



Universiteit  
Leiden  
The Netherlands

## **A Framework for Reconstructing Archaeological Networks Using Exponential Random Graph Models**

Amati, V.; Mol, A.A.A.; Shafie, T.; Hofman, C.L.; Brandes, U.

### **Citation**


Amati, V., Mol, A. A. A., Shafie, T., Hofman, C. L., & Brandes, U. (2019). A Framework for Reconstructing Archaeological Networks Using Exponential Random Graph Models. *Journal Of Archaeological Method And Theory*. doi:10.1007/s10816-019-09423-z

Version: Publisher's Version  
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)  
Downloaded from: <https://hdl.handle.net/1887/84203>

**Note:** To cite this publication please use the final published version (if applicable).



# A Framework for Reconstructing Archaeological Networks Using Exponential Random Graph Models

Viviana Amati<sup>1</sup>  · Angus Mol<sup>2</sup> · Termeh Shafie<sup>3</sup> · Corinne Hofman<sup>4</sup> · Ulrik Brandes<sup>1</sup>

Published online: 19 August 2019  
© The Author(s) 2019

## Abstract

Reconstructing ties between archaeological contexts may contribute to explain and describe a variety of past social phenomena. Several models have been formulated to infer the structure of such archaeological networks. The applicability of these models in diverse archaeological contexts is limited by the restricted set of assumptions that fully determine the mathematical formulation of the models and are often articulated on a dyadic basis. Here, we present a general framework in which we combine exponential random graph models with archaeological substantiations of mechanisms that may be responsible for network formation. This framework may be applied to infer the structure of ancient networks in a large variety of archaeological settings. We use data collected over a set of sites in the Caribbean during the period AD 100–400 to illustrate the steps to obtain a network reconstruction.

**Keywords** Early Ceramic Age · Caribbean networks · Exponential random graph models · Network reconstruction

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10816-019-09423-z>) contains supplementary material, which is available to authorized users.

---

✉ Viviana Amati  
viviana.amati@gess.ethz.ch

<sup>1</sup> Department of Humanities, Social and Political Sciences, Social Networks Lab, ETH Zurich, Weinbergstrasse 109, 8092 Zurich, Switzerland

<sup>2</sup> Leiden University Centre for Digital Humanities, Nonnensteeg 1-3, 2311 VJ Leiden, The Netherlands

<sup>3</sup> Department of Social Statistics, Mitchell Centre for Social Network Analysis, University of Manchester, Humanities Bridgeford Street, Manchester M139PL, UK

<sup>4</sup> Faculty of Archaeology, Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands

## Introduction

Empirical and theoretical studies of networks based on archaeological data are on a rapid rise (Brughmans 2013; Collar et al. 2015; Knappett 2011). However, relatively few studies attempt to infer the structure of past networks. Ties between archaeological contexts (*e.g.* sites) or attributes of these contexts (*e.g.* site assemblages and artefacts) may contribute to the analysis of the structural characteristics of networks in the past and the evaluation of their impact on a variety of past social phenomena. Examples include the diffusion of innovations (Mol 2013, 2014; Roux and Manzo 2018), direct contact (Boomert 2000; Hofman et al. 2014), exchange of goods (Hofman et al. 2007; Knippenberg 2007), central place redistribution (Crock 2000), the mobility of groups of people or individuals (Laffoon et al. 2017; Rouse 1992) and their interrelations—*e.g.* human mobility and the exchange of goods and ideas (Hofman et al. 2010).

Two main approaches have been adopted to infer the structure of past networks (Östborn and Gerding 2014). One relies on the assumption that similarity in site assemblages is a proxy for the existence of ties (Coward 2010, 2013; Mills et al. 2013; Mol 2014; Habiba et al. 2018) so that “the broken links of a ruined network [are inferred] from observable distributions and patterns of association in the archaeological record” (Sindbaek 2013, p. 71). The other one focusses on the processes that might have created ties in the past and consists in specifying a model from which plausible networks are generated. The formulation of this model hinges on the assumptions archaeologists have about the formation of the relation(s) of interest (Bevan and Wilson 2013; Knappett et al. 2008; Terrell 1977). Here, we focus on the second approach and, in particular, on tie-based models, the definition of which depends on assumptions articulated at the tie level, as opposed to agent-based models, which are formulated based on propositions articulated at the node level (Graham 2006; Wurzer et al. 2015).

Several tie-based models have been used for reconstructing archaeological networks, among them maximum distance networks (Evans et al. 2012), proximal point analysis (Broodbank 2002; Terrell 1977), gravity models (Conolly and Lake 2006; Hodder 1974; Johnson 1977) and ariadne (Rivers et al. 2011). The applicability of these models in diverse archaeological contexts is limited by their mathematical formulation which is fully determined by the propositions those models entail. Different sets of tie formation assumptions require the formulation of new generative models.

Maximum distance networks, proximal point analysis, gravity models and ariadne also postulate that the existence of a tie between any two archaeological contexts  $i$  and  $j$  depends on entity attributes measured at the node level (*e.g.* size) or at the dyadic level (*e.g.* geographical location), as well as the constraints imposed on them. Therefore, they assume that the mechanisms that might have generated a network act only on a dyadic base. The assumption of tie independence overlooks more complex processes of network formation suggesting that networks are the outcomes of interdependent interactions embedded in a certain environment—rather than outcomes of interactions taking place in a vacuum of dyadic relations.

Archaeological propositions concerning the formation of ties among diverse archaeological entities (*e.g.* sites, households, cities or regions) that, more or less explicitly, embody the idea of tie dependence cannot be represented using the models mentioned above. Examples of archaeological propositions implying tie dependence relate to

transitivity, and its opposite, indirect exchange in trade networks. Given three archaeological contexts  $i, j$  and  $k$ , transitivity implies that if  $i$  exchanges goods with  $j$  and  $j$  with  $k$ , it is likely that  $i$  will start exchanging goods with  $k$  as well. Contrary to transitivity, indirect exchange indicates that it is less likely that  $i$  will start exchanging goods with  $k$  as well (see, e.g. Blake (2014), for more details and discussion).

In this paper, we propose to use standard statistical models for the analysis of networks to reconstruct ancient networks. In particular, we consider exponential random graph models (ERGMs) (Lusher et al. 2012; Robins et al. 2007; Wasserman and Pattison 1996). These models have already been used to reconstruct structurally efficient networks of contacts, *i.e.* plausible network scenarios where contacts between archaeological contexts were regulated by broker sites (Amati et al. 2018). Building on this previous work, we demonstrate that the applicability of ERGMs is not limited to the reconstruction of structurally efficient networks; rather, ERGMs constitute a flexible family of models that allows researchers to reconstruct networks given a variety of assumptions, contexts and relations. The derivation of the mathematical formulation of ERGMs does not depend on the context and the particular assumptions about the mechanisms regulating the formation of ties in the past. Moreover, ERGMs subsume maximum distance networks, proximal point analysis and gravity models as special cases and therefore have the potential of providing a unified model for inferring ties among entities in the past.

In this paper, we present a general framework in which we combine ERGMs with archaeological theories of mechanisms that may be responsible for network formation. Given a set of archaeological contexts, a particular relation (*e.g.* contact or exchange) and a set of propositions describing how ties formed, the described framework permits researchers to generate plausible network scenarios. To illustrate the steps of the procedure, we used data collected over a group of sites in the Caribbean during the period AD 100–400. This period falls under the (Middle) Early Ceramic Age, but should rather be seen as a new phase in which previous ways of life (more prevalent in the so-called Archaic period) were fully transformed into what is referred to as the Ceramic Age. To reflect this dynamic, as well as to somewhat neutralize the awkward term “Archaic”, we will here refer to the specific AD100–400 period as the Archaic-Early Ceramic Interface (AECI) period.

The remainder of the paper is organised as follows. In the “[Exponential Random Graph Models](#)” section, ERGMs are introduced; the steps necessary to reconstruct archaeological networks are presented in the “[ERGMs for Network Reconstruction](#)” section. Illustrative examples of the proposed framework are provided in the “[Example: an Application to the Pre-colonial Caribbean](#)” section. The framework and its application are further discussed in the “[Conclusion](#)” section.

## Exponential Random Graph Models

### Definition

Let  $\mathcal{N} = \{1, \dots, n\}$  be a set of archaeological contexts, hereafter referred to as sites, and  $R: \mathcal{N} \times \mathcal{N} \rightarrow [0,1]$  be a binary relation between them. We represent a network as an adjacency matrix  $x$ , whose cell  $x_{ij}$  takes value 1 if there is a

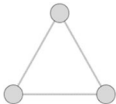

relationship between sites  $i$  and  $j$ , and 0 otherwise. When  $x_{ij} = 1$ , we say that there is a tie between  $i$  and  $j$ , or that  $i$  is tied to  $j$ . Ties are undirected (e.g. contact or connectedness) when the existence of a tie from  $i$  to  $j$  implies the existence of a tie from  $j$  to  $i$  (i.e.  $x_{ij} = 1$  implies  $x_{ji} = 1$ ). Ties are directed (e.g. exchange) when the existence of a tie from  $i$  to  $j$  does not imply the existence of a tie from  $j$  to  $i$  (i.e.  $x_{ij} = 1$  does not imply  $x_{ji} = 1$ ). Let  $v$  be the case by variable matrix containing the attributes of the archaeological contexts (e.g. size and location) and  $w$  be an array containing dyadic information (e.g. distance between contexts).

Exponential random graph models (ERGMs) are a class of statistical models for networks. Their mathematical formulation was originally derived by Frank and Strauss (1986) using concepts from spatial statistics, and subsequently extended by Wasserman and Pattison (1996). Later derivations based on notions of other disciplines have been proposed. In theoretical physics, and more specifically statistical mechanics, scholars have linked the behaviour and the arrangements of particles in a system to those of entities in a network and derived ERGMs from the Gibbs-Boltzmann distribution (Newman 2018a). Related to this derivation, by employing information theory, researchers have deduced the formulation of ERGMs using the maximum entropy principle (Shannon 1948; Jaynes 1957a, 1957b). More recently, Butts (2009) and Mele (2017) derived the formulation of some specifications of ERGMs from principles of game theory. For a more detailed discussion about the derivation of ERGMs, the sources cited above are recommended.

ERGMs implicitly assume that the global structure of an observed network emerges from dynamic processes acting simultaneously and representing the rules describing the formation of ties in a network. Hereafter, we refer to these dynamic processes as generative mechanisms. Examples of generative mechanisms include transitivity and homophily, suggesting that ties are formed between entities that are both tied to one or more common entities and between entities that have similar characteristics, respectively. Traces of these mechanisms in an observed network are particular patterns of ties referred to as local configurations. Table 1 displays the local configurations corresponding to the mechanisms mentioned above. Intuitively, a large number of triangles and homophilic dyads in the observed network indicates that the structure of the observed network emerges from transitivity and homophily mechanisms.

