

Nonparametric Network Bootstrap

Kamila Nowakowicz

October 4, 2024

Abstract

We propose a bootstrap procedure for network data based on a nonparametric linking function estimator. We characterise when the nonparametric linking function estimator is uniformly consistent. We prove bootstrap consistency in the sense that a Wasserstein distance between the bootstrap network generating distribution and the true network generating distribution goes to zero in probability. We provide conditions under which distributions of a class of functions related to U-statistics on the bootstrapped networks consistently replicate the distributions of the corresponding statistics on the original network. Monte Carlo simulations show good confidence interval coverage for a wider class of network functions than are accounted for by our theory. An application to the data from Banerjee, Chandrasekhar, Duflo, and Jackson (2013) not only replicates their findings, but also shows that our method can be applied under weaker assumptions and smaller sample sizes. We propose an alternative specification of their model which takes advantage of our linking function estimator and may be of interest independently of our bootstrap procedure.

1 Introduction

Many papers in economics studying information diffusion (e.g. Banerjee, Chandrasekhar, Duflo, and Jackson (2013), Alatas et al. (2016), Conley and Udry (2010)), spillover effects (e.g. Carter, Laajaj, and Yang (2021)) and other network models run into the issue that it is very difficult to conduct statistical inference on complex, interconnected data structures represented by networks. In this paper, we propose a solution: a bootstrap procedure which does not impose strong assumptions on the form of the network-generating function and can be applied to a wide range of network statistics.

The default approach when analysing the behaviour of statistics is finding an asymptotic approximation to their distribution, a technique well-suited for simple models and data with limited dependence. Unfortunately, the models we build on networks are often complex and the networks themselves tend to exhibit a deeply interconnected structure. All individuals in a network are closely related: the concept of “six degrees of separation” shows that nearly all users of social media platforms like Facebook or Twitter are at most six connections away from each other, while the average distance is below four. At the same time, the rate at which we learn about their connections grows more slowly than the network size. This phenomenon is known as sparsity and can be illustrated by the fact that, during their peak growth periods, social media platforms gained new users at a faster rate than individual users gained new connections. Because of the issues of strong connectedness, sparsity, and complicated functional forms of network statistics, asymptotic theory for network statistics tends to be complicated, specialised to certain classes of estimators, and, in many cases, still underdeveloped.

For similar reasons, standard bootstrap techniques are not valid for network data: we need a specialised bootstrap procedure based on understanding and replicating the dependence structure in a network. The few existing methods for bootstrapping network data suffer from either limited applicability (they tend to focus on specific classes of network statistics and cannot be easily extended to e.g. regressions controlling for the dependence structure defined by the network) or restrictive parametric functional form for the components of the network-generating process. We address both of those concerns.

Our bootstrap procedure takes a given network and generates new networks with similar structure. We provide guarantees that the bootstrapped networks are drawn from a distribution closely resembling that of the original network, allowing them to be treated as independent samples from the original network’s data-generating process. If we are interested in the distribution of a particular statistic of the original network, we can approximate resampling from this distribution by

estimating the same statistic on the bootstrapped networks.

We assume a general form of a network-generating process in which the observed network is determined by an unknown distribution over types of individuals and an unknown function determining the the probability of a link between any pair of individuals. We assume that the network we observe is formed from a group of people, each characterised by a set of (possibly unobserved) features that are independent of the features of others. We assume that the observed links were drawn independently with probabilities determined by a binary linking function which takes the features of any two individuals as inputs and outputs the probability of a link between them.

If we knew the linking function, we could generate networks similar to the observed one by firstly resampling from the original set of individuals and then adding links based on probabilities determined by the linking function. However, the linking function is unknown, and to make matters worse, it depends on possibly unobserved inputs. To address this challenge, we propose a nonparametric method to estimate the linking function which takes advantage of the information provided by the set of observed connections.

We start by borrowing a distance from Auerbach (2022) (see also Zhang, Levina, and Zhu (2017)), which can be intuitively summarised as: people with similar sets of friends are similar to each other. If we observe two people with similar sets of neighbours, it likely happened because the linking function gave similar probabilities for their links with other individuals. This way we can identify people similar to any person i . We can then determine what proportion of these ‘counterfactuals’ of i are linked to a person j , providing an estimate of the link probability between i and j . Similarly, by swapping the roles of i and j , we can find what proportion of individuals similar to j are linked to i . Assuming that links are symmetric, we take the average of these two estimates as an estimate of the link probability between i and j . Whenever we sample i and j as nodes of the bootstrap network we draw a link between them with this estimated probability.

Having developed a method to generate bootstrap networks, we characterise the conditions under which our procedure works well. This paper makes three main contributions. The first contribution is to the literature on the linking function estimation: we propose a linking function estimator and provide conditions under which it achieves uniform consistency. We also develop a cross-validation procedure for choosing an optimal bandwidth parameter for our estimator. The second contribution is to the network bootstrap literature: we develop a new bootstrap procedure and provide conditions under which our bootstrap is consistent. We show this in two ways: we borrow a notion of Wasserstein distance between network generating distributions from Levin and Levina (2019) and we show that the distance between the bootstrap network generating process

and the true network generating process goes to zero in probability as we increase the sample size. Unfortunately, this is not sufficient to ensure that the distribution of any statistic on a bootstrap network replicates the corresponding distribution of that statistic on the original network. We show this directly for a class of statistics which are closely related to U-statistics. The motivation for this is twofold: this is a wide class of functions and includes some estimators we may be directly interested in, for example the density of connections within a network. Additionally, this class includes motif densities, i.e. the densities of different patterns on subgraphs of the adjacency matrix. These are sometimes referred to as “network moments” because they characterise the network generating distribution: if two networks match on densities of all possible patterns, they come from the same distribution. Hence proving that our bootstrap procedure correctly recovers the distributions of all motif densities implicitly shows that the bootstrap networks share the same asymptotic network generating distribution as the original network.

While we don’t currently have explicit asymptotic theory for more general network functions, our simulations suggest that our method is more widely applicable. We also provide an illustration on the data from Banerjee, Chandrasekhar, Duflo, and Jackson (2013), which allows us to replicate their findings for estimates of a diffusion model over a network, provide estimates using a subsample of the data too small for the original inference method to apply, and propose an alternative model specification which relies on our linking function estimator and which could be of interest independently of the bootstrap procedure.

In Section 2 we summarise the related literature. The setup of the model is described in Section 3, where we also provide definitions of our estimators and the bootstrap procedure. Section 4 includes the statements of our main results: the uniform consistency of the linking function estimator in Theorem 1, a bootstrap consistency result for a specific class of estimators related to U-statistics in Theorem 3, and a Wasserstein distance convergence in Theorem 2. Section 5 shows results of Monte Carlo simulations and Section 6 describes an application to the data from Banerjee, Chandrasekhar, Duflo, and Jackson (2013). Section 7 concludes. The appendices start with a list of all notation. Appendix A includes all proofs, Appendix B provides the codes, additional tables, plots for simulations, Appendix C includes extensions.

2 Related literature

2.1 Network Bootstrap

There are a few existing bootstrap procedures for different functions on networks. Most of the literature focuses on bootstrapping a class of network functions closely related to U-statistic, or

their subset, motif densities. Our procedure can be applied to a much wider class of functions.

Green and Shalizi (2022) propose two types of bootstrap: the empirical bootstrap, in which they resample individuals and put a link between them if they were linked in the original graph, and a parametric histogram bootstrap. The empirical bootstrap can be seen as a special case of ours, it is simple and computationally attractive, but it suffers from a few types of bias: whenever an individual gets resampled more than once, these copies are not linked (as there were no self-links in the original graph), and they share the same link patterns with all other individuals (there is correlation between link formation in the bootstrap graph which was not present in the original graph). Green and Shalizi (2022) prove that this dependence and bias is asymptotically negligible for motif densities, but this is not necessarily true for other functions. Our simulations show that their procedure does not perform well e.g. for eigenvalues other than the highest one.

Levin and Levina (2019) assume a specific functional form of the linking function¹. They propose two methods: one in which they directly estimate a U-statistic and one in which they generate a full network that can be used for estimating more general functions of a network, including eigenvalues and measures of small-world behaviour. This is the only paper we are aware of which provides results for functions of the entire network: they show that the entire bootstrapped network converges to an independent copy of the original network in terms of a new notion of Wasserstein network distance they define. Under our more general nonparametric specification we are able to show convergence in terms of the same distance (see Theorem 2).

Lin, Lunde, and Sarkar (2020) propose a computationally efficient multiplier bootstrap for motif densities, based on approximating the first (for large sparse graphs) or first and second (for smaller denser graphs) order terms of a Hoeffding decomposition of the U-statistic. Their method is specific to this class and, unlike ours, it cannot be extended to other types of network functions. They show higher-order accuracy of their quadratic bootstrap using an Edgeworth expansion. The theory of Edgeworth expansion for motif densities is developed Zhang and Xia (2022) who show higher order correctness of a studentised version of the empirical bootstrap of Green and Shalizi (2022). We believe similar methods could be used to show higher-order accuracy of our method, but we do not pursue this direction in the current work.

Shao and Le (2024) provide a parametric bootstrap in a setting different from all the previously mentioned papers, where the nodes are non-exchangeable. In our notation this corresponds to a situation in which the ξ and the matrix of link probabilities are fixed, $h_{0,n}$ takes a known parametric form, and the randomness comes only from η . Their analysis focuses on quantifying the bias and

1. They assume a random dot product graph with a linking function: $h_{0,n}(\xi_i, \xi_j) = \xi_i' \xi_j$ where ξ_i is a vector of latent positions which can be interpreted as characteristics of individual i .

providing bias-corrected bootstrap procedure for motif densities.

The network setup is a special case of an exchangeable array². Papers which propose bootstrap for exchangeable arrays include Davezies, D’Haultfœuille, and Guyonvarch (2021), whose method in our setting is identical to empirical bootstrap, and Menzel (2021), who proposes a new wild bootstrap procedure based on splitting the statistic of interest into orthogonal components, estimating them by sample analogues, and resampling each component with appropriate scaling. Menzel (2021) also points out that, depending on the dependence structure, the limiting distribution may be nonstandard. Their LLN and CLT results apply to functions of finite k -dimensional clusters which take a form of U- or V-statistics (including degenerate cases), their smooth functionals and Z-estimators. Their methods are local, they can be applied to functions of finite subgraphs and cannot account for dependence over the whole adjacency matrix, as in the case of eigenvalues or some centrality measures covered by our method.

In terms of the allowed level of sparsity, we impose a stronger requirement for acyclic motifs than both models in Green and Shalizi (2022) as well as Lin, Lunde, and Sarkar (2020), but our requirement is the same as for cycles in Lin, Lunde, and Sarkar (2020) and is weaker than that for general motifs for Green and Shalizi (2022) empirical graphon. In comparison with Green and Shalizi (2022) histogram graphon, our sparsity condition for general motifs becomes weaker only when $m > 4$, and we also impose weaker conditions than L-Lipschitz on the linking function. Levin and Levina (2019) only include sparsity considerations in one result, for acyclic motifs and cycles. Their assumption is weaker than ours, which is not surprising given their model is parametric.

In terms of alternative ways of estimating distributions of network statistics, Bickel, Chen, and Levina (2011) provide asymptotic theory for motif densities. Subsampling methods have been proposed by Bhattacharyya and Bickel (2015), who give results for motif densities, and Lunde and Sarkar (2022), who provide consistency results for general functions and specify them to two classes: motif densities and eigenvalues of graphons of finite rank. Their methods require minimal assumptions and allow for sparser graphs than ours.

2.2 Linking function estimation

The first step in our bootstrap procedure involves the estimation of a linking function, for which we propose a nonparametric estimator and provide conditions for its uniform consistency. Our idea for the nonparametric linking function estimator was inspired by Auerbach (2022), who provides a way

2. Using notation from Davezies, D’Haultfœuille, and Guyonvarch (2021), our model is a special case of an exchangeable and dissociated array with $k = 2$ and U_{ij} corresponding to the randomness due to Bernoulli trials $\tau(u_i, u_j, u_{ij}) = \mathbb{1}(h(u_i, u_j) \leq u_{ij})$. The kernels of U-statistics on networks can be represented as higher-dimensional ($2 \leq k < \infty$) exchangeable arrays.

of controlling for a network-dependent latent covariate in a partially linear regression setting. Our estimator has been previously proposed, but not analysed by Zelenev (2020). Our contribution to the literature on linking function estimation most closely resembles Zhang, Levina, and Zhu (2017), who propose a similar procedure to ours, with the difference that their distance is defined in terms of the maximum norm and their estimator is a nearest-neighbour estimator instead of a kernel estimator.

One key contribution we make is providing a numerical procedure for selecting an optimal value for a tuning parameter. In Zhang, Levina, and Zhu (2017) the equivalent parameter is motivated theoretically and derived from the rate of convergence. We find that our procedure with the numerically chosen optimal parameter values outperforms Zhang, Levina, and Zhu (2017) with the theoretically best optimal values when used as a first step in our bootstrap procedure. We believe this is largely due to the bandwidth choice.

2.3 Other

In the proofs that our bootstrap procedure is reliable we use a framework inspired by Politis et al. 1999. Our results can be seen as an extension of Bickel and Freedman (1981), the classic paper providing conditions for consistency of bootstrap for U-statistics. We extend their analysis to the case where the objective function becomes a U-statistic only after taking expectation conditional on a vector of unobserved characteristics and after substituting the true linking function for its estimator as the input to the kernel function. We show that our linking function estimator converges to the true linking function in a sense which is sufficient for the bootstrap equivalent to converge weakly in probability to the same limiting distribution as the object of interest in the original sample. Because of the additional levels of approximation, we achieve a weaker notion of convergence than convergence weakly almost surely in Bickel and Freedman (1981), but our result is still sufficient to provide asymptotically correct bootstrap confidence intervals.

For our empirical application, we use the data and some of the codes from Banerjee, Chandrasekhar, Duflo, and Jackson (2013). We confirm their results using our method and we repeat a part of their analysis under weaker assumptions: where the original paper performs the estimation aggregating over many villages, we are able to provide estimates and confidence intervals on individual village level. We are also able to run a related model based on the strength of connections between household rather than the less informative binary information on presence or lack of connection. We find that removing this one level of approximation has a significant effect on our conclusions.

3 Model: setup, definitions and the bootstrap procedure

3.1 Setup

We follow the standard setup in the literature known as the latent space model.

We observe an adjacency matrix A which corresponds to an undirected, unweighted graph on n nodes (also referred to as individuals) indexed by $i \in \{1, 2, \dots, n\}$. The matrix is symmetric, has zeros on the main diagonal and ones in positions corresponding to edges in the graph ($A_{ij} = 1$ if and only if there is an edge between nodes i and j). For a vector of index numbers $\iota = (\iota_1, \dots, \iota_m)'$ with $\iota_i \in \{1, 2, \dots, n\}$ we let $A(\iota)$ denote the corresponding submatrix defined on nodes in ι (i.e. A from which we remove rows and columns not in ι). Each node i is characterised by a vector of unobserved features³ ξ_i , drawn independently from their common distribution F_0 with support $Supp(\xi_i)$. We denote the vector of all $\{\xi_i\}_{i=1}^n$ by ξ and we let $\xi(\iota) = (\xi_{\iota_1}, \dots, \xi_{\iota_m})$. We assume that the distribution has no point mass, i.e. for $\xi_i, \xi_j \sim F_0$ we have $P_{F_0}(\xi_i = \xi_j) = 0$ ⁴. We impose more assumptions on F_0 in Assumption 1.2⁵.

Let $h_{0,n} : Supp(\xi_i) \times Supp(\xi_i) \rightarrow [0, 1]$ be a symmetric, measurable linking function⁶ which can be decomposed as:

$$h_{0,n}(u, v) = \rho_n w_0(u, v) \tag{1}$$

where $\int w_0(u, v) dF_0(u) dF_0(v) = 1$.

For each pair of nodes i, j , $h_{0,n}(\xi_i, \xi_j)$ maps their unobserved characteristics ξ_i, ξ_j into the probability of a link (edge) between them, i.e. the probability with which $A_{ij} = 1$. For concise notation, we use $h_{0,n}(\xi(\iota))$ to mean the collection of pairwise linking probabilities between the elements of $\xi(\iota)$. We treat the linking function as unknown, making minimal assumptions on its properties in Assumption 1.2: we require that for each input there is a neighbourhood of sufficiently large measure in which the behaviour of the function remains similar. Importantly, we do not require a specific form (e.g. random dot product structure: $h_{0,n}(\xi_i, \xi_j) = \xi_i' \xi_j$ like in Levin and Levina (2019)), we do not impose any shape constraints (e.g. that the function is strictly increasing in its inputs).

The decomposition into ρ_n and w_0 is without loss of generality and can be seen as a normalisa-

3. This corresponds to the vector of latent positions X_i in Levin and Levina (2019).

4. This is without loss of generality: if we had a distribution with a point mass we could define a new support of ξ and a new F_0 in which the point mass would be replaced by a region of ξ of total measure equal to the probability at the original point.

5. The assumptions are implicit and would be implied by F_0 bounded above and separated away from zero with $h_{0,n}$ piecewise Lipschitz.

6. The linking function has been referred to as the coupling function $g(.,.)$ in Zelenev (2020) and the graphon function in Green and Shalizi (2022).

tion which allows us to interpret ρ_n as the expected edge density (the marginal probability of an edge between two nodes). We assume $\rho_n \rightarrow 0$ as $n \rightarrow \infty$, which captures the common feature of real economic networks known as sparsity. Intuitively, it says that the number of expected friends grows at a slower rate than the size of the network: no matter how large the potential pool of friends is, people tend to have a fairly small friendship group. This causes issues for estimation because, even as the size of the network grows at a rate n , the amount of information about the links of a specific node i grows at a slower rate of $\rho_n n$. In the extreme case of $\rho_n n$ being bounded we can't hope to get consistency of our estimates. In our results we specify bounds on the rate at which ρ_n approaches zero which still allow us to reliably estimate parameters and their distributions. For the linking function estimator we require that the density decreases at a slower rate than $\sqrt{\frac{\log(n)}{n}}$ (see Assumption 1.1), while in other sections we may strengthen this requirement, e.g. in Theorem 3 the allowed level of sparsity depends on how complicated the statistic we are estimating is.

w_0 is the underlying linking/graphon function after accounting for sparsity. While w_0 cannot be interpreted directly as a probability, it has similar properties, e.g. it is bounded⁷This is the function which determines the data generating process and the function the statistics of which we want to analyse. Although in a sample of size n we encounter its rescaled version $h_{0,n}$, for any asymptotic results we need to remove the effect of sparsity and we look at normalisations which are function of $\frac{h_{0,n}}{\rho_n}$.

To capture the way in which the linking function $h_{0,n}$ is translated into the observed links in A we introduce a random noise parameter: for $1 \leq i \leq j \leq n$ let $\eta_{ij} \stackrel{ind}{\sim} \mathcal{U}[0, 1]$ be independent of ξ . We denote the vector of η_{ij} by η . We assume⁸:

$$A_{ij} = A_{ji} = \mathbf{1}(h_{0,n}(\xi_i, \xi_j) \geq \eta_{ij}) \quad (2)$$

$$A_{ii} = 0. \quad (3)$$

Note that $E(A_{ij}|\xi_i, \xi_j) = P(A_{ij} = 1|\xi_i, \xi_j) = h_{0,n}(\xi_i, \xi_j) = \rho_n w_0(\xi_i, \xi_j)$. To distinguish between adjacency matrices based on the true and estimated/simulated inputs we sometimes explicitly write A as a function: $A(h_{0,n}(\xi), \eta)$.

7. This is a common assumption in the literature, though it is sometimes relaxed to allow $w_0(u, v) \in \mathbb{R}_+$ and let $h_{0,n}(u, v) = \min\{w_0(u, v), 1\}$. This affects the interpretation of ρ_n as the density and makes it more difficult to infer $h_{0,m}$ from $h_{0,n}$. Our results could be generalised to allow for unbounded w_0 at the expense of more complicated proofs and additional assumptions on bounded moments of w_0 or its functions.

8. This is one specific way of achieving:

$$A_{ij}|\xi = A_{ji}|\xi \stackrel{ind}{\sim} \text{Bernoulli}(h_{0,n}(\xi_i, \xi_j))$$

$$A_{ii} = 0$$

3.2 The object of interest

Suppose we are interested in $f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$, or $f_n(A)$ in short, which is a function of the observed adjacency matrix A based on Bernoulli trials with probabilities determined by a linking function $h_{0,n}$ of i.i.d observations ξ from the distribution F_0 , on the sparsity ρ_n , and on the distribution F_0 itself. More precisely, let the distribution of the statistic of interest be:

$$J_n(t, h_{0,n}, F_0) = P(f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0) \leq t). \quad (4)$$

Example. To fix ideas, suppose we are interested in the density $\rho_n = E(A_{ij})$. We may want to test if it takes a specific value predicted by our theory, or we may wish to test if two networks (or perhaps the same network at two points in time) have the same density level. We can estimate the density using the density estimator from the observed adjacency matrix A :

$$\hat{\rho}_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} A_{ij}. \quad (5)$$

We could use $f_n(A) = \hat{\rho}_n$ directly, or we could recentre and normalise the above expression:

$$f_n^{\rho_n}(A(h_{0,n}(\xi), \eta), \rho_n, F_0) = \frac{\sqrt{n}}{\binom{n}{2} \rho_n} \sum_{1 \leq i < j \leq n} A_{ij} - E_{F_0}(h_{0,n}(\xi_i, \xi_j)) \quad (6)$$

to get a statistic which has a well-defined asymptotic distribution. The results in Theorem 3 imply that $\hat{\rho}_n$ is consistent for ρ_n and $f_n^{\rho_n}(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$ is asymptotically normal. The finite-sample distribution is non-trivial and depends on F_0 .

Our goal is to find a good approximation to this finite-sample distribution, e.g. in order to form confidence intervals for $f_n(A)$. We do it by defining estimators \hat{h}_n of the linking function $h_{0,n}$; \hat{F}_n of the distribution of ξ ; and $\hat{\rho}_n$ of the density parameter. We use these estimates to form B bootstrap adjacency matrices $A(\hat{h}_n(\xi_b^*), \eta_b^*)$, where the b th bootstrap adjacency matrix is evaluated using \hat{h}_n based on ξ_b^* from \hat{F}_n .

We evaluate $f_n(A(\hat{h}_n(\xi^*), \eta^*), \hat{\rho}_n, \hat{F}_n)$ for B bootstrap samples to get the simulated distribution:

$$\hat{J}_{n,B}(t, \hat{h}_n, \hat{F}_n) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}(f_n(A(\hat{h}_n(\xi_b^*), \eta_b^*), \hat{\rho}_n, \hat{F}_n) \leq t). \quad (7)$$

For B large enough this provides an arbitrarily good approximation to $J_n(t, \hat{h}_n, \hat{F}_n)$ ⁹ and can be

9.

$$P_{F_0} \left(\sup_t |\hat{J}_{n,B}(t, \hat{h}_n, \hat{F}_n) - J_n(t, \hat{h}_n, \hat{F}_n)| > \varepsilon \right) \leq 4\sqrt{2}e^{-2B\varepsilon^2},$$

used to approximate $J_n(t, h_{0,n}, F_0)$.

In the remainder of this section we define all the estimators and the bootstrap procedure.

3.3 Distance: definition and estimator

Based on the observed matrix A , we want to estimate the linking probability for any pair of nodes. We start by defining a distance between individuals i and j , taking the measure from Auerbach (2022)¹⁰. Intuitively, if two people have similar friendship groups, they should be similar to each other: they likely ended up with similar friendship groups because their linking functions were similar. We let $\varphi(\xi_i, \xi_t) = E(w_0(\xi_i, \xi_s)w_0(\xi_t, \xi_s)|\xi_i, \xi_t) = E\left(\frac{A_{is}}{\rho_n} \frac{A_{ts}}{\rho_n} \middle| \xi_i, \xi_t\right)$ be a function measuring the probability of a common friend between i and t , normalised to remove the effect of sparsity. Similarly, $\varphi(\xi_j, \xi_t)$ gives a normalised measure of the probability of common friends between j and t . To measure the similarity in friendship groups between i and j we look at the expected difference $\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t)$ for any individual t . This motivates the definition of distance between i and j :

$$d_{ij} = \sqrt{E\left(\left(\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t)\right)^2 \middle| \xi_i, \xi_j\right)} \quad (8)$$

$$= \sqrt{E\left(\left(E(w_0(\xi_t, \xi_s)(w_0(\xi_i, \xi_s) - w_0(\xi_j, \xi_s)) \middle| \xi_i, \xi_j, \xi_t)\right)^2 \middle| \xi_i, \xi_j\right)} \quad (9)$$

$$= \sqrt{E\left(E\left(\frac{A_{ts}}{\rho_n} \left(\frac{A_{is}}{\rho_n} - \frac{A_{js}}{\rho_n}\right) \middle| \xi_i, \xi_j, \xi_t\right)^2 \middle| \xi_i, \xi_j\right)}. \quad (10)$$

After an appropriate normalisation by the sparsity level ρ_n , we get an expression in terms of the linking function $h_{0,n} = \rho_n w_0$ at sample size n :

$$\rho_n^2 d_{ij} = \sqrt{E\left(\left(E(h_{0,n}(\xi_t, \xi_s)(h_{0,n}(\xi_i, \xi_s) - h_{0,n}(\xi_j, \xi_s)) \middle| \xi_i, \xi_j, \xi_t)\right)^2 \middle| \xi_i, \xi_j\right)} \quad (11)$$

$$= \sqrt{E\left(E\left(A_{ts}(A_{is} - A_{js}) \middle| \xi_i, \xi_j, \xi_t\right)^2 \middle| \xi_i, \xi_j\right)}. \quad (12)$$

Eq. (11) highlights the close relation between the normalised distance and the similarity between the linking functions of i and j at sample size n : a low value of $\rho_n^2 d_{ij}$ means i and j are similar to each other in the sense that their $h_{0,n}(\xi_i, \cdot)$ and $h_{0,n}(\xi_j, \cdot)$ are close. We exploit this when defining an estimator for $h_{0,n}$. The normalised expression is also attractive because the sample equivalent

see references on p.5 of Politis et al. (1999) for more details.

