Our study focuses on the sites and routes, so-called hollow ways, in the Khabur Valley, Mesopotamia, during the Bronze and Iron Age as one of the most intensively surveyed areas worldwide. We were able to predict additional sites that were missing from the record as well as develop an innovative hybrid approach to complement the partly preserved hollow way system by integrating a manual and computational procedure. The set of methods we used can be adapted to significantly enhance the description of many other cases, and with appropriate extensions successfully tackle almost any archaeological region.

Keywords Hollow ways · Mesopotamia · Archaeological networks · Algorithms · Computational archaeology

✉ Deborah Priß
deborah.priss@durham.ac.uk

[1] Department of Geography, Durham University, Durham, UK

[2] Department of Archaeology, Durham University, Durham, UK

[3] Department of Cultural Geography, Rijksuniversiteit Groningen, Groningen, Netherlands

[4] AAI Scientific Cultural Services Ltd., Nicosia, Cyprus

Published online: 03 January 2025

Introduction

Analysing and understanding the connectivity of human and non-human entities in socio-ecological systems can offer new perspectives on the functioning of human societies, present and past (Bell, 2020). Networks as a representation of connectivity and network analysis as a technique to quantify connectivity have become increasingly popular in archaeology to enhance our understanding of the archaeological record (Brughmans & Peeples, 2022; Brughmans *et al*., 2023). However, the potential for network analysis to be used in archaeology is often constrained by missing data in the archaeological record. Complete data sets can give unique insights into how and why people moved through the landscape, thereby revealing—or at least allowing interpretations of—the social processes hidden behind their physical remains. The overarching aim of this study is therefore to develop approaches to resolve the issue of missing archaeological data, *i.e.* fill the missing gaps, based on the physical network structure.

The hollow ways in Northern Mesopotamia are one of the best-preserved route systems worldwide and formed more than 5000 years ago. The temporal scope of the study covers the periods from the Early Bronze Age to the Iron Age (c. 3000–600 BCE). Together with the settlements they connect, the hollow ways form a network that can be analysed with graph-theoretical tools or statistical models to improve our understanding of the past.

In the Khabur Valley in Mesopotamia, a region with fairly homogenous elevation, a route network of so-called hollow ways represents the human movement of the Bronze and Iron Age (de Gruchy & Cunliffe, 2020; Ur, 2003, 2009; Wilkinson, 1993). Unlike the Chacoan, Inca or Roman roads, they not only provide information about trade and exchange but also about the daily and regular short-distance movement of people (Ur, 2009; Wilkinson, 1993). The hollow ways, as mere depressions on the surface, are prone to attenuation and destruction through geomorphological processes and land-use changes. Although the Mesopotamian hollow ways are still one of the least fragmented ancient road networks globally, their visible remains are fragmented, posing limitations on the application of network-analytical methods.

Previous attempts of route reconstruction in the Khabur Valley include both, network approaches (Menze & Ur, 2012; Palmisano & Altaweel, 2015) and optimal route models (de Gruchy, 2016). Menze and Ur (2012) converted the hollow way system, digitised from satellite imagery, into edges and use the sites as nodes. However, they do not provide a description about how they defined if a hollow way connects two sites or if and how they filled the gaps between the hollow ways. Palmisano and Altaweel (2015) developed a model to calculate traffic between sites in the Khabur Valley and weighted traffic that coincided with hollow ways higher. Although they were able to define factors that impact hollow way location, they did not attempt to fill the gaps between them. De Gruchy (2016) calculated optimal paths based on variables that are known to affect route choice and quantified their influence on route choice according to their overlap with hollow ways. This approach allowed the significance of distinct factors on hollow way formation to be determined but did not involve reconstruction of the complete hollow way system. Reconstructing the complete network opens a new suite of analytical tools to allow for more advanced analysis of network properties and controls on network formation and is therefore not only a necessary first step in the analysis but a valuable addition to the archaeological toolbox.

Road Reconstruction and Network Analysis in Archaeology

Roads, tracks and paths between settlements are physical evidence of the repeated movement of people, resources and material culture, and are one type of network edges that can be represented as a graph. The ancient road infrastructure can give invaluable insights into the structure and functioning of the social, economic and political networks it is embedded in. Studies about ancient infrastructure include research on mediaeval Russian waterway networks (Pitts, 1965), the Roman roads in Britain (Dicks, 1972) and the Chacoan road network in the US Southwest (Ebert & Hitchcock, 1980). Later studies have focused on road networks where constructed features such as (partial) pavements, bridges and retaining walls are easily recognisable in the landscape, *e.g.* Roman roads in Europe (Dicks, 1972; Graham, 2006; Isaksen, 2008; Orengo & Livarda, 2016; Verhagen *et al.*, 2019; Lewis, 2021) and the Inka roads in South America (Jenkins, 2001) as well as combining observed evidence of roads with (least-cost path) reconstructions in study areas with reasonably well documented settlement systems (Amati *et al.*, 2020; Bevan & Wilson, 2013; Brughmans *et al.*, 2014; Ducke & Suchowska, 2021; Groenhuijzen & Verhagen, 2017; Jiménez & Chapman, 2002; Tsirogiannis & Tsirogiannis, 2016).

Approaches to reconstruct fragmented ancient route systems, if there are any remains of the routes at all, include least-cost paths or optimal route models, based on topography but also including other physical and cultural variables (Lewis, 2023; McLean & Rubio-Campillo, 2022). A major challenge for investigating route networks in flat landscapes is that least-cost paths often rely on topography, especially slope. However, this focus on a single parameter (steepness) is unsuitable in relatively flat areas such as the Khabur Valley. A combination of factors such as vegetation, surface roughness or micro-terrain might be able to provide plausible reconstructions for ancient route systems in flat landscapes, but that information is not available for the Khabur Valley in the Bronze and Iron Age.

Other forms of edges in archaeological networks, *i.e.* not based on the physical evidence of roads, are usually either reconstructed by implementing spatial assumptions (*e.g.* geodesic distance or gravity models; see Jiménez & Chapman, 2002; Rivers *et al.*, 2013) or correlating node attributes (*e.g.* similarity of material culture; see Cochrane & Lipo, 2010; Mills *et al.*, 2013; Östborn & Gerding, 2014) but not utilising physical connections such as paths. Therefore, the approach presented in this paper takes advantage of the nature of the networks themselves, *i.e.* defining the physical hollow ways as edges, which differs from previous research in which edges between sites are estimated without using physical evidence. Instead of overestimating connections based on assumption about proximity and interaction, we hence probably underestimate connections due to the attenuation of the hollow way record over time.

In archaeological networks, missing data can concern not only the visibility and incompleteness of edges (*e.g.* hollow ways) but also missing nodes (*e.g.* sites), their attributes or any combination of them and each type of missing data requires different treatment when trying to fill the gaps. Missing attributes can be, for example, dating of sites, population size, site size, demographic data or material remains. Often, nodes and their attributes are missing for a significant number of sites (*e.g.* Lucas, 2012; Peeples & Mills, 2016; Perreault, 2019). Although there are suggestions on how to impute missing nodes or edges (Krause, 2019; Smith *et al.*, 2022; van Buuren, 2018), the imputation of node or edge attributes is much more difficult, if not impossible and has not been addressed yet, in particular for archaeological networks (Brughmans & Peeples, 2022).

One of the most popular methods for archaeological site prediction has always been logistic regression that uses the presence and absence of sites as input, with environmental factors as independent variables, and calculates the likelihood of site presence based on these data (Judge and Sebastian, 1988; Wachtel *et al.*, 2018; Hazra, 2020; Li *et al.*, 2022). However, the simulation of absence data or lack of true absence ("no site") data has the potential of distorting the results significantly. The absence of archaeological evidence does not necessarily mean that there never existed a site at a specific location—the evidence might also have been destroyed or simply not been discovered yet. Hence, we can confirm the presence of a site if there are material remains, but we cannot prove its absence with certainty. Therefore, the approach currently deemed most accurate and reliable for archaeological site prediction is maximum entropy which accounts for the uncertainty and incompleteness of archaeological data by introducing pseudo-absence data, thereby overcoming the lack of true-absence data (Li *et al.*, 2022; Wachtel *et al.*, 2018; Yaworsky *et al.*, 2020).

White and Barber (2012) provide a different approach with their "From Everywhere to Everywhere" (FETE) model which uses Dijkstra's popular shortest path algorithm to predict travel probabilities. They suggest that the model might also be used to predict settlement locations by identifying junctions of the modelled trade routes. Using the FETE model, Crabtree *et al.* (2021)predicted optimal "superhighways" of the first peopling of Sahul. Similar to White and Barber (2012), they suggest that high-traffic areas that have been frequently travelled potentially indicate the locations of settlements. The least-cost paths created by the FETE model could therefore be used as input for machine learning or linear regression models and significantly improve those methods for site prediction.

In order to achieve our aim to enhance the archaeological record in the Khabur Valley by predicting the locations of missing/unobservable network nodes (sites) from the observed network edges (hollow ways) and complementing the fragmented edges, we address the following two objectives: (1) to develop an algorithm that uses ancient roads to predict the locations of unrecorded archaeological settlements; and (2) to develop an approach to enhance the fragmented hollow way datasets, by connecting the path segments to retrieve the long-distance routes they once formed.

Materials and Methods

Empirical Data

To add missing nodes (sites) to the archaeological network and complement the fragmented record of the edges (hollow ways), it was first necessary to compile available empirical datasets for hollow ways and settlement data for the study area.

Settlement Data

The settlement data compiled consist of the location of archaeological sites (Fig. 1) with information about their dating and size, where available. Datasets were compiled from the Fragile Crescent Project (FCP; (FCP, 2013)), the Ancient Near Eastern placemarks project (ANE; (Jones, 2018)) and Kalaycı (Kalaycı, 2013).