Formally, ERGMs are a family of probability distributions over the space  $x$  of all the possible networks defined on the set of nodes  $\mathcal{N}$  and have the form

**Table 1** Example of the correspondence between generative mechanisms and dynamic processes for transitivity and homophily. The colours of the nodes represents different values taken by an entity attribute

Generative mechanism	Transitivity	Homophily
Local configuration		

$$P(X = x) = \frac{1}{\kappa} \exp\left(\sum_k \theta_k s_k(x, v, w)\right), \tag{1}$$

where  $X$  is the network random variable,  $x$  denotes the value taken by  $X$ ,  $\theta_k$  represents a parameter,  $s_k(x, v, w)$  is a statistic and  $\kappa = \sum_{x' \in \mathcal{X}} \exp\left(\sum_k \theta_k s_k(x', v, w)\right)$  is a normalising constant.

The linear combination  $\sum_k \theta_k s_k(x, v, w)$  in Eq. (1) is the mathematical representation of the assumption that the structure of an observed network is the outcome of dynamic processes acting simultaneously. More specifically, the statistic  $s_k(x, v, w)$  counts the number of local configurations of type  $k$  which are the traces of the  $k$ th generative mechanism. The sum over  $k$  expresses the idea that more than one mechanism may have generated the network  $x$ : Ties may occur due to the presence/absence of other ties, as well as the attributes of the entities or pairs of entities. This fact is mathematically represented by the arguments of the statistic  $s_k$ , *i.e.* the network  $x$ , the entity attribute  $v$  and the dyadic information  $w$ . We provide a list of the most common statistics and their connection to the mechanisms they entail in the next section.

The parameter  $\theta_k$  measures the importance of the local configurations of type  $k$  in determining the global structure of the network, in other words, the relative importance of a mechanism to the formation of ties. A positive (negative) value of the parameter  $\theta_k$  indicates that a tie is more (less) likely to occur when its presence increases (decreases) the value of the statistic  $s_k(x, v, w)$ , thereby providing evidence for (against) the corresponding mechanism. This interpretation stems from the fact that ERGMs can be regarded as log-linear models for the binary tie random variables  $X_{ij}$  the collection of which generates the random network  $X$ . According to this conceptualisation, the logarithm of the ratio between the probability of a tie being present and the probability of a tie being absent, conditional on the other ties in the network, is expressed as follows:

$$\log \left[ \frac{P(X_{ij} = 1 | x_{ij}^c; \theta)}{P(X_{ij} = 0 | x_{ij}^c; \theta)} \right] = \sum_k \theta_k [s_k(x^{+ij}, v, w) - s_k(x^{-ij}, v, w)], \tag{2}$$

where  $x^{+ij}$  and  $x^{-ij}$  denote the networks with the tie  $x_{ij}$  present (*i.e.*  $x_{ij} = 1$ ) and absent (*i.e.*  $x_{ij} = 0$ ), respectively, whilst  $x_{ij}^c$  represents the set of all the tie variables except  $x_{ij}$ .










For those familiar with logistic regression models, Eq. (2) indicates that the parameters of an ERGM can be interpreted in a similar way to the parameters of a logistic regression model (Shennan 1997; Agresti and Kateri 2011). If  $\theta_k$  is positive and if the presence of a tie between two nodes  $i$  and  $j$  leads to an increase in the value of the statistic  $s_k(x, v, w)$ , then the tie  $X_{ij}$  is more likely to be present than absent, whilst keeping all the other statistics fixed. However, it should be noted that this interpretation has only a heuristic value since the existence or the deficiency of a tie might change the value of multiple statistics at the same time. For instance, the presence of a tie between  $i$  and  $j$  increases the number of both the edges and triangles when the ties between  $i$  and  $h$ , as well as  $j$  and  $h$  exist.

## Statistics















Many statistics have been defined to specify ERGMs (Morris et al. 2008; Wang et al. 2009). Table 2 and Table 3 show the most common local configurations along with the mechanisms of tie formation that they represent. These tables also contain references to theories that have been developed in archaeology and can be easily encoded into combinations of ties.

Statistics for directed ties are extensions of statistics for undirected ties. Therefore, in the following sections, we mainly focus on the statistics for undirected relations, and we briefly describe those for the directed case. We relate these statistics to both the network properties and the archaeological propositions they represent. We distinguish between endogenous statistics—associated with propositions concerning the existence of ties in reaction to the existence of other ties—and exogenous statistics—encoding the propositions expressing the presence of ties based on site characteristics.

**Table 2** Network properties, local configurations, interpretation and references to corresponding archaeological theories for undirected relations. Node and dyadic attributes are represented by the colour of the nodes and dotted lines, respectively

Network property	Local configuration		Tendency for/against	References
Density	Edges		presence of ties	Hofman and Hoogland (2011)
Degree	Stars		being in contact with multiple partners	Hofman et al. (2011)
Connectedness	Two-stars		presence of intermediaries regulating contacts	Knippenberg (2007)
	Alternating k-two-paths			
Cohesion	Triangles		sites to share contacts	Blake (2014); Boomert (2000); Keegan (2007)
	Alternating k-triangles			
Node covariate	Covariate activity		sites with a certain attribute to have more contacts (e.g. producer sites are more likely to have more contacts than consumer sites)	Hofman et al. (2014)
	Homophily		sites having the same attribute (e.g. cultural affiliation) to be in contact	Rouse (1992); Keegan (2004); Hardy (2008); Hofman et al. (2011); Keegan and Hofman (2017)
Dyadic covariate	Edge covariate		sites with a certain dyadic attribute to be in contact (e.g. sites close to each other are more likely to be in contact)	Cody (1990); Rodriguez Ramos (2011)

**Table 3** Network properties, local configurations, interpretation and references to corresponding archaeological theories for directed relations. Node and dyadic attributes are represented by the colour of the nodes and dotted lines, respectively

Network property	Local configuration		Tendency for/against	References
Density	Arcs		presence of ties	Hofman and Hoogland(2011)
In-degree	In-stars		sites to be popular	Hofman et al. (2011)
Out-degree	Out-stars		sites to be active	Hofman et al. (2011)
Reciprocity	Reciprocal dyads		reciprocal interaction between sites	Renfrew (1975); Siegel (1992)
Connectedness	Two-stars			
	Alternating k-two-paths		presence of intermediaries	Renfrew (1975)
Cohesion	Triangles			
	Alternating k-triangles		the interaction partner of a site's interaction partner to be also the interaction partner of that site	Hage and Harary(1991); Mol(2014)
	3-cycles			
Node covariate			indirect reciprocity	Hage and Harary(1991)
	Covariate popularity		sites having a certain attribute to be more popular	Polanyi (1963); Renfrew (1975); Fitzpatrick (2004); Hofman et al. (2014)
	Covariate activity		sites having a certain attribute to be more active	
	Homophily		sites having a same attribute (e.g. cultural affiliation) to be in contact	Rouse (1992); Keegan (2004); Hardy (2008); Hofman et al. (2011); Keegan and Hofman(2017)
Dyadic covariate	Edge covariate		sites with a certain dyadic covariate to be in contact (e.g. sites close to each other are more likely to be in contact)	Evans et al. (2009)

### Endogenous Statistics

Endogenous statistics model the dependence of ties on the existence of other ties. The constituent elements of a network are ties, thus the most elementary endogenous



statistic counts the number of edges and pertains to the network density. The *edge* statistic describes the propensity of nodes to form ties. However, some nodes might be more prone to form ties than others. This tendency is captured by the degree distribution, with the degree being the number of ties incident to a node. The degree distribution is described by the *k-star* statistic, which counts the number of nodes connected to other  $k$  nodes. The number  $k$  ranges between 1 and  $n-1$ , albeit only 2-stars and 3-stars are used in practice. A statistic accounting for all the  $k$ -stars simultaneously is the *alternating-k-star* statistic, defined as a weighted sum of the  $k$ -star statistics. The  $k$ -star and alternating- $k$ -star statistics capture the tendency of sites to be in contact with multiple partners and are measures of centralisation. For the directed case (e.g. exchange or flow of goods), a distinction between incoming ties and outgoing ties is operated, leading to the *in-degree* and *out-degree* statistics.

Connectedness refers to the existence of paths between any pair of nodes and assesses (direct and indirect) reachability. The *two-path* statistic is used to model this network property. It counts the number of two-paths, sequences of two ties connecting two nodes through an intermediary node. To account for multiple intermediaries between any pair of nodes, the *alternating-k-two-path* statistic is used. This statistic is a weighted sum of counts of  $k$ -two-paths, with  $k$  being the number of intermediaries between two nodes. Two-paths are used, e.g. to account for the “middle-man” assumption, supposing the presence of broker or intermediary sites that mediate the contacts between other sites. For the directed case, two-paths are defined as a sequence of a tie from node  $i$  to node  $j$  and a tie from node  $j$  to a third node  $h$ , with  $j$  being the intermediary node.

Cohesion is another important network property. It refers to the tendency of ties to cluster together. The simplest configuration representing clustering is a triangle. In many networks, triangles tend to cluster as well, forming clumps modelled by the *alternating-k-triangle* statistic. Whilst the *triangle* statistic simply counts the triangles in a network (i.e. the number of times two connected sites are tied to a same other node), the *alternating-k-triangle* statistic accounts for the number of neighbours shared by two connected nodes. The use of these statistics is based on the proposition that ties are formed between sites that are jointly connected to at least a third site. An example of this assumption is that sites tend to form ties within social groups, where the site's contacts are connected to each other.

For directed relations, several types of triangles can be defined, among them *transitive triads* and *three-cycles*, which both describe the extent to which existing two-paths in a network are closed. Transitive triads refer to the tendency of nodes that are indirectly connected through a third node to directly connect. For instance, they might be used to represent the proposition that the exchange partner of a site's exchange partner is also the exchange partner of that site. Three-cycles are an undirected form of reciprocity. The assumption that the flow from site  $i$  to site  $j$  is returned through a third site  $h$  translates into a three-cycle configuration. Due to this interpretation, the transitive triad and three-cycle statistics are usually used jointly to reconstruct networks characterised by hierarchy differences among nodes. Indeed, the tendency towards transitive triads and against three-cycles indicates that certain nodes have more prominent positions than others.

## Exogenous Statistics

Exogenous statistics model the dependence of ties on monadic or dyadic site attributes.

The *covariate-activity* statistic allows to control for the dependence of the degree on node attributes. In the directed case (*e.g.* exchange), this statistic can be used to model the tendency of supplier sites to have more outgoing ties than consumer sites do.