10. Auerbach (2022) refers to $\varphi_{\xi_i}(\tau) = \varphi(\xi_i, \tau)$ as the codegree function of agent i . In his model there is no sparsity: $\rho_n = 1$, which is why he does not need the normalisation by $\frac{1}{\rho_n}$.

of its representation in Eq. (12) provides us with an estimate of $\rho_n^2 d_{ij}$:

$$\rho_n^2 \hat{d}_{ij} = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n A_{ts} (A_{is} - A_{js}) \right)^2}. \quad (13)$$

Remark. If we needed to estimate \hat{d}_{ij} without the normalisation we could substitute the estimated density $\hat{\rho}_n$ (Eq. (5)) for the unknown ρ_n . However, in practice the way we use the distance is with a normalisation by a bandwidth parameter a_n (chosen by the researcher), we look at functions of: $\frac{\rho_n^4 \hat{d}_{ij}^2}{a_n} \equiv \frac{\hat{d}_{ij}^2}{b_n}$, and we can think of the ρ_n as being absorbed into the renormalised bandwidth b_n .

Remark. We could also consider estimators of related distances from Zeleneev (2020), who took it from Zhang, Levina, and Zhu (2017): $\rho_n^2 \hat{d}_{ij}^{(\infty)} = \left(\max_{t \neq i, j} \left| \frac{1}{n} \sum_{s \neq i, j, t} A_{ts} (A_{is} - A_{js}) \right| \right)^{\frac{1}{2}}$ and from Lovász (2012) (section 13.4): $\rho_n^2 \hat{d}_{ij}^{(1)} = \left(\frac{1}{n} \sum_{t=1}^n \left| \frac{1}{n} \sum_{s \neq i, j, t} A_{ts} (A_{is} - A_{js}) \right| \right)^{\frac{1}{2}}$. All of these distances are based on the same idea but average $\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t)$ using L_2 , L_∞ and L_1 distances, respectively.

3.4 Linking function estimator

We are now ready to define an estimator \hat{h}_n for the linking function $h_{0,n}$. We rely on a kernel approximation: let $K(\cdot)$ be a kernel function (for properties see Assumption 1.3), let a_n be a bandwidth parameter (for its rates of convergence see Assumption 1.4, see Section 3.5 for a method of choosing an optimal bandwidth). We can estimate $\hat{h}_n(\xi_i, \xi_j)$ as:

$$\hat{h}_n(\xi_i, \xi_j) = \frac{\tilde{h}_n(\xi_i, \xi_j) + \tilde{h}_n(\xi_j, \xi_i)}{2} \quad (14)$$

where

$$\tilde{h}_n(\xi_i, \xi_j) = \frac{\sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right) A_{tj}}{\sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right)}. \quad (15)$$

$\tilde{h}_n(\xi_i, \xi_j)$ is a local weighted average which puts the highest weights on the individuals most similar to i . The bandwidth controls the required level of similarity beyond which we use zero weights. Each person t with \hat{d}_{it} sufficiently close to zero can be seen as a counterfactual to i , someone with a very similar linking function (i.e. a small $h_{0,n}(\xi_i, \cdot) - h_{0,n}(\xi_t, \cdot)$). The proportion of people similar to i who are linked to j gives an estimate of the link probability between i and j . The observation with $t = j$ is excluded because we assume there are no self-link: $A_{jj} = 0$ is not

defined in terms of $h_{0,n}$. Adding a non-zero weight on this observation would introduce bias.

To get $\hat{h}_n(\xi_i, \xi_j)$ we take advantage of the symmetry of links, we repeat the estimation swapping the roles of i and j and take an average of the two estimates.

We show that this estimator is uniformly consistent for $h_{0,n}$ in Theorem 1.

Remark. *This estimator has been proposed, but not analysed by Zelenev (2020). We could also use a related estimator:*

$$\hat{h}_n^{(K2)}(\xi_i, \xi_j) = \frac{\sum_{t=1}^n \sum_{s=1, s \neq t}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right) K\left(\frac{\rho_n^4 \hat{d}_{js}^2}{a_n}\right) A_{ts}}{\sum_{t=1}^n \sum_{s=1, s \neq t}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right) K\left(\frac{\rho_n^4 \hat{d}_{js}^2}{a_n}\right)}.$$

Or one by Zhang, Levina, and Zhu (2017), which uses the nearest-neighbour idea:

$$\hat{h}_n^{(NN1)}(\xi_i, \xi_j) = \frac{\tilde{h}_n^{(NN1)}(\xi_i, \xi_j) + \tilde{h}_n^{(NN1)}(\xi_j, \xi_i)}{2} \text{ where } \tilde{h}_n^{(NN1)}(\xi_i, \xi_j) = \frac{\sum_{t \in \mathcal{N}_i} A_{tj}}{\|\mathcal{N}_i\|}$$

where \mathcal{N}_i denotes the set of neighbours of i . They show that the optimal size of the neighbourhood, $\|\mathcal{N}_i\|$, should grow at the rate of $(n \ln(n))^{1/2}$. Or, we could use a nearest-neighbour approach in both inputs simultaneously (again, mentioned, and this time analysed, by Zelenev (2020)):

$$\hat{h}_n^{(NN2)}(\xi_i, \xi_j) = \frac{\sum_{t \in \mathcal{N}_i} \sum_{s \in \mathcal{N}_j} A_{ts}}{\|\mathcal{N}_i\| \|\mathcal{N}_j\|}.$$

In simulations all of these estimators perform similarly.

3.5 Optimal bandwidth

The linking function estimator relies on a bandwidth parameter chosen by the researcher. We propose a cross-validation procedure which allows choosing the bandwidth in an optimal way.

The idea is to choose a bandwidth for which \hat{h}_n best explains the observed network A , if we leave out A_{ij} when estimating A_{ij} . The reason for leaving out A_{ij} is that if we don't, we are trying to estimate A_{ij} using a set of observations which include A_{ij} , hence we can estimate it perfectly. We just need to choose $a_n = \infty$, this puts weight one on A_{ij} and zero on all other observations, leading to a perfect prediction of A but a poor choice of bandwidth. This issue of overfitting can be avoided by removing the observation A_{ij} from the model predicting A_{ij} .

We firstly define a leave-one-out version of \hat{h}_n :

$$\begin{aligned}\tilde{h}_n^-(\xi_i, \xi_j) &= \frac{\sum_{\substack{t=1 \\ t \neq i, j}}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right) A_{tj}}{\sum_{\substack{t=1 \\ t \neq i, j}}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right)} \\ \hat{h}_n^-(\xi_i, \xi_j) &= \frac{\tilde{h}_n^-(\xi_i, \xi_j) + \tilde{h}_n^-(\xi_j, \xi_i)}{2}.\end{aligned}$$

and then use it to obtain an estimate for the log-likelihood:

$$\ell(A, a_n) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} \log\left(\hat{h}_n^-(\xi_i, \xi_j)\right) + (1 - A_{ij}) \log\left(1 - \hat{h}_n^-(\xi_i, \xi_j)\right). \quad (16)$$

We choose a_n which maximises the above expression to be our optimal bandwidth:

$$a^{(opt)} = \max_{a_n} \ell(A, a_n). \quad (17)$$

3.6 Empirical distribution function estimator

The formation of matrix A is determined by an initial sample of ξ from F_0 and a linking probability between any pair of elements from ξ . We have defined a way to estimate the linking probabilities, but we still need a way to recreate the formation of ξ . This follows a very standard procedure, with one twist. Since the elements of ξ are i.i.d. from F_0 , we should be able to use a standard bootstrap (resample from the values from the original sample, with replacement) to create a bootstrap equivalent. The non-standard part is that ξ_i are unobserved. We get around it by resampling not directly from the set of ξ_i , but from the set of original nodes: we let each bootstrap node correspond to one of the original nodes and we assign the set of characteristics of a bootstrap node (ξ_i^*) to be equal to the set of characteristics of the resampled original node. The resulting distribution is an empirical distribution function \hat{F}_n defined as the CDF which corresponds to the probability mass function¹¹:

$$P(\xi_i^* = x) = \begin{cases} \frac{1}{n} & \text{if } x \in \{\xi_1, \dots, \xi_n\} \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

11. We would like to use the standard definition of an empirical distribution function:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\xi_i < x)$$

but unfortunately ξ_i are not observed, and may in general not be scalar, hence this notation doesn't apply.

Each bootstrap node corresponds to one of the original nodes and inherits its characteristics. The result is the same as if we formed the set of bootstrap characteristics ξ^* by resampling from the original set of characteristics ξ , with replacement.

3.7 Nonparametric network bootstrap procedure

We now describe the bootstrap procedure for $f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$. Just for this section, we introduce simplified notation for the matrix of estimated distances (D) and the matrix of estimated linking probabilities (H).

1. Calculate the distance between each pair of nodes $i, j \in \{1, 2, \dots, n\}$:

$$D_{ij} = \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n A_{ts} (A_{is} - A_{js}) \right)^2.$$

2. Calculate the optimal bandwidth parameter $a^{(opt)}$ as described in Eq. (17).
3. Calculate the probability of a link between each pair of nodes $i, j \in \{1, 2, \dots, n\}$:

$$\hat{h}_n(\xi_i, \xi_j) = \frac{1}{2} \left(\frac{\sum_{t \neq j}^n K\left(\frac{D_{it}}{a^{(opt)}}\right) A_{tj}}{\sum_{t \neq j}^n K\left(\frac{D_{it}}{a^{(opt)}}\right)} + \frac{\sum_{t \neq i}^n K\left(\frac{D_{jt}}{a^{(opt)}}\right) A_{ti}}{\sum_{t \neq i}^n K\left(\frac{D_{jt}}{a^{(opt)}}\right)} \right)$$

4. Calculate the density estimate of the original graph: $\hat{\rho}_n$ as described in Eq. (5).
5. For each $b = 1, \dots, B$:

- (a) draw an i.i.d. sample $\{\xi_{b,i}^*\}_{i=1}^n$ of size n from \hat{F}_n , i.e. resample from the original set of nodes $\{1, 2, \dots, n\}$ with equal probabilities and with replacement, then assign the unobserved characteristics of the bootstrap node to be the same as the unobserved characteristics of its corresponding original node. Let $\xi_b^* = (\xi_{b,1}^*, \dots, \xi_{b,n}^*)'$.
- (b) draw $\eta_{b,ij}^* \stackrel{iid}{\sim} \mathcal{U}[0, 1]$ for $1 \leq i \leq j \leq n$.
- (c) form the bootstrap adjacency matrix A_b^* :

$$A_{b,ij}^* = A_{b,ji}^* = \mathbf{1} \left(\hat{h}_n(\xi_{b,i}^*, \xi_{b,j}^*) \geq \eta_{b,ij}^* \right) \quad (19)$$

$$A_{b,ii}^* = 0. \quad (20)$$

(e.g. if $\xi_{b,i}^* = \xi_t$, $\xi_{b,j}^* = \xi_s$ then $\hat{h}_n(\xi_{b,i}^*, \xi_{b,j}^*) = \hat{h}_n(\xi_t, \xi_s)$).

(d) calculate the object of interest on the bootstrap adjacency matrix:

$$f_n(A_b^*) \equiv f_n \left(A \left(\hat{h}_n(\xi_b^*), \eta_b^* \right), \hat{\rho}_n, \hat{F}_n \right). \quad (21)$$

6. Form a $(1-\alpha)\%$ confidence interval for $f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$ by taking the interval between $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of $\{f_n(A_b^*)\}_{b=1}^B$.

For a description of how we used this procedure in simulations see Section 5. For the codes used in simulation see Appendix B.1.

3.8 Appropriate notions of convergence

The aim of the remainder of this section is to introduce definitions used in the proofs that the bootstrap procedure works, in the sense that the asymptotic distribution of the bootstrap statistic is the same as the asymptotic distribution of the original statistic and, as a result, the bootstrap confidence intervals asymptotically achieve correct coverage.

We choose f_n that has a distribution limit (usually a normal random variable), i.e. we assume $J_n(t, h_{0,n}, F_0) \Rightarrow J(t, w_0, F_0)$ for some non-degenerate distribution $J(t, w_0, F_0)$, where “ \Rightarrow ” denotes weak convergence. One convenient way to characterise weak convergence is through the following distance between measures: let P and Q be probability measures on a common metric space S equipped with a distance d_S and let

$$f(S) = \left\{ f : S \rightarrow \mathbb{R} : |f(x) - f(y)| \leq d_S(x, y), \sup_{x \in S} |f(x)| \leq 1 \right\}$$

be the set of (Lipschitz) continuous and bounded real-valued functions on S , then:

$$d_W(P, Q) \equiv \sup_{f \in f(S)} \left| \int f(x) dP(x) - \int f(x) dQ(x) \right|.$$

It can be shown¹² that $P_n \Rightarrow P$ if and only if $d_W(P_n, P) \rightarrow 0$ as $n \rightarrow \infty$.

In order to prove consistency of the bootstrap procedure we would like to show that the distribution $J_n(t, \hat{h}_n, \hat{F}_n)$, the bootstrap equivalent of $J_n(t, h_{0,n}, F_0)$, achieves the same asymptotic distribution $J(t, w_0, F_0)$. Unfortunately, the concept of weak convergence cannot be applied directly to the bootstrap statistic because both the bootstrap distribution \hat{F}_n and the estimator \hat{h}_n are random functions depending on the realisation of ξ , hence $J_n(t, \hat{h}_n, \hat{F}_n)$ and $d(J_n(t, \hat{h}_n, \hat{F}_n), J(t, w_0, F_0))$ are also random. To proceed, we define two new concepts which generalise weak convergence to

¹². See e.g. Proposition (M) in Chapter I of Hahn (1993)

account for this randomness:

Definition 1. We say that P_n converges weakly to P almost surely, denoted by $P_n \xrightarrow{a.s.} P$, if $d_W(P_n, P) \xrightarrow{a.s.} 0$. For $X_n \sim P_n$, $X \sim P$ we write $X_n \xrightarrow{d} X$ almost surely.

Analogously,

Definition 2. we say that P_n converges weakly to P in probability, denoted by $P_n \xrightarrow{P} P$, if $d_W(P_n, P) \xrightarrow{P} 0$. For $X_n \sim P_n$, $X \sim P$ we write $X_n \xrightarrow{d} X$ in probability.

Although this is a weaker requirement than regular convergence in distribution, Gine and Zinn (1990), who introduced the concept of weak convergence in probability, show that this notion is sufficient for the construction of asymptotically correct confidence intervals.

3.9 Distributions of intermediate terms

Since the elements of A exhibit dependence (e.g. A_{ij} and A_{jk} both depend on ξ_j), it is relatively difficult to work with $f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$ directly. We can instead consider its expectation taken with respect to the Bernoulli trials with probabilities determined by $h_{0,n}(\xi)$:

$$\tilde{f}_n(h_{0,n}(\xi), \rho_n, F_0) \equiv E_{h_{0,n}}(f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)|\xi) \quad (22)$$

where we have taken the expectation over η and the remaining object becomes a function of the i.i.d. ξ_i .

Let \tilde{J}_n denote the distribution of \tilde{f}_n :

$$\tilde{J}_n(t, h_{0,n}, F_0) = P_{F_0}(\tilde{f}_n(h_{0,n}(\xi), \rho_n, F_0) \leq t). \quad (23)$$

The limit of \tilde{J}_n is easier to find than that of J_n , and the limits coincide if we can show that $f_n - \tilde{f}_n$ is negligible.

Remark. We often work with conditional expectations and switch between variables that follow different distributions (e.g. the true distribution F_0 and the estimated empirical distribution \hat{F}_n). When we think it is beneficial to clarify, we add subscripts to the expectation operator indicating with respect to which distribution we are taking the expectation. For example, $E_{h_{0,n}}(f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)|\xi) = \int f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)d\eta$ indicates that we are taking expectation with respect to the independent Bernoulli trials with probabilities determined by $h_{0,n}$ while $E_{h_{0,n}, F_0}(f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0))$ denotes the expectation with respect to both the Bernoulli trials and the true distribution of ξ . The latter can also be written as $E_{F_0}(\tilde{f}_n(h_{0,n}(\xi), \rho_n, F_0))$,

where $\tilde{f}_n(h_{0,n}(\xi), \rho_n, F_0)$ has already been integrated over the Bernoulli trials, hence its randomness only comes from F_0 , the true distribution of ξ .

To illustrate the need for these intermediate terms we introduce an important class of statistics for which \tilde{f}_n take the form of a U-statistic.

Definition 3. Let ι be a set of m unique nodes, ξ be an n -dimensional vector of i.i.d. draws from a distribution F and η an n -dimensional vector of independent draws from $\mathcal{U}[0, 1]$, and $\rho \in [0, 1]$ be the sparsity level. Denote the adjacency matrix on the subgraph with nodes in ι by $A(h(\xi(\iota)), \eta(\iota))$.

Let $g(A(\iota)) : \{0, 1\}^{\binom{m}{2}} \rightarrow \mathbb{R}$ be a non-degenerate symmetric function from a subgraph on $m < \infty$ nodes to the real line such that $E_{h_{0,n}, F_0}(g(A(\iota))) = \theta$, where θ is a parameter of interest. We can estimate θ on the whole network A by

$$\hat{\theta} = \frac{1}{\binom{n}{m}} \sum_{1 \leq \iota_1 < \iota_2 < \dots < \iota_m \leq n} g(A(\iota)). \quad (24)$$

To get the corresponding $f_n(A(h(\xi), \eta), \rho, F)$ with a well-defined distribution we recentre and normalise the above expression:

$$f_n^U(A(h(\xi), \eta), \rho, F) = \frac{\sqrt{n}}{\binom{n}{m} \rho^{\tau(g)}} \sum_{1 \leq \iota_1 < \iota_2 < \dots < \iota_m \leq n} (g(A(h(\xi(\iota)), \eta(\iota))) - E_{h, F}(g(A(h(\xi(\iota)), \eta(\iota)))) \quad (25)$$

and

$$\tilde{f}_n^U(h(\xi), \rho, F) = \frac{\sqrt{n}}{\binom{n}{m} \rho^{\tau(g)}} \sum_{1 \leq \iota_1 < \iota_2 < \dots < \iota_m \leq n} (\tilde{g}(h(\iota)) - E_F(\tilde{g}(h(\iota)))) \quad (26)$$

where $\tilde{g}(h(\xi(\iota))) \equiv E_h(g(A(h(\xi(\iota)), \eta(\iota))) | \xi(\iota))$ and we choose $\tau(g)$ to get a normalisation for which there exists a non-degenerate bounded function \tilde{g} such that

$$\frac{\tilde{g}(h_{0,n}(\iota))}{\rho_n^{\tau(g)}} = \tilde{g}(w_0(\iota)) + O(\rho_n).$$

The choice of $\tau(g)$ is quite simple: it is the smallest number of ones such that $g(\cdot)$ evaluated at a vector of $\tau(g)$ ones and $\binom{m}{2} - \tau(g)$ zeros is non-zero. More importantly, the normalisation is not important for practical applications. We introduce it in the definition because it is necessary to get a well-defined asymptotic distribution (see Theorem 3, without the normalisation the limiting value of the $\hat{\theta}$ would be 0), but we do not need it if our interest is in constructing a confidence interval for θ . To see this, suppose θ and $\hat{\theta}$ are as above and let the bootstrap equivalent of the estimator be $\hat{\theta}_b^* = \frac{1}{\binom{n}{m}} \sum_{1 \leq \iota_1 < \dots < \iota_m \leq n} g(A_b^*(\iota))$. We can calculate the estimator for B bootstrap adjacency matrices and find a confidence interval for θ as $[\hat{\theta}_L^*, \hat{\theta}_U^*]$ where $\hat{\theta}_L^* = q_{\frac{\alpha}{2}} \left(\left\{ \hat{\theta}_b^* \right\}_{b=1}^B \right)$ and

$\hat{\theta}_U^* = q_{1-\frac{\alpha}{2}} \left(\left\{ \hat{\theta}_b^* \right\}_{b=1}^B \right)$. This way we get:

$$\begin{aligned} 1 - \alpha &\simeq P \left(\hat{\theta}_L^* < \theta < \hat{\theta}_U^* \right) \\ &= P \left(\frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \hat{\theta}_U^* \right) < \frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \theta \right) < \frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \hat{\theta}_L^* \right) \right) \\ &\simeq P \left(\frac{\sqrt{n}}{\hat{\rho}_n^{*\tau(g)}} \left(\hat{\theta} - \hat{\theta}_U^* \right) < \frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \theta \right) < \frac{\sqrt{n}}{\hat{\rho}_n^{*\tau(g)}} \left(\hat{\theta} - \hat{\theta}_L^* \right) \right). \end{aligned}$$

The above confidence interval for θ is equivalent to a confidence interval for $\frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \theta \right)$ of the form $\left[\frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \hat{\theta}_U^* \right), \frac{\sqrt{n}}{\hat{\rho}_n^{\tau(g)}} \left(\hat{\theta} - \hat{\theta}_L^* \right) \right]$ and is closely approximated by $\left[\frac{\sqrt{n}}{\hat{\rho}_n^{*\tau(g)}} \left(\hat{\theta} - \hat{\theta}_U^* \right), \frac{\sqrt{n}}{\hat{\rho}_n^{*\tau(g)}} \left(\hat{\theta} - \hat{\theta}_L^* \right) \right]$ (we could also use the quantiles of $f_n^U \left(A \left(\hat{h}_n \left(\xi^* \right), \eta^* \right), \hat{\rho}_n, \hat{F}_n \right)$ or $f_n^U \left(A \left(\hat{h}_n \left(\xi^* \right), \eta^* \right), \hat{\rho}_n^*, \hat{F}_n \right)$ directly). The consistency of $\hat{\rho}_n$ and $\hat{\rho}_n^*$ for ρ_n follows from the proof of Theorem A.2.

Example. We now show the relation between f_n and \tilde{f}_n on an example. Suppose $m = 3$, e.g. $\iota = (1, 2, 3)$, and the function g only depends on two entries in A^{13} : $g(A(\iota)) = g(A_{1,2}, A_{2,3})$.