The settlement data set from the FCP (FCP, 2013) was collected from surveys and excavations in the ancient Near East and was compiled in a comprehensive database (Galiatsatos *et al*., 2009). Data from the FCP used in this study consist of settlement locations, dating information and estimates of settlement size based on surveys (partly supplemented by excavations and remote sensing) and cover the Bronze Age to Iron Age (c. 3000–600 BCE). Those periods were chosen because they provide abundant archaeological evidence for sites and their trajectories over time as

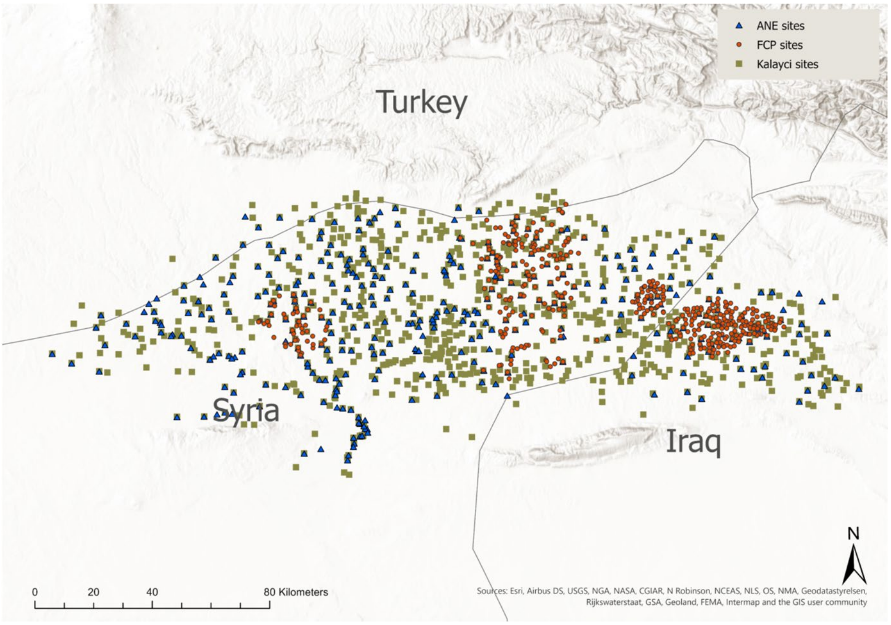

Fig. 1 Overview of site data employed in this study. Data sets were compiled by the Fragile Crescent Project (FCP), the Ancient Near Eastern placemarks (ANE) and by Kalaycı

extensive surveys have been carried out to uncover those material remains. The FCP data are the most complete and reliable data used in this study because they are based on fieldwork informed by remote sensing. For the relevant periods and region, the results of four surveys are included in the analysis: the North Jazira Survey (NJS) (Wilkinson & Tucker, 1995), the Leilan Regional Survey (LLN) (Weiss, 2014), the Tell Hamoukar Survey (THS) (Ur, 2010) and the Tell Beydar Survey (TBS) (Ur & Wilkinson, 2008). For an overview of the basic survey information, see Table 1.

The data within the FCP reflect the research aims and methods of the individual projects: the NJS focused on a 475 m^2 area in north-west Iraq which was about to undergo significant development caused by irrigation methods (Wilkinson & Tucker, 1995). The LLN started as a survey in a radius of 15 km around Tell Leilan and in its last phase, was extended to include a 1650 km^2 area on a transect from the Turkish to the Iraqi border (Ristvet, 2005; Weiss, 1985, 2014). The THS covers an area of 125 m^2 just a few kilometres northwest of the NJS and was primarily initiated to test new methodologies (Ur, 2010). The TBS targeted Tell Beydar and its hinterland within a radius of 12 km, covering 450 km^2 (Ur & Wilkinson, 2008; Wilkinson, 2000). All four surveys are hence spatially restricted (see distribution of FCP sites in Fig. 1) and the space between them has not been intensively surveyed yet (see Lyonnet (1996) and Meijer (1986) for large but low intensity surveys in the region), which introduces a high amount of missing data paired with the high-resolution data gathered within the surveys.

Further information about site locations was drawn from the ANE project (Jones, 2018; Pedersen, 2012), gathered using survey and excavation reports and plans as well as aerial photos and a variety of maps (Pedersen *et al*., 2010). Place names (modern and/or ancient) are available for most of the sites but no dating or sizes. Hence, the data need to be used with caution because sites may belong to any period and site names might be incorrect. This data set is therefore used primarily to add site names to the FCP data, if appropriate, and to include sites in the data set for this study for which dating and/or size can be added from the literature and/or remote sensing.

The Kalaycı (2013, 2022) data used CORONA satellite images to identify ancient settlements, and estimate the visible extent of ancient settlement mounds. This approach has some limitations: first, the settlement sizes are based on the observable extent of sites which may not represent the ancient settlement size in specific periods. Second, site names and dating are not available from remote sensing which makes it difficult to relate the sizes to the respective ancient site. Therefore, the data

Table 1 Overview of the surveys recorded in the Fragile Crescent Project and used in this study

Survey name	Abbreviation	Field seasons	Area	Site count
North Jazira Survey	NJS	1986–1990	475 km^2	184
Leilan Regional Survey	LLN	1984, 1987, 1995, 1997	1650 km^2	335
Tell Hamoukar Survey	THS	1999–2001	125 km^2	60
Tell Beydar Survey	TBS	1997–1998	450 km^2	83

Table 2 Overview of attributes available in the complete data sets from the Fragile Crescent Project (FCP), Ancient Near Eastern placemarks (ANE) and Kalaycı

Data Set	Total sites	Dating known	Size known
FCP	489	390	351
ANE	359	0	0
Kalaycı	910	0	910

Table 3 Definition of periods for this study

Period	Abbreviation	Period dates
Early Bronze Age 1	EBA 1	c. 3000–2500 BCE
Early Bronze Age 2	EBA 2	c. 2500–2000 BCE
Middle Bronze Age	MBA	c. 2000–1600 BCE
Late Bronze Age	LBA	c. 1600–1200 BCE
Iron Age 1	IA 1	c. 1200–900 BCE
Iron Age 2	IA 2	c. 900–600 BCE

is used, with a certain level of caution, to add size information to FCP sites but not to add additional sites to the data set. Size information is available for all sites in Kalaycı's data set and is therefore a valuable supplement for the FCP data, where site sizes are often missing.

An overview of the settlement data in the three data sets listed above is given in Table 2. The ANE and Kalaycı data sets cover a region much larger than the focal area of this study and have been spatially subsampled to the extent of the hollow ways. Figure 1 shows the complete data sets as illustrated in Table 2 while only the FCP sites are used in this study, complemented by information from the two other data sets.

In order to enhance the site attribute data, the site extent was added when it was possible to gauge it from satellite images, providing an estimate for the maximum extent of sites. However, sites might have been smaller at any point in time; hence, those estimates are a rough proxy for site size. Dating information was added when it could be determined from the literature (*e.g.* Mallowan (1936), McMahon (2009), McMahon *et al.* (2001) for Chagar Bazar; Pfälzner (1990) for Tell Bderi; Lebeau (1993) for Tell Melebyia; Bieliński (1992) for Tell Rad Shaqrah; or Hole (1999) for Tell Ziyade, amongst other excavation reports).

A total of 489 sites were available from the FCP, for which we defined the centroids of each site. Settlement size and names in the FCP data are only available for sites that were sufficiently investigated to reveal their extent or that were mentioned in historic documents. The FCP data was therefore supplemented with names from ANE and sizes from Kalaycı. The time frame provided by the dating of the sites consists of various sub-periods, defined by regional chronologies. Therefore, to make the data more comparable, those sub-periods were merged into six broader periods (Table 3). Periods are derived from pottery collected at the sites by surveyors and excavators and are tied into standard local sequences (see Ur, 2010 for the most recent version).

Hollow Ways Data

Data for the hollow ways are from Ur (2008). Although the Mesopotamian hollow ways have not been in continuous use since their formation, likely in the Early Bronze Age or Late Chalcolithic, some of them have survived and can still be observed from satellite images as broad (50–120 m) and shallow (up to 0.5 m) depressions (Ur, 2009; Wilkinson *et al.*, 2010). Their distinctive appearance as dark lines with light borders makes it possible to distinguish them from other features such as recent roads or canals (Casana, 2013; Ur, 2009; Wilkinson *et al.*, 2010). The hollow ways were digitised from CORONA satellite images taken in the 1960s (Ur, 2003, 2010) and therefore represent their modern fragmented remains (Fig. 2). It should be noted that human-induced bias during the visual interpretation of the imagery could have impacted the resulting hollow way dataset. Since the 1960s, the region has changed significantly due to increased development and the introduction of mechanised agriculture and large-scale irrigation measures which destroyed a considerable amount of archaeological remains (de Gruchy & Cunliffe, 2020). Therefore, more recent satellite and aerial images generally lack the rich and detailed archaeological landscape that was still observable on the CORONA images. Conversely, CORONA images have a lower resolution than most modern imagery and are greyscale versions of images collected using the visible spectrum which impedes the use of more complex computational enhancements available through current multi-band images.

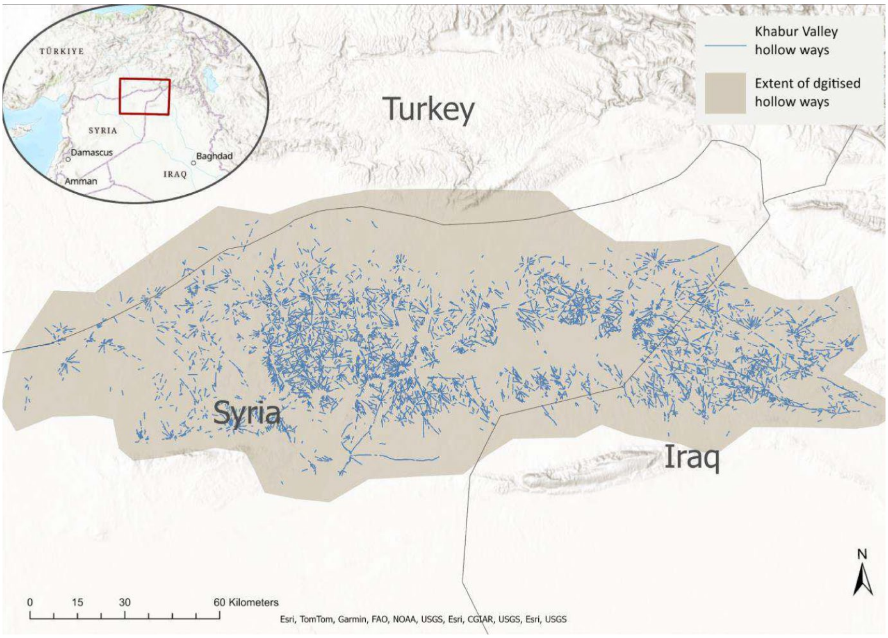

Fig. 2 Research area with mapped out hollow ways (Ur, 2018)

Bronze and Iron Age hollow ways in Mesopotamia have been discovered in three regions, with several studies describing and investigating them: Northwest Syria (Casana, 2013), Southern Mesopotamia (Jotheri *et al*., 2019) and Northern Mesopotamia (de Gruchy, 2016; de Gruchy & Cunliffe, 2020; Ur, 2003, 2009; Wilkinson, 1993). Here, we focus on the hollow ways in the Greater Khabur Valley in Northern Mesopotamia, bounded by what is today the Syrian-Turkish border to the north, the Syrian-Iraqi border to the east, the Jebel Sinjar to the south and the Khabur river to the west.

The hollow ways in Northern Mesopotamia exhibit a characteristic behaviour: they radiate from the settlements which can help to identify archaeological sites based on the hollow ways and gives an indication of their age. Hollow ways are assumed to be at least as old as the latest phase of the settlements they radiate from which can serve as a *terminus post quem* (Wilkinson *et al*., 2010). The hollow ways can be broadly separated into two categories: short ones that fade out after a few kilometres and long-distance routes (Ur, 2009; Wilkinson, 1993; Wilkinson *et al*., 2010). The short hollow ways represent the daily movement of the farmers and herders that went to the adjacent fields or led their livestock to pastures beyond the agricultural area (Ur & Wilkinson, 2008; Wilkinson, 1993). The long-distance routes emerged through regular interaction such as trade and exchange between sites (Ur & Wilkinson, 2008; Wilkinson, 1993). However, the record is patchy and the routes are segmented into 6531 fragments of which many can be identified by eye as parts of potential long-distance routes, sometimes interrupted by hundreds of metres.