Homophily, a mechanism referring to the similarity of connected nodes, is another classic example (McPherson et al. 2001). Two sets of statistics are available to model homophily: one counting the number of ties between nodes having the same characteristics and another counting the number of ties between pairs of nodes having different characteristics. Both statistical counts are suitable, for instance, to model the assumption that sites with the same cultural affiliation are more likely to be in contact.

Another set of statistics depends on dyadic attributes, *i.e.* the characteristics of pairs of nodes, such as geographical proximity. The corresponding statistic is a sum of ties weighted by the distance between the nodes and is used to account for the role that distance plays in regulating the existence of ties. Therefore, a widespread, because intuitive, proposition that ties between closer sites are more likely, can be modelled using this statistic.

Many other statistics can be defined as interaction effects among the statistics. An example is the interaction of two-paths with geographical distance. This interaction, as shown in one of the illustrative examples in the “[Example: an Application to the Pre-colonial Caribbean](#)” section, allows to account for the assumption that intermediary sites act on a local scale.

## ERGMs for Network Reconstruction

In this section, we describe the steps needed to reconstruct archaeological networks using ERGMs.

Given a certain relation and a set of theories concerning the mechanisms regulating the formation of ties, the first step of the procedure consists of fully specifying the model, *i.e.* choosing the statistics and the values of the corresponding parameters. The choice of the statistics requires matching the archaeological propositions to the corresponding local configurations. The previous section, jointly with Table 1 and Table 2, provides examples of this correspondence.

The values of the parameters determine how strong the tendency towards or against a specific proposition is. In general, positive (negative) values of a parameter lead to networks with high (low) values of the corresponding statistic, thereby indicating tendencies for (against) the associated proposition. However, certain combinations of parameter values (*e.g.* large positive or negative values) might lead to unrealistic network reconstructions corresponding to almost complete or empty networks. This phenomenon has been investigated in several disciplines, and it is referred to as *near-degeneracy* in statistics (Chatterjee and Diaconis 2013) and in the social science (Handcock et al. 2003; Snijders 2002; Snijders et al. 2006), and *phase transition* in physics (Newman 2003, 2018a). It follows that the choice and the calibration of the

parameters are fundamental in order to avoid near-degeneracy and reconstruct networks of archaeological interest.

To avoid the pitfall of specifying a degenerate distribution concentrated only on the empty or the complete network, the following procedure was used. The parameter values of all the statics were fixed on values derived from the network literature, and then tuned so that the simulated networks have structural characteristics coherent with the archaeological evidence. A similar method was used by Amati et al. (2018), but the initial values of the parameters were obtained by estimating a fully specified ERGM on a network reconstructed using one of the previous models (e.g. gravity model or ariadne), and then tuning the values obtained according to the available archaeological information and the density of the resulting networks.

The second step of the procedure aims at determining a plausible network reconstruction. For a fully specified ERGM, which is assumed to be a good representation of the processes thought to have generated networks in the past, it is natural to look for the most likely network(s) as a plausible reconstruction. Thus, the second step of the procedure consists of finding the network(s) that maximises the linear combination of statistics and parameters defined by  $\sum_k \theta_k s_k(x, v, w)$ .

Due to the large number of networks that can be defined on the set of nodes  $\mathcal{N}$ , the maximisation of this function is difficult except in some trivial cases (e.g. when all the parameters are positive or negative). The solution of the maximisation problem can be approximated by using simulated annealing (Metropolis et al. 1953; Kirkpatrick et al. 1983), an algorithm that applies to the optimisation of a function on a finite set with very large size and owes its name to the annealing process in metallurgy. This algorithm avoids the trapping attraction of the local maxima of the function that needs to be maximized by scaling the function by a parameter  $T$ , named temperature.

Given an initial value of the temperature  $T > 0$  and an initial network  $x$ , at each step of the algorithm, a tie  $x_{ij}$  is selected uniformly at random and a change is suggested: If the tie  $x_{ij}$  is present in the network, the proposed change is the termination of the tie. Conversely, if the tie  $x_{ij}$  is absent, the proposed change is the creation of the tie. We denote the network resulting from the suggested change by  $x'$ . If the proposed change increases the value of  $\sum_k \theta_k s_k(x, v, w)$ , the change is accepted and the new state of the network is  $x'$ ; otherwise, it is accepted with a probability  $p(x, x')$  that is proportional to

$$p(x, x') \propto \exp \left\{ \frac{\sum_k \theta_k [s_k(x', v, w) - s_k(x, v, w)]}{T} \right\}. \quad (3)$$

At each step, the temperature is decreased by a small factor and the algorithm stops when the value of  $T$  is close to 0. The last network is the most likely network (or one of the most likely networks) according to the specified ERGM. The intuition behind Eq. (3) is that the smaller the change in the linear combination of the statistics and parameters,  $\sum_k \theta_k [s_k(x', v, w) - s_k(x, v, w)]$ , and the higher the temperature, the more likely it is for the algorithm to accept the proposed network  $x'$  as the next state, even though  $x'$  is a worse solution than  $x$ . This procedure allows the algorithm to jump to different regions of the network space and therefore to search for the optimum in the

entire space. When  $T$  is decreased, the acceptance probability decreases; therefore, the search for the optimum concentrates in a more localised region.

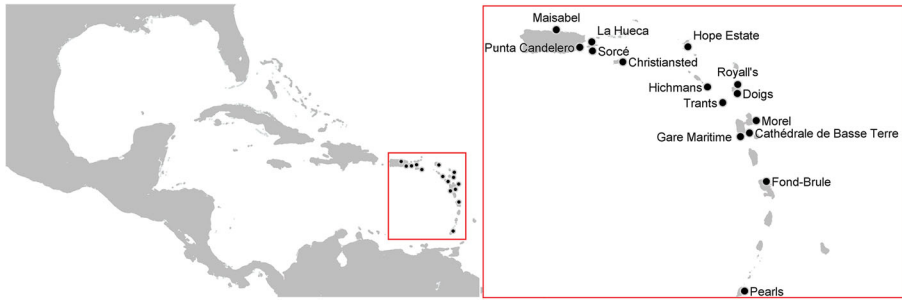
For some ERGM specifications, the rationale behind choosing the most likely network does not rely merely on probabilistic theory, but it is motivated by a micro-foundation of ERGMs rested on principles of game theory (Butts 2009; Mele 2017). According to this derivation, a network is the outcome of a network formation game (Goyal 2012; Jackson 2010) in which pairs of nodes decide to create or sever ties based on a pay-off expressing the reward of the ties. The pay-off is defined as a linear combination of statistics, counting the number of configurations involving the tie considered, and parameters, representing the trade-off between the costs and benefits of a tie when it is part of the network configuration corresponding to that parameter. Under certain conditions (Monderer and Shapley 1996; Vega-Redondo 2003; Butts 2009; Mele 2017), the limiting distribution of the network formation game is the ERGM and the most likely networks are in equilibrium, that is they are networks in which none of the nodes would like to sever an existing tie or create a non-existing tie. Thus, the reconstructed networks can be thought of as attractive network configurations that would have arisen if sites had striven to form rewarding ties according to the specified ERGM.

### Example: an Application to the Pre-colonial Caribbean

To illustrate the framework we have described in the previous sections, we used data collected over a group of sites located in the north-western Greater Antilles and in the southern Lesser Antilles and we considered archaeological propositions concerning connectivity, inter-cultural contacts and exchange among those sites. The networks were generated using the simulated annealing algorithm described in the “[ERGMs for Network Reconstruction](#)” section. The initial parameters of temperature  $T$  was fixed at 6 and its value was decreased at each step by a factor of 0.9. The steps were repeated until  $T < 0.00001$ . Several runs of the algorithm were performed to check that the obtained networks were the most likely networks under the specified ERGM distribution.

The considered data set is composed of a list of 15 sites that have been (i) suggested to be places of permanent habitation as based on their excavation reports and (ii) securely dated to the period AD 100–400 as based on a regional standard for chronometric hygiene (Fitzpatrick 2006). Figure 1 illustrates the geographical region and the location of the 15 sites. We refer to Hofman et al. (2014) and Hofman et al. (2019) for an in-depth description of those sites.

The period AD 100–400, hereafter referred to as the Archaic-Early Ceramic Interface (AECI) period, marks the end of a transitional process in Caribbean culture history, notably in the northern Lesser Antilles. Traditionally speaking, this marked the end of the so-called Archaic Age lifestyle, mobile hunter-gatherers living in small social units, and the advent of the Ceramic Age. However, the neolithisation process in the Caribbean, like elsewhere in the world, happened in a much more gradual way and nearly every aspect that was once argued to be introduced as part of the Ceramic Age package, from plant management to clearly articulated ceremonial life and the production of ceramics, already existed during the Archaic Age (Hofman et al. 2014, 2018,



**Fig. 1** Map illustrating the geographical region and location of the sites in the data set considered. The 15 sites have been (i) suggested to be places of permanent habitation as based on their excavation reports and (ii) securely dated to the period AD 100–400 as based on a regional standard for chronometric hygiene (Fitzpatrick 2006). The red square denotes the area that appears in the subsequent figures

2019). In short, the exact timing of the introduction of certain materials and practices remains hotly debated based on two models: (i) migration by which the incoming people displaced the former inhabitants and (ii) a pan-Caribbean network in which newcomers interacted with the original inhabitants and made use of existing networks. These models are furthermore based on two common traits: (i) some form of transition took place during the AECl and (ii) social networks—or their inverse (deliberate) disconnectedness—as conduits for the shift in social and cultural practices. A major touchstone in this debate is the exchange of chert and semi-precious stones raw materials, half-fabricates or finished objects (Boomert 2000; Cody 1990, 1993, Hofman et al. 2007, 2014, 2019; Keegan 2007; Rodríguez Ramos 2007).

The 15 sites selected for this study all date to the AECl. They are all characterized by ceramics of the Saladoid, Huecoid or a mix of both series (Rouse 1992) and of one or more of the named chert or semi-precious materials. The corresponding data set (see Table 1 in the supplementary material) includes information on several attributes of these sites, among them the location, the cultural affiliation and the role played in the distribution of lithic material.

The location of sites is determined by the latitude and longitude and used to compute the geographical distances among the sites. The 15 sites are located between Puerto Rico in the north-western Greater Antilles and Grenada in the southern Lesser Antilles. This area is partitioned into three sub-regions: the northern sub-region, with sites on Puerto Rico (Maisabel and Punta Candelero), Vieques (Sorcé and La Hueca), the US Virgin Island (Christiansted) and Saint Martin (Hope Estate); the central eastern sub-region with sites on Nevis (Hichmans), Antigua and Barbuda (Royall's and Doigs), Montserrat (Trants) and Guadeloupe (Morel, Gare Maritime and Cathédrale de Basse Terre); and the southern sub-region with sites on Martinique (Fond Brule) and Grenada (Pearls). We refer to Table 1 in the supplementary material and Fig. 1 for more details.