Conditional on ξ , the Bernoulli trials that determine the entries of A are independent. Hence, for example, $P(A_{ij} = 1, A_{jk} = 1 | \xi) = P(A_{ij} = 1 | \xi) P(A_{jk} = 1 | \xi) = h_{0,n}(\xi_i, \xi_j) h_{0,n}(\xi_j, \xi_k)$. It follows that

$$g(A_{1,2}, A_{2,3}) | \xi = \begin{cases} g(0, 0) & \text{with probability } (1 - h_{0,n}(\xi_1, \xi_2))(1 - h_{0,n}(\xi_2, \xi_3)) \\ g(0, 1) & \text{with probability } (1 - h_{0,n}(\xi_1, \xi_2))h_{0,n}(\xi_2, \xi_3) \\ g(1, 0) & \text{with probability } h_{0,n}(\xi_1, \xi_2)(1 - h_{0,n}(\xi_2, \xi_3)) \\ g(1, 1) & \text{with probability } h_{0,n}(\xi_1, \xi_2)h_{0,n}(\xi_2, \xi_3). \end{cases}$$

The conditional expectation is a function of $h_{0,n}(\xi(\iota))$:

$$\begin{aligned} E(g(A_{1,2}, A_{2,3}) | \xi) &\equiv \tilde{g}(h_{0,n}(\xi_1, \xi_2), h_{0,n}(\xi_2, \xi_3)) \\ &= g(0, 0)(1 - h_{0,n}(\xi_1, \xi_2))(1 - h_{0,n}(\xi_2, \xi_3)) + g(0, 1)(1 - h_{0,n}(\xi_1, \xi_2))h_{0,n}(\xi_2, \xi_3) \\ &\quad + g(1, 0)h_{0,n}(\xi_1, \xi_2)(1 - h_{0,n}(\xi_2, \xi_3)) + g(1, 1)h_{0,n}(\xi_1, \xi_2)h_{0,n}(\xi_2, \xi_3). \end{aligned}$$

If $g(0, 0) \neq 0$, the first term on the right is $O(1)$ and dominates over the next terms. In this case we choose $\tau(g) = 0$. If $g(0, 0) = 0$ but $g(0, 1) \neq 0$ or $g(1, 0) \neq 0$, the dominating term is proportional

13. For simplicity in this example we used a function which is not necessarily symmetric. Before plugging it into Eq. (25) we should symmetrise it in the following way:

$$\bar{g}(A(\iota)) = \frac{g(A_{1,2}, A_{2,3}) + g(A_{1,2}, A_{1,3}) + g(A_{1,3}, A_{2,3}) + g(A_{2,3}, A_{1,2}) + g(A_{1,3}, A_{1,2}) + g(A_{2,3}, A_{1,3})}{6}.$$

to $(1 - h_{0,n}(\xi_i, \xi_j))h_{0,n}(\xi_j, \xi_k) = (1 - \rho_n w_0(\xi_i, \xi_j))\rho_n w_0(\xi_j, \xi_k) = O(\rho_n)$, hence we choose $\tau(g) = 1$ to normalise it. If $g(0,0) = g(0,1) = g(1,0) = 0$ but $g(1,1) \neq 0$, the dominating term is proportional to $h_{0,n}(\xi_i, \xi_j)h_{0,n}(\xi_j, \xi_k) = \rho_n^2 w_0(\xi_i, \xi_j)w_0(\xi_j, \xi_k) = O(\rho_n^2)$, hence the correct normalisation is $\tau(g) = 2$. When $\rho_n \rightarrow 0$, only the dominating term influences the limiting behaviour. We call this dominating term \tilde{g} . In this example:

$$\tilde{g}(h_{0,n}(\xi_1, \xi_2), h_{0,n}(\xi_2, \xi_3)) = \begin{cases} g(0,0) & \text{if } g(0,0) \neq 0 \\ g(0,1)h_{0,n}(\xi_2, \xi_3) + g(1,0)h_{0,n}(\xi_1, \xi_2) & \text{if } g(0,0) = 0, g(0,1) \neq 0, g(1,0) \neq 0 \\ g(1,1)h_{0,n}(\xi_1, \xi_2)h_{0,n}(\xi_2, \xi_3) & \text{if } g(0,0) = g(0,1) = g(1,0) = 0, g(1,1) \neq 0. \end{cases}$$

In all cases we have:

$$\frac{\tilde{g}(h_{0,n}(\xi_1, \xi_2), h_{0,n}(\xi_2, \xi_3))}{\rho_n^{\tau(g)}} = \tilde{g}(w_0(\xi_1, \xi_2), w_0(\xi_2, \xi_3)) + O(\rho_n)$$

where the first term is $O(1)$ and does not depend on the sample size n .

We will later do Taylor expansion of \tilde{g} , for which it is interesting to note that regardless of whether g is a nice function (continuous, differentiable, etc) of A , \tilde{g} is (infinitely many times) continuously differentiable in $h_{0,n}$ and has bounded derivatives. Taking a derivative of \tilde{g} with respect to $h_{0,n}$ lowers the power on the $h_{0,n}$ terms by one, hence if $\tilde{g} \sim \rho_n^{\tau(g)}$ then $\tilde{g}' \sim \rho_n^{\tau(g)-1}$ and $\frac{\tilde{g}'}{\rho_n^{\tau(g)}} = \frac{1}{\rho_n}$.

Then

$$\tilde{f}_n(h_{0,n}(\xi), \rho_n, F_0) = \frac{\sqrt{n}}{\binom{n}{m} \rho_n^{\tau(g)}} \sum_{1 \leq \iota_1 < \iota_2 < \dots < \iota_m \leq n} (\tilde{g}(h_{0,n}(\iota)) - E_{F_0}(\tilde{g}(h_{0,n}(\iota)))) \quad (27)$$

We can think of the above as a function of the i.i.d. ξ . If g is symmetric, so is $\tilde{g}(h_{0,n}(\cdot))$, and the $\tilde{f}_n(h_{0,n}(\xi), \rho_n, F_0)$ takes the form of a (normalised) U -statistic, for which we have results such as LLN and CLT. Hence it's much easier to work with than the original $f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$ (which was not a U -statistic due to the dependence in A).

Remark. It is usually not the case that \tilde{f}_n has the same form as f_n with A_{ij} replaced with $h_{0,n}(\xi_i, \xi_j)$, but it can happen in some special cases. One such example are motif densities, for which the g function is a product of terms of the form A_{ij} (if the motif has an edge between nodes i and j) and $(1 - A_{ij})$ (if the edge is supposed to be missing). For example, if the motif of interest is a triangle, we have $g(A_{ij}, A_{jk}, A_{ki}) = A_{ij}A_{jk}A_{ki}$. This is 1 if all inputs are equal to 1 and 0 in

all other cases, hence $\tilde{g}(h_{0,n}(\xi_i, \xi_j), h_{0,n}(\xi_j, \xi_k), h_{0,n}(\xi_k, \xi_i)) = h_{0,n}(\xi_i, \xi_j)h_{0,n}(\xi_j, \xi_k)h_{0,n}(\xi_k, \xi_i)$.

Remark. Because in the proof of Theorem 3 we rely on a CLT for U-statistics applied to \tilde{f}_n instead of f_n , the normalisation by $\frac{1}{\rho_n^{\tau(g)}}$ is chosen to balance the rate of growth of the variance of \tilde{g} rather than g . Take a simple example of $g(A_{ij}) = A_{ij}$. Then $\tau(g) = 1$ and:

$$E \left(\left(\frac{g(A_{ij})}{\rho_n^{\tau(g)}} \right)^2 \right) = \frac{E(h_{0,n}(\xi_i, \xi_j))}{\rho_n^{2\tau(g)}} = \frac{1}{\rho_n^{\tau(g)}} \rightarrow \infty$$

$$E \left(\left(\frac{E(g(A_{ij})|\xi)}{\rho_n^{\tau(g)}} \right)^2 \right) = \frac{(E(h_{0,n}(\xi_i, \xi_j)))^2}{\rho_n^{2\tau(g)}} = O(1).$$

4 Main results

In this section we state our main results which characterise the conditions under which the linking function estimator and our entire bootstrap procedure are consistent.

4.1 Consistency of the linking function estimator

We start by showing the uniform consistency of the linking function estimator.

Assumption 1 (The Assumptions for Uniform Consistency of the Linking Function Estimator).

We make the following assumptions:

1.1 $\frac{1}{\rho_n} = o \left(\sqrt{\frac{n}{\log(n)}} \right)$.

1.2 Let $N(\xi_j, \delta) = \{\xi_k : \sup_{\xi_t} |w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta\}$ denote the neighbourhood of ξ_j of size δ and let $\omega(\delta) = \inf_{\xi_j \in \text{Supp}(\xi_j)} P(\xi_k \in N(\xi_j, \delta) | \xi_j)$. There exist some $\alpha, C > 0$ such that $\omega(\delta) \geq \left(\frac{\delta}{C}\right)^\alpha$ for all $\delta > 0$.

1.3 $K(\cdot)$ is a kernel function which is

- a continuous bounded probability density function (non-negative: $K(u) \geq 0$, integrates to 1: $\int K(u)du = 1$),
- non-zero on a bounded support: there exists a $D \in \mathbb{R}$ such that $\forall |u| > D : K(u) = 0$,
- positive close to 0: there exist positive constants C_1, C_2 such that $K(u) \geq C_1$ whenever $|u| \leq C_2$,
- Lipschitz continuous: there exists $C > 0$ such that $|K(u) - K(v)| \leq C|u - v|$.

1.4 The bandwidth can be written as $a_n = \rho_n^4 b_n$ for some $b_n = o(1)$ and $\frac{1}{b_n} = o \left(\left(\frac{n\rho_n^2}{\log(n)} \right)^{\frac{\alpha}{1+2\alpha}} \right)$.

We now discuss the assumption and give some intuition.

Assumption 1.1 is our sparsity assumption. It gives a lower bound on how sparse a model can be for our estimator to remain consistent. Intuitively, the only informative observations are the links, and their number grows at a slower rate than the sample size: we expect on average $n\rho_n$ links in a sample of n individuals. Our model works well if the number of links increases at a rate faster than $\sqrt{n \log(n)}$. This is analogous to the assumption in Zhang, Levina, and Zhu (2017), with the exception that they model the increasing difficulty in estimation with n by allowing $\delta(n) \rightarrow 0$ instead of having a sparsity parameter $\rho_n \rightarrow 0$.

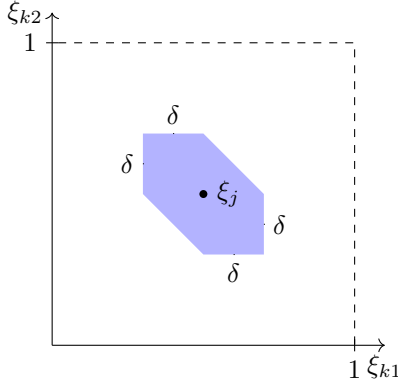
Assumption 1.2 ensures that the neighbourhoods for all observations are sufficiently large. We can think of it as a “continuity” condition for w_0 , analogous to that assumed by Auerbach (2022): for all $\delta > 0$:

$$\inf_{\xi_j \in \text{Supp}(\xi_j)} P_{\xi_k \sim F_0} \left(\sup_{\xi_t} |w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta \mid \xi_j \right) \geq \left(\frac{\delta}{C} \right)^{\frac{1}{\alpha}}.$$

i.e., for each $\xi_j \in \text{Supp}(\xi_j)$ there exists a sufficiently large positive measure of ξ_k with very similar friendship groups: such that $|w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta$ holds for all ξ_t . The consequences of this assumption are similar to those of the Piecewise-Lipschitz assumption in Definition 2 of Zhang, Levina, and Zhu (2017) (in the proof of Theorem 2 we show that under Assumption 1.2 $\forall \varepsilon > 0 \exists K < \infty$ such that $\text{Supp}(\xi_i)$ can be split into K disjoint regions, each of size at least $(\frac{\varepsilon}{C})^{\frac{1}{\alpha}}$, such that for any two points u, v which fall in the same region we have $\sup_{\xi_t} |w_0(\xi_t, u) - w_0(\xi_t, v)| \leq 2\varepsilon$ and there exists a set of points a_k , each from a different region, such that if $k \neq j$ we have $\sup_{\xi_t} |w_0(\xi_t, a_k) - w_0(\xi_t, a_j)| > \varepsilon$).

The example below shows that we can also think of this assumption as ensuring that the distribution of ξ_i is bounded away from zero while w_0 is sufficiently smooth:

Example. For an example of Assumption 1.2 and some intuition on what α means, suppose that $\xi_i \in [0, 1]^2$ and $w_0(\xi_i, \xi_j) = \xi_i \cdot \xi_j$. ξ_k satisfies $\sup_{\xi_t} |w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta$ if it falls within a region in $[0, 1]^2$ centred at ξ_j with a radius proportional to δ and area proportional to δ^2 . If ξ_j is at least δ away from all boundaries of the support, the region has the following shape:



and area $3\delta^2$. If the point ξ_j is closer to the boundary, we get the subset of this region which overlaps with $[0, 1]^2$. The smallest area possible is $\frac{\delta^2}{2}$. If the distribution of ξ_i is uniformly bounded away from zero, the measure of ξ_j that satisfy the condition is at least proportional to δ^2 , hence $\omega(\delta) \geq \left(\frac{\delta}{C}\right)^2$. A sufficient condition for our assumption is if the distribution of ξ_i is uniformly bounded away from zero and w_0 is piecewise Lipschitz. If w_0 is a well-behaved function and $\text{Supp}(\xi_i) \subset \mathbb{R}^d$, we would expect $\frac{1}{\alpha} = d$. We can think of $\frac{1}{\alpha}$ as a measure of complexity of the feature space: the more complex ξ_i are the harder the estimation.

Assumption 1.3 gives a list of fairly standard assumptions on the form of the kernel function. These are quite strong, but the kernel is chosen by the researcher and many of the standard kernels (e.g. the Epanechnikov kernel: $K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}(|u| < 1)$ or the triangular kernel: $K(u) = (1 - |u|)\mathbf{1}(|u| < 1)$) satisfy all the requirements.

Assumption 1.4 specifies the range of bandwidths for which the method performs well. The bandwidth a_n is a product of ρ_n^4 , which cancels out the normalisation in $\left(\rho_n^2 \hat{d}_{it}\right)^2$, and $b_n \rightarrow 0$ which ensures that $\frac{\rho_n^4 \hat{d}_{it}^2}{a_n} = \frac{\hat{d}_{it}^2}{b_n} \rightarrow \infty$ for all $i \neq t$. ρ_n can be estimated and b_n is chosen by the researcher. As the effective dimension of the support of ξ_i increases, i.e. α decreases, the estimation becomes more difficult and we need b_n to go to zero at a slower rate.

Theorem 1. *Under Assumption 1:*

$$\max_{i,j} \left| \frac{\hat{h}_n(\xi_i, \xi_j) - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| \xrightarrow{p} 0.$$

Remark. Notice that we can write $h_{0,n}(\xi_i, \xi_j) = \rho_n w_0(\xi_i, \xi_j)$, decomposing the linking function into a bounded function w_0 which does not depend on n and the sparsity $\rho_n \rightarrow 0$. Without the normalisation by $\frac{1}{\rho_n}$, the difference $\hat{h}_n - h_{0,n}$ would trivially go to zero because both components go to zero at the rate ρ_n . In the statement of Theorem 1 we normalise by $\frac{1}{\rho_n}$ to show that, even after removing the trend to zero, the estimate of the linking function approaches its true value.

Remark. Note on notation: we use $\max_{i,j}$ to refer to maximising over indices in a specific sample of size n : it is a shorthand notation for $\max_{i,j \in \{1,2,\dots,n\}}$. We alter use \max_{ξ_i} which refers to maximising over all $\xi_i \in \text{Supp}(\xi_i)$, i.e. all possible values in the support, not the set of realised values in a specific sample.

Remark. Theorem 1 shows uniform convergence of $\hat{h}_n(\xi_i, \xi_j)$ to $h(\xi_i, \xi_j)$, where “uniform” refers to convergence over all pairs of nodes in the original graph. The reason why we do not look at uniformity over all general points in the underlying sample space of ξ_i ¹⁴ is because our estimator of distance $\hat{d}_{ij}^{(2)}$ is defined in terms of similarity of friendship groups, hence it can only be estimated for one of the observed individuals. In our procedure we never estimate ξ_i directly, we don’t put strong assumptions on the space it comes from, and we don’t have a way of estimating \hat{d} , and hence \hat{h}_n , at a general point (u, v) outside of our realised set of observed individuals.

However, the results we show later in Theorem 2 can be seen as an extension of Theorem 1 to the whole support of ξ_i : under the assumption Assumption 1.2, for any $\xi_i \in \text{Supp}(\xi_i)$, if n is high enough, with high probability we can observe ξ_j similar enough to ξ_i that \hat{h}_n evaluated at ξ_j provides a good approximation to ξ_i and the frequencies with which we observe different values of \hat{h}_n is representative of the frequencies of similar values of the true $h_{0,n}$ over the support of ξ_i .

4.2 Consistency of the bootstrap procedure

4.2.1 Convergence of bootstrap network in terms of Wasserstein distance

To show that the distribution of the bootstrap network approaches that of the original network we follow the approach from Levin and Levina (2019) (see their Section 4 for more details and motivation). We start by defining an appropriate notion of convergence between network distributions. Firstly, let the graph matching distance be the proportion of edges that differ between two graphs after their vertices have been aligned to minimise the number of such differences:

Definition 4 (Graph matching distance). Let A_1, A_2 be two $n \times n$ adjacency matrices, Π_n be the set of $n \times n$ permutation matrices and let $\|A\|_{1,1} = \sum_{i=1}^n \sum_{j=1}^n |A_{ij}|$. The graph matching distance is:

$$d_{GM}(A_1, A_2) = \min_{P \in \Pi_n} \binom{n}{2}^{-1} \frac{\|A_1 - PA_2P'\|_{1,1}}{2}. \quad (28)$$

Equipped with a distance between graphs we can define a distance between two distributions over graphs by using the Wasserstein distance:

14. For example, if $\xi_i \sim \mathcal{U}[0, 1]$ we could be interested in showing $\sup_{u,v \in [0,1]} \left| \frac{\hat{h}_n(u,v) - h_{0,n}(u,v)}{\rho_n} \right| \xrightarrow{p} 0$.

Definition 5. Let A_1, A_2 be the adjacency matrices of two random graphs on n vertices and let $\Gamma(A_1, A_2)$ be the set of all couplings of A_1 and A_2 (i.e. all joint distributions with marginal distributions matching those of A_1 and A_2). For $p \geq 1$ the Wasserstein p -distance is given by:

$$W_p(A_1, A_2) = \inf_{\nu \in \Gamma(A_1, A_2)} \left(\int d_{GM}^p(A_1, A_2) d\nu \right)^{\frac{1}{p}}. \quad (29)$$

Theorem 2. Let A be the observed adjacency matrix, let H be another adjacency matrix drawn independently from the distribution of A and let A^* be a bootstrap adjacency matrix derived from A . Under Assumption 1:

$$W_p^p(A^*, H) = o_p(\rho_n). \quad (30)$$

The graph matching distance is an upper bound on the cut metric, which in turn metrises convergence of subgraph densities, hence Theorem 2 implies that all subgraph densities of A^* converge to the same limit as those of H , proving that the bootstrap network distribution converges to the original network's distribution.

Remark. As noted by Levin and Levina (2019), this notion of convergence is not sufficient to ensure that $f(A^*)$ converges to the same distribution as $f(H)$ for a general function $f(\cdot)$.

4.2.2 Consistency of bootstrap for U-statistics

Because the result of Theorem 2 is not sufficient to guarantee that the distribution of a function of A^* is close to the distribution of the same function of A , we show this directly for an important class of functions with known limiting distributions.

Throughout this argument we use stars to denote the bootstrap equivalent, e.g. $\xi_i^* \sim \hat{F}_n$.

Definition 6. Set \mathcal{M}_m : Let \mathcal{M}_m be the set of all possible multisets¹⁵ of cardinality m with elements from $\{1, 2, \dots, m\}$.

That is, \mathcal{M}_m contains all possible combinations of index numbers from 1 to m that are of length m and can be all unique or have any value repeated any number of times.

Theorem 3. Let $f_n^U(A(h(\xi), \eta), \rho, F)$ be as in Eq. (25). There exists a normalisation $\tau(g)$ ¹⁶ and a function $\tilde{g}: \text{Supp}(\xi)^m \rightarrow \mathbb{R}$ such that $\frac{\tilde{g}(h_{0,n}(\xi(\iota)))}{\rho_n^{\tau(g)}} = \tilde{g}(w_0(\xi(\iota))) + O(\rho_n)$ and $0 < E(|\tilde{g}(w_0(\xi(j)))|) <$

15. A multiset is like a set but allows for repeated elements.

16. For $m = 2$, if $g(0) \neq 0$ we set $\rho_n^{-\tau(g)} = 1$, $\tilde{g}(w_0(\xi_i, \xi_j)) = g(0)$ and if $g(0) = 0$ but $g(1) \neq 0$ we set $\rho_n^{-\tau(g)} = \frac{1}{\rho_n}$ and $\tilde{g}(w_0(\xi_i, \xi_j)) = g(1)w_0(\xi_i, \xi_j)$. More generally, for $m \geq 2$, $\rho_n^{-\tau(g)} = \frac{1}{\rho_n^k}$ where k is the smallest number of ones such that $g(\cdot)$ evaluated at a vector of k ones and $\binom{m}{2} - k$ zeros is non-zero.

∞ for all $j \in \mathcal{M}_m$. If Assumption 1 holds and:

$$\begin{aligned} \text{Var}_{F_0}(E_{F_0}(\tilde{g}(w_0(\xi(t)))|\xi_{L_1})) &\equiv \sigma_1^2 > 0 \\ \frac{n}{\binom{n}{m}\rho_n^{\tau(g)}} &\rightarrow 0 \end{aligned}$$

then

1. $f_n^U(A(h_{0,n}(\xi), \eta), \rho_n, F_0) \xrightarrow{d} N(0, m^2\sigma_1^2)$ and
 $f_n^U(A(\hat{h}_n(\xi^*), \eta^*), \hat{\rho}_n, \hat{F}_n) \xrightarrow{d} N(0, m^2\sigma_1^2)$ in probability.
2. $\sup_t \left| P\left(f_n^U(A(\hat{h}_n(\xi^*), \eta^*), \hat{\rho}_n, \hat{F}_n) \leq t\right) - P\left(f_n^U(A(h_{0,n}(\xi), \eta), \rho_n, F_0) \leq t\right) \right| \xrightarrow{P} 0$.

Theorem 3 shows that for statistics which can be represented as in Definition 3 and are non-degenerate (i.e. $\sigma_1^2 \neq 0$): 1. the limiting distribution in probability of the bootstrap statistic is asymptotically normal and the same as the limiting distribution of the original statistic and 2. the bootstrap is consistent, in the sense that the finite-sample distribution of the bootstrap statistic approaches the finite-sample distribution of the original statistic as the sample size increases.

In the above theorem we add two new assumptions on top of those in Assumption 1. The first one, $\sigma_1^2 \neq 0$, restricts our attention to non-degenerate U-statistics. Levin and Levina (2019) claim that in the case of degenerate U-statistics the approximation error is of a comparable size to the leading term, implying that their bootstrap cannot recover the distributions of degenerate U-statistics. As explained in Serfling (2009) section 5.5, in the degenerate case the correct normalisation would be of the form $\frac{n^{\frac{c}{2}}}{\binom{n}{m}\rho_n^2}$ for some $c \geq 2$ and we would expect a more complicated limiting distribution than normal. In that case $f_n - \tilde{f}_n = O\left(\sqrt{\frac{n^c}{\binom{n}{m}\rho_n^2}}\right)$, which could go to zero sufficiently fast to remain negligible. We suspect that recovering distributions of degenerate U-statistics could still be possible with our method but we leave the detailed analysis for future work.

The other condition: $\frac{n}{\binom{n}{m}\rho_n^{\tau(g)}} \rightarrow 0$ gives a restriction on the allowed level of sparsity. We require $\frac{1}{\rho_n} = o\left(n^{\frac{m-1}{\tau(g)}}\right)$, where $\tau(g) \in \{0, 1, \dots, \binom{m}{2}\}$. For sufficiently large $\tau(g)$ this condition may be stronger than Assumption 1.1. This is not surprising: large $\tau(g)$ means that the function g takes non-zero values only for very rare events, and these events are even less common in sparser graphs. Hence to be able to maintain consistency we need to restrict the allowed level of sparsity.

This condition is needed to ensure that the $f_n - \tilde{f}_n$ term does not affect the limiting distribution. If $\frac{n}{\binom{n}{m}\rho_n^{\tau(g)}} = O(1)$, the limit of this term would affect the resulting distribution and the overall limit would be the current one plus the limit of this adjustment term. If $\frac{n}{\binom{n}{m}\rho_n^{\tau(g)}} \rightarrow \infty$ this adjustment term would dominate the asymptotic behaviour. In that case we would need to use normalisation by $\frac{n}{\binom{n}{m}\rho_n^{\frac{\tau(g)}{2}}}$. The currently dominating term under the new normalisation would go to zero. Deriving

the distribution of the new dominating term is difficult due to a high level of dependence between the elements of A .

Remark. Green and Shalizi (2022) specify the maximal allowed level of sparsity in two cases: when the motif is acyclic they assume $\frac{1}{\rho_n} = o(n)$ and for a general motif they require $\frac{1}{\rho_n} = o\left(n^{\frac{1}{2m}}\right)$ for the empirical graphon and the weaker condition of $\frac{1}{\rho_n} = o\left(n^{\frac{2}{m}}\right)$ for a general linking function estimator, e.g. their histogram graphon.

These conditions are weakly stronger than our $\frac{1}{\rho_n} = o\left(n^{\frac{m-1}{\tau(g)}}\right)$: when $g(\cdot)$ corresponds to an acyclic motif we have $\tau(g) \leq m - 1$ ¹⁷, hence $\frac{m-1}{\tau(g)} \geq 1$; when $g(\cdot)$ corresponds to a general motif we have $\tau(g) \leq \binom{m}{2}$ ¹⁸, hence $\frac{m-1}{\tau(g)} \geq \frac{2}{m} \geq \frac{1}{2m}$.

However, for consistency of our linking function estimator we require $\frac{1}{\rho_n} = o\left(\sqrt{\frac{n}{\log(n)}}\right)$, which is stronger than the $\frac{1}{\rho_n} = o(n)$ condition for acyclic motifs. For general motifs, our condition is always weaker than the condition needed for the empirical graphon. In comparison with $\frac{1}{\rho_n} = o\left(n^{\frac{2}{m}}\right)$ for histogram graphon, our condition is weaker when $m > 4$ and stronger for $m \leq 4$.