Methods for Site Prediction and Hollow Way Reconstruction

Settlements—Predicting Missing Site Locations (Site Algorithm)

To fulfil objective 1, *i.e.* predicting the location of missing sites, we developed an algorithm to computationally find junctions of multiple hollow ways, indicating the location of an archaeological site. To do this, the end points of every hollow way are selected to create a cone-shaped search area facing away from the hollow way in which to identify potential settlements (Fig. 3). The width of the cone is defined by the offset angle, a, on either side of the hollow way direction, and the length, d, is the radius of the cone. A number of trials were carried out using a split-sample approach and optimised to capture already known sites. It was found that the optimal parameter values are $a=30°$ and $d=800$ m (see supplements Fig. 11 – Fig. 15).

We use the Woods-Saxon shape (WSS) to define a decay function (Woods & Saxon, 1954). The WSS has been used widely in nuclear physics to model the density of a nucleus, *i.e.* the probability of finding a nucleon (the constituent of the nucleus) at a given location from the centre of that nucleus (Jones, 1970). The WSS shape can be translated to the context of this study: the centre of the nucleus is the end point of the hollow way and we are looking for a nucleon (a site) in the vicinity of the end point. The probability of finding a site is constant near the end point and then decreases with increasing distance from the end point and with increasing offset from the orientation of the hollow way. Hence, the WSS value provides an

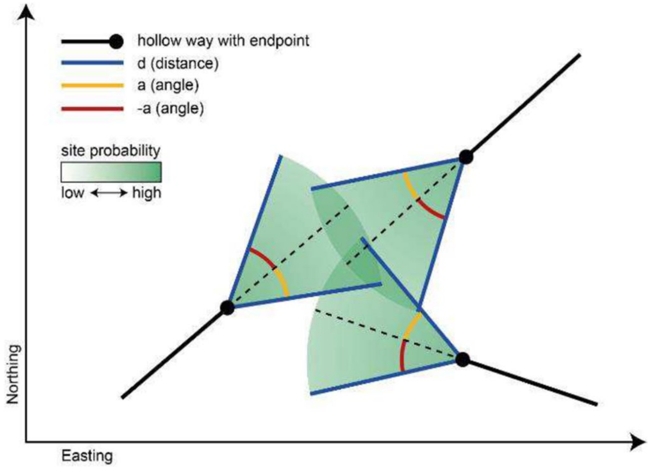

Fig. 3 Schematic drawing of the search area to find missing sites: from the end points of each hollow, a conical search area is defined with radius d and offset a from the orientation of the hollow way

estimate of finding a settlement within the search area. This can be computed for any point and in case of an overlap of WSS cones, the sum of WSS values is computed which can therefore increase and pass an appropriately chosen threshold. The higher the value of a grid cell, the higher the likelihood of a settlement in this cell.

We use a bootstrapping approach to define a threshold for deciding whether a settlement is identified: we apply the method for hollow ways around known settlements and set the parameters of the WSS and the threshold for identification of a settlement to correctly identify the known settlements. We assign a value based on the decay function outlined in Eq. 1 (with the parameters fixed by bootstrapping) to every grid cell in the search area.

The WSS is controlled by the decay length R which is the distance at which the potential is half of the value at the centre (Jones, 1970). The function used to calculate the likelihood value, Z, for each grid point is a multiplication of two Woods-Saxon functions, where the first one considers the distance of the grid point and the second one the angle.

$$Z(\theta, r) = \left(\frac{1}{1 + e^{\frac{r-R}{b}}} \right) * \left(\frac{1}{1 + e^{\frac{|\theta| - \Theta}{\beta}}} \right) \tag{1}$$

where:

Z is the calculated value of the product of the two WSS form factors at the point within the cone defined by r and θ;

θ is the angle between the hollow way direction and the line connecting the grid point to the hollow way end point, in the range [-a a] (see Fig. 3);

r is the distance between the grid point and the hollow way end point in the range [0 d] (see Fig. 3);

R and Θ are constants, defining the range of constant probability in the radial and angular variation, respectively;

b and β are constants, defining the decay rate of the WSS for large values of r and θ respectively.

We will use conventional units, *i.e.* meter (m) for length (variable r and constants *R* and *b*) and degree (°) for angle (variable θ, constants Θ and β and range limit a). When only hollow ways are used as input, there is no preferred angular direction; therefore, only the modulus of the deviation from the hollow way direction is relevant; typical estimates for optimal values for Θ are found to be much higher than the range of θ values we used here (up to 30°), which implies that as a first approximation the angular WSS can be removed (replaced by unity).

We kept the second (angular) WSS in Eq. 1 to indicate how the method can be extended to reinforce the influence of known settlements. The angular WSS can also be effective if an iterative approached is used, to enhance the influence of putative settlements identified by our method in a previous iteration. The radial fixed parameters we determined, by bootstrapping optimisation (see supplements Fig. 16 – Fig. 22), are $R=250$ and $b=750$. Those values seem reasonable: initially, the WSS has a high and fairly constant value, which decreases as r becomes comparable to R, reducing to half its initial value when $r=R$. It then decreases more rapidly, becoming nearly a quarter of its original value about 1 km away from the hollow way end.

The areas for potential sites are identified by determining a threshold that describes the statistically significant minimum *Z*-value. For this purpose, we focus on non-zero values of *Z*, excluding zero values because they represent areas without estimations. We calculate the mean and standard deviation of these non-zero values to establish the baseline statistics of our dataset. We then quantify how much each point deviates from the mean in terms of standard deviations. Setting a *p*-value at 0.05, we apply the Bonferroni correction by dividing this *p*-value by the number of non-zero values, which adjusts our threshold to minimise the risk of false positives. Using this corrected *p*-value, we determine the threshold and identify regions where *Z* exceeds this threshold. In order to visually highlight the results, we plot a circle centred around the point with the highest *Z*-value in each significant region, filling it with the colour corresponding to the *Z*-value at that point.

Hollow Ways—Connecting the Fragments

Due to the incomplete preservation of the hollow ways, evaluating which settlements were connected by them is difficult. Hence, it is necessary to connect the individual route fragments. In the following sections, we describe procedures to fill the gaps in the complete hollow way data set both manually and computationally. To connect the fragmented hollow ways systematically, we assumed the following: first, distance is important, *i.e.* only hollow ways that are within a certain distance of each other have the potential to be connected. Second, orientation is important, *i.e.* only hollow ways that share a similar orientation have the potential to be connected.

The manual approach, whereby hollow ways are connected by manually digitising missing links, benefits from the ability of the digitiser to make an informed decision (based on an understanding of the system) about whether it is reasonable (or not) for two hollow ways to be connected. For example, if two hollow way fragments are separated by agricultural fields with a higher distance than given in the parameter specifications for the computational procedure, it can be assumed that they once formed a route that was partially destroyed by agricultural activities such as ploughing (Fig. 4a). On the other hand, if two fragments are close and have (roughly) the same orientation but connecting them would result in a sharp bend, it is not very likely that those fragments were part of the same route because humans naturally avoid too sharp turns and bends (Fig. 4b).

Conversely, the computational approach benefits from a lack of human bias and computationally links fragmented hollow ways according to the pre-defined rules. We also explored a third approach—a combination of both the manual and

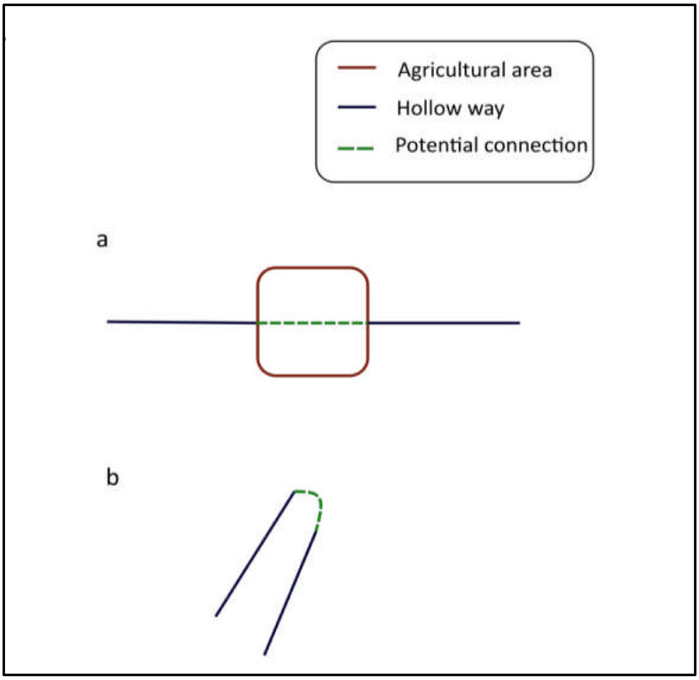

Fig. 4 Examples for connecting hollow ways. (**a**) Two hollow way fragments are separated by an obstruction such as agricultural fields and it would make sense to connect them although the distance might be higher than defined in the parameters; (**b**) two hollow way fragments fulfil the parameter requirements but a connection would be unreasonable in terms of human movement because humans in general avoid sharp bends and turns

computational approaches, with the intention of benefitting from the "human knowledge" component of the manual approach, alongside the automated and systematic benefits of the computational approach.

Manual Approach to Connect Hollow Ways

Settlements were connected to their adjacent hollow ways as well as to neighbouring sites, both within a distance of 2000 m. This distance was chosen to account for slight shifts in site location over time and varying site sizes, because settlement coordinates represent the centroid of the settlement rather than their extent. The radius of 2000 m represents the highest distance between the end points of radiating hollow ways and the centroids of the sites. Neighbouring sites within this radius were assumed to be in regular contact due to their proximity.

As hollow ways possibly also run through (abandoned) sites, hollow ways were connected even if they cross a settlement. The maximum distance to connect two hollow ways was set to 3000 m as space between hollow ways might be more disturbed in the flat landscape than the area in or around settlements, leading to larger gaps between hollow way fragments than between hollow way fragments and sites. However, in some cases (less than 1% of all manual connections), a connection was reasonable although the distance between fragments exceeded 3000 m (see, for example, Fig. 4a). Hence, the actual maximum distance for manual connections is 5500 m with 99% of connections being less than 3000 m long.

Extracting networks from the hollow way dataset for each period was done manually as well. The sites for the respective period are projected on top of the hollow way network, and for every site, the next neighbours are recorded in the form of an edgelist. Neighbours are defined as sites to which the current one is directly connected by a hollow way.

Computational Approach to Connect Hollow Ways (Hollow Way Algorithm)

In addition to the manual approach, a computational procedure was developed to reduce potential bias and make the approach reproducible. Two threshold distances were chosen to connect the hollow ways automatically. Comparing those two sets of parameter values enables us to assess which of them results in a more realistic network with regard to the archaeological record. The first one is based on the initial parameter for distance in the manual network (3000 m, computational network 2), and the second on the actual maximum distance in the manual network (5500 m, computational network 1). The offset from the orientation of the hollow way (angle) was set to a high value for small distances and a low value for large distances. Hence, with increasing distance, the offset between two hollow ways needs to be small in order for them to be connected. Those parameter values were chosen because the higher distance increases the covered area significantly if the angles are the same.