To establish the role played by the sites in material distribution, five semi-precious stones and cherts that circulated in the considered area were taken into account: Long Island flint, amethyst, serpentinite, carnelian and Saint Martin greenstone. Conditional on the presence of a lithic material, a site was then classified as a supplier (site with lithic workshops), a supplier/intermediate, a

consumer/intermediate or a consumer (site without evidence of stone working) based on the information deriving from studies of the lithic assemblage (Knippenberg 2007; Rodríguez Ramos 2007). Other information comes from excavation of the sites (see Hofman et al. (2014) and Hofman et al. (2019)). According to the quantity of finds, sites were classified into three categories: sites with a small amount (3 sites), a medium amount (6 sites) and a large amount (6 sites). Following Rouse's classification (1992) and the composition of the ceramic assemblages, the cultural affiliation of each site has been coded into four categories: Saladoid (5 sites), Huecoid (2 sites), Saladoid and Huecoid (4 sites), Huecoid and Saladoid (4 sites).

To illustrate how ERGMs can be used to reconstruct networks between the 15 sites mentioned above, we present three different model specifications and we provide a qualitative assessment of the reconstructed networks. This assessment evaluates the coherence of the structure of the generated networks with archaeological evidence that is not directly accounted for by the model specification. In particular, we test whether certain model specifications are able to explain the presence of sites known to have functioned as hubs, *i.e.* major community gathering sites expected to be well (directly) connected to the other sites. Consequently, we consider the degree centrality of a site, *i.e.* the number of ties incident to that site, as a measure of connectedness. We expect that, in a plausible network reconstruction, the sites having higher degree must correspond to the hubs.

### Proximity Model

Many of the propositions aiming to explain connectedness underline the importance of geographical space in both formation and maintenance of network ties. Certain ways of thinking consider islands as more bounded spaces, in which connectivity between sites that are close to each other are more likely. An example of this is the “island-hopping” model (Rouse 1992) according to which movements across the Caribbean were based on a sequence of short journeys between islands. Other lines of thought consider the Caribbean Sea to have functioned as a more unbounded connector where single journeys from one destination to another took place, affording higher connectivity to all sites in the region (Keegan and Hofman 2017; Torres and Rodríguez Ramos 2008; Watters 1997). It should be noted that none of the existing theories hold extreme positions on the subject, *i.e.* either complete boundedness or complete connectivity.

Table 4 illustrates the model specification for reconstructing connectivity based on the assumption described above and on the statistics in the first and last rows of Table 2. The edge statistic counts the number of ties in the network. Large and negative values of the corresponding parameter  $\theta_1$  generate sparse networks, whilst large and positive values of  $\theta_1$  generate dense networks. Propositions concerning geographical proximity are modelled by the edge covariate statistic since distance between sites is a dyadic attribute. The edge covariate statistic measures the total distance spanned by all the edges presented in the network. The corresponding parameter  $\theta_2$  measures the impact of distance on the existence of ties: positive

**Table 4** ERGM specification for reconstructing connectivity. In the formulas,  $x_{ij} = 1$  if there is a tie between  $i$  and  $j$ , and 0 otherwise;  $d_{ij}$  denotes the distance between sites  $i$  and  $j$

Local configuration	Parameter ( $\theta_k$ )	Statistic ( $s_k$ )
Edges	$\theta_1$	$\sum_{ij} x_{ij}$
Edge covariate (distance)	$\theta_2$	$\sum_{ij} x_{ij} \log(d_{ij})$

values of  $\theta_2$  indicate that long-distance ties are more likely to occur as the log-odds for the model in Table 4 are equal to  $\theta_1 + \theta_2 \cdot \log(d_{ij})$ .

The ERGM specification in Table 4 corresponds to a spatial inhomogeneous Bernoulli random graph model (Butts 2002) which assumes that the likelihood of a tie depends on the distance according to an attenuated power law function having the form

$$P(X_{ij} = x_{ij} | D_{ij} = d_{ij}) = \frac{1}{1 + e^{-\theta_1 - \theta_2 \log(d_{ij})}} = \frac{1}{1 + \alpha d_{ij}^\gamma}, \quad (4)$$

with  $\alpha = e^{-\theta_1}$  and  $\gamma = -\theta_2$ . This model assumes dyadic independence and is a generalization of maximum distance networks (MDNs) (Evans et al. 2012). In MDNs, a tie is present if the distance between two sites is less than a threshold distance  $D$  far apart. Thus, the probability of a tie conditional on the distance is either 0 or 1 and is described by a degenerate distribution. The model in Table 4, in contrast, assigns a probability to each tie according to a decreasing function of the distance.

The left-hand side of Fig. 2 shows the most likely networks generated according to the model specification above and different values of the parameter  $\theta_2$ . The initial values of the parameters ( $\theta_1 = 10.98$  and  $\theta_2 = -1.91$ ) were determined by imposing that the probability of a tie between sites 100 km (maximum daily travelling distance) far apart was equal to 0.9, whilst the probability of a tie between sites 1000 km far apart was small and equal to 0.1. The parameter  $\theta_2$  was tuned to represent the archaeological propositions concerning site proximity and reachability. An increase of  $\theta_2$  leads to networks characterized by the presence of long-distance ties,  $\theta_1$  being equal. Therefore, when  $\theta_2$  is large in absolute value and negative, the most likely networks are in line with the insularity and isolationism scenario. When  $\theta_2$  is positive or small in absolute value and negative, the inter-island interaction scenario is more likely. These results are justified by the fact that high values of  $\theta_2$  increase the likelihood of long-distance ties as described by Eq. (4) and visualized in the networks on the right-hand side of Fig. 2. In these networks, the colour and the size of the edges represent the probability of ties. The darker and the thicker an edge, the more likely the connectivity between two sites.

Due to the diverse structures of the reconstructed networks, Fig. 2a–c on the left-hand side differ also in the presence of hubs as suggested by differences in the size of the nodes (that are proportional to the number of incident ties) across the networks. In particular, in Fig. 2a, the hubs are the sites that belong to the largest connected component and are geographically closer to many other sites, *i.e.* Royall's, Doigs and Trants. In Fig. 2b, when the assumption on distance is relaxed, Hope Estate is the most central site since it is geographically closer to many other sites. In Fig. 2c, none of the

sites has a prevalent role with respect to the others since the resulting network is almost a complete network.

One of the hypotheses related to connectivity concerns the presence of intermediary sites leading to sequences of contacts as described by the “down the line” model (Knippenberg 2007). Generating networks coherent with this proposition requires the use of an interaction effect between two-paths and distance, so that intermediary sites act on a local scale, as shown by the model specification reported in Table 5. The local two-path statistic expresses the idea that the existence of ties between sites far apart depends on the presence of other ties connecting intermediary sites, thereby implying tie dependence. This statistic counts the number of two-paths such that connected sites are located in the same neighbourhood. For this example, we defined the neighbourhood of a site as the set of sites that are less than 200 km far apart. Large and positive values of the corresponding parameter  $\theta_3$  generate networks coherent with the down the line model, whilst large and negative values of  $\theta_3$  generate networks with no intermediary sites.

An example of network coherent with the down the line model is shown in Fig. 3. Compared to the network in Fig. 2b, the presence of intermediary sites bridging the connections between far-away sites reduces the number of long-distance ties. For instance, there are no more direct ties from Hope Estate and Gare Maritime to Morel and Cathédrale de Basse Terre due to the presence of the intermediary sites Trants, Royall’s and Doigs. In this network, the most central sites (e.g. Hope Estate, Trants and Morel) are located in the central eastern sub-region as denoted by the size of the corresponding nodes.

### Proximity and Inter-cultural Model

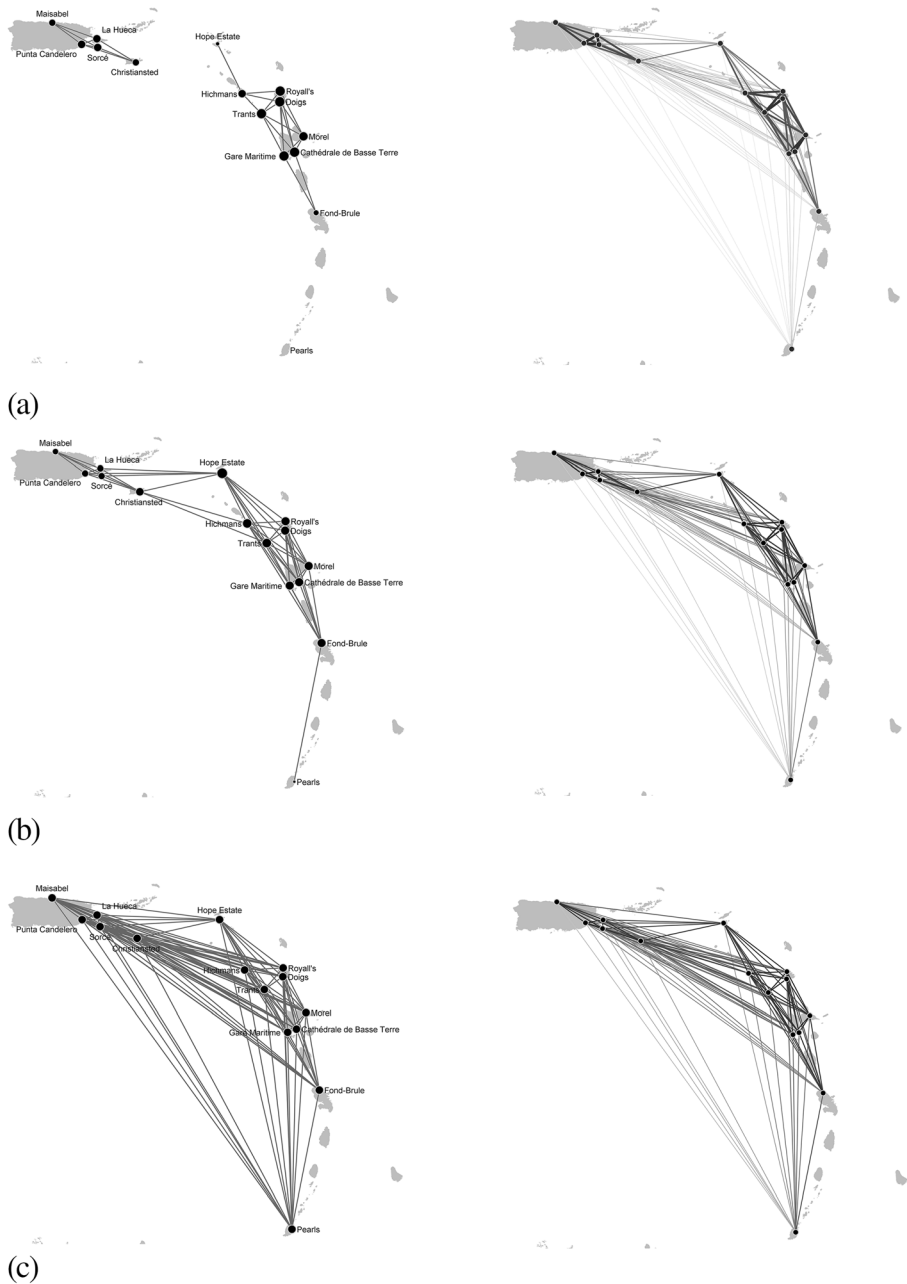
In Caribbean Archaeology, the type of theories that argue for lower or no cohesion between sites focusses on the role of cultural boundaries constricting contacts in AECI networks. Different pottery decorations and ceramic assemblages are used as evidence for the presence of different groups of people and cultures, who competed with and culturally supplanted each other. Long-range network contacts would only exist with groups with the same culture or, in Rouse’s (1992) terminology, within the same “people”.