One of the motivations for looking at this class of functions on networks is that it contains subgraph densities, which can be viewed as ‘network moments,’ in the sense that if two networks match on the densities of all subgraphs they come from the same network generating distribution. Theorem 3 implicitly shows that the bootstrap network distribution converges to the distribution of the original network. However, as the subgraphs become more complicated we need to impose stronger conditions on sparsity, meaning that full convergence of all subgraphs would only follow for dense models in which ρ_n does not go to 0.

There are other linking function estimators (e.g. Zhang, Levina, and Zhu (2017)) and alternative ways to resample nodes. The next result characterises the conditions needed for consistency of the class of functions considered in Theorem 3 when we replace the (\hat{h}_n, \hat{F}) in our procedure with alternative estimators of $(h_{0,n}, F_0)$.

Lemma 1. *Theorem 3 holds for any estimators (h_n, F_n) of $(h_{0,n}, F_0)$ which satisfy:*

1. $E_{F_n} \left(\left(\frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*)) \right)^2 \right) \xrightarrow{p} 0$.
2. $E_{F_n} (f(\xi^*(\iota))) \xrightarrow{p} E_{F_0} (f(\xi(\iota)))$ for all $f : \text{Supp}(\xi)^k \rightarrow \mathbb{R}$ such that $E_{F_0} (|f(\xi(\iota))|) < \infty$ for all $\iota \in \mathcal{M}_k$, for any $k \leq 2m - 1$.

In the proofs in Appendix A we restate Theorem 3 in a generalised way which incorporates the conditions given in Lemma 1.

¹⁷. $\tau(g)$ corresponds to the number of edges in the motif and m denotes the number of vertices. The maximal number of edges in an undirected acyclic graph on m nodes is $m - 1$.

¹⁸. The maximal number of edges in an undirected graph on m nodes is $\binom{m}{2}$.

A consequence of Theorem 3 is that the bootstrap procedure can consistently recover critical values and asymptotically valid confidence intervals, as stated in Corollary 1. In order to be able to define the confidence intervals and comment on their coverage we need an inverse of the bootstrap distribution, but $J_n(t, \hat{h}_n, \hat{F}_n)$ may not necessarily be continuous or strictly increasing in t , hence it may not be invertible in the standard sense. Because of this we define the inverse of a distribution in the following way:

Definition 7. *Let:*

$$J^{-1}(\alpha, h, F) \equiv \inf \{t : J(t, h, F) \geq \alpha\} \quad (31)$$

be the α th quantile of the distribution $J(t, h, F)$.

Corollary 1. *Under the conditions of Theorem 3:*

1. $J_n^{-1}\left(1 - \alpha, \hat{h}_n, \hat{F}_n\right) \xrightarrow{P} c_{1-\alpha}$, where $c_{1-\alpha}$ is the $1 - \alpha$ critical value from $N(0, m^2\sigma_1^2)$ ¹⁹.
2. If F_0 doesn't enter the function f_n^U directly but only through a parameter θ ²⁰: $f_n^U(A(h_{0,n}(\xi), \eta), \rho_n, \theta)$, then the $(1 - \alpha)$ confidence interval for θ constructed as:

$$CI_n\left(1 - \alpha, A, \hat{h}_n, \hat{F}_n\right) = \left\{ \theta : J_n^{-1}\left(\frac{\alpha}{2}, \hat{h}_n, \hat{F}_n\right) \leq f_n^U(A, \hat{\rho}_n, \theta) \leq J_n^{-1}\left(1 - \frac{\alpha}{2}, \hat{h}_n, \hat{F}_n\right) \right\} \quad (32)$$

is asymptotically valid:

$$P_{h_{0,n}, F_0}\left(\theta \in CI_n\left(1 - \alpha, A, \hat{h}_n, \hat{F}_n\right)\right) \xrightarrow{P} 1 - \alpha. \quad (33)$$

In defining the bootstrap statistic and forming confidence intervals we do not normalise the $f_n(\cdot)$ by the estimated variance. This is in part because the variance estimators may not be readily available, and even when they are, they tend to be complicated (e.g. for the subclass of motif densities Green and Shalizi (2022) Lemma 2 gives an expression for an estimator of variance. It is derived combinatorially by considering all motifs that can be achieved by merging two copies of the motif of interest on partially overlapping sets of nodes). Another reason is that, as pointed out by Hahn (1993), convergence weakly in probability ensures convergence of moments over the set of bounded and Lipschitz continuous functions, which does not include $f(x) = x^2$, meaning that weak convergence in probability of our bootstrap estimator does not guarantee the consistency of

19. I.e. $\Phi\left(\frac{c_{1-\alpha}}{m^2\sigma_1^2}\right) = 1 - \alpha$ where $\Phi(\cdot)$ denotes the CDF of $N(0, 1)$.

20. For example in equation (25) we have $\theta = E_{h_{0,n}, F_0}(g(A(t)))$.

its variance. When properties of the variance estimate are unknown, it is safer to use the percentile method for the construction of confidence intervals.

However, when a reliable variance estimate is known, normalising the statistic of interest by the estimate of its standard deviation could improve the performance of the bootstrap procedure. While we do not analyse the rates theoretically, the logic should be close to the case of standard bootstrap, where Edgeworth expansion arguments show that a normalised bootstrap with a pivotal limiting distribution can achieve a faster rate of convergence, see e.g. Hansen (2014) sections 10.8-10.11.

5 Simulations

We test the performance of our procedure using Monte Carlo simulations. We simulate the true adjacency matrices for $\xi_i \stackrel{iid}{\sim} \mathcal{U}[0, 1]$ and one of the following linking functions:

1. dot product function: $h(\xi_i, \xi_j) = \rho_n \xi_i \xi_j$.
2. horseshoe function²¹: $h(\xi_i, \xi_j) = \frac{\rho_n}{2} \left(e^{-200(\xi_i - \xi_j^2)^2} + e^{-200(\xi_j - \xi_i^2)^2} \right)$.
3. high-density function:

$$h(\xi_i, \xi_j) = \frac{\rho_n}{0.975} \left(1 - \mathbb{1} \left(\left| \frac{1}{2} - \xi_i \right| \leq 0.05 \right) \mathbb{1} \left(\left| \frac{1}{2} - \xi_j \right| \leq 0.05 \right) \right) \left(1 - \frac{1}{2} \left(\left| \frac{1}{2} - \xi_i \right| + \left| \frac{1}{2} - \xi_j \right| \right) \right)$$

The plots of these functions are included in Appendix B.3.

In the estimation procedure we use the normal kernel: $K(u) = e^{-\frac{u^2}{2}}$ and the bandwidth $a^{(opt)}$ chosen by maximising $\ell(A, a_n)$, as described in Section 3.5.

We test the performance of the algorithm for a few different statistics of interest:

- density: $f_n(A) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} A_{ij}$, i.e. the number of edges divided by the number of possible edges;
- triangle density: $f_n(A) = \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} A_{ij} A_{jk} A_{ik}$, i.e. the proportion of all subsets of 3 nodes that are fully connected;
- transitivity: $f_n(A) = 3 \frac{\#triangles}{\#triads} = \frac{tr(A^3)}{\sum_{i=1}^n \sum_{j=1}^n (A^2)_{ij} - tr(A^2)}$, i.e. the ratio of fully connected triples to connected triples;
- k th largest eigenvalue of the adjacency matrix: $f_n(A) = \lambda_k(A)$;
- maximal betweenness centrality: $f_n(A) = \max_i \sum_{j,k} \frac{\# \text{shortest paths between } j \text{ and } k \text{ through } i}{\# \text{shortest paths between } j \text{ and } k}$;

21. This function was also used by Green and Shalizi (2022). They borrow it from Wang (2016), who described it as “a challenging example for graphon estimation.”

- modularity of the Louvain community detection algorithm: for a given partition of nodes into communities, modularity is defined as the proportion of edges within communities minus the proportion if the edges were distributed at random. The Louvain community detection algorithm aims to find a partition which maximises modularity by iteratively moving nodes to communities and aggregating communities until no further improvement is possible. We used functions ‘louvain_communities’ and ‘modularity’ from the Python networkx package, see their documentation for precise definitions.

In order to speed up the simulations we obtain the confidence interval coverage using the WARP procedure from Giacomini, Politis, and White (2013):

1. Generate S true adjacency matrices on n nodes using the same true linking function (usually $S = 1000$, n between 25 and 1000).
2. For each true adjacency matrix A_s :
 - (a) Find the optimal bandwidth $a_s^{(opt)} = \max_a \ell(A_s, a)$.
 - (b) Calculate the matrix $\hat{h}_{n,s}$ based on A_s with bandwidth $a_s^{(opt)}$.
 - (c) Resample n nodes of A_s to form the nodes of the bootstrap graph.
 - (d) Generate a single bootstrap adjacency matrix $A_{s,1}^*$ by adding an edges between nodes with probabilities determined by $\hat{h}_{n,s}$.

Note that the number of bootstrap replications is $B = 1$.

3. Estimate the true value by the average of the statistic evaluated for the true graphs: $f_n^{(true)} = \frac{1}{S} \sum_{s=1}^S f_n(A_s)$ (or use the theoretical true value, if known).
4. Calculate the deviation of the statistic in the bootstrapped graph from the statistic evaluated for the corresponding true graph: $f_n(A_{s,1}^*) - f_n(A_s)$ for all $s \in \{1, \dots, S\}$. Denote the α th quantile of the empirical distribution of this set by $\hat{q}_\alpha \left(\{f_n(A_{s,1}^*) - f_n(A_s)\}_{i=1}^S \right)$.
5. Calculate the confidence intervals as: $CL_s = [CL_s^l, CL_s^u]$, where

$$CL_s^l = f_n(A_s) - \hat{q}_{1-\frac{\alpha}{2}} \left(\{f_n(A_{s,1}^*) - f_n(A_s)\}_{i=1}^S \right)$$

$$CL_s^u = f_n(A_s) - \hat{q}_{\frac{\alpha}{2}} \left(\{f_n(A_{s,1}^*) - f_n(A_s)\}_{i=1}^S \right).$$

6. Store the empirical coverage, i.e. the proportion of confidence intervals which cover the true value: $1 - \alpha^{(emp)} = \frac{1}{S} \sum_{s=1}^S \mathbf{1} \left(CL_s^l \leq f_n^{(true)} \leq CL_s^u \right)$.

The following plots and tables show results of some of our simulations. The code used in the simulations can be found in Appendix B.1 and tables with more results are in Appendix B.2.

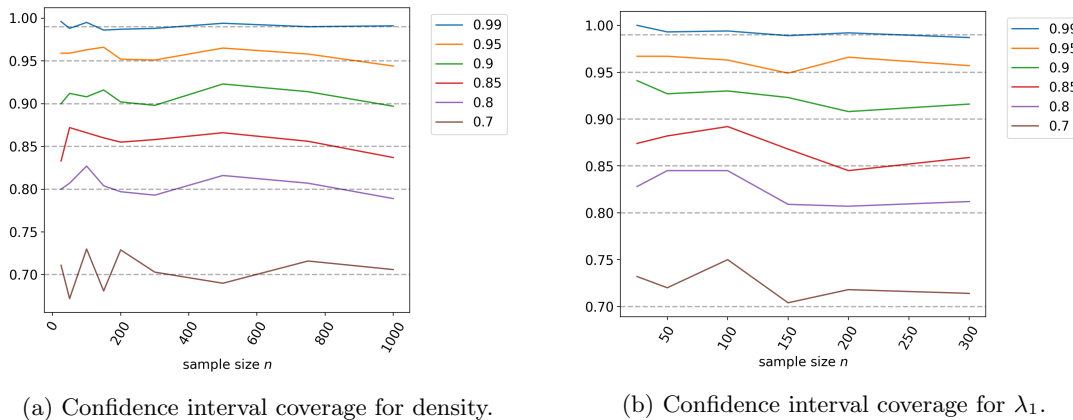


Figure 1: Confidence interval coverage for different sample sizes n based on Monte Carlo simulations using the product generating function and $\rho_n = 0.1875$.

We start by looking at the confidence interval coverage for different values of n , ρ_n and α . From Fig. 1 we can see that performance at different α is quite similar, but larger values have proportionally larger deviations and allow us to see the trend more clearly. Hence although in practice we tend to be most interested in the 95% confidence intervals, we present results for 70% or 80% in most of our plots. From Fig. 2 we can see that at a constant density level (no sparsity, ρ_n doesn't decrease with n) the performance improves with sample size. For $n \geq 250$ all statistics achieve good confidence interval coverage levels, although not always perfect: we may get coverage of e.g. 60% instead of 70%. As the sparsity level increases ($\rho_n \rightarrow 0$), the performance tends to get worse, but remains close to desired for statistics which are easier to estimate (e.g. density, triangle density, or the highest eigenvalue) but gets significantly worse for more complicated statistics (e.g. λ_{10}). The bottom two panels check performance for sparsity levels at which our theoretical results don't give any performance guarantees: as expected, the performance is poor in those cases.

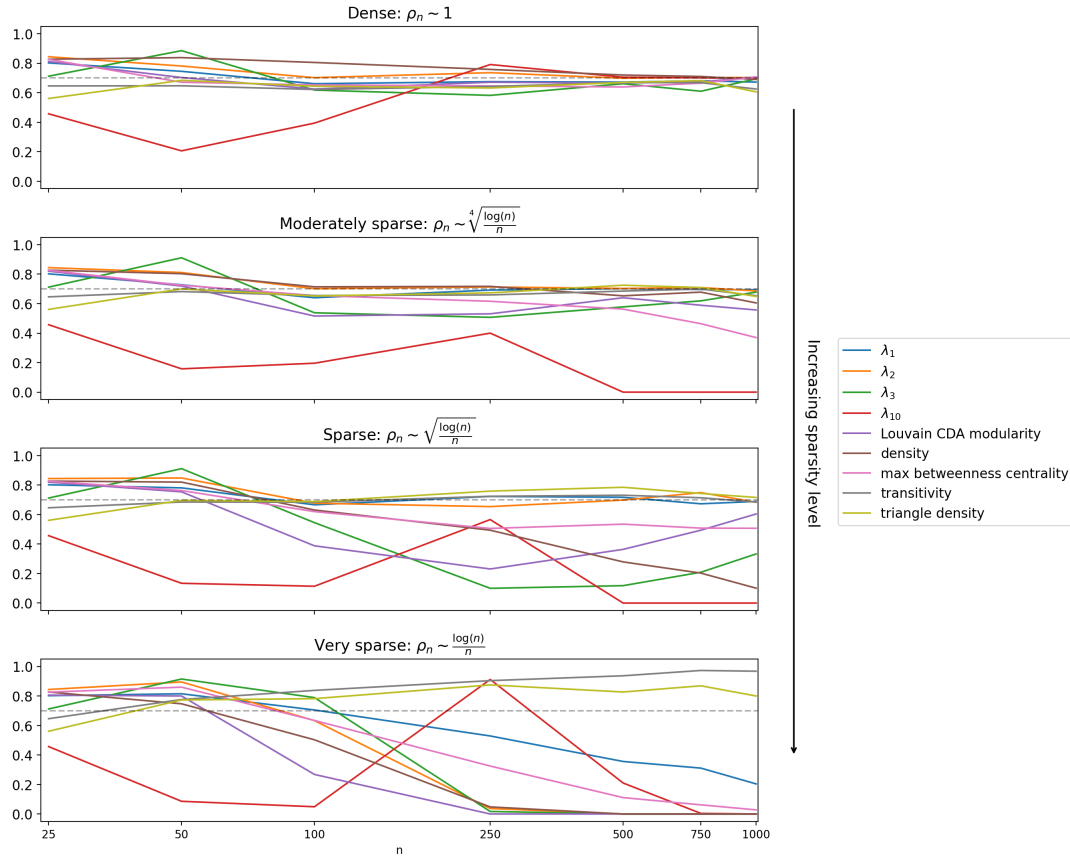
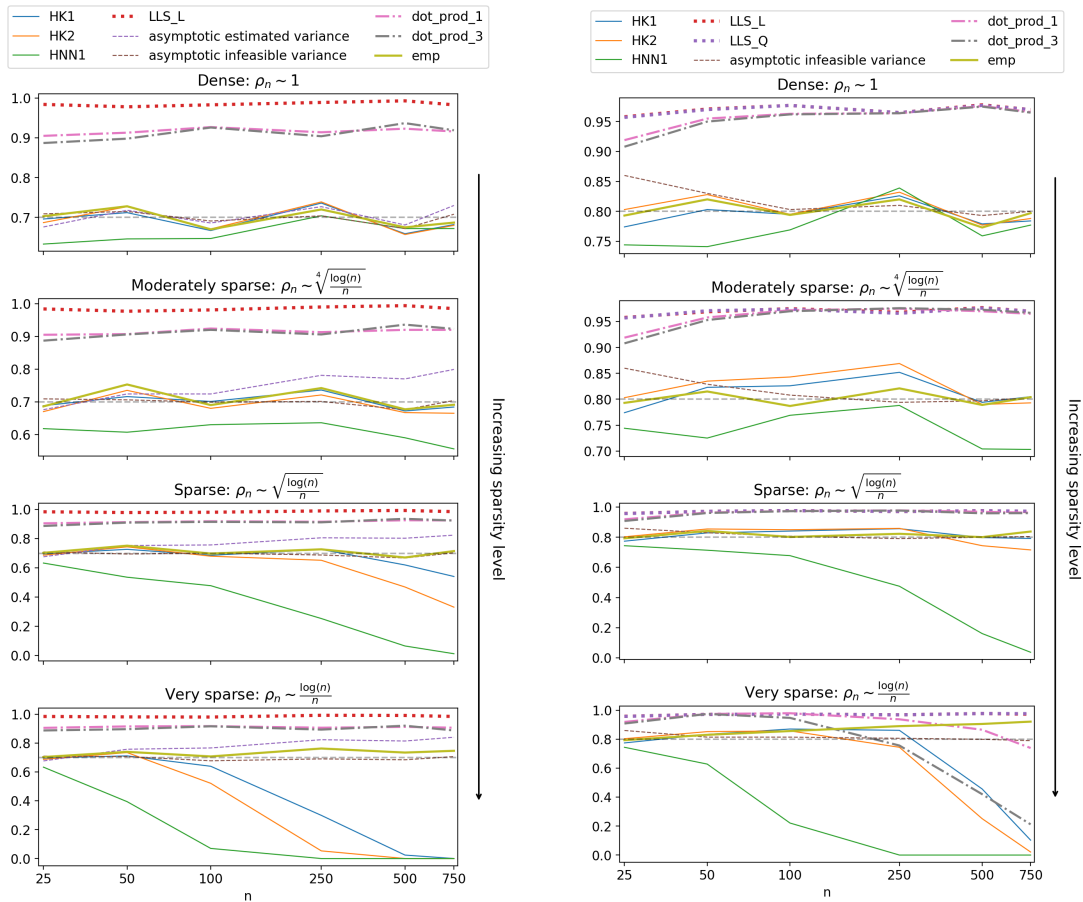


Figure 2: 70% confidence interval coverage for a range of statistics for the horseshoe generating function at different levels of sparsity.

Next we compare the performance of different variations of our method and some of the existing competitor methods. Fig. 3a shows the performance comparison between: HK1 (our main method based on $\hat{h} \equiv \hat{h}^{(K1)}$ with $a^{(opt)}$), HK2 (our bootstrap method but using the linking function estimator $\hat{h}^{(K2)}$ with $a^{(optK2)}$ based on $\hat{h}^{(K2)}$), HNN1 (our bootstrap method but but using the linking function estimator $\hat{h}^{(NN1)}$ from Zhang, Levina, and Zhu (2017) with their optimal choice of neighbourhood size), emp (empirical bootstrap from Green and Shalizi (2022)), dot_prod_k (the bootstrap method from Levin and Levina (2019) based on assuming a k -dimensional ξ_i), asymptotic estimated variance (the asymptotic distribution from Bickel, Chen, and Levina (2011) with variance estimated according to the formula in Green and Shalizi (2022)), asymptotic infeasible variance (the asymptotic distribution from Bickel, Chen, and Levina (2011) with the true theoretical variance), LLS.L and LLS.Q (the linear and quadratic methods from Lin, Lunde, and Sarkar (2020)).



(a) Density: 70% confidence interval coverage across different methods.

(b) Triangle density: 80% confidence interval coverage across different methods.

Figure 3: Confidence interval coverage for different methods using the product generating function.

HK1 and HK2 perform very similarly: we choose to use HK1 as the main method since HK2 is more computationally intensive and HK1 shows slight advantage for sparse graphs. HNN1 performs slightly worse and its performance drops more significantly for sparser graphs. We believe this is mostly due to the choice of bandwidth (HNN1 relies on the theoretical optimal bandwidth instead of our numerically chosen $a^{(opt)}$). For U-statistics the empirical bootstrap performs very well and remains good even at sparsity levels our methods can't handle. This suggests that for sparse graphs we may want to consider lower bandwidth choice than $a^{(opt)}$, which would make our method more similar to the empirical bootstrap. The dot product bootstrap of Levin and Levina (2019) is presented for the correctly specified case of $k = 1$, as well as for $k = 3$. This method should have an advantage over the other ones since it is parametric and for $k = 1$ it is based on a correctly specified functional form of the linking function. However, it achieves coverage which is too high, over 90% instead of the required 70%. The linear and quadratic methods from Lin, Lunde, and Sarkar (2020) are very similar to each other and suffer from the same issue of giving confidence

intervals with higher coverage than desired.

Most of the competitor methods can only be applied to U-statistics. An exception is the empirical bootstrap which can be seen as a limiting case of our procedure with bandwidth close to 0. From Fig. 4 and Fig. 5 we can see that the empirical bootstrap would perform very poorly for statistics such as lower eigenvalues, max betweenness centrality and Louvain CDA modularity. There is also a version of the dot product bootstrap Levin and Levina (2019) which applies to more general functions, but, probably due to our bad coding or relatively small sample sizes, we were getting many estimated probabilities outside of $[0, 1]$ which led to very bad performance, likely not representative of the quality of their method, and is not presented here.

In the final set of simulations we compare the confidence interval coverage with different bandwidths of the form $ca^{(opt)}$. Fig. 4 shows an example when all statistics perform well for our default choice with $c = 1$. For most statistics using a smaller bandwidth is not an issue: they do reasonably well for $c \leq 1$, even as small as $c = 0.01$, although they don't perform as well as for c close to 1. However, this is not universally true: some statistics, such as λ_{10} and maximal betweenness centrality, have very poor coverage outside of the region of $0.9 \leq c \leq 1.25$. The coverage for all statistics gets significantly worse when we use wider bandwidths ($c \geq 10$).

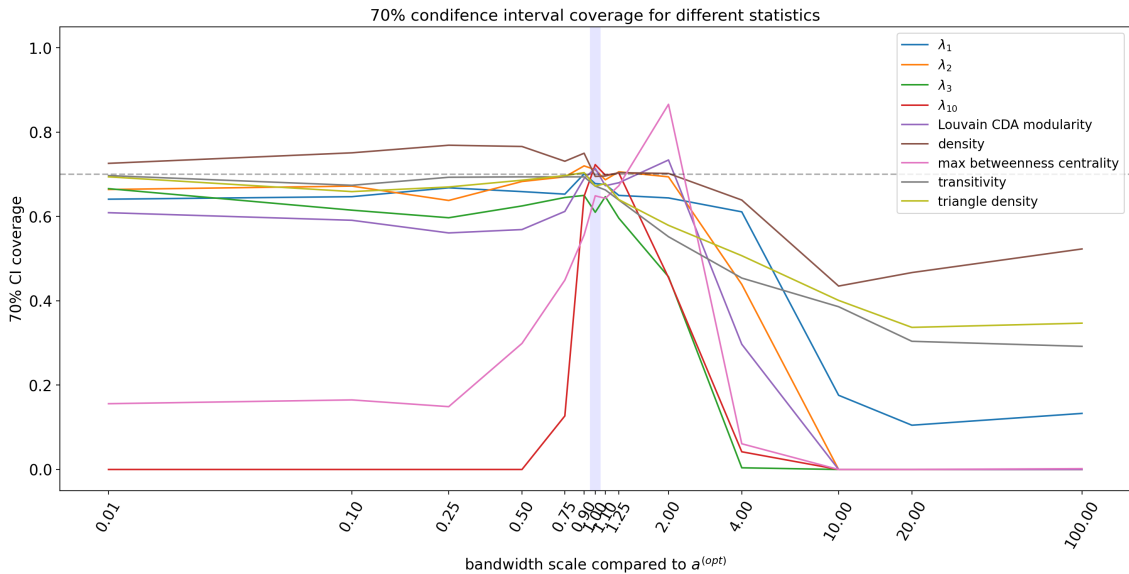


Figure 4: Confidence interval coverage for different bandwidths $c \times a^{(opt)}$: comparison of different statistics at $\alpha = 0.3$, $n = 500$ and $\rho_n = 0.1125$ based on Monte Carlo simulations using the horseshoe generating function.