Each iteration consists of three steps: the connection of sites and hollow ways, the connection of hollow ways with small distance and large angle and the connection of hollow ways with large distance and small angle (Fig. 5). In the first step, a new line

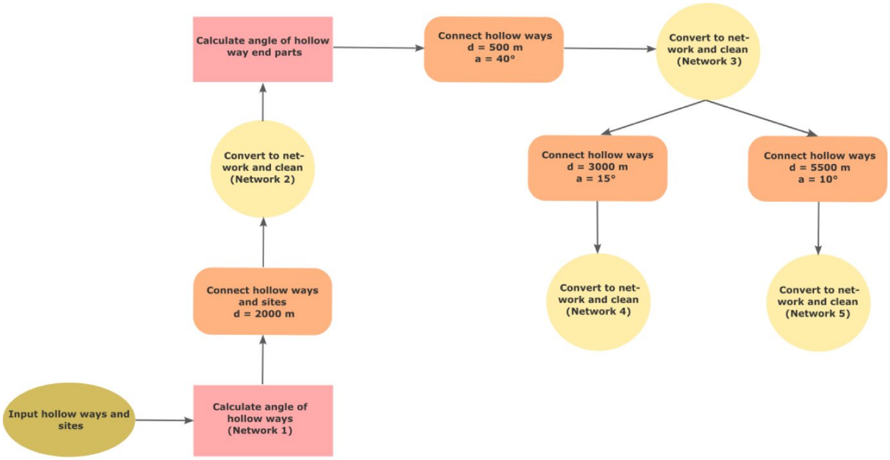

Fig. 5 Flow diagram of the algorithm procedure, including parameter values (d, distance; a, angle)

is drawn between a site and a hollow way if the distance parameter ($d< =$2000 m) is met, which increases the length of the hollow way and makes the site the new endpoint of this hollow way. In the second step, the orientation of each hollow way is calculated and a conical search area was selected around each endpoint of every hollow way with the values for angle (a) and distance (d) and new lines were drawn between the current hollow way and all hollow ways within the search area that had the same orientation. In the third step, the procedure of step 2 was repeated but with different parameter values for angle (a) and distance (d). For each iteration, the network generated from the previous iteration was used.

After each iteration, the network was cleaned; *i.e.* multiple edges (edges that connect the same two nodes), self-loops (edges that connect one node with itself) and pseudo-nodes (nodes that are redundant because they are located on an edge between two sites but are not a site themselves) were removed and intersection structures compressed into single nodes. The creation of the edgelist for the individual period networks was automated as well. First, all simple paths between all sites of the individual periods were extracted, with the length of paths restricted to avoid infinite paths. Then, the nodes of each path were added to an edgelist so that direct and indirect paths are equally captured. Removing of duplicate entries in the edgelist is necessary because all simple paths between all pairs of nodes are recorded, producing the same connections several times.

Hybrid Networks

The hybrid networks were constructed by combining the manual and the computational networks. The computational network where d = 3000 m and a = 30° was chosen to generate the hybrid network as it better reflected the archaeological record in terms of the representation of important sites and had less superfluous connections (see the "Results" section). The hybrid networks were produced by adding the adjacency matrices of the manual and the computational network. In total, 24 networks were generated, one for every of the six periods for (a) the manual, (b) the two computational and (c) the hybrid approach.

Comparing the Manual and Computational Approach to Connect the Hollow Ways

To evaluate how well the manual and computational approaches replicate the network structure, we chose two different methods: first, the sizes of the manual and computational networks (number of nodes and number of edges) were compared and the percentage of matching edges between the manual and the two computational networks were assessed to evaluate how much the networks differ. Furthermore, local network metrics for a selection of well-studied sites, for which detailed information about their dating, development over time and size are available, were evaluated. Those sites are large and well-understood sites which have been mostly excavated and which provide evidence for significant changes in status over the study period, in contrast to smaller sites that have only been surveyed but never excavated. We have a clear grasp of what we think these sites should be doing in relation to the network. Hence, the trajectories of those sites can be determined for most, if not all, of the relevant periods for this study, thereby providing the basis to evaluate the approaches. The selected sites are Tell Leilan (LLN) (see Ristvet, 2005), Mohammed Diyab (LLN) (see Ristvet, 2005), Tell Beydar (TBS) (see Ur & Wilkinson, 2008), Tell Hamoukar (THS) (see Ur, 2010), Khirbet al- ‘Abd (THS) (see Ur, 2010) and Tell Hawa (NJS) (see Wilkinson & Tucker, 1995).

For the comparison of the selected sites, degree and betweenness centrality for each site in each period were calculated as these metrics represent the number of edges of a site (degree) and the flows that go through it, relative to all other nodes in the network (betweenness). Our assumption is that if a site was a centre (*i.e.* the largest site in the region) in one period, its degree and betweenness centrality should be high, indicating that the site had a central function in the network. We assessed if betweenness centrality and degree are in a quartile that reflects its importance. Hence, if a site was the main centre of the survey region, its degree and betweenness centrality should be in the 4th quartile, and if a site was a large site or major centre (together with other sites), its betweenness centrality and degree should be in the 4th or 3rd quartile. For every site, its importance (based on the archaeological record) was determined and cross-checked with its betweenness centrality and degree. The data for the sites and their importance was drawn from the surveys utilised in this study (Ristvet, 2005; Ur, 2010; Ur & Wilkinson, 2008; Wilkinson & Tucker, 1995).

Results

Verifying Predicted Sites

The results of the Site Algorithm can be illustrated as a heatmap with higher values of Z indicating a higher likelihood of a site. From this heatmap, centroid coordinates of areas with values higher than the threshold are extracted and cross-checked with recent and historic aerial and satellite images. By cross-checking the locations with CORONA and Google Earth images from the last 40 years (NASA/USGS Landsat 7 satellite images), we can confirm the location of a site.

The resulting heatmap of the Site Algorithm is presented in Fig. 6. Despite using the centroid as a proxy for an area, matching the 176 predicted sites with the recorded ones was straightforward because the offset between them was very low, in most cases below 100 m with a maximum of c. 300 m.

The site algorithm detected 176 sites in the study area. The area of the predicted sites was cropped to match the area of the FCP data to render the evaluation more comparable; hence, only 128 predicted sites were included for the comparison with the FCP data. In summary, 17.2 to 64.2% can be matched with known sites (Table 4). Around 4.5 to 17.0% of known sites have been detected by the algorithm. Only 39 of the predicted sites could not be confirmed by either matching them with the known sites or by remote sensing and we were able to add 19 previously unrecorded sites that were visible on CORONA and/or Google Earth satellite images.

Hollow Ways (Edges)

The manually and computationally reconstructed hollow way networks are inherently different (Fig. 7). The manual network contains fewer connections between the same two hollow way fragments or between sites and the same hollow way fragments and fewer duplicated lines, lines that turn and bend unreasonably or zigzag (superfluous connections), and the routes are more easily recognisable. The computational network 2 with d = 3000 m and a = 30° is denser than the

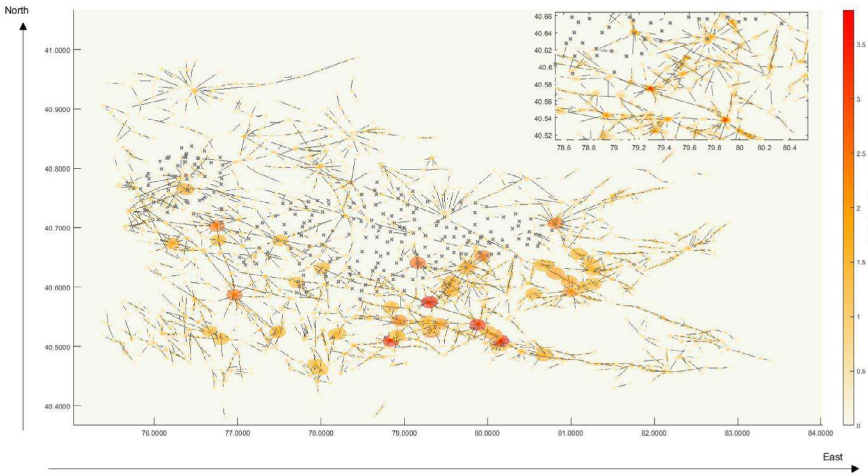

Fig. 6 Result of the Site Algorithm for a subset of the hollow way system. Higher values indicate higher likelihood of a site. Areas with medium to high likelihood are enhanced in the main figure for better visibility while the inset illustrates the initial results. Known settlements in the region are marked by a grey “x”

Table 4 Overview of the results of the Site Algorithm: Percentage of matches between known sites from either of the utilised data sets (FCP, ANE, Kalaycı) and the sites predicted by the Site Algorithm, percentage of predicted sites that could be confirmed and added as new sites and predicted sites that could not be confirmed (false positive) as well as the percentage of matches in the individual FCP surveys. Note that for the evaluation of matches with the FCP data, the area of the predicted sites was cropped to the area of the FCP data; hence, less predicted sites are included

	Match	No match	Total predicted sites	Percentage matched
Percentage of predicted sites that match known sites				
FCP	22	30	52	42.31
ANE	61	115	176	34.66
Kalaycı	113	63	176	64.20
Percentage of predicted confirmed and unconfirmed sites				
Confirmed sites (additions)		19	176	10.80
Unconfirmed sites (potential false positives)		39	176	22.16
Percentage of known sites that match predicted sites				
FCP	22	467	489	4.50
ANE	61	298	359	16.99
Kalaycı	113	797	910	12.42
Percentage of FCP sites detected for individual FCP surveys				
NJS	8	174	182	4.40
LLN	5	196	201	2.49
TBS	7	45	52	13.46
THS	2	52	54	3.70
Sum	22	467	489	4.50

computational network 1 with d = 5500 m and a = 10°; *i.e.* more connections were generated with smaller distance and larger angle parameters.

We produced period-specific edgelists for each of the 18 networks (six period-specific manual networks and six period-specific networks for both of the computational approaches) that represent only direct connections between nodes (see Fig. 8 for the MBA which will be used to illustrate the results).

The number of nodes and edges varies between the manual and computational approaches with the manual networks usually having fewer edges and more nodes (Fig. 8, Table 5). In general, there are on average 15% fewer nodes in the computational networks but their relative numbers are similar; *i.e.* the MBA network is always the one with most nodes while EBA1 always has least nodes. On average, between 30 and 40% of the edges in the manual and computational networks overlap. Furthermore, the edges in the computational networks are mainly from short-distance routes while long-distance connections are sparse (for the remainder of this section, distance will be used in terms of physical geodesic distance, not as graph-theoretic distance, *i.e.* path length). In the manual networks, both kinds of links are present but with fewer long-distance connections.