Several archaeologists explained the presence of bounded groups with different migration waves into the Caribbean islands (see, for instance, Veloz Maggiolo (1991)

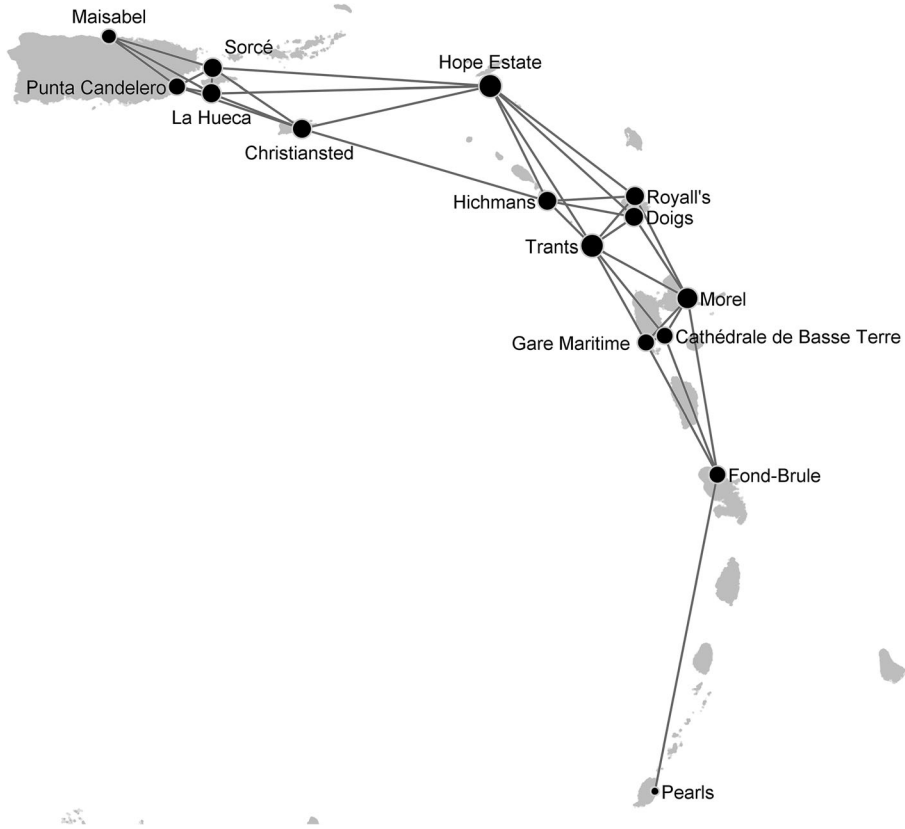
**Table 5** ERGM specification for reconstructing connectivity in networks with intermediaries and cohesive sub-groups of sites. In the formulas,  $d_{ij}$  denotes the distance between sites  $i$  and  $j$ , whilst  $I_{\{i,j \text{ neighbours}\}}$  is an indicator function, taking value 1 if  $i$  and  $j$  are in the same neighbourhood, and 0 otherwise. We defined the neighbourhood of a site as the set of sites that are less than 200 km far apart from that site

Local configuration	Parameter ( $\theta_k$ )	Statistic ( $s_k$ )
Edges	$\theta_1$	$\sum_{ij} x_{ij}$
Edge covariate (distance)	$\theta_2$	$\sum_{ij} x_{ij} \log(d_{ij})$
Local two-path	$\theta_3$	$\sum_{ijh} x_{ij} I_{\{i,j \text{ neighbours}\}} x_{jh} I_{\{j,h \text{ neighbours}\}}$





**Fig. 2** Networks reconstructed using the proximity model in Table 4. Parameter values: **(a)**  $\theta_1 = 10.98$ ,  $\theta_2 = -2.20$ ; **(b)**  $\theta_1 = 10.98$ ,  $\theta_2 = -1.91$ ; **(c)**  $\theta_1 = 10.98$ ,  $\theta_2 = -1.60$ . The size of the nodes is proportional to the node degree, *i.e.* the number of incident ties to a node. The initial value of temperature  $T$  was fixed at 6 and decreased by a factor of 0.9 at each step. The step of the simulated annealing algorithm was repeated until  $T < 0.00001$



**Fig. 3** Network reconstructed using the model in Table 5. Parameter values:  $\theta_1 = 9.0$ ,  $\theta_2 = 0.5$ ,  $\theta_3 = -1.6$ . The size of the nodes is proportional to the node degree, *i.e.* the number of incident ties to a node. The initial value of temperature  $T$  was fixed at 6 and decreased by a factor of 0.9 at each step. The step of the simulated annealing algorithm was repeated until  $T < 0.00001$

and Zucchi (1990)). At the core of this is the idea that one or more peoples migrated into the northern Lesser Antilles in the first millennium BC and their movement was limited by the social distance represented by diverse cultural affiliations. This social distance led to the formation of strictly bounded social groups, as reflected in ceramic assemblages, with their own regional and inter-regional contact networks. These only gradually converged in the centuries after the migration had taken place.

In opposition to these older ideas are several theories that consider the AECI to be a period of networks operating at a pan-Caribbean scale with a high degree of inter-regional mobility and multi-cultural communities, despite differences in culture affiliation (*e.g.* Hofman et al. 2010, 2011; Rodríguez Ramos et al. 2010). Whilst these theories are probably better suited to understand an archaeological record that showcases a unity in diversity (Mol 2014), many are vested in archaeologically unobservable or unspecified effects, such as family ties or the unspecified concept of “network” (Hardy 2008; Keegan 2004; Keegan and Hofman 2017).

None of the previous models for network reconstruction account for presence or absence of cultural boundaries, and therefore cannot be used in this context.

**Table 6** An example of ERGM specification for reconstructing inter-cultural networks. The node attribute  $v$  represents the cultural affiliation. In the formulas,  $d_{ij}$  denotes the distance between sites  $i$  and  $j$ , whilst  $I_{\{v_i=a, v_j=b\}}$  is an indicator function, taking value 1 if the cultural affiliation  $v_i$  of site  $i$  is  $a$  and the cultural affiliation  $v_j$  of site  $j$  is  $b$

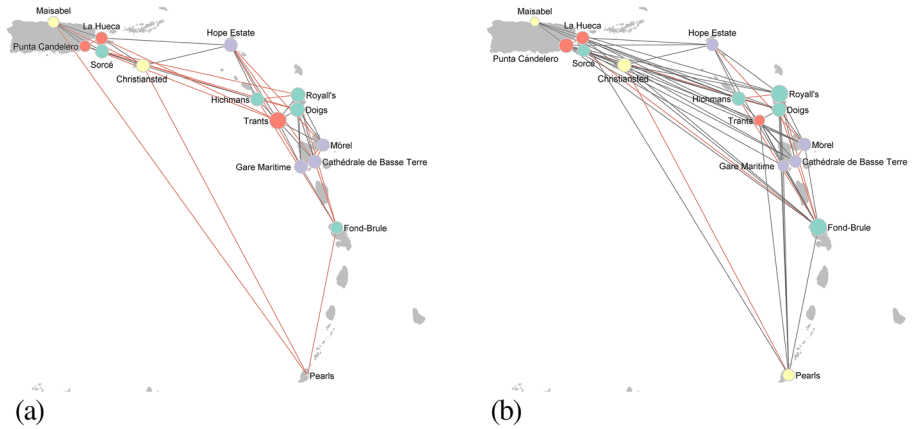
Local configuration	Parameter ( $\theta_k$ )	Statistic ( $s_k$ )
Edges	$\theta_1$	$\sum_{ij} x_{ij}$
Edge covariate (distance)	$\theta_2$	$\sum_{ij} x_{ij} \log(d_{ij})$
Homophily (cultural similarity)	$\theta_3$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid}, v_j=\text{Saladoid and Huecoid}\}}$
	$\theta_4$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid}, v_j=\text{Huecoid and Saladoid}\}}$
	$\theta_5$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid}, v_j=\text{Huecoid}\}}$
	$\theta_6$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid and Huecoid}, v_j=\text{Huecoid and Saladoid}\}}$
	$\theta_7$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid and Huecoid}, v_j=\text{Huecoid}\}}$
	$\theta_8$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Huecoid and Saladoid}, v_j=\text{Huecoid}\}}$

Conversely, ERGMs provide a set of node covariate statistics to model a more general notion of (dis) similarity in cultural affiliation. Those statistics are added to those present in Table 4 to control for geographically proximity. Thus, this example demonstrates that ERGMs can account for more propositions at the same time.

Let  $v$  be the variable describing the cultural affiliation of the sites. This variable takes four categories: Saladoid, Huecoid, Saladoid and Huecoid, Huecoid and Saladoid. Table 6 reports one possible model specification to represent networks coherent with the assumptions on cultural homophily and distance. Using the heuristic interpretation of ERGMs, the corresponding parameters  $\theta_3, \dots, \theta_8$  can be interpreted as the change in the log-odds of a tie being present between sites having a different cultural affiliation and sites having the same cultural affiliation. Thus, a model based on theory of cultural boundaries is characterized by negative values of those parameters, whilst a model coherent with the theory of inter-cultural contacts is characterized by positive values of those parameters. In fact, positive values of the parameters  $\theta_3, \dots, \theta_8$  lead to a positive contribution to the log-odds in Eq. (2), thereby suggesting that ties between sites having different cultural affiliation are more likely to be present than absent.