This great performance around $c = 1$ is not guaranteed. In Fig. 5 we consider a smaller graph, for which confidence interval coverage for simple statistics such as density, triangle density or the largest eigenvalue is not too sensitive to the choice of bandwidth: from $0.01a^{(opt)}$ to $4a^{(opt)}$ the

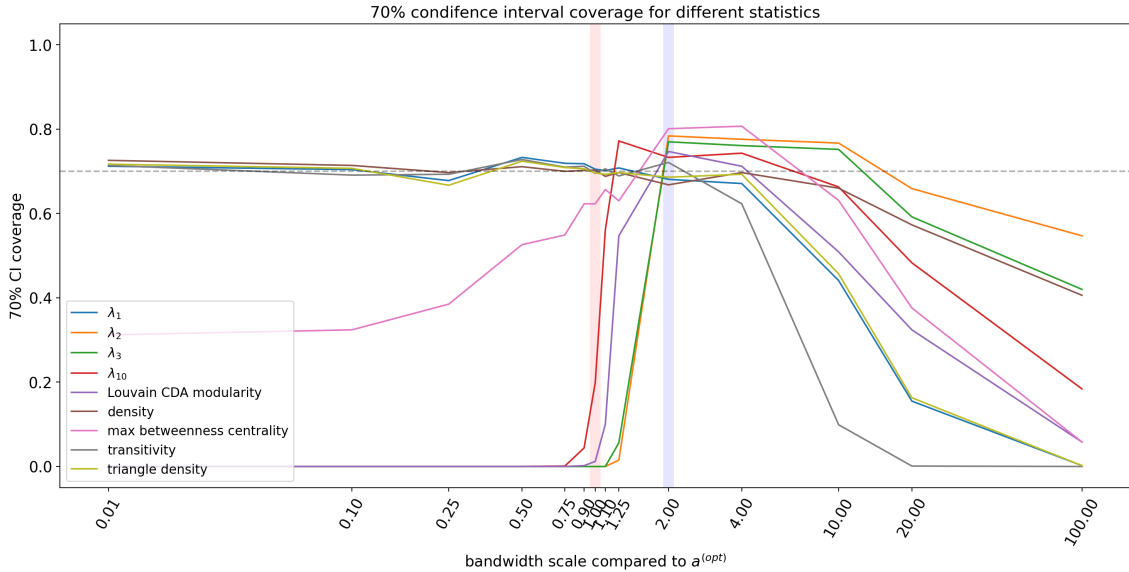
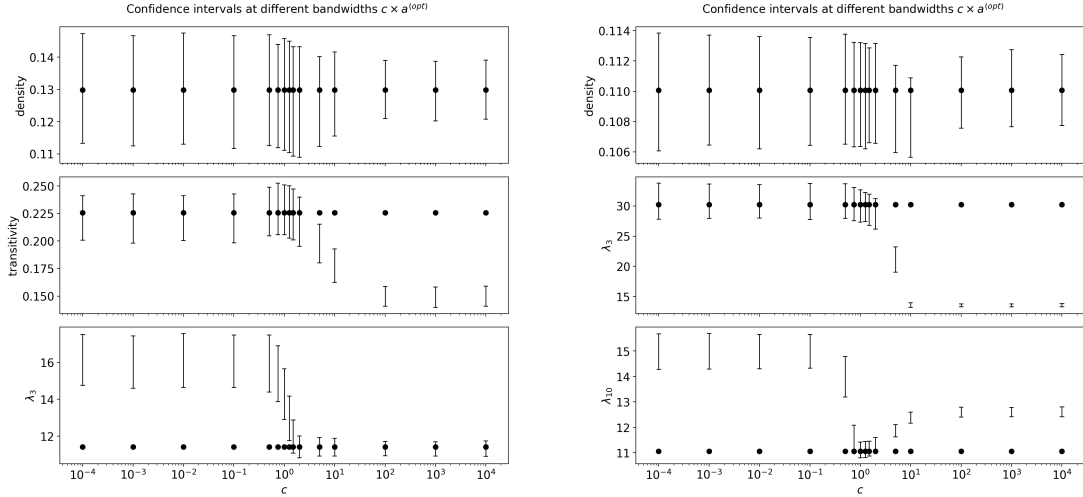


Figure 5: Confidence interval coverage for different bandwidths $c \times a^{(opt)}$: comparison of different statistics at $\alpha = 0.3$, $n = 300$ and $\rho_n = 0.1875$ based on Monte Carlo simulations using the product generating function.

performance remains good. However, other statistics, such as eigenvalues below the largest one (e.g. λ_2 , λ_3 and λ_{10} in Fig. 5) have zero coverage for low values of c (worryingly, in some cases including $c = 1$) but perform relatively well for $c \simeq 2$.

When the graph is sufficiently large and dense, c close to 1 gives good performance of most statistics we have checked. However, the performance does depend on the bandwidth choice and some statistics may be estimated poorly with the default bandwidth, especially when the graph is relatively sparse and the statistic is more complicated. In those cases experimenting with different values of c could give more reliable results. This is one advantage of our method over the empirical bootstrap (which can be seen as a limiting case of our model with $c = 0$): while at the default setting both algorithms may be bad at estimating confidence intervals for lower eigenvalues, the performance of our method can be improved by selecting a larger bandwidth (oversmoothing) while the empirical bootstrap does not depend on any parameters that could be tweaked in a similar way.

The possibility of a poor performance at $a^{(opt)}$ raises a question: how can we choose the best bandwidth in an application, when we only have one observed network and no way to run a Monte Carlo simulation confirming the coverage? Luckily, there is an easy way to verify our choice. Fig. 6 shows an example of bootstrap confidence intervals formed from $B = 1000$ bootstrap replications for different statistics of a specific single true graph A estimated using different bandwidths. We can use it in the following way: if the statistic estimate from the original graph is in the middle



(a) Confidence intervals for different statistics for the product generating function at $n = 300$ and $\rho_n = 0.125$.

(b) Confidence intervals for different statistics for the horseshoe generating function at $n = 500$ and $\rho_n = 0.1125$

Figure 6: Confidence intervals for different statistics and for bandwidths $c \times a^{(opt)}$ based on $B = 1000$ bootstrap graphs.

of the confidence interval formed by bootstrapped graphs, the choice of the bandwidth is good. In Fig. 6a we see that density is always estimated relatively well, transitivity remains well-estimated for smaller than optimal bandwidths but is underestimated when the bandwidth is too large, and λ_3 is overestimated for smaller than optimal bandwidths but remains well estimated for larger than optimal bandwidth. In simulations we have noticed a pattern that when the true value is above the estimated confidence interval lowering the bandwidth tends to improve the performance, while when the value from the original graph is below the confidence interval increasing the bandwidth often solves the problem. However, this is not always true: the lowest panel in Fig. 6b shows that λ_{10} is overestimated when bandwidth is either too low or too high compared to the optimal one. The middle panel also shows that the choice of a statistic doesn't determine the behaviour: for the horseshoe function λ_3 is better estimated for lower bandwidths and underestimated for higher ones, which is a different pattern than that of λ_3 from the product generating function in the bottom of Fig. 6a.

6 Application: the Diffusion of Microfinance

For our application, we use the data from Banerjee, Chandrasekhar, Duflo, and Jackson (2013), a paper which analyses how information about microfinance spreads through social networks in 43 villages in India.

Prior to the introduction of a microfinance program they surveyed households in these villages

and formed a network of connections based on 12 binary signals indicating if households knew each other (e.g. did they visit each other's homes, lend each other money, etc.). Each of these variables could be seen as a our A matrix, or we could combine them by taking a union, like in the original paper, to get an overall adjacency matrix. This is compatible with our framework: the closer two households are, the higher the probability that they will report each other as connected, hence we can view the reporting of a connection ($A_{ij} = 1$) as a signal from a Bernoulli distribution with probability of success proportional to the closeness of their friendship ($h_{0,n}(\xi_i, \xi_j)$).

Once the microfinance program entered the villages, they observed a set of first-informed villagers (injection points, chosen because they were village leaders who tend to be well-connected) and subsequent participation by households over a number of years.

One of the goals of the paper is to understand how the information about the program was spreading through the villages. This is modelled by a parametric diffusion model. Firstly, the probability p_{it} of household i with characteristics X_i participating when first informed is estimated from the logistic function:

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = X_i'\beta.$$

The parameter $\hat{\beta}$ is estimated using the information about the leaders only. The aim is to estimate the probability of transferring information about a microfinance program by people who participate in it themselves (q^P) and by those who know about it but do not participate (q^N). This is done by simulating the information spreading over time discretised into T periods (trimesters). In each period the newly informed decide whether to participate ($m_{it} = 1$) with probability p_{it} , then each informed household spreads the information to its neighbours with probability $m_{it}q^P + (1 - m_{it})q^N$.

For each village v and for each set of discretised parameter values (q^P, q^N) they simulate the spread of information and adoption decisions, and then calculate moments $m_{sim,v}(q^P, q^N)$ based on the final set of participating households (e.g. the fraction of households that have no participating neighbours but participate themselves, or the covariance of households participating in the program with the share of second neighbours that are participating). The average of these simulated moments across S simulations is compared to the observed empirical moments for the given village, $m_{emp,v}$. They then choose the parameter values which minimise the average of a

function of deviation of simulated moments from empirical moments across all villages:

$$(\hat{q}^P, \hat{q}^N) = \arg \min_{q^P, q^N} \left(\frac{1}{43} \sum_{v=1}^{43} m_{sim,v}(q^P, q^N) - m_{emp,v} \right)' \hat{W} \left(\frac{1}{43} \sum_{v=1}^{43} m_{sim,v}(q^P, q^N) - m_{emp,v} \right).$$

where $\hat{W} = \frac{1}{43} \sum_{v=1}^{43} (m_{sim,v}(\tilde{q}^P, \tilde{q}^N) - m_{emp,v}) (m_{sim,v}(\tilde{q}^P, \tilde{q}^N) - m_{emp,v})'$ for a first-stage estimates \tilde{q}^P, \tilde{q}^N obtained by using I as the weighting matrix. To form confidence intervals they use bootstrap which resamples whole villages. The resulting estimates are shown as the ‘‘Original’’ ones in Fig. 7.

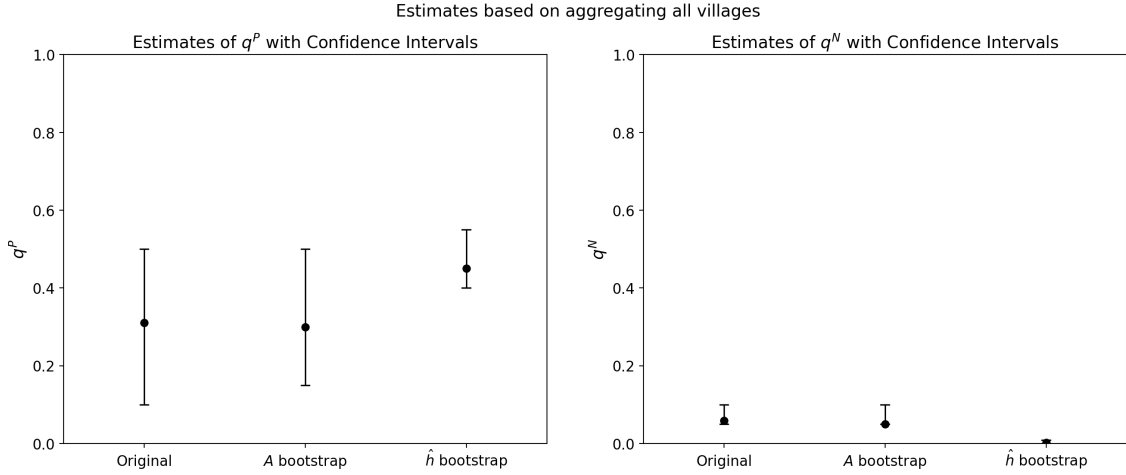


Figure 7: Estimates of q^P (left) and q^N (right) with 95% confidence intervals based on aggregating all villages: a comparison of the original result from Banerjee, Chandrasekhar, Duflo, and Jackson (2013) and our two methods.

The original paper considers a few variations of the model, including one which allows for endorsement effects. We only consider the information model without endorsement because it’s less computationally demanding (the parameters are identified using a grid search and increasing the dimension of the parameter space by one leads to an exponential growth in the number of required simulations) and the original paper did not find evidence of a significant endorsement effect.

In our replication we use the same procedure for finding the parameters but we use our bootstrap to form confidence intervals. Instead of resampling whole villages we can estimate the matrix \hat{h}_n for each of the villages²² and use it to generate $B = 1000$ new sets of 43 villages with structures similar to the original ones. We can then repeat the whole estimation procedure for each new set of villages and obtain bootstrap estimates $(\hat{q}_b^{*P}, \hat{q}_b^{*N})$. The confidence intervals are formed by taking the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles for the distributions of \hat{q}_b^{*P} and \hat{q}_b^{*N} . These estimates are presented in

22. The villages are assumed to be independent due to relatively large geographical distances between them. If they were not independent we could treat all households as belonging to one larger network.

Fig. 7 as the “ A bootstrap”. They are very similar, to the original estimates, with slightly narrower confidence intervals²³.

The replication of the original result indicates that our method performs well, though it would not be advised in this situation because it is much more computationally demanding than the original bootstrap. However, with our setup we can do more. It’s reasonable to assume that the spread of information is more likely between households which have a stronger connection (higher $h_{0,n}$), i.e. we can view the current model as an approximation to the true model in which the probability of spreading information depends not on the binary connection status A_{ij} , but on the actual strength of connection $h_{0,n}(\xi_i, \xi_j)$. Since \hat{h}_n estimated as part of our procedure is a consistent estimate of $h_{0,n}$, we can repeat the simulations using a diffusion model based on \hat{h}_n rather than on A (the imperfect signal about $h_{0,n}$). We assume that in any period $t \in \{1, \dots, T\}$ the informed individual i spreads the information to another individual j with probability $\hat{h}_n(\xi_i, \xi_j) (m_{it}q^P + (1 - m_{it})q^N)$. The rest of the estimation procedure remains unchanged. The resulting estimates are reported as the “ \hat{h} bootstrap” estimates in Fig. 7. The confidence intervals are now much narrower than in the previous two cases and the conclusions differ as well: q^P is estimated to be higher (point estimate 0.05, 95% confidence interval [0.35, 0.65]) while q^N is essentially zero (point estimate 0.004, 95% confidence interval [0, 0.05]).

This contrasts with the findings of Banerjee, Chandrasekhar, Duflo, and Jackson (2013) who highlight the importance of non-participants in the diffusion process by showing that constraining q^N to be equal to zero leads to simulated participation dropping from 20.0% to 13.97%²⁴. Our model shows that if we use a diffusion model based on \hat{h}_n instead of A the estimated value of q^N is not distinguishable from zero (point estimate 0.002, 95% confidence interval [0, 0.009]) and the simulated participation drops from 18.46% at the optimal values to 18.21% when we restrict q^N to 0, a drop of one seventy-fifth instead of one third. Using the more realistic assumption that the likelihood of spreading information depends on how well the households know each other removes the need for information spreading by non-participants.

Another extension made possible by our model is performing the analysis on individual village level. So far we have assumed that the parameters are common across all villages, and with the original bootstrap resampling whole villages there was no way to form confidence intervals on at the village-level. With our bootstrap method, instead of minimising a (weighted) average of deviations of simulated moments from empirical moments across all villages, we can minimise them for each

²³. Note that the confidence intervals for q^N can’t get much narrower because of the discretisation of the parameter space.

²⁴. The actual observed participation rate was 19.38%. It is not used as one of the moments matched in the parameter estimation.

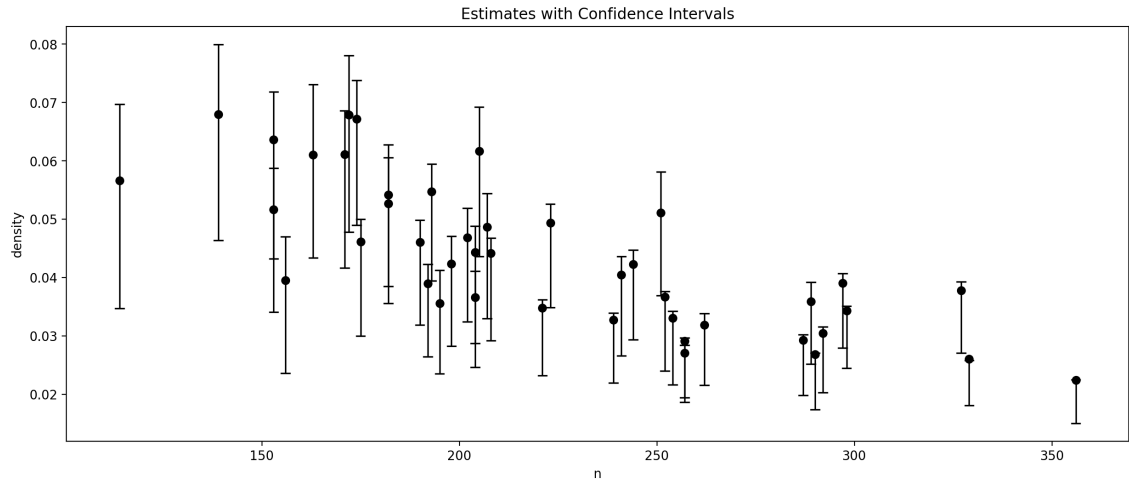


Figure 8: Densities of all 43 villages with bootstrap confidence intervals plotted against village size on the horizontal axis.

individual village. This allows us to:

1. **Check if all the villages come from the same network generating distribution.** We bootstrap each village separately, find confidence intervals for some network statistics (e.g. density, or the largest eigenvalue) and see if these intervals overlap for all villages.
2. **Estimate the q^P and q^N parameters (and their confidence intervals) for each village separately, see if there are systematic differences between villages.** If there are differences, see if the hypotheses ($q^N > 0$ and $q^P > q^N$) hold in each village.

We firstly look at densities of all the 43 villages. Our bootstrap method allows us to not only obtain their point estimates but also add confidence intervals to see if the villages are systematically different from each other. Fig. 8 shows that the villages become more sparse as their size increases (consistent with the assumption that $\rho_n \rightarrow 0$ as $n \rightarrow \infty$). The confidence intervals in this graph are formed using the same bootstrapped villages that were used for estimating the model parameters.

Moving on to the model parameters, we have repeated the estimation using the original diffusion model based on the adjacency matrix A (Fig. 9) and the new diffusion model based on the linking probabilities \hat{h}_n (Fig. 10). We can see that the two methods produce similar though not identical results. For some villages the estimation is very imprecise, leading to very wide confidence intervals. At least half of the villages have q^N precisely estimated to be zero, even in the model based on A : this may suggest that it's the imprecisely measured villages which drive the aggregate estimate to be positive.

In the last panels we can see that one of the predictions of the model, $q^P - q^N > 0$, cannot be concluded for most of the villages as we can't reject the hypothesis that $q^P - q^N = 0$ (mostly due

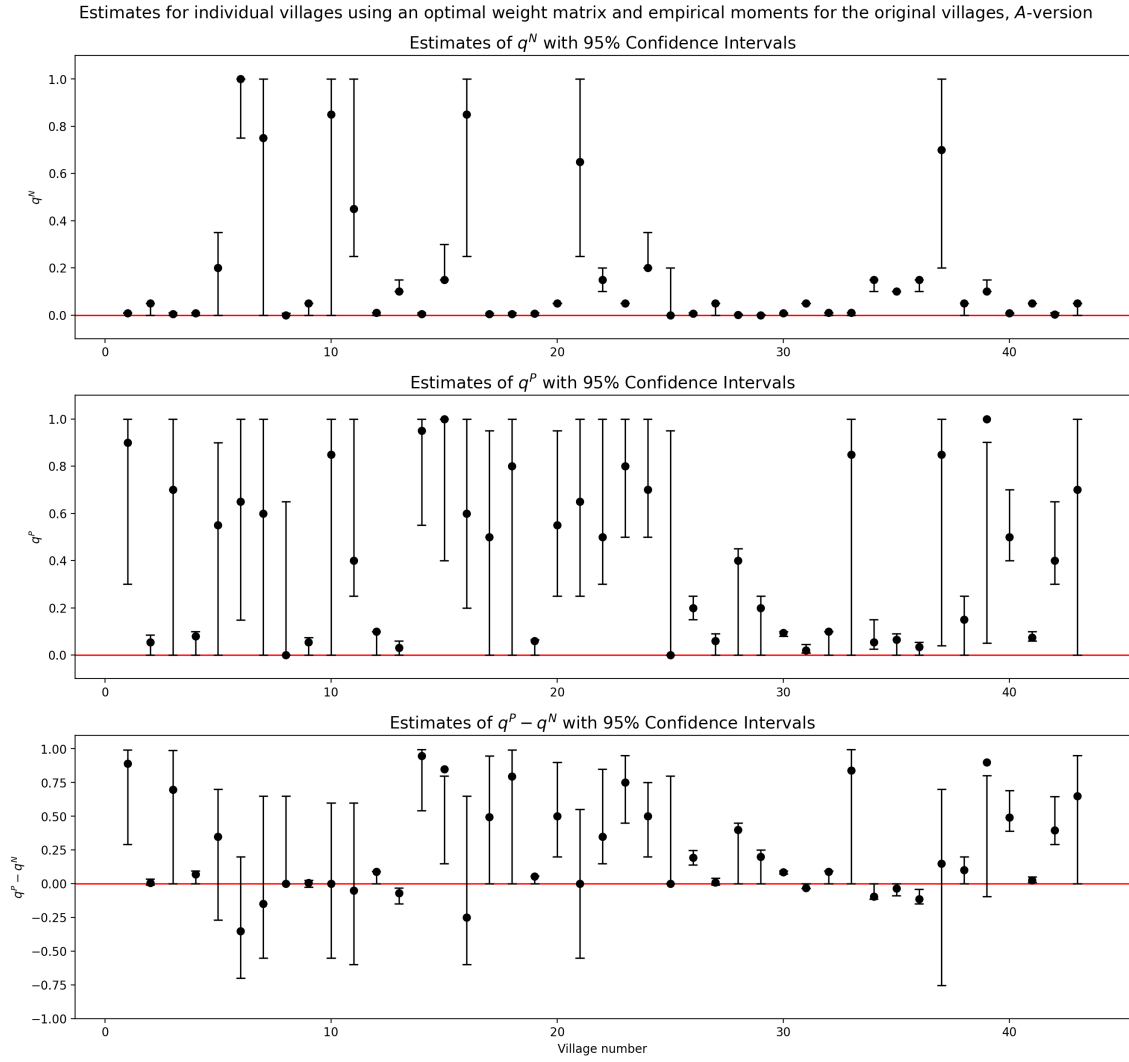


Figure 9: Estimates of q^N (top), q^P (middle) and $q^N - q^P$ (bottom) with 95% confidence intervals for individual villages using an optimal weight matrix and empirical moments for the original villages, A-version.

to imprecise measurements).

A practical extension of the current analysis would be to identify the village characteristics which help predict the estimated ranges of q^P and q^N . This would help policymakers choose villages in which the microfinance programs would have the highest chance of success or personalise the way in which the initial group of informed leaders is chosen depending on the information transmission characteristics in a given village.

7 Conclusion and Extensions

In this paper we have proposed a nonparametric linking function estimator and a related network bootstrap procedure and we have provided conditions under which both achieve some notions of

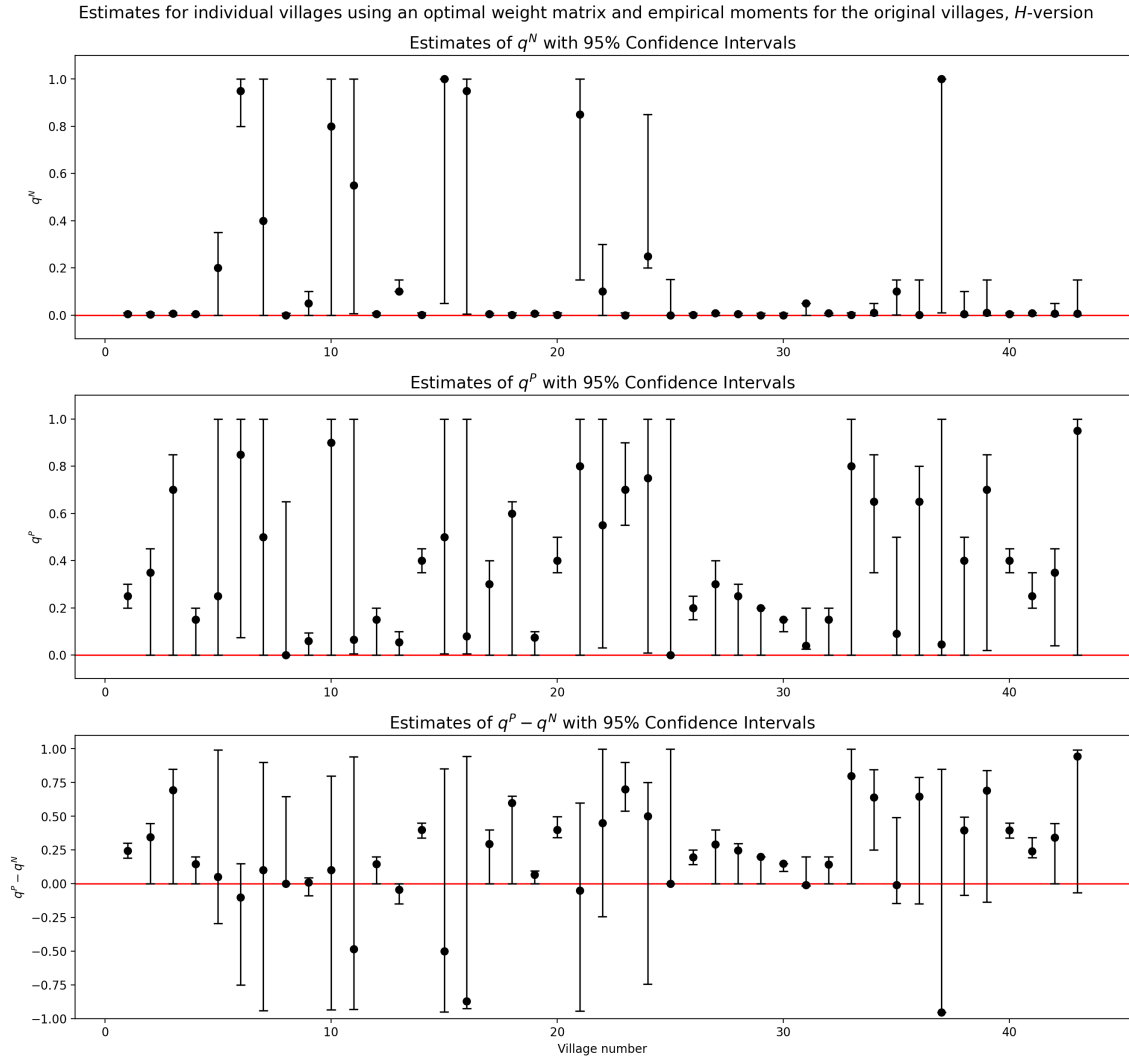


Figure 10: Estimates of q^N (top), q^P (middle) and $q^N - q^P$ (bottom) with 95% confidence intervals for individual villages using an optimal weight matrix and empirical moments for the original villages, H -version.

consistency.