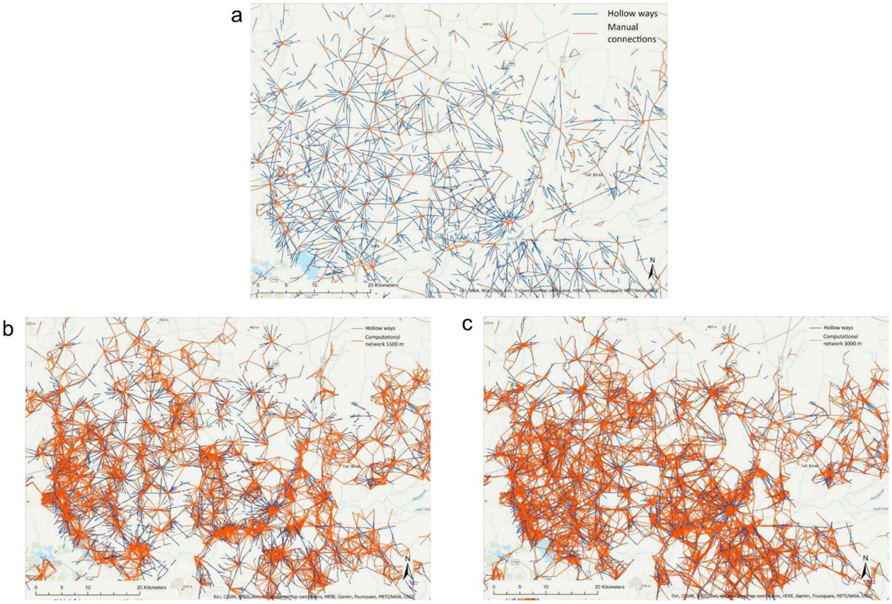

Fig. 7 Results for **a** manually connected network, **b** computational network 1 with d = 5500 m and a = 10° and **c** computational network 2 with d = 3000 m and a = 30°

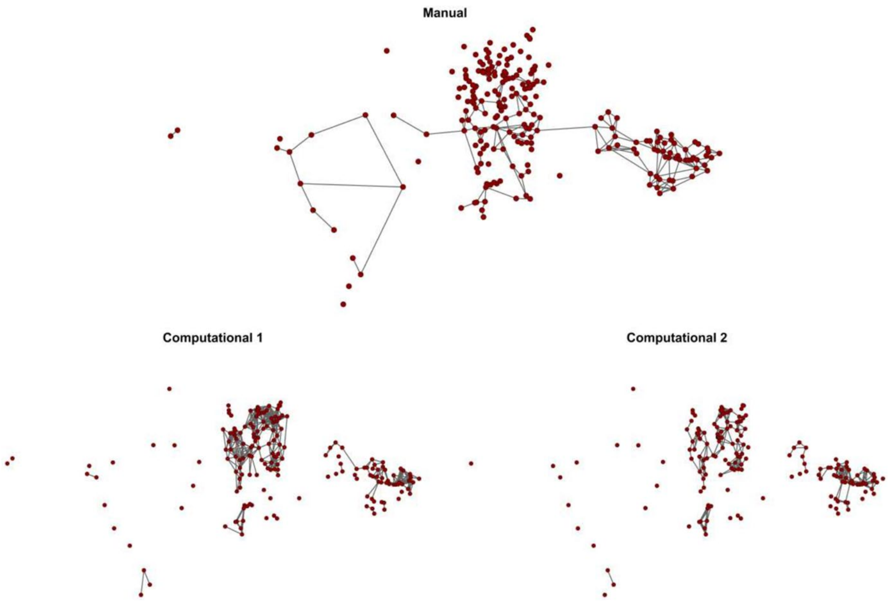

Fig. 8 Graphs produced from the respective edgelists for the manual, computational 1 (d = 5500 m, a = 10°) and computational 2 (d = 3000 m, a = 30°) network

Table 5 Comparison of sizes of networks derived from manual (man) and computational (comp) approaches. The "match" columns show how many of the edges were present in both networks and the percentage they represent in the respective manual and computational networks

Period	Man		Comp 1 (5500 m/10°)		Comp 2 (3000 m/30°)		Match Man – Comp 1			Match Man – Comp 2		
	Edges	Nodes	Edges	Nodes	Edges	Nodes	Edges	Man (%)	Comp (%)	Edges	Man (%)	**Comp (%)**
EBA1	257	136	299	121	240	116	104	40.47	34.78	95	36.96	**39.58**
EBA2	214	165	261	139	149	134	91	42.52	34.87	60	28.04	**40.27**
MBA	304	255	534	188	560	189	146	48.03	27.34	129	42.43	**23.04**
LBA	181	143	153	121	164	121	56	30.94	36.6	67	37.02	**40.85**
IA1	190	145	173	130	148	127	41	21.58	23.7	72	37.89	**48.65**
IA2	286	166	364	164	422	158	137	47.9	37.63	141	49.30	**33.41**
Mean	***239***	***168***	***297***	***144***	***281***	***141***	***82***	***38.57***	***32.49***	***94***	***38.61***	***37.63***

In the manual networks, between 3 and 14% of the nodes are isolates (except MBA with 20% isolates). Isolates are sites that coincide with a point on a hollow way and are therefore included in the network but that are not connected to any other site. In the computational networks, the number of isolates is within a range of 18–23% and 15–36%, respectively (Table 6).

The comparison of the selected sites shows large variation between the approaches. We exemplify the results for site importance with Tell Beydar, a site that was occupied in most of the periods analysed here (Table 7). Tell Beydar was one of the largest sites in the TBS region during EBA1 and EBA2. It was destroyed in the MBA with a long hiatus afterwards with barely any evidence of occupation. Occupation resumed in the LBA and Tell Beydar remained a small site during this period, regaining importance in the Iron Age. Hence, its degree and betweenness centrality should be in the 4th quartile in EBA1 and EBA2, the site should be not in the network in the MBA and degree and betweenness centrality should be in the 2nd or 3rd quartile in the LBA and in the 4th or 3rd quartile in IA1 and IA2.

In the manual network, the importance of Tell Beydar is reasonably well captured, although betweenness in EBA1 and EBA2 is only in the 2nd quartile and too high (4th quartile) in the LBA. The computational network 2 produced values that diverged more than the manual network from our expectations of how the site would behave, with degree and betweenness being 0 in EBA1 and EBA2 and degree 1 and betweenness 0 in IA2. The importance of Tell Beydar in the LBA and IA1 is well captured. In the computational network 1, only the LBA is captured reasonably well. In all three networks, the site is excluded in the MBA, reflecting the hiatus after its destruction. Results for the other sites are similar and presented in the supplements (Table 9 – Table 19).

The hybrid network was generated by adding the matrices of the manual and computational network 2. In those matrices, 1 indicates an edge between two sites and 0 indicates that there is no edge between two sites. Therefore, in the resulting matrix, 2 shows that the edge between the respective sites exists in both the manual and computational network and 1 represents an edge that is derived from only one of those networks. The matrix for the hybrid network was then converted into a network graph. The resulting graph for the MBA is presented in Fig. 9 alongside the graphs of the manual and computational networks. Graphs for the other periods are available in the supplements (Fig. 23 – Fig. 28).

Table 6 Number and percentage of isolates in the manual network, computational network 1 (d = 5500 m, a = 10°) and computational network 2 (d = 3000 m, a = 30°)

Period	*Manual*	*Comp 1*	*Comp 2*
EBA1	15 (11%)	28 (23%)	42 (36%)
EBA2	18 (11%)	28 (20%)	46 (34%)
MBA	52 (20%)	35 (19%)	29 (15%)
LBA	20 (14%)	25 (21%)	37 (31%)
IA1	14 (10%)	26 (20%)	43 (34%)
IA2	5 (3%)	30 (18%)	35 (22%)

Table 7 Results for degree and betweenness centrality for the site Tell Beydar. Green fields indicate a good fit between archaeological evidence and network metrics, yellow fields a reasonable (yet not perfect) fit and red fields imply that the network metrics do not reflect the importance of the site at all

Period	Manual		Computational 2		Computational 1	
	Degree	Betweenness	Degree	Betweenness	Degree	Betweenness
EBA1	5 (4th quartile)	59 (2nd quartile)	0 (2nd quartile)	0 (3rd quartile)	1 (1st quartile)	0 (3rd quartile)
EBA2	5 (4th quartile)	21 (2nd quartile)	0 (2nd quartile)	0 (3rd quartile)	2 (2nd quartile)	22 (4th quartile)
MBA	na	na	na	na	na	na
LBA	4 (4th quartile)	24 (3rd quartile)	0 (2nd quartile)	0 (3rd quartile)	1 (1st quartile)	0 (3rd quartile)
IA1	5 (4th quartile)	37 (3rd quartile)	6 (4th quartile)	47 (4th quartile)	0 (1st quartile)	0 (3rd quartile)
IA2	6 (4th quartile)	876 (4th quartile)	1 (2nd quartile)	0 (2nd quartile)	1 (1st quartile)	0 (2nd quartile)

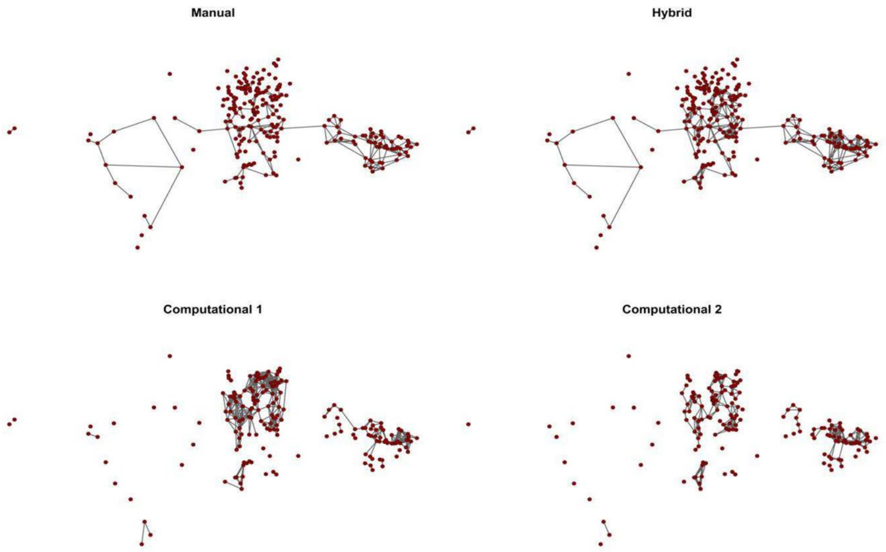

Fig. 9 Network graphs for the manual, hybrid and computational approaches

The results for the importance of selected sites are captured reasonably well, not as good as in the manual network but better than in the computational network. The results for the hybrid network are generally slightly worse than for the manual network but have improved compared to the computational network. Table 8 shows the degree and betweenness centrality for Tell Beydar; the results for the other sites can be accessed in the supplements (Table 9—Table 19). Degree and betweenness centrality for all periods in the hybrid network are captured not perfectly but reasonably well, similar to the manual network. The only difference to the manual network is degree in EBA1 which is slightly less well captured in the hybrid network (2nd instead of 3rd quartile).

Discussion

Site Prediction

The sites predicted by the Site Algorithm had only a small offset to known sites with a maximum of 300 m, which suggests that the predictions are accurate and that using proxy centroid coordinates is reasonable. Between 34 and 64% of the recorded sites could be predicted by the algorithm (Table 4).