Reconstructed networks for different values of the parameters  $\theta_3, \dots, \theta_8$  are shown in Fig. 4. Both networks are characterized by the presence of short-distance ties due to the negative values of the distance parameter. However, whilst Fig. 4a provides a picture of a network coherent with Rouse's (1992) hypothesis of the presence of cultural boundaries, Fig. 4b shows a plausible network characterized by absence of cultural boundaries. The proportion of ties between sites having a different cultural affiliation (grey ties in Fig. 4) is indeed 0.41 for the network in Fig. 4a, and 0.77 for the network in Fig. 4b.

The two networks differ also in their structures. The network in Fig. 4b is denser and characterized by the presence of long-distance ties. In Fig. 4a, Trants, Royall's and Doigs are the most central sites since they are located close to other sites having a similar cultural affiliation. Conversely, Fund-Brule, Royall's and Punta Candelerero are the most important sites in Fig. 4b obtained by assuming cultural affiliation heterophily.



**Fig. 4** Network reconstructed using the inter-cultural model in Table 3. Parameter values: **(a)**  $\theta_1 = 2, \theta_2 = -0.3, \theta_3 = -0.3, \theta_4 = -0.6, \theta_5 = -0.9, \theta_6 = -0.3, \theta_7 = -0.6, \theta_8 = -0.6$ ; **(b)**  $\theta_1 = 2.0, \theta_2 = -0.3, \theta_3 = 0.1, \theta_4 = 0.2, \theta_5 = 0.3, \theta_6 = 0.1, \theta_7 = 0.2, \theta_8 = 0.2$ . The colour of the nodes represents their cultural affiliation: Saladoid (green), Saladoid and Huecoid (yellow), Huecoid and Saladoid (violet), Huecoid (pink). The colour of the ties indicates whether a tie exists between two sites having the same cultural affiliation (red) or having a different cultural affiliation (grey). The size of nodes is proportional to the node degree, *i.e.* the number of incident ties to a node. The initial value of temperature  $T$  was fixed at 6 and decreased by a factor of 0.9 at each step. The step of the simulated annealing algorithm was repeated until  $T < 0.00001$

### Proximity, Inter-cultural and Exchange Model

We consider now the reconstruction of exchange networks to provide an example of ERGM specification for directed relations and introduce effects that have not been considered so far. The distribution of semi-precious stones and cherts that circulated in the considered area is the outcome of exchange relations from one site, referred to as the “sender”, to another site, referred to as the “receiver”. Therefore, exchange relations are characterized by directionality.

A simple model for reconstructing exchange is specified by assuming that exchange ties depend on the proximity between sites (Fitzpatrick 2004), similarity in material culture and the role played by the sites in the distribution of the lithic material. In particular, Hofman et al. (2014) suggested that supplier sites (*i.e.* sites with lithic workshops) are more likely to be senders, whilst consumer sites (*i.e.* sites without evidence of stone working) are more likely to be receivers.

These assumptions are included in the model by considering the covariate activity and covariate popularity statistics as shown in Table 7. Let  $u$  and  $z$  be the variables describing the number of lithic sources of which a site is a supplier and a consumer, respectively. The covariate activity and covariate popularity statistics measure the number of outgoing ties for supplier sites and the number of incoming ties for consumer sites, respectively. Thus, positive values of  $\theta_3$  indicate that supplier sites tend to have more outgoing ties than consumer sites, and positive values of  $\theta_4$  indicate that consumer sites tend to have more incoming ties than supplier sites.

To account for the presence of intermediary sites and the assumption that those sites might have both outgoing and incoming ties, we defined a variable  $w$  describing the

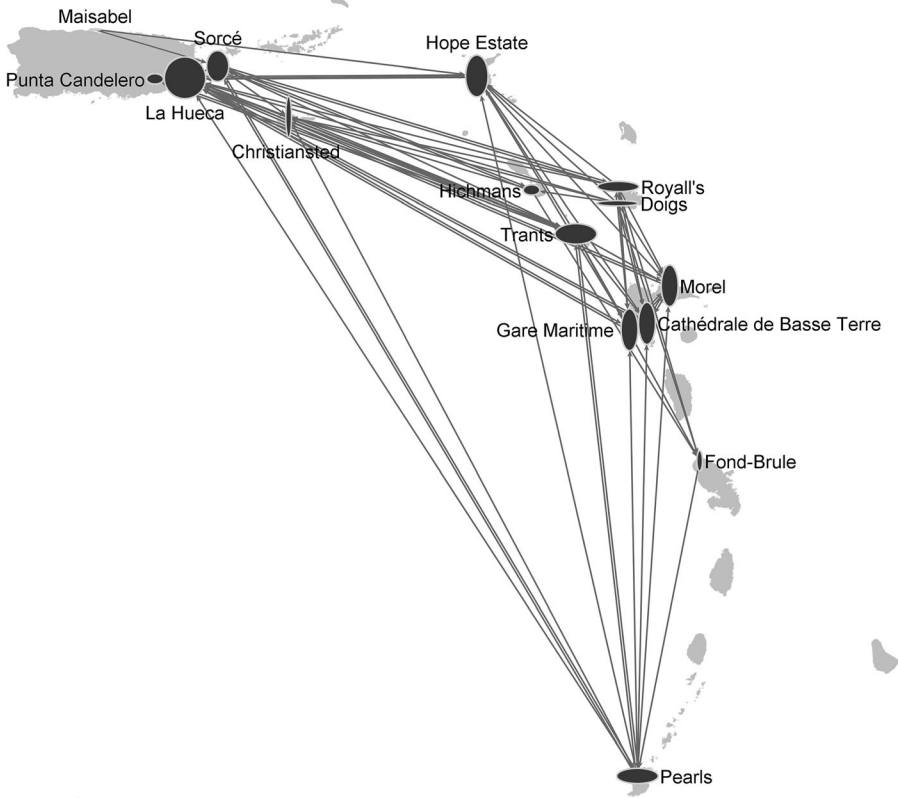
**Table 7** An example of ERGM specification for reconstructing exchange networks based on distance ( $d_{ij}$ ), material culture ( $v$ ), the role of sites in the redistribution of lithic material ( $z$  for supplier,  $u$  for consumer and  $w$  for intermediate), the quantity of finds in a site ( $q$ ) and the directionality of the exchange from supplier to consumer sites ( $s_{ij}$ )

Local configuration	Parameter ( $\theta_k$ )	Statistic ( $s_k$ )
Edges	$\theta_1$	$\sum_{ij} x_{ij}$
Edge covariate (distance)	$\theta_2$	$\sum_{ij} x_{ij} \log(d_{ij})$
Covariate activity	$\theta_3$	$\sum_{ij} x_{ij} z_i$
Covariate popularity	$\theta_4$	$\sum_{ij} x_{ij} u_i$
Covariate activity	$\theta_5$	$\sum_{ij} x_{ij} w_i$
Covariate popularity	$\theta_6$	$\sum_{ij} x_{ij} w_i$
Covariate activity	$\theta_7$	$\sum_{ij} x_{ij} q_i$
Covariate popularity	$\theta_8$	$\sum_{ij} x_{ij} q_j$
Edge covariate (source)	$\theta_9, \dots, \theta_{13}$	$\sum_{ij} x_{ij} s_{ij}$
Homophily (cultural similarity)	$\theta_{14}$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid}, v_j=\text{Saladoid and Huecoid}\}}$
	$\theta_{15}$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid}, v_j=\text{Huecoid and Saladoid}\}}$
	$\theta_{16}$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid}, v_j=\text{Huecoid}\}}$
	$\theta_{17}$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid and Huecoid}, v_j=\text{Huecoid and Saladoid}\}}$
	$\theta_{18}$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Saladoid and Huecoid}, v_j=\text{Huecoid}\}}$
	$\theta_{19}$	$\sum_{ij} x_{ij} I_{\{v_i=\text{Huecoid and Saladoid}, v_j=\text{Huecoid}\}}$

number of lithic sources of which a site is an intermediary. We then included the corresponding covariate activity and covariate popularity statistics in the model in Table 7. Following the same reasoning, we added to the model the covariate activity and covariate popularity statistics for the variable  $q$  describing the quantity of finds in a site. The corresponding parameters  $\theta_5$  and  $\theta_7$  are interpreted as the parameter  $\theta_3$ , whilst the parameters  $\theta_6$  and  $\theta_8$  are interpreted as the parameter  $\theta_4$ .

Finally, to account for the exchange flow, we assumed that ties between suppliers and consumers of the same lithic material were more likely than ties between sites being both suppliers or both consumers of the same lithic material. This assumption was incorporated in the model by using an edge covariate statistic for each lithic material. The edge covariate  $s_{ij}$  for a specific source takes value 1 if site  $i$  is a supplier and site  $j$  is a consumer, and 0 otherwise. This definition of the edge covariate allows to account for the directionality of the exchange from supplier to consumer sites. Positive values of parameters  $\theta_9, \dots, \theta_{13}$  suggest that ties from supplier to consumer sites of the same lithic material were more likely.

The most likely network coherent with the propositions of Hofman et al. (2014) is depicted in Fig. 5. The network is quite dense and characterized by long-distance ties suggesting that the Caribbean Sea has functioned as an unbounded connector where single journeys from one destination to another took place, affording higher connectivity to all sites in the region. The size of the node indicate the role played by the sites in the exchange network. Mainly supplier sites (*i.e.* sites that are suppliers of multiple lithic materials, such as Trants and Royall's suppliers of Long Island flint and Carnelian) have more outgoing ties than incoming ties and therefore are represented by nodes



**Fig. 5** Exchange network reconstructed using the model in Table 7. Parameter values:  $\theta_1 = -4.0$ ,  $\theta_2 = -0.1$ ,  $\theta_3 = 1.0$ ,  $\theta_4 = 1.0$ ,  $\theta_5 = 0.6$ ,  $\theta_6 = 0.6$ ,  $\theta_7 = 0.1$ ,  $\theta_8 = 0.1$ ,  $\theta_9 = \theta_{10} = \theta_{11} = \theta_{12} = \theta_{13} = 1$ ,  $\theta_{14} = -1.5$ ,  $\theta_{15} = -3$ ,  $\theta_{16} = -4.5$ ,  $\theta_{17} = -1.5$ ,  $\theta_{18} = -3$ ,  $\theta_{19} = -1.5$ . The width and the height of a node are proportional to its out-degree (*i.e.* the number of outgoing ties), and in-degree (*i.e.* the number incoming ties), respectively. The initial value of temperature  $T$  was fixed at 6 and decreased by a factor of 0.9 at each step. The step of the simulated annealing algorithm was repeated until  $T < 0.00001$

with width greater than height. On the contrary, mainly consumer sites (*i.e.* sites that are consumers of multiple lithic materials, such as Hope Estate—consumer of Long Island flint, amethyst, serpentinite and carnelian—and Morel—consumer of Long Island flint, amethyst, carnelian and Saint Martin greenstone) have more incoming ties than outgoing ties and therefore are represented by nodes with height greater than width. Rounded nodes denote sites that are both suppliers and consumers without any prevalence. The computation of the degree of the nodes (*i.e.* the sum of the in-degree and out-degree) suggests that the network is characterized by three main hubs La Hueca, Trants and Pearls which play the role of the three major community gathering sites in the northern, central-eastern and southern sub-regions.