In the future projects (or the future versions of this project) we aim to provide a theoretical justification for consistency of our bootstrap method over a wider class of statistics, which is suggested by the promising results in our simulations. Most importantly, we would like to extend our results to regression models in which the outcome depends, possibly in a complicated way, on the entire adjacency matrix (e.g. spillover effects from neighbours). Unfortunately, it looks like in these cases the behaviour is not well approximated by that of an average of i.i.d. random variables, which makes deriving asymptotic results tricky. This is both an obstacle in proving bootstrap consistency and a reason why bootstrap methods are particularly needed when it comes to strongly dependent data structures such as networks.

One way in which we may be able to get around this issue is by looking for another notion of network distance than the Wasserstein metric proposed by Levin and Levina (2019). More specifically, one which would be sufficient for proving that the convergence is preserved after a transformation.

A less closely related future project inspired by this paper could be formulating a fast numerical procedure which allows for the selection of an objective-specific optimal bandwidth in two-step procedures. As in our setup, suppose we have a two-step estimation procedure where in the first step we estimate some parameter dependent on a tuning parameter (like \hat{h}_n based on a_n), which we then plug into a (possibly random) second-step estimation (e.g. density in the resulting network). We ultimately care about the result of the second step (for us, the bootstrap confidence interval coverage of the statistic estimated in the second step) and we wish to optimise its performance by choosing an optimal tuning parameter in the first step. In this paper we have relied on optimising the first-step estimation ($a^{(opt)}$ was chosen to optimise the estimation of \hat{h}_n), but this was shown not to provide the best results in the second step across all second-step statistics. We have considered some algorithms which do rely on the second step performance (e.g. the approximate average MSE approach), but which are too slow to be useful in practice.

Finally, we believe that forming models which use \hat{h}_n as a better proxy for the strength of connections between individuals than the adjacency matrix A opens up exciting opportunities for empirical research.

References

- Alatas, Vivi, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken. 2016. "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia." *American Economic Review* 106 (7): 1663–1704.
- Auerbach, Eric. 2022. "Identification and Estimation of a Partially Linear Regression Model Using Network Data." *Econometrica* 90 (1): 347–365.
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341 (6144): 1236498.
- Bhattacharyya, Sharmodeep, and Peter J Bickel. 2015. "Subsampling bootstrap of count features of networks." *The Annals of Statistics* 43 (6): 2384–2411.
- Bickel, Peter J, Aiyou Chen, and Elizaveta Levina. 2011. "The method of moments and degree distributions for network models."
- Bickel, Peter J., and David A. Freedman. 1981. "Some Asymptotic Theory for the Bootstrap." *The Annals of Statistics* 9 (6): 1196–1217.
- Billingsley, P. 1995. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley.
- Carter, Michael, Rachid Laajaj, and Dean Yang. 2021. "Subsidies and the African Green Revolution: direct effects and social network spillovers of randomized input subsidies in Mozambique." *American Economic Journal: Applied Economics* 13 (2): 206–229.
- Conley, Timothy G, and Christopher R Udry. 2010. "Learning about a new technology: Pineapple in Ghana." *American economic review* 100 (1): 35–69.
- Davezies, Laurent, Xavier D'Haultfœuille, and Yannick Guyonvarch. 2021. "Empirical process results for exchangeable arrays." *The Annals of Statistics* 49 (2).
- Giacomini, Raffaella, Dimitris N. Politis, and Halbert White. 2013. "A WARP-speed method for conducting monte carlo experiments involving bootstrap estimators." *Econometric Theory* 29 (3): 567–589.
- Gine, Evarist, and Joel Zinn. 1990. "Bootstrapping General Empirical Measures." *The Annals of Probability* 18 (2): 851–869.

- Green, Alden, and Cosma Rohilla Shalizi. 2022. “Bootstrapping exchangeable random graphs.” *Electronic Journal of Statistics* 16 (1): 1058–1095.
- Hahn, Jinyong. 1993. “Three essays in econometrics” [in English]. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-02-20. PhD diss.
- Hansen, Bruce E. 2014. *Econometrics*. University of Wisconsin Department of Economics.
- Levin, Keith, and Elizaveta Levina. 2019. “Bootstrapping networks with latent space structure.” *arXiv preprint arXiv:1907.10821*.
- Lin, Qiaohui, Robert Lunde, and Purnamrita Sarkar. 2020. “Trading off Accuracy for Speedup: Multiplier Bootstraps for Subgraph Counts.” *arXiv preprint arXiv:2009.06170*.
- Lovász, László. 2012. *Large networks and graph limits*. Vol. 60. American Mathematical Soc.
- Lunde, Robert, and Purnamrita Sarkar. 2022. “Subsampling sparse graphons under minimal assumptions.” *Biometrika* 110 (1): 15–32.
- Menzel, Konrad. 2021. “Bootstrap With Cluster-Dependence in Two or More Dimensions.” *Econometrica* 89 (5): 2143–2188.
- Politis, D.N., J.P. Romano, M. Wolf, P. Diggle, and S. Fienberg. 1999. *Subsampling*. Springer Series in Statistics. Springer New York.
- Pólya, Georg. 1920. “Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem.” *Mathematische Zeitschrift* 8:171–181.
- Serfling, R.J. 2009. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley.
- Shao, Zhixuan, and Can M Le. 2024. “Parametric Bootstrap on Networks with Non-Exchangeable Nodes.” *arXiv preprint arXiv:2402.01866*.
- Wang, Lawrence. 2016. “Network Comparisons using Sample Splitting.” *PhD thesis, Carnegie Mellon University*.
- Zelenev, Andrei. 2020. “Identification and estimation of network models with nonparametric unobserved heterogeneity.” *Department of Economics, Princeton University*.

Zhang, Yuan, Elizaveta Levina, and Ji Zhu. 2017. “Estimating network edge probabilities by neighbourhood smoothing.” *Biometrika* 104 (4): 771–783.

Zhang, Yuan, and Dong Xia. 2022. “Edgeworth expansions for network moments.” *The Annals of Statistics* 50 (2): 726–753.

List of all notation

The notation in this file can get a bit heavy so we provide this list for reference.

- n – sample size, number of individuals in the network.
- A – an $n \times n$ adjacency matrix. Binary, symmetric, observed.
- A_{ij} – i, j th entry of the matrix A : 1 if i, j are connected (are neighbours), 0 if they are not.
- i, j, k, s, t – usually used to refer to one of the n individuals.
- ξ_i – vector of characteristics of individual i , enters the linking function.
- F_0 – distribution of ξ_i .
- $h_{0,n}$ – linking function, takes characteristics ξ_i, ξ_j as inputs and outputs the probability with which individuals i and j are linked. If the inputs are vectors $\xi^{(\iota)} = (\xi_{\iota 1}, \xi_{\iota 2}, \dots, \xi_{\iota m})$ of characteristics of multiple individuals it outputs the matrix of linking probabilities.
- ρ_n – density/sparsity parameter. Density in the sense that it is the expected edge density, sparsity in the sense that as $n \rightarrow \infty$ the density of edges decreases: $\rho_n \rightarrow 0$.
- w_0 – underlying linking probability before accounting for sparsity: $\rho_n w_0(\xi_i, \xi_j) = h_{0,n}(\xi_i, \xi_j)$.
- $\varphi(\xi_i, \xi_t) = E\left(\frac{A_{is}A_{ts}}{\rho_n^2} \mid \xi_i, \xi_t\right)$ – a function measuring the probability of a common friend between i and j normalised by the sparsity level.
- $d_{ij} = \sqrt{E\left(E(w_0(\xi_t, \xi_s)(w_0(\xi_i, \xi_s) - w_0(\xi_j, \xi_s)) \mid \xi_i, \xi_j, \xi_t)^2 \mid \xi_i, \xi_j\right)}$ – theoretical distance between i and j .
- $\hat{d}_{ij} = \frac{1}{\rho_n^2} \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n A_{ts}(A_{is} - A_{js})\right)^2}$ – estimated distance between i and j .
- D_{ij} – shorthand notation for $\rho_n^4 \hat{d}_{ij}^2$ used in the description of the bootstrap procedure.
- \hat{h}_n – estimated linking function:

$$\hat{h}_n(\xi_i, \xi_j) = \frac{\tilde{h}_n(\xi_i, \xi_j) + \tilde{h}_n(\xi_j, \xi_i)}{2} \quad \text{where} \quad \tilde{h}_n(\xi_i, \xi_j) = \frac{\sum_{t \neq j}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right) A_{tj}}{\sum_{t \neq j}^n K\left(\frac{\rho_n^4 \hat{d}_{it}^2}{a_n}\right)}$$

- H_{ij} – shorthand notation for $\hat{h}_n(\xi_i, \xi_j)$ evaluated at $a^{(opt)}$ used in the description of the bootstrap procedure.
- K – kernel function used in estimating linking probability.

- a_n – a bandwidth parameter, chosen by the researcher.
- \hat{F}_n – the empirical distribution function of ξ_i ; assigns equal probability to each of the original observations.
- $*$ – a bootstrap equivalent, e.g. $\xi_i^* \sim \hat{F}_n$ is the bootstrap version of $\xi_i \sim F_0$.
- $\hat{}$ – an estimate.
- $\max_{i,j} \equiv \max_{i,j \in \{1,2,\dots,n\}}$ – maximum over indices in a specific sample of size n .
- $\max_{\xi_i} \equiv \max_{\xi_i \in \text{Supp}(\xi_i)}$ – maximum over all $\xi_i \in \text{Supp}(\xi_i)$.
- $N(\xi_j, \delta) = \{\xi_k : \sup_{\xi_t} |w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta\}$ – the neighbourhood of ξ_j of size δ .
- $\omega(\delta) = \inf_{\xi_j \in \text{Supp}(\xi_j)} P(\xi_k \in N(\xi_j, \delta) | \xi_j)$ – the infimum over all possible ξ_i of the measures of their neighbourhoods of size δ .
- $b_n = \frac{a_n}{\rho_n^4}$ – a bandwidth parameter normalised by sparsity; the effective bandwidth size after accounting for the rate at which density goes to zero.
- \hat{h}_n^- – leave-one-out version of \hat{h}_n , evaluated in the same way as \hat{h}_n but without the observations $t = i, j$. Used for numerically choosing the optimal bandwidth.
- $\ell(A, a_n)$ – log-likelihood used for numerically choosing the optimal bandwidth. Defined in Eq. (16).
- $a^{(opt)}$ – numerically chosen optimal bandwidth. Defined in Eq. (17).
- B – number of bootstrap replications.
- $f_n(A_n(h_{0,n}(\xi), \eta), \rho_n, F_0)$ – a function whose distribution we are interested in.
- η – a vector of random variables which together with the linking function determine the realised links in A . We assume $\eta_{ij} \stackrel{ind}{\sim} \mathcal{U}[0, 1]$ for $1 \leq i \leq j \leq n$ and η independent of ξ .
- $\tilde{f}_n(h_{0,n}(\xi), \rho_n, F) \equiv E(f_n(A_n(h_{0,n}(\xi), \eta), \rho_n, F_0) | \xi)$ – a function whose distribution we are interested in after averaging out the variation due to observing A instead of $h_{0,n}$.
- $E_{h_{0,n}}$, e.g. in $E_{h_{0,n}}(f_n(A_n(h_{0,n}(\xi), \eta), \rho_n, F_0) | \xi) = \int f_n(A_n(h_{0,n}(\xi), \eta), \rho_n, F_0) d\eta$ – expectation taken with respect to the independent Bernoulli trials with probabilities determined by $h_{0,n}$.
- $E_{h_{0,n}, F_0}$ – expectation taken with respect to both the Bernoulli trials and the true distribution of ξ .

- ι – a vector of m nodes from $\{1, \dots, n\}$.
- $A(\iota)$ – the adjacency matrix of the subgraph with nodes ι (i.e. A from which we remove $n - m$ rows and columns not in ι).
- m – usually denotes the size of a subgraph or an order of U-statistic.
- g – a kernel function (in the U-statistic sense); a function of a subset of A .
- \tilde{g} – a kernel function (in the U-statistic sense); a function of a subset of $h_{0,n}$. Equal to g after averaging out the variation due to observing A instead of $h_{0,n}$.
- $\tau(g)$ – a normalisation chosen to ensure $\frac{E_{h_{0,n}, F_0}(g(A_n(\iota)))}{\rho_n^{\tau(g)}} = O_p(1)$.
- $\tilde{\tilde{g}}$ – the leading term in the normalised \tilde{g} ; a function of a subset of w_0 . $O_p(1)$.
- \tilde{J}_n – the distribution of \tilde{f}_n .
- J_n – the distribution of f_n .
- $\hat{J}_{n,B}$ – an estimate of the distribution of f_n based on B bootstrap samples.
- J – limiting distribution of J_n as $n \xrightarrow{\infty} J_n(t, h_{0,n}, F_0) \Rightarrow J(t, w_0, F_0)$.
- \Rightarrow – weak convergence.
- $\xrightarrow{a.s.}$ – weak convergence almost surely, see Definition 1.
- \xrightarrow{P} – weak convergence in probability, see Definition 2.
- d_W – distance between measures which metrises weak convergence.
- $f(S) = \{f : S \rightarrow \mathbb{R} : |f(x) - f(y)| \leq d_S(x, y), \sup_{x \in S} |f(x)| \leq 1\}$ – the set of Lipschitz continuous and bounded real-valued functions on a metric space S equipped with distance d_S .
- $C_{w,F,\rho}$ – the set of non-random sequences of pairs of functions and distributions $\{(h_n, F_n)\}_{n=1}^{\infty}$ which satisfy a set of conditions on convergence of moments, see Definition A.1.
- CI_n – a bootstrap confidence interval as defined in Eq. (32).
- M_m – the set of all possible multisets of cardinality m with elements from $\{1, 2, \dots, m\}$
- $\hat{\rho}_n$ – estimator of density; the density of the observed adjacency matrix A .
- $\lambda_k(A)$ or λ_k – the k th largest eigenvalue of matrix A .
- \hat{q}_α – the estimate of α th quantile.

- q^P – in the application, the probability of transferring information about a microfinance program by program’s participants
- q^N – in the application, the probability of transferring information about a microfinance program by those who don’t participate the program themselves.
- C – generic positive constant, its value may change between different expressions in which it is used.
- C_ε – a positive constant which depends on $\varepsilon > 0$. Its value may change between different expressions in which it is used.
- T_n – a remainder term used in the proof of Theorem 1.
- M_w – an upper bound on the value of w_0 : $\sup_{\xi_i, \xi_j} |w_0(\xi_i, \xi_j)| \leq M_w$.
- $r_n(i) = E\left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i\right)$ – the shorthand notation for the expected kernel weights based on the distance between i and other individuals used in the estimation of $\hat{h}_n(\xi_i, \xi_j)$.
- $\hat{r}_n(i) = \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{d_{it}^2}{b_n}\right)$ – the estimate of $r_n(i)$.
- $r_n = \inf_{\xi_i} r_n(i)$ – the smallest possible expected kernel weight. We need to ensure it’s not too small or we would not be able to successfully estimate $h_{0,n}(\xi_i, \xi_j)$.

A Appendix: proofs

A.1 Proof of consistency of the linking function estimator

Proof of Theorem 1. Throughout this argument we use C_ε to denote a positive constant which depends on $\varepsilon > 0$. The value of C_ε may change between different expressions in which it is used.

By definition,

$$\begin{aligned}
\max_{i,j} \left| \frac{\hat{h}_n(\xi_i, \xi_j) - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| &= \max_{i,j} \left| \frac{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{d_{it}^2}{b_n}\right) \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n}\right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{d_{it}^2}{b_n}\right)} \right| \\
&= \max_{i,j} \left| \frac{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{d_{it}^2}{b_n}\right) \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n}\right)}{E\left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i\right) + \left(\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{d_{it}^2}{b_n}\right) - E\left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i\right)\right)} \right| \\
&\leq \left(\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n}\right) \right| \right) \left(1 + \max_{i,j} \left| 1 - \frac{E\left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i\right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K\left(\frac{d_{it}^2}{b_n}\right)} \right| \right) \quad (34)
\end{aligned}$$

The inequality follows from $\max \left| \frac{a}{b+c} \right| \leq \max \left| \frac{a}{b} \right| + \max \left| \frac{ac}{b(b+c)} \right| \leq (\max \left| \frac{a}{b} \right|) \left(1 + \max \left| \frac{c}{b+c} \right| \right)$, where $b = E \left(K \left(\frac{d_{it}^2}{b_n} \right) \mid \xi_i \right)$. In Lemma A.1 we show that the second factor converges almost surely to one.

We now focus on the first factor. Since $K(\cdot)$ is Lipschitz continuous with a Lipschitz constant C (by Assumption 1.3):

$$K \left(\frac{\hat{d}_{it}^2}{b_n} \right) \leq K \left(\frac{d_{it}^2}{b_n} \right) + \left| K \left(\frac{\hat{d}_{it}^2}{b_n} \right) - K \left(\frac{d_{it}^2}{b_n} \right) \right| \leq K \left(\frac{d_{it}^2}{b_n} \right) + C \left| \frac{\hat{d}_{it}^2 - d_{it}^2}{b_n} \right|.$$

It follows that

$$\begin{aligned} & \max_{i,j} \left| \frac{1}{(n-1)r_n(i)} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{\hat{d}_{it}^2}{b_n} \right) \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right) \right| \\ & \leq \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right) + C \left| \frac{\hat{d}_{it}^2 - d_{it}^2}{b_n r_n(i)} \right| \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right) \right| \\ & \leq \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right) \right| \\ & \quad + C \max_{i,j,t,t \neq j} \left| \frac{\hat{d}_{it}^2 - d_{it}^2}{b_n r_n(i)} \right| \underbrace{\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right|}_{\leq 2M_w + \underbrace{T_n}_{\xrightarrow{a.s.} 0}}. \end{aligned}$$

$M_w < \infty$ is defined in Lemma A.1. In Lemma A.2 we define T_n and show that the last factor in

the last expression is almost surely bounded, hence

$$\begin{aligned}
& \max_{i,j} \left| \frac{1}{(n-1)\rho_n r_n(i)} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{\hat{d}_{it}^2}{b_n} \right) (A_{tj} - h_{0,n}(\xi_i, \xi_j)) \right| \\
& \leq \max_{i,j} \left| \frac{1}{(n-1)\rho_n r_n(i)} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right) (A_{tj} - h_{0,n}(\xi_t, \xi_j) + h_{0,n}(\xi_t, \xi_j) - h_{0,n}(\xi_i, \xi_j)) \right| \\
& \quad + C(2M_w + T_n) \frac{1}{b_n r_n} \left(\max_{i,j} |\hat{d}_{ij}^2 - d_{ij}^2| \right) \\
& \leq \max_{i,j} \left| \frac{\sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right) (A_{tj} - h_{0,n}(\xi_t, \xi_j))}{(n-1)\rho_n r_n(i)} \right| + \max_{i,j} \left| \frac{\sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right) (h_{0,n}(\xi_t, \xi_j) - h_{0,n}(\xi_i, \xi_j))}{(n-1)\rho_n r_n(i)} \right| \\
& \quad + C(2M_w + T_n) \frac{1}{b_n r_n} \left(\max_{i,j} |\hat{d}_{ij}^2 - d_{ij}^2| \right).
\end{aligned}$$

where $T_n \xrightarrow{a.s.} 0$. We complete the proof by showing that the three terms go to zero in probability in Lemma A.3, Lemma A.4 and Lemma A.5. \square

Lemma A.1. *Under the assumptions of Theorem 1, for any $\varepsilon > 0$:*

$$\begin{aligned}
& \sum_{n=3}^{\infty} P \left(\left| \max_{i,j} \left| 1 - \frac{E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right)} \right| \right| > \varepsilon \right) \\
& = O \left(\sum_{n=3}^{\infty} n^2 \exp(-nr_n C_\varepsilon) + \sum_{n=3}^{\infty} n^4 \exp(-nb_n^2 r_n^2 \rho_n^2 C_\varepsilon) \right) = O(1).
\end{aligned}$$

hence

$$\max_{i,j} \left| 1 - \frac{E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right)} \right| \xrightarrow{a.s.} 0.$$

Proof. Take any $\varepsilon > 0$. We start by using a union bound:

$$P \left(\max_{i,j} \left| 1 - \frac{E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right)} \right| > \varepsilon \right) \leq n^2 P \left(\left| 1 - \frac{E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i, \xi_j \right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right)} \right| > \varepsilon \right)$$

Let $r_n(i) = E\left(K\left(\frac{d_{it}^2}{b_n}\right) \mid \xi_i\right) \geq r_n \geq 0$ and $\hat{r}_n(i) = \frac{1}{n-1} \sum_{t=1, t \neq j}^n K\left(\frac{d_{it}^2}{b_n}\right)$. We have:

$$\begin{aligned}
P\left(\left|1 - \frac{r_n(i)}{\hat{r}_n(i)}\right| > \varepsilon\right) &\leq P(|\hat{r}_n(i) - r_n(i)| > \varepsilon |\hat{r}_n(i)|) \\
&\leq P\left(|\hat{r}_n(i) - r_n(i)| > \varepsilon |\hat{r}_n(i)| \text{ and } |\hat{r}_n(i)| \geq \frac{r_n(i)}{2}\right) \\
&\quad + P\left(|\hat{r}_n(i) - r_n(i)| > \varepsilon |\hat{r}_n(i)| \text{ and } |\hat{r}_n(i)| < \frac{r_n(i)}{2}\right) \\
&\leq P\left(|\hat{r}_n(i) - r_n(i)| > \varepsilon \frac{r_n(i)}{2}\right) + P\left(|\hat{r}_n(i)| < \frac{r_n(i)}{2}\right) \\
&\leq P\left(\left|\frac{\hat{r}_n(i)}{r_n(i)} - 1\right| > \frac{\varepsilon}{2}\right) + P\left(\left|\frac{\hat{r}_n(i)}{r_n(i)} - 1\right| > \frac{1}{2}\right)
\end{aligned}$$

where the last line follows from:

$$\begin{aligned}
P\left(|\hat{r}_n(i)| < \frac{r_n(i)}{2}\right) &\leq P\left(\hat{r}_n(i) < \frac{r_n(i)}{2}\right) = P\left(\hat{r}_n(i) - r_n(i) < -\frac{r_n(i)}{2}\right) \\
&= P\left(r_n(i) - \hat{r}_n(i) > \frac{r_n(i)}{2}\right) \leq P\left(|r_n(i) - \hat{r}_n(i)| > \frac{r_n(i)}{2}\right) \\
&\leq P\left(\left|\frac{\hat{r}_n(i)}{r_n(i)} - 1\right| > \frac{1}{2}\right).
\end{aligned}$$

We use the above derivation and the law of iterated expectations to get an upper bound of the form:

$$\begin{aligned}
P\left(\max_{i,j} \left|1 - \frac{E\left(K\left(\frac{d_{it}^2}{b_n}\right) \mid \xi_i\right)}{\frac{1}{n-1} \sum_{t=1, t \neq j}^n K\left(\frac{d_{it}^2}{b_n}\right)}\right| > \varepsilon\right) &\leq n^2 E\left(P\left(\left|\frac{1}{n-1} \sum_{t=1, t \neq j}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} - 1\right| > \frac{\varepsilon}{2} \mid \xi_i, \xi_j\right)\right) \\
&\quad + n^2 E\left(P\left(\left|\frac{1}{n-1} \sum_{t=1, t \neq j}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} - 1\right| > \frac{1}{2} \mid \xi_i, \xi_j\right)\right).
\end{aligned}$$

The last two terms are identical, up to the value of ε . We use $K(\cdot)$ Lipschitz continuous and separate

out the terms with $t = i, j$, so that the remaining average is of i.i.d terms that only depend on t .

$$\begin{aligned}
& n^2 E \left(P \left(\left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K \left(\frac{\hat{d}_{it}^2}{b_n} \right)}{r_n(i)} - 1 \right| > \frac{\varepsilon}{2} \middle| \xi_i, \xi_j \right) \right) \\
& \leq n^2 E \left(P \left(\left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} - 1 + \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n C \left| \frac{\hat{d}_{it}^2 - d_{it}^2}{r_n(i)b_n} \right| \right| > \frac{\varepsilon}{2} \middle| \xi_i, \xi_j \right) \right) \\
& \leq n^2 E \left(P \left(\left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} - 1 \right| > \frac{\varepsilon}{4} \middle| \xi_i, \xi_j \right) \right) \\
& \quad + n^2 E \left(P \left(\left| \max_{i,t} \left(\frac{\hat{d}_{it}^2 - d_{it}^2}{r_n(i)b_n} \right) \right| > \frac{\varepsilon}{4C} \middle| \xi_i, \xi_j \right) \right) \\
& \leq n^2 E \left(P \left(\left| \frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} - 1 \right| > \frac{\varepsilon}{4} - \frac{2C}{(n-2)r_n} \middle| \xi_i, \xi_j \right) \right) + O(n^4 \exp(-nb_n^2 r_n^2 \rho_n^2 C_\varepsilon))
\end{aligned}$$

where the last rate follows from Lemma A.5. For the first term we apply Bernstein's inequality: conditional on ξ_i, ξ_j , $\frac{1}{r_n(i)} K \left(\frac{d_{it}^2}{b_n} \right) - 1$ are i.i.d., mean zero, bounded by $\frac{2C}{r_n}$, with variance $O\left(\frac{1}{r_n}\right)$:

$$\begin{aligned}
\text{Var} \left(\frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} - 1 \middle| \xi_i \right) & \leq \frac{E \left(\left(K \left(\frac{d_{it}^2}{b_n} \right) \right)^2 \middle| \xi_i \right)}{r_n(i)^2} \leq \frac{CE \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right)}{\left(E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right) \right)^2} \\
& = \frac{C}{E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right)} \leq \frac{C}{r_n} = O\left(\frac{1}{r_n}\right),
\end{aligned}$$

hence for any $\varepsilon > 0$:

$$\begin{aligned}
& n^2 P \left(\left| \frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{K \left(\frac{d_{it}^2}{b_n} \right)}{r_n(i)} - 1 \right| > \frac{\varepsilon}{4} - \frac{2C}{(n-2)r_n} \middle| \xi_i, \xi_j \right) \\
& \leq 2n^2 \exp \left(- \frac{(n-2) \left(\frac{\varepsilon}{4} - \frac{2C}{(n-2)r_n} \right)^2}{2 \left(O\left(\frac{1}{r_n}\right) + \frac{1}{3} \frac{C}{r_n} \left(\frac{\varepsilon}{4} - \frac{2C}{(n-2)r_n} \right) \right)} \right) \\
& \leq n^2 \exp(-nr_n C_\varepsilon).
\end{aligned}$$

where $C_\varepsilon > 0$ is some constant dependent on ε . Note that the final value does not depend on the choice of ξ_i, ξ_j , hence it does not change when we take expectation over ξ_i, ξ_j .