The high correspondence between the FCP data and the predicted sites (42.3%) is caused by the data source: the FCP compiled data from survey and excavation projects which were intensive and recovered as complete as possible record of the four survey regions. Hence, it can be assumed that a high percentage of sites has been discovered, in particular large tells which are often associated with hollow ways.

Table 8 Results for degree and betweenness centrality for the site Tell Beydar. Green fields indicate a good fit between archaeological evidence and network metrics, yellow fields a reasonable (yet not perfect) fit and red fields imply that the network metrics

Period	Manual		Computational 2		Hybrid	
	Degree	Betweenness	Degree	Betweenness	Degree	Betweenness
EBA1	5 (3rd quartile)	59 (2nd quartile)	0 (2nd quartile)	0 (3rd quartile)	5 (2nd quartile)	59 (2nd quartile)
EBA2	5 (4th quartile)	21 (2nd quartile)	0 (2nd quartile)	0 (3rd quartile)	5 (3rd quartile)	21 (2nd quartile)
MBA	na	na	na	na	na	na
LBA	4 (4th quartile)	24 (3rd quartile)	0 (2nd quartile)	0 (3rd quartile)	4 (3rd quartile)	47 (3rd quartile)

However, if the sites have no radiating hollow ways around them, the algorithm was not able to predict them, which explains that more than half of the FCP sites could not be predicted.

The high correspondence with predicted sites and those in Kalaycı's data set (64.2%) can be explained by the data collection via remote sensing and extensive spatial coverage. Settlement mounds, or *tells*, are distinctive features, clearly visible on satellite or aerial images (Fig. 10). Most of the tells, in particular the largest ones, date to the Early Bronze Age (Ristvet, 2005; Wilkinson, 2000), when the hollow ways are assumed to have been in use. This relationship with the hollow ways explains the high correspondence with the predicted sites.

The data collection of the ANE sites is the reason for the 34.7% matching between them and the predicted sites: as with the FCP data base, the ANE placemarks mainly rely on surveyed and published data, but from different sources, so that the two data sets are different. The FCP data have been subset to only include sites from the relevant periods (Bronze Age to Iron Age) which is not possible for the ANE sites because no dating information are available (except those added in this study). Therefore, although the Site Algorithm detected more ANE than FCP sites, the uncertainty is much higher. In summary, the Site Algorithm to predict potential locations of archaeological sites performs generally well in areas where hollow ways are present and we were able to add 19 more sites to the data set compiled from the FCP, ANE and Kalaycı's data.

With the introduction of more and more computational tools in archaeology, machine and deep learning have become a means to train models for site prediction (Bachagha *et al.*, 2023; Ben-Romdhane *et al.*, 2023; Garcia-Molsosa *et al.*, 2021; Orengo *et al.*, 2020). The sparseness and incompleteness of the archaeological record provide a good case study for those approaches because those models can be trained to identify data gaps on the basis of observed data. However, for machine and deep learning to be successful and meaningful for archaeological site prediction, large training data sets of site locations in various environments are necessary, but those are still rare globally. A combination of LCP modelling using the

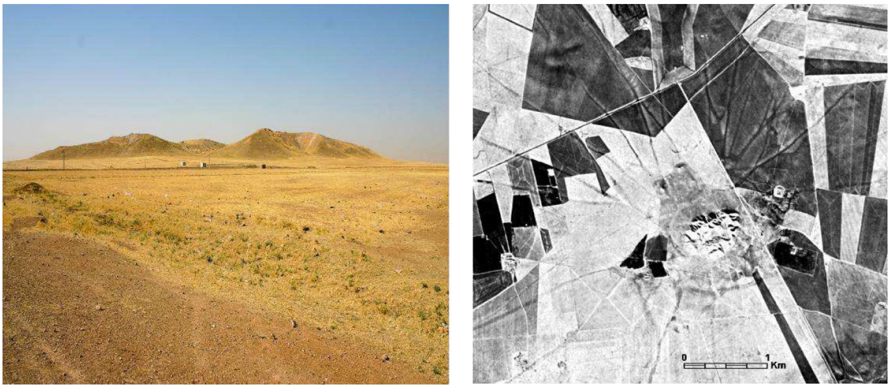

Fig. 10 The Bronze Age settlement Tell Brak from the ground (left) and on CORONA images (right)

FETE model as presented by White and Barber (2012) and Crabtree *et al*. (2021) and machine learning or linear regression would be a more efficient and less computationally demanding alternative. The approach presented here adds further to this line of research by including the partial evidence from known route segments and predicting the most probable site locations based on route intersections. The combination of remote sensing and machine learning is promising as well but also has its limitations such as the resolution of satellite and LiDAR images (Ben-Romdhane *et al*., 2023), the preservation of archaeological features (Orengo *et al*., 2020) or the difficulty of defining rules for the machine learning model that are precise enough so that they can be distinguished from other landscape features.

All those methods are accurate to varying degrees but they all have some common disadvantages: they rely on additional data such as satellite images and therefore the precision of the models is restricted to the resolution of these data. Moreover, although the models include environmental factors like slope, soil characteristics or vicinity to water, they usually ignore other landscape features such as paths, canals or other landmarks that could give further insights into the distribution of settlements.

The computational procedure for site prediction presented in this paper therefore uses ancient routes to predict sites in contrast to previous approaches that use sites to estimate the routes between them. Hence, we employ features created by the past societies themselves: the paths they moved on and that represent their approaches to wayfinding and the cultural, social and economic factors that influenced their route choice. Results from this study demonstrate that this method offers a more realistic and accurate way for the prediction of archaeological sites that can be used in any region of the world where characteristic landscape features of past societies are preserved and exhibit specific patterns.

Filling the Gaps in the Hollow Way System

The manual connecting of the hollow ways is prone to human error introduced by subjective perception—two persons might have different views about which fragments form a longer route. The resulting network therefore inevitably relies on the researcher constructing it and their perception. Hence, it is impossible to develop a computational procedure that incorporates all subjective perceptions and there simply might not be a "correct" answer when manually connecting hollow ways. Algorithms on the other hand lack the human component. Therefore, although the manual connecting has the major disadvantage of being subjective and not exactly reproducible, it might be the better approach to reflect how past humans moved while the computational procedure adds connections that the human might have missed.

Another issue with the manual approach is the possibility to miss connections that might not seem relevant on a small scale but contribute to the network on the large scale, as we are not always able to see the whole picture, especially when the data are messy. We tend to focus on smaller subsets which leads to a bias towards short-distance routes. However, this bias might be mitigated by the connections from one node to another via intermediate nodes (indirect paths), so that we can capture long-distance routes as well.

The computational approaches have on average fewer nodes than the manual approach. In particular, c. 15% of the archaeological sites are excluded due to the specifications of the computational procedure to connect the hollow way fragments with the sites: if a site has no connections to hollow ways, it is not included in the network (see Fig. 5 Step 3 “Connect hollow ways and sites” and Step 4 “Convert to network and clean” and code available on GitHub). The computational extraction of the network graph from the spatial object containing hollow way and site data (shapefile) required a site to coincide with a point on a (connected) hollow way and offsets of more than a 100 m between sites and points on hollow ways caused the sites to be excluded.

There are 10–20% more isolates in the computational networks compared to the manual network. Those results imply that the computationally connected hollow ways are generally less complete than the manual ones. Finding the ideal parameter values for the Hollow Way Algorithm might be an option to overcome this issue but those time- and resource-intensive tests were outside the scope of this study.

Both sets of parameters for the Hollow Way Algorithm introduce a high number of superfluous connections, unlike the manual approach (Fig. 7) which illustrates the most important difference in the two approaches: when manually connecting the hollow ways, only those connections are made that are perceived as forming a route. Those connections are based on the factors distance and direction as well as on the perception that it “looked right”. However, we are unable to explain what “look right” means in scientific terms which is one of the challenges when developing a computational approach. Although archaeological evidence was not explicitly used to connect the hollow ways—only points and lines were displayed when drawing the lines, without any further information—it might have influenced the process unconsciously. The Hollow Way Algorithm on the other hand creates new lines whenever the parameter thresholds are met, introducing superfluous connections, *i.e.* duplicated lines, lines that turn and bend unreasonably or zigzag. In general, there are less edges in the manual networks than in the computational ones.

The decision for the computational network 2 to combine the manual network into a hybrid one was based on the results described above: computational network 2 captures the importance of the selected sites better than network 1 (see Table 7) and produces fewer superfluous connections (Fig. 7). There are slightly more matching edges between the computational network 2 and the manual network (Table 5) which indicates that the parameter values are better suited to reproduce the human perception but might be more accurate as the human error is eliminated. Furthermore, there are more isolates in the computational than in the manual networks with much more isolates in computational network 1 than in computational network 2. Therefore, a combination of the manual and computational network 2 enhances the number of nodes while adding potentially more accurate edges. Another way might be to use the computational results as a starting point to manually connect the hollow ways. However, in that case, one would ideally use the computationally connected hollow ways as presented in Fig. 7 as those are the original ones resulting from the computational procedure. As they are quite messy and hard to distinguish, it might not be desirable to do it as it would probably take more time than doing the manual and computational connections separately and then merging the two.

Site importance is generally better captured in the manual network, and computational network 2 returns results that align better with the archaeological record than computational network 1. The hybrid network naturally provides results that present a trade-off between the manual and computational network 2; *i.e.* they are slightly worse than those of the manual but better than those of the computational approach. However, site importance is a vague proxy to evaluate the ability of an approach to reflect the archaeological record because a site might be a centre at the beginning of a period and lost its importance by the end of it. This intra-period variation cannot be recognised properly due to the broad chronological periods and even sub-periods. Another limitation is the incomplete data which probably causes a distortion for the investigated metrics degree and betweenness centrality. The results show that at least the manual, the computational network 2 and consequently the hybrid network capture site importance reasonably well, justifying site importance as a plausible way to evaluate the credibility of the networks.

Based on the results of this exploratory analysis, we conclude that the hybrid approach is the archaeologically most plausible and accurate one to represent the real-world network in the Khabur Valley in the Bronze and Iron Age. It combines the strengths of the manual and computational approaches while reducing their weaknesses and produces reasonably exact results with regard to the archaeological record. However, the hybrid network is still incomplete to a large and unknown degree and additional data would enhance the results of this approach significantly.

This paper fills the gap evident in research concerned with movement, connectivity and settlement dynamics in the Khabur Valley by offering a method to reconstruct the route system represented by the hollow ways. This method has the potential to enhance the archaeological record of paths and is not restricted to a specific area. However, a more extensive analysis to define the ideal parameter values is necessary.

Conclusion

In this paper, we presented manual and computational approaches to handle missing data in archaeological networks. More precisely, we offer computational methods to fill the gaps in two archaeological data sets to retrieve a more complete network: incomplete edges (*i.e.* routes or paths) and nodes (*i.e.* sites). Although the research area, periods and data are specific for this research project, the algorithms can be used for similar data with some adjustments. Any data set of features with a distinctive characteristic behaviour (such as the radiating patterns of the hollow ways) and any fragmented line data can be used as input for the algorithms and the parameters can be adjusted to the specific case study and research question.