A more complex model can be specified to account for hierarchy among the sites. As discussed in the “Statistics” section, networks coherent with this assumption can be obtained by adding the transitive triads and the 3-cycles statistics to the model specified in Table 7.

## Conclusion

The archaeological literature provides a large variety of assumptions concerning interaction mechanisms between archaeological contexts. Those assumptions have been used to infer the structure of past networks by specifying models that reflect the available archaeological knowledge.

In this paper, we considered tie-based models, specifically exponential random graph models (ERGMs) which offer a general framework that may be applied to infer the structure of ancient networks in diverse archaeological settings. Compared to previous models, the formulation of ERGMs does not hinge on the specific assumptions, the time period, the geographical area or the type of relation considered. Moreover, ERGMs enable the reconstruction of networks based on a large variety of propositions, ranging from assumptions based on dyadic independence to those assuming tie dependence and accounting for node or dyadic attributes. Thus, the application of ERGMs opens up the investigation of scenarios that cannot be explored by previous models.

The application of ERGMs to reconstruct archaeological networks is a probabilistic approach and contrasts with the deterministic approach of maximum distance networks, proximal point analysis, and gravity models. In those models, the presence of a relationship between two sites is defined by a rule stating which ties exist; therefore, the outcome of those models is fixed. The probabilistic approach of ERGMs, in contrast, has the advantage of assigning probabilities to the generate networks and thus to partially control for the incomplete information derived from the data.

The probabilistic approach of ERGMs requires two fundamental steps: (i) the specification of the model and (ii) the generation of plausible networks.

The model specification consists of choosing the statistics and the parameter values. The selection of the statistics is based on the archaeological assumptions and their encoding into local network configurations, *i.e.* graphical representations of nodes and ties. In Table 2 and Table 3, we provided a summary of the most common network configurations along with the corresponding ERGM statistics, and related them to some of the archaeological hypotheses that have been formulated when analysing relations among different archaeological contexts. The choice of the parameter values was based on both archaeological knowledge and the tuning procedure described in the “[ERGMs for Network Reconstruction](#)” section. Whilst the archaeological assumptions provide information on the sign of the parameter values (positive if there is a tendency for the tie mechanism implied by the assumption, but negative otherwise), the tuning avoids the generation of uninformative network structures, such as empty and full networks.

Given a fully specified ERGM, possible network scenarios are generated by maximising the corresponding probability distribution. Due to the high number of networks that can be defined over a set of nodes, the maximisation problem is difficult and thus an optimisation algorithm based on simulated annealing is used. The choice of considering the most likely network as a plausible reconstruction is justified by statistical principles and the derivation of the ERGM distribution from concepts of game theory. For some specifications, ERGMs are the limiting distribution of a process

in which nodes form beneficial ties as quantified by a pay-off function measuring the trade-off between the costs and benefits of ties. The maxima of this limiting distribution are desirable networks, *i.e.* configurations where none of the nodes would form a non-existing tie or sever an existing tie. Thus, the reconstructed networks can be thought of as attractive network configurations that would have arisen if sites had striven to form beneficial ties according to the specified ERGM.

Although, ERGMs provide a flexible method that can be applied to many archaeological contexts, the illustrated framework has some limitations.

Firstly, the inferred networks provide only one picture of the network in the past. Even if the generated networks can be interpreted as desirable network configurations emerging from a process in which sites form and sever ties according to their costs and benefits, the outcome of the framework is essentially static. Therefore, this approach cannot be used to investigate network evolution or the diffusion of practices and innovations. For this purpose, agent-based, network diffusion and dynamic network models are more suitable than ERGMs. Moreover, due to the difficulty of the optimisation problem, the networks generated are binary networks, indicating only if a tie was present or absent, and they do not provide any information on the strength of the ties among the sites.

Secondly, the procedure illustrated allows to compute the probability of a tie only in some trivial cases. For instance, in the illustrative example, we demonstrated that the likelihood of a tie can be computed using an attenuated power law function when connectivity is determined only by distance. However, when the specification of the ERGM includes triadic effects, such as the model accounting for intermediary sites, then the best we can do is compute tie probabilities conditional on the reconstructed network. We cannot provide the unconditional probability of a tie as we can in models for tie independence.

Finally, as with all the other models for network reconstruction, the networks resulting from ERGMs are sensitive to missing data. The structure of a reconstructed network might indeed vary when new sites are considered. The use of Bayesian procedures might offer a better approach to dealing with the incompleteness of the archaeological data. Bayesian statistics have been already used to model partially observed networks (Handcock and Gile 2010; Koskinen et al. 2010), to impute missing data in network studies (C. Wang et al. 2016), and to infer links given noisy or proxy data (Newman 2018b, 2018c). Due to the high level of uncertainty and incompleteness of the archaeological data, those approaches, albeit promising, need to be further developed to be used to reconstruct past networks.

We illustrated the applicability of the framework by reconstructing networks between 15 sites located between Puerto Rico, in the north-western Greater Antilles, and Grenada, in the southern Lesser Antilles, during the period AD 100–400, here referred to as the Archaic-Early Ceramic Interface period (AECI period). In particular, we considered several relations and some model specifications to demonstrate the operation of the framework and the flexibility of the model in a period and cultural context of probable inter-cultural contact. The networks generated indicate that the assumptions about tie formation and, consequently, the specified model largely influence the structure of the reconstructed network; thus, these networks provide a variety of different scenarios that were qualitatively assessed by evaluating the coherence of the structure of the generated networks with archaeological evidence that was not directly



accounted for by the model specification. In this paper, we tested whether the considered model specifications were able to explain the presence of sites known to have functioned as hubs but other criteria might be used according to the available archaeological information.

All the networks generated for this study provide credence to a balanced view of Caribbean inter-community interactions in the AECI. However, the networks generated using only propositions related to distance and cultural homophily are not plausible network reconstructions since they do not reflect the archaeological evidence on the importance of some sites in the past. Specifically, none of these networks points to the existence of major community gathering sites located in the northern, central-eastern and southern sub-regions of the area considered. Adding information on the distribution of lithic materials and the quantity of finds provides a more faithful reconstruction of the network of contacts between the considered sites. In fact, the resulting network underlines previous ideas on tight local lithic networks combined with a moderate amount of connectivity at the level of the region and supports the archaeological evidence that La Hueca, Trants and Pearls were the three major gathering sites during the AECI. This finding indicates that the structure of networks in the AECI period was determined by multiple interdependent mechanisms which go beyond hypotheses about distance and cultural homophily. Moreover, the coupling of archaeological provenance data, such as the AECI distribution of local lithic raw materials, with specific archaeological theories of network effects is profitable in our illustrative example and easy to implement within the ERGM framework. Regardless, we should keep in mind that this is only one data source; therefore, the results of the model may not be robust when confronted with expanded and new provenance data.

The illustrative example also indicates where the value of ERGMs for archaeological network reconstruction resides. Regardless of region, time period or data set quality and quantity, ERGMs require a formal exploration of theories of network foundation and development. This requirement goes against the grain of “informal archaeological network studies” in the Caribbean and elsewhere, where cultural histories have been built on at best tacitly understood links between material culture and social networks or at worst fuzzy ones. The ERGM framework (1) necessitates the formation of theories based on well-understood and clearly communicable network effects, such as geographic proximity, cultural homophily or the(re-)distribution of raw materials, and (2) allows for the exploration of these effects, as well as their interdependency, which can then be used to scaffold data-driven theories. In short, this study advocates for the adoption or creation of more formalised network theories, as well as data in archaeology at large, and underscores their value. Doing so in the case of the Caribbean may provide new and more specific insights into connectivity in the AECI and other periods.

**Funding** This research is part of the project NEXUS 1492, which has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 319209, and the Island Networks project, grant no. 360-62-060 financed by the Netherlands Organisation for Scientific Research (NWO).

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Agresti, A., & Kateri, M. (2011). Categorical data analysis. In *International encyclopedia of statistical science* (pp. 206–208). Springer.
- Amati, V., Shafie, T., & Brandes, U. (2018). Reconstructing archaeological networks with structural holes. *Journal of Archaeological Method and Theory*, 25(1), 226–253.
- Bevan, A., & Wilson, A. (2013). Models of settlement hierarchy based on partial evidence. *Journal of Archaeological Science*, 40(5), 2415–2427.
- Blake, E. (2014). Dyads and triads in community detection: a view from the Italian Bronze Age. *Les nouvelles de l'archéologie*, 135, 28–32.
- Boomert, A. (2000). Trinidad, Tobago, and the lower Orinoco interaction sphere: An archaeological/ethnohistorical study. Cairi.
- Broodbank, C. (2002). An island archaeology of the early Cyclades. Cambridge University Press.
- Brughmans, T. (2013). Thinking through networks: a review of formal network methods in archaeology. *Journal of Archaeological Method and Theory*, 20(4), 623–662.
- Butts, C. T. (2002). *Spatial models of large-scale interpersonal networks* (PhD Thesis). Carnegie Mellon University Pittsburgh.
- Butts, C. T. (2009). A behavioral micro-foundation for cross-sectional network models. UC Irvine. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1008326](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1008326)
- Chatterjee, S., & Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5), 2428–2461.
- Cody, A. (1990). *Prehistoric patterns of exchange in the Lesser Antilles: materials, models, and preliminary observations* (unpublished Master of Arts thesis). San Diego State University.
- Cody, A. (1993). Distribution of exotic stone artifacts through the Lesser Antilles: their implications for prehistoric interaction and exchange. In *Proceedings of the Fourteenth International Congress of the International Association for Caribbean Archaeology* (pp. 22–28).
- Collar, A., Coward, F., Brughmans, T., & Mills, B. J. (2015). Networks in archaeology: phenomena, abstraction, representation. *Journal of Archaeological Method and Theory*, 22(1), 1–32.
- Conolly, J., & Lake, M. (2006). Geographical information systems in archaeology. Cambridge University Press.
- Coward, F. (2010). Small worlds, material culture and ancient Near Eastern social networks. *Social brain, distributed mind*, 449–79.
- Coward, F. (2013). Grounding the net: social networks, material culture and geography in the Epipalaeolithic and Early Neolithic of the Near East (~ 21,000–6,000 cal BCE). In C. Knappett (Ed.), *Network Analysis in Archaeology: New Approaches to Regional Interaction* (pp. 125–150). Oxford University Press.
- Crock, J. G. (2000). *Interisland interaction and the development of chiefdoms in the Eastern Caribbean* (PhD Thesis). University of Pittsburgh Pittsburgh, PA.
- Evans, T., Knappett, C., & Rivers, R. (2009). Using statistical physics to understand relational space: a case study from Mediterranean prehistory. In *Complexity perspectives in innovation and social change* (pp. 451–480). Springer.
- Evans, T., Rivers, R., & Knappett, C. (2012). Interactions in space for archaeological models. *Advances in Complex Systems*, 15(01n02).
- Fitzpatrick, S. M. (2004). *Quo vadis Caribbean archaeology? The future of the discipline in an international forum*. UNIV PUERTO RICO, COLLEGE ARTS SCIENCES, MAYAGUEZ, PR 00680 USA.
- Fitzpatrick, S. M. (2006). A critical approach to 14 C dating in the Caribbean: using chronometric hygiene to evaluate chronological control and prehistoric settlement. *Latin American Antiquity*, 17(4), 389–418.

- Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832–842.
- Goyal, S. (2012). *Connections: An introduction to the economics of networks*. Princeton University Press.
- Graham, S. (2006). Networks, agent-based models and the Antonine itineraries: Implications for Roman archaeology. *Journal of Mediterranean Archaeology*, 19(1), 45–64.
- Habiba, Athenstädt, J. C., Mills, B. J., & Brandes, U. (2018). Social networks and similarity of site assemblages. *Journal of Archaeological Science*, 92, 63–72.
- Hage, P., & Harary, F. (1991). *Exchange in Oceania: a graph theoretic analysis*. Oxford University Press.
- Handcock, M. S., & Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1), 5–25.
- Handcock, M. S., Robins, G., Snijders, T., Moody, J., & Besag, J. (2003). Assessing degeneracy in statistical models of social networks. Citeseer.
- Hardy, M. D. (2008). *Saladoid economy and complexity on the Arawakan frontier* (PhD Thesis). Florida State University.
- Hodder, I. (1974). Some marketing models for Romano-British coarse pottery. *Britannia*, 5, 340–359.
- Hofman, C. L., & Hoogland, M. L. (2011). Unravelling the multi-scale networks of mobility and exchange in the pre-colonial circum-Caribbean. In *Communities in contact. Essays in archaeology, Ethnohistory and Ethnography of the Amerindian circum-Caribbean* (pp. 14–44).
- Hofman, C. L., Bright, A. J., Boomert, A., & Knippenberg, S. (2007). Island rhythms: the web of social relationships and interaction networks in the Lesser Antillean archipelago between 400 BC and AD 1492. *Latin American Antiquity*, 18(3), 243–268.
- Hofman, C. L., Bright, A. J., Ramos, R. R., de Utuado, R., & de Ciencias Sociales, P. (2010). Crossing the Caribbean Sea: towards a holistic view of pre-colonial mobility and exchange. *Journal of Caribbean Archaeology*, 3, 1–18.
- Hofman, C. L., Boomert, A., Bright, A. J., Hoogland, M. L., Knippenberg, S., & Samson, A. V. (2011). Ties with the homelands: archipelagic interaction and the enduring role of the South and Central American mainlands in the Pre-Columbian Lesser Antilles. In *Islands at the crossroads: migration, seafaring, and interaction in the Caribbean* (pp. 73–86). Tuscaloosa: University of Alabama Press.
- Hofman, C. L., Mol, A., Ramos, R. R., & Knippenberg, S. (2014). Networks set in stone: Archaic-Ceramic interaction in the early pre-colonial northeastern Caribbean. *Archéologie Caraïbe*, 119.
- Hofman, C. L., Rodríguez Ramos, R., & Pagan, J. (2018). The neolithisation of the northeastern Caribbean: mobility and social interaction. In *The Archaeology of Caribbean and Circum-Caribbean Farmers (6000 BC-AD 1500)* (pp. 99–125). Routledge.
- Hofman, C. L., Borck, L., Slayton, E., & Hoogland, M. L. (2019). Archaic age voyaging, networks, and resource mobility around the Caribbean Sea. In *Early Settlers of the Insular Caribbean Dearchaizing the Archaic*. Sidestone Press Academics.
- Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.
- Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630.
- Jaynes, E. T. (1957b). Information theory and statistical mechanics. II. *Physical Review*, 108(2), 171.
- Johnson, G. A. (1977). Aspects of regional analysis in archaeology. *Annual Review of Anthropology*, 6(1), 479–508.
- Keegan, W. F. (2004). Islands of chaos. In A. Delpuech & C. L. Hofman (Eds.), *Late Ceramic Age societies in the eastern Caribbean*. Archaeopress.
- Keegan, W. F. (2007). *Taino Indian myth and practice: the arrival of the Stranger King*. University Press of Florida USA.
- Keegan, W. F., & Hofman, C. L. (2017). *The Caribbean before Columbus*. Oxford University Press.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.
- Knappett, C. (2011). *An archaeology of interaction: network perspectives on material culture and society*. Oxford University Press Oxford.
- Knappett, C., Evans, T., & Rivers, R. (2008). Modelling maritime interaction in the Aegean Bronze Age. *Antiquity*, 82(318), 1009–1024.
- Knippenberg, S. (2007). *Stone artefact production and exchange among the Lesser Antilles* (Vol. 13). Amsterdam University Press.
- Koskinen, J. H., Robins, G. L., & Pattison, P. E. (2010). Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. *Statistical Methodology*, 7(3), 366–384.
- Laffoon, J. E., Sonnemann, T. F., Shafie, T., Hofman, C. L., Brandes, U., & Davies, G. R. (2017). Investigating human geographic origins using dual-isotope ( $^{87}\text{Sr}/^{86}\text{Sr}$ ,  $\delta^{18}\text{O}$ ) assignment approaches. *PLoS One*, 12(2), e0172562.

- Lusher, D., Koskinen, J., & Robins, G. (2012). Exponential random graph models for social networks: theory, methods, and applications. Cambridge University Press.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444.
- Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3), 825–850.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092.
- 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. (2013). Transformation of social networks in the late pre-Hispanic US southwest. *Proceedings of the National Academy of Sciences*, 110(15), 5785–5790.
- Mol, A. A. (2013). Studying Pre-Columbian interaction networks. In *The Oxford Handbook of Caribbean Archaeology*. Oxford University Press.
- Mol, A. A. (2014). The connected Caribbean: a socio-material network approach to patterns of homogeneity and diversity in the pre-colonial period. Sidestone Press.
- Monderer, D., & Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, 14(1), 124–143.
- Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of Statistical Software*, 24(4), 1548–7660.
- Newman, M. (2003). Mixing patterns in networks. *Physical Review E*, 67(2), 026126.
- Newman, M. (2018a). *Networks* (2nd ed.). Oxford university press.
- Newman, M. (2018b). Network structure from rich but noisy data. *Nature Physics*, 14(6), 542–545.
- Newman, M. (2018c). Network reconstruction and error estimation with noisy network data. *arXiv preprint arXiv:1803.02427*.
- Östborn, P., & Gerding, H. (2014). Network analysis of archaeological data: a systematic approach. *Journal of Archaeological Science*, 46, 75–88.
- Polanyi, K. (1963). Ports of trade in early societies. *The Journal of Economic History*, 23(01), 30–45.
- Renfrew, C. (1975). Trade as action at a distance: questions of integration and communication. *Ancient Civilization and Trade*, 3, 59.
- Rivers, R., Knappett, C., & Evans, T. (2011). Network models and archaeological spaces. *Computational Approaches to Archaeological Spaces*.
- Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph ( $\mathbb{S}^p$ ) models for social networks. *Social Networks*, 29(2), 173–191.
- Rodríguez Ramos, R. (2007). *Puerto Rican precolonial history etched in stone* (PhD Thesis). University of Florida.
- Rodríguez Ramos, R. (2010). What is the Caribbean? An archaeological perspective. *Journal of Caribbean Archaeology*, 3, 19–51.
- Rodríguez Ramos, R. (2011). The circulation of jadeite across the Caribbeanscape. *Communities in Contact: Essays in Archaeology, ethnohistory and ethnography of the Amerindian circum-Caribbean*, 117–136.
- Rouse, I. (1992). *The Tainos: rise and decline of the people who greeted Columbus*. Yale University Press.
- Roux, V., & Manzo, G. (2018). Social boundaries and networks in the diffusion of innovations: a short introduction. *Journal of Archaeological Method and Theory*, 25(4), 967–973.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.
- Shennan, S. (1997). *Quantifying archaeology*. University of Iowa Press.
- Siegel, P. E. (1992). *Ideology, power, and social complexity in prehistoric Puerto Rico*. Binghamton University.
- Sindbaek, S. M. (2013). Broken links and black boxes: material affiliations and contextual networks synthesis in the Viking world. In C. Knappett (Ed.), *Network analysis in archaeology*. Oxford University Press.
- Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1–40.
- Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1), 99–153.
- Terrell, J. (1977). Human biogeography in the Solomon Islands. *Fieldiana. Anthropology*, 1–47.
- Torres, J. M., & Rodríguez Ramos, R. (2008). *The Caribbean: a continent divided by water. Archaeology and geoinformatics: case studies from the Caribbean* (pp. 13–29). Tuscaloosa: University of Alabama Press.
- Vega-Redondo, F. (2003). *Economics and the theory of games*. Cambridge university press.
- Veloz Maggiolo, M. (1991). Panorama histórico del Caribe precolombino. *Quinto Centenario del Descubrimiento de América Banco Central de la República Dominicana*.
- Wang, P., Robins, G., & Pattison, P. (2009). PNet: Program for the estimation and simulation of  $p^*$  exponential random graph models, User Manual.

- Wang, C., Butts, C. T., Hipp, J. R., Jose, R., & Lakon, C. M. (2016). Multiple imputation for missing edge data: a predictive evaluation method with application to add health. *Social Networks*, *45*, 89–98.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p\*. *Psychometrika*, *61*(3), 401–425.
- Watters, D. R. (1997). Maritime trade in the prehistoric Eastern Caribbean. The indigenous people of the Caribbean, 88–99.
- Wurzer, G., Kowarik, K., & Reschreiter, H. (2015). Agent-based modeling and simulation in archaeology. Springer.
- Zucchi, A. (1990). La serie Meillicoide y sus relaciones con la cuenca del Orinoco. In *Proceedings of the Eleventh Congress of the International Association for Caribbean Archaeology, La Fundación Arqueológica, Antropológica e Histórica de Puerto Rico, San Juan* (pp. 272–286).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.