It remains to show that the following expression:

$$\begin{aligned} & \sum_{n=3}^{\infty} P \left(\max_{i,j} \left| 1 - \frac{E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right)}{\frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n K \left(\frac{d_{it}^2}{b_n} \right)} \right| > \varepsilon \right) \\ & \leq O \left(\sum_{n=3}^{\infty} n^2 \exp(-nr_n C_\varepsilon) + \sum_{n=3}^{\infty} n^4 \exp(-nb_n^2 r_n^2 \rho_n^2 C_\varepsilon) \right) \end{aligned}$$

is bounded. The last sum is bounded by arguments shown in Lemma A.5. For the first sum we have:

$$\sum_{n=3}^{\infty} n^4 e^{-nr_n C_\varepsilon} = \sum_{n=3}^{\infty} n^2 e^{-nr_n C_\varepsilon \log(n) \frac{1}{\log(n)}} = \sum_{n=3}^{\infty} n^2 \left(e^{\log(n)} \right)^{-C_\varepsilon \frac{nr_n}{\log(n)}} = \sum_{n=3}^{\infty} n^{2-C_\varepsilon \frac{nr_n}{\log(n)}}.$$

It remains to show $\frac{nr_n}{\log(n)} \rightarrow \infty$. We start by showing $r_n \geq C b_n^{\frac{1}{2\alpha}}$:

$$\begin{aligned} r_n &= \inf_{\xi_i} r_n(i) = \inf_{\xi_i} E \left(K \left(\frac{d_{it}^2}{b_n} \right) \middle| \xi_i \right) \geq C_1 \inf_{\xi_i} P \left(\frac{d_{it}^2}{b_n} \leq C_2 \middle| \xi_i \right) \\ &\geq C_1 \inf_{\xi_i} P \left(E_s \left((w_0(\xi_i, \xi_s) - w_0(\xi_t, \xi_s))^2 \middle| \xi_i, \xi_t \right) \leq \frac{C_2}{M_w^2} b_n \middle| \xi_i \right) \\ &\geq C_1 \inf_{\xi_i} P \left(\xi_t \in N \left(\xi_i, \sqrt{\frac{C_2}{M_w^2} b_n} \right) \middle| \xi_i \right) = C_1 \omega \left(\sqrt{\frac{C_2}{M_w^2} b_n} \right) \geq C b_n^{\frac{1}{2\alpha}}. \end{aligned}$$

In the first inequality we use the part of Assumption 1.3 which says the kernel is separated from 0 for input values sufficiently close to 0. The second inequality comes from:

$$\begin{aligned} d_{ij}^2 &= E_t \left((E_s (w_0(\xi_t, \xi_s) (w_0(\xi_i, \xi_s) - w_0(\xi_j, \xi_s)) \middle| \xi_i, \xi_j, \xi_t))^2 \middle| \xi_i, \xi_j \right) \\ &\leq E_t \left(E_s \left(w^2(\xi_t, \xi_s) (w_0(\xi_i, \xi_s) - w_0(\xi_j, \xi_s))^2 \middle| \xi_i, \xi_j, \xi_t \right) \middle| \xi_i, \xi_j \right) \\ &\leq M_w^2 E_s \left((w_0(\xi_i, \xi_s) - w_0(\xi_j, \xi_s))^2 \middle| \xi_i, \xi_j \right) \leq M_w^4 < \infty \end{aligned}$$

where the first inequality is due to Jensen's inequality and the second follows from the fact that for any $\xi_i, \xi_j \in \text{Supp}(\xi)$ $w_0(\xi_i, \xi_j)$ is bounded; we denote the bound by $M_w < \infty$. To see this, recall that $\rho_n w_0(u, v) = h_{0,n}(u, v) \in [0, 1]$, hence we have $w_0(u, v) \in \left[0, \frac{1}{\rho_n}\right]$ for all $n \in \mathbb{N}$. Then also $w_0(u, v) \in \bigcap_{n=1}^{\infty} \left[0, \frac{1}{\rho_n}\right] \subset \left[0, \frac{1}{\sup_n \rho_n}\right]$. $\sup_n \rho_n$ exists since ρ_n , which can be interpreted as the marginal probability of an edge, is bounded above by 1. Let $M_w = \frac{1}{\sup_n \rho_n}$ denote the upper bound on the size of w_0 , i.e. for any $\xi_i, \xi_j \in \text{Supp}(\xi)$ we have $|w_0(\xi_i, \xi_j)| \leq M_w$.

The third inequality follows from the fact that if for some ξ_t we have $\sup_{\xi_s} |w_0(\xi_t, \xi_s) - w_0(\xi_i, \xi_s)| < \delta$, then $E_s \left((w_0(\xi_i, \xi_s) - w_0(\xi_t, \xi_s))^2 \middle| \xi_i, \xi_t \right) < \delta^2$, i.e. $\xi_t \in N(\xi_i, \delta)$. For the final

steps we use Assumption 1.2. The required divergence follows from Assumption 1.4:

$$\frac{nr_n}{\log(n)} \geq C \frac{nb_n^{\frac{1}{2\alpha}}}{\log(n)} \rightarrow \infty.$$

□

Lemma A.2. *Under the assumptions of Theorem 1, there exists a sequence of random variables T_n such that*

$$\sum_{n=3}^{\infty} P(|T_n| > \varepsilon) = O\left(\sum_{n=3}^{\infty} n^2 \exp(-n\rho_n C_\varepsilon)\right) = O(1) \quad \text{hence} \quad T_n \xrightarrow{a.s.} 0$$

and

$$\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| \leq 2M_w + T_n.$$

We add and subtract $h_{0,n}(\xi_t, \xi_j)$, use triangle inequality and the fact that w_0 is bounded by M_w .

$$\begin{aligned} & \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| \\ & \leq \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| + \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{h_{0,n}(\xi_t, \xi_j) - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| \\ & \leq \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| + \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n w_0(\xi_t, \xi_j) \right| + \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n w_0(\xi_i, \xi_j) \right| \\ & \leq \max_{i,j,i \neq j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| + \max_i \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq i}}^n \frac{A_{ti} - h_{0,n}(\xi_t, \xi_i)}{\rho_n} \right| + 2M_w = 2M_w + T_n. \end{aligned}$$

In the second to last step we split the cases of $i \neq j$ and $i = j$. We apply union bound and Bernstein's theorem to the averages. For the first one, we separate out the term with $t = i$ (we later condition on ξ_i, ξ_j , we want the remaining terms in the sum to be i.i.d. after conditioning).

$\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n}$ for $t \neq i, j$ are, conditional on ξ_i, ξ_j , independent, zero mean:

$$\begin{aligned} E\left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \middle| \xi_i, \xi_j\right) &= E\left(E\left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \middle| \xi_i, \xi_j, \xi_t\right) \middle| \xi_i, \xi_j\right) \\ &= E\left(E\left(\frac{A_{tj}}{\rho_n} \middle| \xi_i, \xi_j, \xi_t\right) - \frac{h_{0,n}(\xi_t, \xi_j)}{\rho_n} \middle| \xi_i, \xi_j\right) = E(w_0(\xi_t, \xi_j) - w_0(\xi_t, \xi_j) \middle| \xi_i, \xi_j) = 0 \end{aligned}$$

and bounded by $\frac{1}{\rho_n}$: since A and $h_{0,n}$ take values in $[0, 1]$, we have $\left|\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n}\right| \leq \frac{1}{\rho_n}$. The second moments are $O\left(\frac{1}{\rho_n}\right)$:

$$\begin{aligned} Var\left(\frac{1}{\rho_n}(A_{tj} - h_{0,n}(\xi_t, \xi_j)) \middle| \xi_i, \xi_j\right) &= E\left(\left(\frac{1}{\rho_n}(A_{tj} - h_{0,n}(\xi_t, \xi_j))\right)^2 \middle| \xi_i, \xi_j\right) \\ &= E\left(\frac{1}{\rho_n^2}(h_{0,n}(\xi_t, \xi_j)(1 - h_{0,n}(\xi_t, \xi_j))) \middle| \xi_i, \xi_j\right) = E\left(w_0(\xi_t, \xi_j)\left(\frac{1}{\rho_n} - w_0(\xi_t, \xi_j)\right) \middle| \xi_j\right) \\ &= O\left(\frac{1}{\rho_n}\right) + O(1) = O\left(\frac{1}{\rho_n}\right). \end{aligned}$$

For any $\varepsilon > 0$:

$$\begin{aligned} &P\left(\max_{i,j,i \neq j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| > \varepsilon\right) \\ &\leq P\left(\max_{i,j,i \neq j} \left| \frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| > \varepsilon - \frac{1}{(n-2)\rho_n}\right) \\ &\leq n(n-1)E\left(P\left(\left|\frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{1}{\rho_n}(A_{tj} - h_{0,n}(\xi_t, \xi_j))\right| > \varepsilon - \frac{1}{(n-2)\rho_n} \middle| \xi_i, \xi_j\right)\right) \\ &\leq 2n(n-1)\exp\left(-\frac{(n-2)\left(\varepsilon - \frac{1}{(n-2)\rho_n}\right)^2}{2\left(O\left(\frac{1}{\rho_n}\right) + \frac{1}{3}\frac{1}{\rho_n}\left(\varepsilon - \frac{1}{(n-2)\rho_n}\right)\right)}\right) \\ &\leq n^2 \exp(-n\rho_n C_\varepsilon). \end{aligned}$$

Similarly,

$$\begin{aligned}
& P \left(\max_i \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq i}}^n \frac{1}{\rho_n} (A_{ti} - h_{0,n}(\xi_t, \xi_i)) \right| > \varepsilon \right) \\
& \leq nE \left(P \left(\left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq i}}^n \frac{1}{\rho_n} (A_{ti} - h_{0,n}(\xi_t, \xi_i)) \right| > \varepsilon \mid \xi_i \right) \right) \\
& \leq 2n \exp \left(-\frac{(n-1)\varepsilon^2}{2 \left(O\left(\frac{1}{\rho_n}\right) + \frac{1}{3} \frac{1}{\rho_n} \varepsilon \right)} \right) \\
& \leq n \exp(-n\rho_n C_\varepsilon).
\end{aligned}$$

This is dominated by the previous term. Combining the above results, for any $\varepsilon > 0$:

$$\begin{aligned}
\sum_{n=3}^{\infty} P \left(\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{A_{tj} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| > 2M_w + \varepsilon \right) & \leq \sum_{n=3}^{\infty} P(|T_n| > \varepsilon) \\
& \leq O \left(\sum_{n=3}^{\infty} n^2 \exp(-n\rho_n C_\varepsilon) \right) < \infty.
\end{aligned}$$

For the last claim, note that under Assumption 1.1 we have $\frac{\log(n)}{\rho_n n} \rightarrow 0$. Then:

$$\sum_{n=3}^{\infty} n^2 e^{-n\rho_n C_\varepsilon} = \sum_{n=3}^{\infty} n^2 e^{-n\rho_n C_\varepsilon \log(n) \frac{1}{\log(n)}} = \sum_{n=3}^{\infty} n^2 \left(e^{\log(n)} \right)^{-C_\varepsilon \frac{\rho_n n}{\log(n)}} = \sum_{n=3}^{\infty} n^{2-C_\varepsilon \frac{\rho_n n}{\log(n)}} < \infty$$

for any $C_\varepsilon > 0$, since $2 - C_\varepsilon \frac{\rho_n n}{\log(n)} \rightarrow -\infty$. Hence $T_n \xrightarrow{a.s.} 0$ and the term of interest is almost surely bounded above by $2M_w$.

Lemma A.3. *Under the assumptions of Theorem 1, for any $\varepsilon > 0$:*

$$\sum_{n=3}^{\infty} P \left(\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| > \varepsilon \right) \leq O \left(\sum_{n=3}^{\infty} n^2 \exp(-nr_n \rho_n C_\varepsilon) \right) < \infty$$

hence

$$\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| \xrightarrow{a.s.} 0.$$

Proof. We start by separating the cases when $i \neq j$ and $i = j$:

$$\begin{aligned} \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| &\leq \max_{i,j,i \neq j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| \\ &+ \max_i \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq i}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{ti} - h_{0,n}(\xi_t, \xi_i)}{\rho_n} \right) \right|. \end{aligned}$$

We split the first sum into the term with $t = i$ and the rest the rest (which is i.i.d. over t), use the triangle inequality and the fact that A_{ij} and $h_{0,n}(\xi_i, \xi_j)$ take values in $[0, 1]$ for any choice of i, j while the kernel function is absolutely bounded.

$$\begin{aligned} &\max_{i,j,i \neq j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| \\ &\leq \max_{i,j,i \neq j} \left| \frac{n-2}{n-1} \frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| \\ &\quad + \max_{i,j,i \neq j} \frac{1}{n-1} \left(\underbrace{\frac{K\left(\frac{d_{ii}^2}{b_n}\right)}{r_n(i)}}_{\leq \frac{C}{r_n}} \underbrace{\left| \frac{A_{ij} - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right|}_{\leq \frac{1}{\rho_n}} \right) \\ &\leq \max_{i,j,i \neq j} \frac{n-2}{n-1} \left| \frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| + \frac{C}{(n-1)r_n\rho_n}. \end{aligned}$$

The expression inside the sum in the first term is bounded by $\frac{C}{r_n\rho_n}$, hence after conditioning on ξ_i, ξ_j we can apply the Bernstein's inequality for bounded i.i.d. random variables. The conditional expectation of that term is zero:

$$\begin{aligned} E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \middle| \xi_i, \xi_j \right) &= E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{1}{\rho_n} E(A_{tj} | \xi_i, \xi_j, \xi_t) - w_0(\xi_t, \xi_j) \right) \middle| \xi_i, \xi_j \right) \\ &= E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} (w_0(\xi_t, \xi_j) - w_0(\xi_t, \xi_j)) \middle| \xi_i, \xi_j \right) = 0 \end{aligned}$$

where the first equality is due to the law of iterated expectations, the second uses the fact that d_{it}^2 , $r_n(i)$ and $h_{0,n}(\xi_t, \xi_j)$ are not random after conditioning on ξ_i, ξ_j, ξ_t . A_{tj} is independent of ξ_i , hence $E(A_{tj} | \xi_i, \xi_j, \xi_t) = E(A_{tj} | \xi_j, \xi_t)$ which by definition equals $h_{0,n}(\xi_t, \xi_j) = \rho_n w_0(\xi_t, \xi_j)$. The

conditional variance is $O\left(\frac{1}{r_n \rho_n}\right)$:

$$\begin{aligned}
& \text{Var} \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \middle| \xi_i, \xi_j \right) \\
&= E \left(\left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \right)^2 E \left(\left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right)^2 \middle| \xi_i, \xi_j, \xi_t \right) \middle| \xi_i, \xi_j \right) \\
&= E \left(\left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \right)^2 \left(w_0(\xi_t, \xi_j) \left(\frac{1}{\rho_n} - w_0(\xi_t, \xi_j) \right) \right) \middle| \xi_i, \xi_j \right) \\
&\leq \frac{M_w}{\rho_n} \underbrace{E \left(\left(K\left(\frac{d_{it}^2}{b_n}\right) \right)^2 \middle| \xi_i \right)}_{=O\left(\frac{1}{r_n}\right)} = O\left(\frac{1}{r_n \rho_n}\right)
\end{aligned}$$

where in the last line we use that the kernel function is bounded ($K(\cdot) \leq C$ by Assumption 1.3) and hence

$$\frac{E \left(\left(K\left(\frac{d_{it}^2}{b_n}\right) \right)^2 \middle| \xi_i \right)}{r_n(i)^2} \leq \frac{CE \left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i \right)}{\left(E \left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i \right) \right)^2} = \frac{C}{E \left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i \right)} \leq \frac{C}{r_n} = O\left(\frac{1}{r_n}\right).$$

By union bound and Bernstein's inequality, for any $\varepsilon > 0$ and $n \geq 3$:

$$\begin{aligned}
& P \left(\max_{\substack{i,j,i \neq j \\ t=1 \\ t \neq j}} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| > \varepsilon \right) \\
&\leq n(n-1) E \left(P \left(\left| \frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right| > \varepsilon - \frac{C}{(n-2)\rho_n r_n} \middle| \xi_i, \xi_j \right) \right) \\
&\leq 2n(n-1) \exp \left(\frac{-(n-2) \left(\varepsilon - \frac{C}{(n-2)\rho_n r_n} \right)^2}{2 \left(O\left(\frac{1}{r_n \rho_n}\right) + \frac{C}{3r_n \rho_n} \left(\varepsilon - \frac{C}{(n-2)\rho_n r_n} \right) \right)} \right) \\
&\leq n^2 \exp(-nr_n \rho_n C_\varepsilon)
\end{aligned}$$

for some $C_\varepsilon > 0$. We can proceed in a very similar way for the case of $i = j$ to get:

$$P \left(\max_i \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq i}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \frac{A_{ti} - h_{0,n}(\xi_t, \xi_i)}{\rho_n} \right| > \varepsilon \right) \leq O(n \exp(-nr_n \rho_n C_\varepsilon)).$$

Combining all the terms gives the required result: for any $\varepsilon > 0$

$$\sum_{n=3}^{\infty} P \left(\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| > \varepsilon \right) \leq O \left(\sum_{n=3}^{\infty} n^2 \exp(-nr_n \rho_n C_\varepsilon) \right) < \infty$$

under Assumption 1.1 and Assumption 1.4 which, by derivation similar to that at the end of the proof of A.1, give:

$$\frac{n\rho_n r_n}{\log(n)} \geq C \frac{n\rho_n b_n^{\frac{1}{2\alpha}}}{\log(n)} \rightarrow \infty.$$

Hence

$$\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{A_{tj} - h_{0,n}(\xi_t, \xi_j)}{\rho_n} \right) \right| \xrightarrow{a.s.} 0.$$

□

Lemma A.4. *Under the assumptions of Theorem 1, for any $\varepsilon > 0$:*

$$\begin{aligned} & P \left(\left\| \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{h_{0,n}(\xi_t, \xi_j) - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right) \right| \right\| > \varepsilon \right) \\ & \leq O(n^2 \exp(-nr_n C_\varepsilon)) + O\left(b_n^{\frac{\alpha^2}{(2\alpha+1)^2}}\right) \rightarrow 0. \end{aligned}$$

hence

$$\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \left(\frac{h_{0,n}(\xi_t, \xi_j) - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right) \right| \xrightarrow{p} 0.$$

Proof. Intuitively, this result holds because as n increases $\frac{d_{it}^2}{b_n}$ becomes large, and hence $K\left(\frac{d_{it}^2}{b_n}\right)$ becomes zero, unless ξ_i and ξ_t are very close to each other in the sense that their $h_{0,n}(\xi_t, \xi_j)$ and $h_{0,n}(\xi_i, \xi_j)$ are similar for all ξ_j .

We start by showing that whenever $h_{0,n}(\xi_t, \xi_j)$ and $h_{0,n}(\xi_i, \xi_j)$ are not close, their distance d_{it} will be separated away from zero.

We follow the ideas from Auerbach (2022)'s proof of Lemma 1 which shows that for any i, t, n and any $\varepsilon > 0$ we can find a $\delta > 0$ such that $\sqrt{E\left(\left(w_0(\xi_i, \xi_j) - w_0(\xi_t, \xi_j)\right)^2 \middle| \xi_i, \xi_t\right)} \geq \varepsilon \implies d_{it} =$

$$\sqrt{E \left((\varphi(\xi_i, \xi_j) - \varphi(\xi_t, \xi_j))^2 \middle| \xi_i, \xi_t \right)} \geq \delta.$$

Our idea is to add an extra step at the beginning: if for given i, j, t there is a $\nu > 0$ for which we have $|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)| \equiv \left| \frac{h_{0,n}(\xi_t, \xi_j) - h_{0,n}(\xi_i, \xi_j)}{\rho_n} \right| > \nu$, then there exists an $\varepsilon > 0$ such that $\sqrt{E \left((w_0(\xi_i, \xi_j) - w_0(\xi_t, \xi_j))^2 \middle| \xi_i, \xi_t \right)} \geq \varepsilon$ (which in turn implies $d_{it} \geq \delta$). In other words, $|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)|$ can be large only if d_{it} is large, in which case the weight placed on that term is small.

In our case, the issue is that in $\sqrt{E \left((w_0(\xi_i, \xi_j) - w_0(\xi_t, \xi_j))^2 \middle| \xi_i, \xi_t \right)}$ we take an expectation with respect to j , but the initial statement is given for a fixed j . To get around it, we replace the fixed j with a random element of a neighbourhood of j , then take an expectation with respect to an element of that neighbourhood, and use an upper bound which takes expectation over all possible values, not just those in the neighbourhood of j .