Although both algorithms seem to be able to enhance the reconstruction of past settlement systems, they have limitations and need to be developed further. The main restrictions are the dependence of the Site Algorithm on the hollow ways, excluding areas without them, and the simplicity of the Hollow Way Algorithm that results in a significant number of superfluous connections. Despite these limitations, the accordance of the resulting network with the archaeological evidence, *i.e.* site importance in the individual periods and the location of the predicted sites, proved to be good, in particular considering the broad chronological classification. Therefore, we conclude that the computational procedures are suitable to fill the gaps in those types of archaeological data and would even improve with additional research on them.

Given the results of the Hollow Way Algorithm and the manual connecting of the hollow ways, a hybrid network, *i.e.* a combination of manual and computational network, is the most accurate and realistic option, although it is also affected by the large amount of missing data. However, our results demonstrate that the combination of manual and computational methods, either from the start or after initially conducting them separately, can significantly enhance the results. This is not only applicable for our data but for archaeological data in general. With current and future advances in computational methods, tools and techniques, we need to integrate them into our research to improve the quality of the inherently sparse and incomplete archaeological record. With a more complete archaeological record, the outcomes of any subsequent analysis will be improved as well and give us more accurate and realistic insights into the life of past humans. Furthermore, deriving site locations from route systems is an innovative way to semi-independently cross-check the results of AI-driven site detection approaches using satellite data, thereby providing an additional way to test the results if fieldwork is not possible.

ArcGIS Pro 3.1 was used to manually connect the hollow ways and generate Figs. 1, 2 and 7. Figures 3, 4 and 5 were created in InkScape and Figs. 8 and 9 were created in R. Figure 6 was produced in MatLab.

Software and Code

Data preparation, analysis and The Hollow Way Algorithm (Hollow way reconstruction) were carried out in R (R Core Team, 2021) with the GUI RStudio, version 4.1.0 (Windows 10). R packages used in for this paper include *sf* (Pebesma, 2018), *nngeo* (Dorman, 2023), *rgeos* (R. S. Bivand & Rundel, 2023), *sp* (R. Bivand *et al.*, 2013), *sfnetworks* (van der Meer *et al.*, 2023), *igraph* (Csardi & Nepusz, 2006; Csardi *et al.*, 2023), *shp2igraph* (Lu *et al.*, 2018) and *network* (Butts, 2008).

The Site Algorithm (site prediction) was developed in MatLab 2019b (The MathWorks Inc., 2019) with the toolboxes Image Processing 11.0 and Statistics and Machine Learning 11.6.

Data and scripts are available on GitHub (https://github.com/dpriss/Filling_the_gaps) and on Zenodo (https://zenodo.org/records/13825544).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10816-024-09688-z.

Author Contribution D.P. designed and conceptualised the study, developed the code in R, conducted the analysis in R and ArcGIS and prepared the figures as well as the manuscript draft. C.K. developed the code in MatLab. J.W. and L.T. reviewed and supported the coding in R. A.I. conceptualised algorithm 1.

All authors were involved in the conceptualisation of the study and the discussion of the results. All authors reviewed the manuscript.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 859937.

Data Availability Data and scripts are available on GitHub: https://github.com/dpriss/Filling_the_gaps and on Zenodo: https://zenodo.org/records/13825544.

Declarations

Competing Interests The authors declare no competing interests.

References

Amati, V., et al. (2020). A framework for reconstructing archaeological networks using exponential random graph models. *Journal of Archaeological Method and Theory, 27*(2), 192–219. https://doi.org/10.1007/s10816-019-09423-z

Bachagha, N., et al. (2023). The use of machine learning and satellite imagery to detect Roman fortified sites: The case study of Blad Talh (Tunisia Section). *Applied Sciences, 13*(4), 2613. https://doi.org/10.3390/app13042613

Bell, M. (2020). *Making one's way in the world: The footprints and trackways of prehistoric people*. Oxbow Books. http://centaur.reading.ac.uk/67596/

Ben-Romdhane, H., et al. (2023). Detecting and predicting archaeological sites using remote sensing and machine learning—Application to the Saruq Al-Hadid site, Dubai, UAE. *Geosciences, 13*(6), 179. https://doi.org/10.3390/geosciences13060179

Bevan, A., & Wilson, A. (2013). Models of settlement hierarchy based on partial evidence. *Journal of Archaeological Science, 40*(5), 2415–2427. https://doi.org/10.1016/j.jas.2012.12.025

Bieliński, P. (1992). The first campaign of excavations on Tell Rad Shaqrah (Hassake Southern Dam Basin). *Polish Archaeology in the Mediterranean, 3*, 77–85.

Bivand, R. S., & Rundel, C. (2023). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. Available at: https://r-forge.r-project.org/projects/rgeos/ https://libgeos.org http://rgeos.r-forge.r-project.org/index.html

Bivand, R., Pebesma, E. J. and Gómez-Rubio, V. (2013). *Applied spatial data analysis with R*. Second edition. New York: Springer (Use R!).

Brughmans, T., Keay, S., & Earl, G. (2014). Introducing exponential random graph models for visibility networks. *Journal of Archaeological Science, 49*, 442–454. https://doi.org/10.1016/j.jas.2014.05.027

Brughmans, T., & Peeples, M. A. (2023). *Network science in archaeology*. Cambridge University Press (Cambridge manuals in archaeology). https://doi.org/10.1017/9781009170659

Brughmans, T., Mills, B. J., Munson, J., & Peeples, M. A. (2023). *The Oxford handbook of archaeological network research* (1st ed). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198854265.001.0001.

Butts, C. T. (2008). network: A package for managing relational data in R. *Journal of Statistical Software, 24*(2). https://doi.org/10.18637/jss.v024.i02 van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed). Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259

Casana, J. (2013). Radial route systems and agro-pastoral strategies in the Fertile Crescent: New discoveries from western Syria and southwestern Iran. *Journal of Anthropological Archaeology, 32*(2), 257–273. https://doi.org/10.1016/j.jaa.2012.12.004

Cochrane, E. E., & Lipo, C. (2010). Phylogenetic analyses of Lapita decoration do not support branching evolution or regional population structure during colonization of Remote Oceania. *Proceedings of the Royal Society B, 365*, 3889–3902.

Crabtree, S. A., et al. (2021). Landscape rules predict optimal superhighways for the first peopling of Sahul. *Nature Human Behaviour, 5*(10), 1303–1313. https://doi.org/10.1038/s41562-021-01106-8

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, p. 1695. https://igraph.org

Csardi, G. *et al.* (2023). 'igraph: Network analysis and visualization in R'. Available at: https://CRAN.R-project.org/package=igraph.

de Gruchy, M. W. (2016). Beyond replication: The quantification of route models in the North Jazira, Iraq. *Journal of Archaeological Method and Theory, 23*(2), 427–447. https://doi.org/10.1007/s10816-015-9247-x

Dicks, T. R. B. (1972). Network analysis and historical geography. *Area, 4*(1), 4–9.

Dorman, M. (2023). 'nngeo: k-nearest neighbor join for spatial data'. Available at: https://michaeldorman.github.io/nngeo/, https://github.com/michaeldorman/nngeo.

Ducke, B., & Suchowska, P. (2022). Exploratory network reconstruction with sparse archaeological data and XTENT. *Journal of Archaeological Method and Theory, 29*, 508–539. https://doi.org/10.1007/s10816-021-09529-3

Ebert, J. I., & Hitchcock, R. J. (1980). Locational modeling in the analysis of the prehispanic roadway system at and around Chaco Canyon, New Mexico. In T. R. Lyons & F. J. Mathien (Eds.), *Cultural Resources Remote Sensing* (pp. 169–207). National Park Service.

FCP .(2013). *The Fragile Crescent: Settlement change during the urban transition*. Available at: https://gtr.ukri.org/projects?ref=AH%2FF010095%2F1 (Accessed: 14 July 2022).

Galiatsatos, N., Wilkinson, T., Donoghue, D., & Philip, G. (2009). The Fragile Crescent Project (FCP): Analysis of settlement landscapes using satellite imagery. Presented at CAA 2009: Making history interactive.

Garcia-Molsosa, A., et al. (2021). Potential of deep learning segmentation for the extraction of archaeological features from historical map series. *Archaeological Prospection, 28*(2), 187–199. https://doi.org/10.1002/arp.1807

Graham, S. (2006). Networks, agent-based models and the Antonine itineraries: Implications for Roman archaeology. *Journal of Mediterranean Archaeology, 19*(1), 45–64. https://doi.org/10.1558/jmea.2006.19.1.45

Groenhuijzen, M. R., & Verhagen, P. (2017). Comparing network construction techniques in the context of local transport networks in the Dutch part of the Roman limes. *Journal of Archaeological Science: Reports, 15*, 235–251. https://doi.org/10.1016/j.jasrep.2017.07.024

de Gruchy, M. and Cunliffe, E. (2020). 'How the hollow ways got their form (and kept it): 5000 years of hollow ways at Tell al-Hawa', pp. 124–143.

Hazra, S. (2020). (2020) 'Prediction of archaeological potential site in middle and lower course of Mayurakshi River Basin, Eastern India using logistic regression model and GIS.' *Heritage: Journal of Multidisciplinary Studies in Archaeology, 8.2*, 875–890.

Hole, F. (1999). Economic implications of possible storage structures at Tell Ziyadeh, NE Syria. *Journal of Field Archaeology, 26*(3), 267–283. https://doi.org/10.2307/530514

Isaksen, L. (2008). The application of network analysis to ancient transport geography: A case study of Roman Baetica. *Digital Medievalist*, 4. https://doi.org/10.16995/dm.20

Jenkins, D. (2001). A network analysis of inka roads, administrative centers, and storage facilities. *Ethnohistory, 48*(4), 655–687. https://doi.org/10.1215/00141801-48-4-655

Jiménez, D., & Chapman, D. (2002). An application of proximity graphs in archaeological spatial analysis. In D. Wheatley, G. Earl, & S. Poppy (Eds.), *Contemporary themes in archaeological computing* (pp. 90–99). Oxbow Books.

Jones, G. A. (1970). The nuclear surface. *Reports on Progress in Physics, 33*(2), 645–689. https://doi.org/10.1088/0034-4885/33/2/304

Jones, C. (2018). 'AWOL - The Ancient World Online: ANE Placemarks for Google Earth', *AWOL - The Ancient World Online*, 16 April. Available at: http://ancientworldonline.blogspot.com/2011/07/ane-placemarks-for-google-earth.html (Accessed: 2 November 2021).

Jotheri, J., et al. (2019). Remote sensing the archaeological traces of boat movement in the marshes of Southern Mesopotamia. *Remote Sensing, 11*(21), 2474. https://doi.org/10.3390/rs11212474

Judge, W. J., & Sebastian, L. (1988). *Quantifying the present and predicting the past: Theory, method and application of archaeological predictive modeling*. Department of the Interior, Bureau of Land Management Service Centre. Available at: https://dn720500.ca.archive.org/0/items/quantifyingprese00judg/quantifyingprese00judg.pdf

Kalaycı, T. (2013). *Agricultural production and stability of settlement systems in Upper Mesopotamia during the Early Bronze Age (Third Millennium BCE)*. PhD Thesis. University of Arkansas. Available at: https://scholarworks.uark.edu/etd/719/.

Kalaycı, T. (2022). Mounded site boundaries [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.7048407

Krause, R. (2019). *Multiple imputation for* missing *network data*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. https://doi.org/10.33612/diss.103522814

Lebeau, M., & Ciavarini-Azzi, E. (1993). *Tell Melebiya: cinq campagnes de recherches sur le Moyen-Khabour (1984-1988)*. Peeters.

Lewis, J. (2021). Probabilistic modelling for incorporating uncertainty in least cost path results: A postdictive Roman road case study. *Journal of Archaeological Method and Theory, 28*, 911–924. https://doi.org/10.1007/s10816-021-09522-w

Lewis, J. (2023). Explaining known past routes, underdetermination, and the use of multiple cost functions. *Journal of Archaeological Method and Theory, 31*, 854–874. https://doi.org/10.1007/s10816-023-09621-w

Li, L., et al. (2022). A prediction study on archaeological sites based on geographical variables and logistic regression—A case study of the Neolithic Era and the Bronze Age of Xiangyang. *Sustainability, 14*(23), 15675. https://doi.org/10.3390/su142315675

Lu, B., et al. (2018). Shp2graph: Tools to convert a spatial network into an igraph graph in R. *ISPRS International Journal of Geo-Information, 7*(8), 293. https://doi.org/10.3390/ijgi7080293

Lucas, G. (2012). *Understanding the archaeological record* (1st ed). Cambridge University Press. https://doi.org/10.1017/CBO9780511845772

Lyonnet, B., & Durand, J-M. (1996). 'La prospection archéologique de la partie occidentale du Haut-Khabur (Syrie du Nord-Est): Méthodes, résultats et questions autour de l'occupation aux IIIe et IIe millénaires av. n. è.', *Amurru 1. Mari, Ébla et les Hourrites dix ans de travaux. Actes du colloque international (Paris, mai 1993)*, pp. 363–376.

Mallowan, M. E. L. (1936). The excavations at Tall Chagar Bazar, and an archaeological survey of the Habur Region, 1934–5. *Iraq, 3*(1), 1–85.

McLean, A., & Rubio-Campillo, X. (2022). Beyond least cost paths: Circuit theory, maritime mobility and patterns of urbanism in the Roman Adriatic. *Journal of Archaeological Science, 138*, 105534. https://doi.org/10.1016/j.jas.2021.105534

McMahon, A., Tunca, Ö., & Bagdo, A.-M. (2001). New excavations at Chagar Bazar, 1999–2000. *Iraq, 63*, 201–222. https://doi.org/10.2307/4200512

McMahon, A., Colantoni, C., Frane, J. E., & Soltysiak, A. (2009) *Once there was a place: Settlement archaeology at Chagar Bazar, 1999 - 2002*. British Institute for the Study of Iraq.

van der Meer, L. *et al.* (2023). 'sfnetworks: Tidy Geospatial Networks'. Available at: https://luukvdmeer.github.io/sfnetworks/.

Meijer, D. J. W. (1986). *A survey in Northeastern Syria*. PIPHANS 58. Nederlands Instituut voor het Nabije Oosten.

Menze, B. H., & Ur, J. A. (2012). Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. *Proceedings of the National Academy of Sciences, 109*(14), E778–E787. https://doi.org/10.1073/pnas.1115472109

Mills, B. J., et al. (2013). Transformation of social networks in the late pre-Hispanic US Southwest. *Proceedings of the National Academy of Sciences, 110*(15), 5785–5790. https://doi.org/10.1073/pnas.1219966110

Orengo, H. A., & Livarda, A. (2016). The seeds of commerce: A network analysis-based approach to the Romano-British transport system. *Journal of Archaeological Science, 66*, 21–35. https://doi.org/10.1016/j.jas.2015.12.003

Orengo, H. A., et al. (2020). Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data. *Proceedings of the National Academy of Sciences, 117*(31), 18240–18250. https://doi.org/10.1073/pnas.2005583117

Östborn, P., & Gerding, H. (2014). Network analysis of archaeological data: A systematic approach. *Journal of Archaeological Science, 46*, 75–88. https://doi.org/10.1016/j.jas.2014.03.015

Palmisano, A., & Altaweel, M. (2015). Landscapes of interaction and conflict in the Middle Bronze Age: From the open plain of the Khabur Triangle to the mountainous inland of Central Anatolia. *Journal of Archaeological Science: Reports, 3*, 216–236. https://doi.org/10.1016/j.jasrep.2015.06.015

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal, 10*(1), 439. https://doi.org/10.32614/RJ-2018-009

Pedersen, O., Sinclair, P., Hein, I., & Anderssen, J. (2010). Cities and urban landscapes in the ancient Near East and Egypt with special focus on the city of Babylon. Studies in Global Archaeology 15. *The Urban Mind : Cultural and Environmental Dynamics*, 113–147.

Pedersen, O. (2012). Ancient near East on Google Earth: Problems, preliminary results, and prospects. In *Proceedings of the 7th International Congress on the Archaeology of the Ancient Near East: 12 April - 16 April 2010, the British Museum and UCL, London, Fielworks and recent research* (pp. 385–393). https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-172607

Peeples, M. A., & Mills, B. J. (2016). 'Analytical challenges for the application of social network analysis in archaeology'. In T. Brughmans, A. Collar, & F. Coward (Eds.), *The connected past: Challenges to network studies in archaeology and history* (pp. 59–84). Oxford University Press. https://doi.org/10.1093/oso/9780198748519.003.0010

Perreault, C. (2019). *The quality of the archaeological record*. The University of Chicago Press. https://doi.org/10.7208/chicago/9780226631011.001.0001

Pfälzner, P. (1990). 'Tell Bderi - The development of a Bronze Age town', in S. Kerner (ed.) *The Near East in Antiquity. German contributions to the archaeology of Jordan, Palestine, Syria, Lebanon and Egypt*. Amman: Heidelberg University Library, pp. 63–79. Available at: https://api.semanticscholar.org/CorpusID:132232147.

Pitts, F. R. (1965). A graph-theoretic approach to historical geography. *The Professional Geographer, 17*(5), 15–20. https://doi.org/10.1111/j.0033-0124.1965.015_m.x

R Core Team .(2021). 'R: A language and environment for statistical computing.' Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Ristvet, L. (2005). *Settlement, economy, and society in the Tell Leilan Region, Syria, 3000–1000 BC*. PhD Thesis. University of Cambridge. https://leilan.yale.edu/sites/default/files/publications/article-specific/ristvet_diss.pdf

Rivers, R., Knappett, C., & Evans, T. (2013). What makes a site important? Centrality, gateways, and gravity. In C. Knappett (Ed.), *Network analysis in archaeology: New approaches to regional interaction* (pp. 125–150). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199697090.003.0006

Smith, J. A., Morgan, J. H., & Moody, J. (2022). Network sampling coverage III: Imputation of missing network data under different network and missing data conditions. *Social Networks, 68*, 148–178. https://doi.org/10.1016/j.socnet.2021.05.002

The MathWorks Inc. (2019). 'MATLAB'. Natick, Massachusetts: The MathWorks Inc. Available at: https://www.mathworks.com.

Tsirogiannis, C., & Tsirogiannis, C. (2016). Uncovering the hidden routes: Algorithms for identifying paths and missing links in trade networks. In T. Brughmans, A. Collar, & F. Coward (Eds.), *The Connected Past: Challenges to network studies in archaeology and history* (pp. 103–120). Oxford University Press.

Ur, J. (2003). CORONA satellite photography and ancient road networks: A Northern Mesopotamian case study. *Antiquity, 77*(295), 102–115. https://doi.org/10.1017/S0003598X00061391

Ur, J., & Wilkinson, T. (2008). 'Settlement and economic landscapes of Tell Beydar and its hinterland'. Edited by M. *Lebeau and a. Suleiman, I*, 305–327.

Ur, J. (2008). 'Hollow ways in Northern Mesopotamia'. Available at: https://www.arcgis.com/home/item.html?id=39d4e9cc633a46bba925d3ce6b0f6b8b (Accessed: 7 January 2021).

Ur, J. (2009). Emergent landscapes of movement in Early Bronze Age northern Mesopotamia. In J. E. Snead, C. L. Erickson, & J. A. Darling (Eds.), *Landscapes of movement: Paths, trails, and roads in anthropological perspective* (pp. 180–203). University of Pennsylvania Press. https://doi.org/10.9783/9781934536537.180

Ur, J., & McGuire, G. (2010). *Tell Hamoukar. Volume 1, Urbanism and cultural landscapes in Northeastern Syria: The Tell Hamoukar survey, 1999–2001*. The Oriental Institute of the University of Chicago.

Verhagen, P., Joyce, J., & Groenhuijzen, M. R. (2019). *Finding the limits of the limes: Modelling demography, economy and transport on the edge of the Roman Empire*. Computational Social Science, Simulating the Past, Springer. https://doi.org/10.1007/978-3-030-04576-0

Wachtel, I., et al. (2018). Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north Israel and north-east China. *Journal of Archaeological Science, 92*, 28–36. https://doi.org/10.1016/j.jas.2018.02.001

Weiss, H. (1985). Rediscovering: Tell Leilan on the Habur Plains of Syria. *The Biblical Archaeologist, 48*(1), 5–34. https://doi.org/10.2307/3209945

Weiss, H. (2014). *Tell Leilan Region Survey | Tell Leilan Project*, *Tell Leilan Region Survey*. Available at: https://leilan.yale.edu/works-progress/tell-leilan-region-survey (Accessed: 6 October 2022).

White, D. A., & Barber, S. B. (2012). Geospatial modeling of pedestrian transportation networks: A case study from precolumbian Oaxaca, Mexico. *Journal of Archaeological Science, 39*(8), 2684–2696. https://doi.org/10.1016/j.jas.2012.04.017

Wilkinson, T. J. (1993). Linear hollows in the Jazira, Upper Mesopotamia. *Antiquity, 67*(256), 548–562. https://doi.org/10.1017/S0003598X00045750

Wilkinson, T. J., et al. (2010). The geoarchaeology of route systems in northern Syria. *Geoarchaeology, 25*(6), 745–771. https://doi.org/10.1002/gea.20331

Wilkinson, T. J., & Tucker, D. J. (1995). *Settlement development in the North Jazira, Iraq: A study of the archaeological landscape*. Published for the British School of Archaeology in Iraq and the Department of Antiquities & Heritage, Baghdad, by Aris & Phillips.

Wilkinson, T. J. (2000). Archaeological survey of the Tell Beydar Region, Syria, 1997: A preliminary report. In K. van Lerberghe, & G. Voet (Eds.), *Tell Beydar: Environmental and technical studies* (pp. 1–37). Subartu 6. Brepols.

Woods, R. D., & Saxon, D. S. (1954). Diffuse surface optical model for nucleon-nuclei scattering. *Physical Review, 95*(2), 577–578. https://doi.org/10.1103/PhysRev.95.577

Yaworsky, P. M., et al. (2020). Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument. *PLOS ONE, 15*(10), e0239424. https://doi.org/10.1371/journal.pone.0239424