Recall from Assumption 1.2 that $N(\xi_j, \delta) = \{\xi_k : \sup_{\xi_t} |w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta\}$ denotes the neighbourhood of ξ_j of size δ . We fix i, j, t, k where $k \in N(\xi_j, \frac{\nu}{3})$. Then

$$\begin{aligned} & \mathbb{1} (|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)| > \nu) \\ &= \mathbb{1} (|w_0(\xi_t, \xi_j) - w_0(\xi_t, \xi_k) + w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k) + w_0(\xi_i, \xi_k) - w_0(\xi_i, \xi_j)| > \nu) \\ &\leq \underbrace{\mathbb{1} \left(|w_0(\xi_t, \xi_j) - w_0(\xi_t, \xi_k)| > \frac{\nu}{3} \right)}_{=0} + \mathbb{1} \left(|w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k)| > \frac{\nu}{3} \right) \\ &\quad + \underbrace{\mathbb{1} \left(|w_0(\xi_i, \xi_k) - w_0(\xi_i, \xi_j)| > \frac{\nu}{3} \right)}_{=0} \\ &= \mathbb{1} \left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 > \frac{\nu^2}{9} \right). \end{aligned}$$

If the above holds for any fixed i, j, t and for any $\xi_k \in N(\xi_j, \frac{\nu}{3})$, it also holds if we take expectation over $\xi_k \in N(\xi_j, \frac{\nu}{3})$. Recall from Assumption 1.2 that $\omega(\delta) = \inf_{\xi_j} P(\xi_k \in N(\xi_j, \delta) | \xi_j)$ and $\omega(\delta) \geq$

$\left(\frac{\delta}{C}\right)^{\frac{1}{\alpha}}$ for all $\delta > 0$. We use $E(X|A) = \frac{E(\mathbb{1}_A X)}{P(A)}$.

$$\begin{aligned}
& \mathbb{1}(|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)| > \nu) \\
& \leq \mathbb{1}\left(E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_k \in N\left(\xi_j, \frac{\nu}{3}\right), \xi_i, \xi_j, \xi_t\right) > \frac{\nu^2}{9}\right) \\
& = \mathbb{1}\left(\frac{E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mathbb{1}\left(\xi_k \in N\left(\xi_j, \frac{\nu}{3}\right)\right) \mid \xi_i, \xi_j, \xi_t\right)}{P\left(\xi_k \in N\left(\xi_j, \frac{\nu}{3}\right) \mid \xi_j\right)} > \frac{\nu^2}{9}\right) \\
& \leq \mathbb{1}\left(E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mathbb{1}\left(\xi_k \in N\left(\xi_j, \frac{\nu}{3}\right)\right) \mid \xi_i, \xi_j, \xi_t\right) > \frac{\nu^2}{9} \omega\left(\frac{\nu}{3}\right)\right) \\
& \leq \mathbb{1}\left(E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_i, \xi_t\right) > \frac{\nu^2}{9} \omega\left(\frac{\nu}{3}\right)\right) \\
& = \mathbb{1}\left(\sqrt{E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_i, \xi_t\right)} > \frac{\nu}{3} \sqrt{\omega\left(\frac{\nu}{3}\right)}\right)
\end{aligned}$$

We set $\varepsilon = \frac{\nu}{3} \sqrt{\omega\left(\frac{\nu}{3}\right)}$, which completes the argument.

Like Auerbach (2022)²⁵, we assume there exist some $\alpha, C > 0$ such that for any δ we have $\omega(\delta) \geq \left(\frac{\delta}{C}\right)^{\frac{1}{\alpha}}$ (this is our Assumption 1.2). Then:

$$\begin{aligned}
\mathbb{1}(|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)| > \nu) & \leq \mathbb{1}\left(\sqrt{E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_i, \xi_t\right)} > \frac{\nu}{3} \sqrt{\left(\frac{\nu}{3C}\right)^{\frac{1}{\alpha}}}\right) \\
& = \mathbb{1}\left(3C^{\frac{1}{2\alpha+1}} \left(E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_i, \xi_t\right)\right)^{\frac{\alpha}{2\alpha+1}} > \nu\right).
\end{aligned}$$

hence

$$|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)| \leq 3C^{\frac{1}{2\alpha+1}} \left(E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_i, \xi_t\right)\right)^{\frac{\alpha}{2\alpha+1}}.$$

Combining with Auerbach (2022)²⁶ result:

$$\sqrt{E\left((w_0(\xi_t, \xi_k) - w_0(\xi_i, \xi_k))^2 \mid \xi_i, \xi_t\right)} \leq 2C^{\frac{1}{4\alpha+2}} d_{it}^{\frac{\alpha}{2\alpha+1}}.$$

we get

$$|w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)| \leq \tilde{C} d_{it}^{\frac{2\alpha^2}{(2\alpha+1)^2}}$$

25. One major difference is that Auerbach (2022) does not allow for sparsity in his model, in his case $\rho_n = 1$. Hence we impose an assumption analogous to his to w_0 , not $h_{0,n}$. If we were to define everything in terms of $h_{0,n}$, we would need $N_n(\xi_j, \delta) = \{\xi_k : \sup_{\xi_t} |h_{0,n}(\xi_t, \xi_k) - h_{0,n}(\xi_t, \xi_j)| < \delta\}$, $\omega_n(\delta) = \inf_{\xi_j} P(\xi_k \in N_n(\xi_j, \delta) \mid \xi_j)$ and $\omega_n(\delta) \geq \left(\frac{\delta}{\rho_n C}\right)^{\frac{1}{\alpha}}$.

26. This is Auerbach (2022) Lemma A1 restated in our notation. To account for the fact that Auerbach (2022) does not allow for sparsity we replace their f , which is equivalent to our $h_{0,n}$, with a w_0 and their δ with our equivalent term d .

for $\tilde{C} = 3 \times 2^{\frac{2\alpha}{2\alpha+1}} \times C^{\frac{3\alpha+1}{(2\alpha+1)^2}}$.

We can now return to the term of interest.

$$\begin{aligned} & \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} (w_0(\xi_t, \xi_j) - w_0(\xi_i, \xi_j)) \right| \leq \max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \tilde{C} d_{it}^{-\frac{2\alpha^2}{(2\alpha+1)^2}} \right| \\ & \leq \tilde{C} \left(\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{-\frac{2\alpha^2}{(2\alpha+1)^2}} - E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{-\frac{2\alpha^2}{(2\alpha+1)^2}} \middle| \xi_i \right) \right| + \right. \\ & \quad \left. + \max_i \left| E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{-\frac{2\alpha^2}{(2\alpha+1)^2}} \middle| \xi_i \right) \right| \right). \end{aligned}$$

The first term goes to zero by the union bound and Bernstein's inequality, where, conditionally on ξ_i, ξ_j and after separating out the term with $t = i$, the terms inside the average are i.i.d., mean zero, bounded by $r_n^{-1} C M_w^{\frac{4\alpha^2}{(2\alpha+1)^2}} = O(r_n^{-1})$ and have variance $O(r_n^{-1})$. For any $\varepsilon > 0$:

$$\begin{aligned} & P \left(\max_{i,j} \left| \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq j}}^n \frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{-\frac{2\alpha^2}{(2\alpha+1)^2}} - E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{-\frac{2\alpha^2}{(2\alpha+1)^2}} \middle| \xi_i \right) \right| > \varepsilon \right) \\ & \leq 2n(n-1) \exp \left(- \frac{(n-2) \left(\varepsilon - \frac{C M_w^{\frac{4\alpha^2}{(2\alpha+1)^2}}}{(n-2)r_n} \right)^2}{2 \left(O(r_n^{-1}) + \frac{C M_w^{\frac{4\alpha^2}{(2\alpha+1)^2}}}{3r_n} \left(\varepsilon - \frac{C M_w^{\frac{4\alpha^2}{(2\alpha+1)^2}}}{(n-2)r_n} \right) \right)} \right) \\ & \quad + 2n \exp \left(- \frac{(n-1)\varepsilon^2}{2 \left(O(r_n^{-1}) + \frac{C M_w^{\frac{4\alpha^2}{(2\alpha+1)^2}}}{3r_n} \varepsilon \right)} \right) \\ & \leq n^2 \exp(-nr_n C_\varepsilon) \rightarrow 0. \end{aligned}$$

The last convergence was shown at the end of the proof of Lemma A.1.

It remains to show that the last term goes to zero too. By Assumption 1.3, there exists a $D \in \mathbb{R}$ such that $\forall |u| > D : K(u) = 0$. If $d_{it} \neq 0$: $\frac{d_{it}^2}{b_n} = O\left(\frac{1}{b_n}\right) \rightarrow \infty$, so eventually, as $n \rightarrow \infty$, $\frac{d_{it}^2}{b_n} > D$

and $K\left(\frac{d_{it}^2}{b_n}\right) = 0$ (and if $d_{it} = 0$ the whole term is identically equal to zero). We have:

$$\begin{aligned}
\max_i \left| E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{\frac{2\alpha^2}{(2\alpha+1)^2}} \middle| \xi_i \right) \right| &= \max_i \left| E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} d_{it}^{\frac{2\alpha^2}{(2\alpha+1)^2}} \mathbb{1}\left(\frac{d_{it}^2}{b_n} \leq D\right) \middle| \xi_i \right) \right| \\
&\leq \max_i \left| (Db_n)^{\frac{\alpha^2}{(2\alpha+1)^2}} E \left(\frac{K\left(\frac{d_{it}^2}{b_n}\right)}{r_n(i)} \middle| \xi_i \right) \right| \\
&= \max_i \left| (Db_n)^{\frac{\alpha^2}{(2\alpha+1)^2}} \underbrace{E \left(K\left(\frac{d_{it}^2}{b_n}\right) \middle| \xi_i \right)}_{=1} \right| \\
&\leq D^{\frac{\alpha^2}{(2\alpha+1)^2}} b_n^{\frac{\alpha^2}{(2\alpha+1)^2}} = O\left(b_n^{\frac{\alpha^2}{(2\alpha+1)^2}}\right) \rightarrow 0.
\end{aligned}$$

The last expression goes to zero by Assumption 1.4. Note however that under Assumption 1.4 the rate of convergence to zero is too slow to ensure almost sure convergence of this term. This is the reason why we only get uniform convergence in probability in Theorem 1 and convergence weakly in probability in Theorem 3. □

Lemma A.5. *Under the assumptions of Theorem 1, for any $\varepsilon > 0$:*

$$\sum_{n=3}^{\infty} P\left(\frac{1}{b_n r_n} \max_{i,j} |\hat{d}_{ij}^2 - d_{ij}^2| > \varepsilon\right) = O\left(\sum_{n=3}^{\infty} n^2 \exp(-nb_n^2 r_n^2 \rho_n^2 C_\varepsilon)\right) = O(1)$$

hence

$$\frac{1}{b_n r_n} \max_{i,j} |\hat{d}_{ij}^2 - d_{ij}^2| \xrightarrow{a.s.} 0.$$

Proof. We follow the same steps as in Lemma B1 in Auerbach (2022). By definition:

$$\begin{aligned}
\hat{d}_{ij} &= \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts}}{\rho_n} \left(\frac{A_{is} - A_{js}}{\rho_n} \right) \right)^2} \\
\tilde{d}_{ij} &= \sqrt{\frac{1}{n} \sum_{t=1}^n (\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2} \\
d_{ij} &= \sqrt{E_t \left((\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2 \middle| \xi_i, \xi_j \right)}.
\end{aligned}$$

Take any $\varepsilon > 0$. We have:

$$\begin{aligned}
& P\left(\frac{1}{b_n r_n} \max_{i,j} \left| \hat{d}_{ij}^2 - d_{ij}^2 \right| > \varepsilon\right) = P\left(\max_{i,j} \left| \hat{d}_{ij}^2 - d_{ij}^2 \right| > \varepsilon b_n r_n\right) \\
& = P\left(\max_{i,j} \left| \hat{d}_{ij}^2 - \tilde{d}_{ij}^2 + \tilde{d}_{ij}^2 - d_{ij}^2 \right| > \varepsilon b_n r_n\right) \\
& \leq P\left(\max_{i,j} \left| \hat{d}_{ij}^2 - \tilde{d}_{ij}^2 \right| > \frac{\varepsilon b_n r_n}{2}\right) + P\left(\max_{i,j} \left| \tilde{d}_{ij}^2 - d_{ij}^2 \right| > \frac{\varepsilon b_n r_n}{2}\right) \tag{35}
\end{aligned}$$

where the last inequality follows from the fact that $|a + b| > \varepsilon$ implies $|a| > \frac{\varepsilon}{2}$ or $|b| > \frac{\varepsilon}{2}$ and hence $P(|a + b| > \varepsilon) \leq P(|a| > \frac{\varepsilon}{2}) + P(|b| > \frac{\varepsilon}{2})$.

For the first term in (35), we plug in the definitions, then use $a^2 - b^2 = (a - b)(a + b)$ and the fact that the second bracket approaches a limit bounded by $4M_w^2$:

$$\begin{aligned}
& P\left(\max_{i,j} \left| \hat{d}_{ij}^2 - \tilde{d}_{ij}^2 \right| > \frac{\varepsilon b_n r_n}{2}\right) \\
& = P\left(\max_{i,j} \left| \frac{1}{n} \sum_{t=1}^n \left(\left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts}}{\rho_n} \left(\frac{A_{is}}{\rho_n} - \frac{A_{js}}{\rho_n} \right) \right)^2 - (\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2 \right) \right| > \frac{\varepsilon b_n r_n}{2}\right) \\
& = P\left(\max_{i,j} \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) + \varphi(\xi_j, \xi_t) - \frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{js}}{\rho_n^2} \right) \times \right. \right. \\
& \quad \left. \left. \times \left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{is}}{\rho_n^2} + \varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t) - \frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{js}}{\rho_n^2} \right) \right| > \frac{\varepsilon b_n r_n}{2}\right) \\
& \leq P\left(\max_{i,j} \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) + \varphi(\xi_j, \xi_t) - \frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{js}}{\rho_n^2} \right) \right| > \frac{\varepsilon b_n r_n}{16M_w^2}\right) + \\
& \quad + P\left(\max_{i,j,t} \left| \frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{is}}{\rho_n^2} + \varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t) - \frac{1}{n} \sum_{s=1}^n \frac{A_{ts} A_{js}}{\rho_n^2} \right| > 8M_w^2\right)
\end{aligned}$$

where the last equality follows from the fact that $ab > \varepsilon$ implies $b \geq M$ or $a > \frac{\varepsilon}{M}$. For the first term, we again note that $|a + b| > \varepsilon$ implies $|a| > \frac{\varepsilon}{2}$ or $|b| > \frac{\varepsilon}{2}$, we split the expression into a part

with terms that only depend on i and a part with terms that only depend on j . We then get:

$$\begin{aligned}
& P \left(\max_{i,j} \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) + \varphi(\xi_j, \xi_t) - \frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{js}}{\rho_n^2} \right) \right| > \frac{\varepsilon b_n r_n}{16M_w^2} \right) \\
& \leq 2P \left(\max_i \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right) \right| > \frac{\varepsilon b_n r_n}{32M_w^2} \right) \\
& \leq 2nE \left(P \left(\max_{t,t \neq i} \left| \frac{1}{n-1} \sum_{s=1}^n \frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right| > \frac{\varepsilon b_n r_n}{32M_w^2} - \frac{1}{(n-1)\rho_n^2} \middle| \xi_i \right) \right) \\
& \leq 2n(n-1)E \left(P \left(\left| \frac{1}{n-2} \sum_{\substack{s=1 \\ s \neq i,t}}^n \frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right| > \frac{\varepsilon b_n r_n}{32M_w^2} - \frac{1}{(n-1)\rho_n^2} - \frac{2}{(n-2)\rho_n^2} \middle| \xi_i, \xi_t \right) \right) \\
& \leq 4n(n-1) \exp \left(\frac{-(n-2) \left(\frac{\varepsilon b_n r_n}{32M_w^2} - \frac{1}{(n-1)\rho_n^2} - \frac{2}{(n-2)\rho_n^2} \right)^2}{2 \left(\frac{C}{\rho_n^2} + \frac{1}{3} \frac{1}{\rho_n^2} \left(\frac{\varepsilon b_n r_n}{32M_w^2} - \frac{1}{(n-1)\rho_n^2} - \frac{2}{(n-2)\rho_n^2} \right) \right)} \right) \\
& \leq n^2 \exp(-nb_n^2 r_n^2 \rho_n^2 C_\varepsilon)
\end{aligned}$$

where the second inequality follows from the union bound applied to \max_i and the fact that $\frac{1}{n} \sum_{t=1}^n x_t \leq \frac{n-1}{n} \frac{1}{n-1} \sum_{\substack{t=1 \\ t \neq i}}^n \max_{t,t \neq i} x_t + \frac{1}{n} x_i = \frac{n-1}{n} \left(\max_{t,t \neq i} x_t + \frac{1}{n-1} x_i \right)$. In this case $x_i = \frac{1}{n} \sum_{s=1}^n \frac{A_{is}^2}{\rho_n^2} - \varphi(\xi_i, \xi_i)$ and $|x_i| \leq \frac{1}{\rho_n^2}$ (since A and $\rho_n^2 \varphi$ both belong to $[0, 1]$). We also use the fact that $\frac{n-1}{n}|a| > \varepsilon$ implies $|a| > \varepsilon$. Next, notice that $\left| a \pm \frac{1}{(n-1)\rho_n^2} \right| > \varepsilon$ implies that either $|a| \geq a > \varepsilon \pm \frac{1}{(n-1)\rho_n^2} > \varepsilon - \frac{1}{(n-1)\rho_n^2}$ or $|a| \geq -a > \varepsilon \pm \frac{1}{(n-1)\rho_n^2} > \varepsilon - \frac{1}{(n-1)\rho_n^2}$, so in either case we get $|a| > \varepsilon - \frac{1}{(n-1)\rho_n^2}$. For the third inequality, we again apply the union bound, this time over $t \neq i$, and separate out the terms with $s = i$ or $s = t$, similarly to the previous step. The final inequality follows from Bernstein's inequality with $\left| \frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right| \leq \frac{1}{\rho_n^2}$ and $\text{Var} \left(\frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right) = E \left(\varphi(\xi_i, \xi_t) \left(\frac{1}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right) \right) \leq \frac{M_w^2}{\rho_n^2}$.

For the second term, take any $0 < \varepsilon < 2M_w^2$ and use similar arguments.

$$\begin{aligned}
& P \left(\max_{i,j,t} \left| \frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{is}}{\rho_n^2} + \varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t) - \frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{js}}{\rho_n^2} \right| > 8M_w^2 \right) \\
& \leq 2P \left(\max_{i,t} \left| \frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{is}}{\rho_n^2} + \varphi(\xi_i, \xi_t) \right| > 4M_w^2 \right) \\
& \leq 2 \left(\underbrace{P \left(\max_{i,t} \left| \frac{1}{n} \sum_{s=1}^n \frac{A_{ts}A_{is}}{\rho_n^2} - \varphi(\xi_i, \xi_t) \right| > \varepsilon \right)}_{\leq 2n^2 \exp \left(\frac{-(n-2) \left(\varepsilon - \frac{2}{(n-1)\rho_n^2} \right)^2}{2 \left(\frac{C}{\rho_n^2} + \frac{1}{3} \frac{1}{\rho_n^2} \left(\varepsilon - \frac{2}{(n-1)\rho_n^2} \right) \right)} \right)} + \underbrace{P \left(\max_{i,t} |2\varphi(\xi_i, \xi_t)| > \underbrace{4M_w^2 - \varepsilon}_{> 2M_w^2} \right)}_{\substack{\leq 2M_w^2 \\ =0}} \right) \\
& \leq n^2 \exp(-n\rho_n^2 C_\varepsilon)
\end{aligned}$$

We show that the second term in (35) goes to zero almost surely by applying the union bound and Bernstein's inequality. We also use that, by definition, $d_{ii} = \tilde{d}_{ii} = 0$.

$$\begin{aligned}
P\left(\max_{i,j} \left| \tilde{d}_{ij}^2 - d_{ij}^2 \right| > \frac{\varepsilon b_n r_n}{2}\right) &\leq P\left(\max_{i,j,i \neq j} \left| \tilde{d}_{ij}^2 - d_{ij}^2 \right| > \frac{\varepsilon b_n r_n}{2}\right) + \underbrace{P\left(\max_i \left| \tilde{d}_{ii}^2 - d_{ii}^2 \right| > \frac{\varepsilon b_n r_n}{2}\right)}_{=0} \\
&\leq n(n-1)E\left(P\left(\left| \tilde{d}_{ij}^2 - d_{ij}^2 \right| > \frac{\varepsilon b_n r_n}{2} \middle| \xi_i, \xi_j\right)\right) \\
&= n(n-1)E\left(P\left(\left|\frac{1}{n} \sum_{t=1}^n \left((\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2 - E\left((\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2 \middle| \xi_i, \xi_j\right)\right)\right| \right. \right. \\
&\quad \left. \left. > \frac{\varepsilon b_n r_n}{2} \middle| \xi_i, \xi_j\right)\right) \\
&\leq n(n-1)E\left(P\left(\left|\frac{1}{n-2} \sum_{\substack{t=1 \\ t \neq i,j}}^n \left((\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2 - E\left((\varphi(\xi_i, \xi_t) - \varphi(\xi_j, \xi_t))^2 \middle| \xi_i, \xi_j\right)\right)\right| \right. \right. \\
&\quad \left. \left. > \frac{\varepsilon b_n r_n}{2} - \frac{2}{(n-2)\rho_n^2} \middle| \xi_i, \xi_j\right)\right) \\
&\leq 2n(n-1) \exp\left(\frac{-(n-2) \left(\frac{\varepsilon b_n r_n}{2} - \frac{2}{(n-2)\rho_n^2}\right)^2}{2 + \frac{2\left(\frac{\varepsilon b_n r_n}{2} - \frac{2}{(n-2)\rho_n^2}\right)}{3}}\right) \leq n^2 \exp(-nb_n^2 r_n^2 C_\varepsilon).
\end{aligned}$$

The conclusion follows since

$$\sum_{n=3}^{\infty} P\left(\frac{1}{b_n r_n} \max_{i,j} \left| \tilde{d}_{ij}^2 - d_{ij}^2 \right| > \varepsilon\right) = O\left(\sum_{n=3}^{\infty} n^2 \exp(-nb_n^2 r_n^2 \rho_n^2 C_\varepsilon)\right) = O(1)$$

The last term is bounded for any $C_\varepsilon > 0$ because under Assumption 1.4 $\frac{\log(n)}{nb_n^2 r_n^2 \rho_n^2} \rightarrow 0$. \square

A.2 Proofs of consistency of the bootstrap procedure

Proof of Theorem 2. We start by constructing a particular coupling in $\Gamma(A^*, H)$. Let $\tilde{\gamma}$ be a particular joint distribution over \hat{F}_n and F_0 , the details of which we specify later in the proof. We use $\tilde{\gamma}$ to construct a coupling between A^* and H : we draw pairs $\{(\xi_i^*, \xi_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} \tilde{\gamma}$. We also independently draw $\{\eta_{ij}\}_{i < j}^n \stackrel{i.i.d.}{\sim} \mathcal{U}[0, 1]$ and set $\eta_{ij}^* = \eta_{ij}$. We denote $(A^*, H) \sim \tilde{\nu}$ and note that this construction gives correct marginal distributions of A^* and H , hence:

$$\begin{aligned} W_p^p(A^*, H) &\leq \int d_{GM}^p(A^*, H) d\tilde{\nu} \leq \int \left(\binom{n}{2}^{-1} \frac{\|A^* - H\|_{1,1}}{2} \right)^p d\tilde{\nu} \\ &\leq \int \binom{n}{2}^{-1} \sum_{i < j} |A_{ij}^* - H_{ij}|^p d\tilde{\nu} = \binom{n}{2}^{-1} \sum_{i < j} \int |A_{ij}^* - H_{ij}| d\tilde{\nu} = \int |A_{ij}^* - H_{ij}| d\tilde{\nu} \end{aligned}$$

where the second inequality is due to the definition of d_{GM} , the third follows from the definition of $\frac{1}{2} \|A^* - H\|_{1,1} = \sum_{i < j} |A_{ij}^* - H_{ij}|$ and Jensen's inequality. The first equality is due to the fact that both adjacency matrices are binary ($1^p = 1$, $0^p = 0$) and the linearity of expectation. The final equality follows from the identity of distribution over all pairs (i, j) . Expanding the final term:

$$\begin{aligned} \int |A_{ij}^* - H_{ij}| d\tilde{\nu} &= \tilde{\nu}(\{A_{ij}^* \neq H_{ij}\}) \\ &= \tilde{\nu}\left(\left\{ \mathbf{1}\left(\hat{h}_n(\xi_i^*, \xi_j^*) \geq \eta_{ij}\right) \neq \mathbf{1}\left(h_{0,n}(\xi_i, \xi_j) \geq \eta_{ij}\right) \right\}\right) \\ &= \int_0^1 \int \int \left| \mathbf{1}\left(\hat{h}_n(\xi_i^*, \xi_j^*) \geq \eta_{ij}\right) - \mathbf{1}\left(h_{0,n}(\xi_i, \xi_j) \geq \eta_{ij}\right) \right| d\tilde{\gamma}(\xi_i^*, \xi_i) d\tilde{\gamma}(\xi_j^*, \xi_j) d\eta_{ij} \\ &= \int \int \left| \hat{h}_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i, \xi_j) \right| d\tilde{\gamma}(\xi_i^*, \xi_i) d\tilde{\gamma}(\xi_j^*, \xi_j) \\ &\leq \int \int \left| \hat{h}_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*) \right| d\tilde{\gamma}(\xi_i^*, \xi_i) d\tilde{\gamma}(\xi_j^*, \xi_j) \\ &\quad + \int \int \left| h_{0,n}(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i, \xi_j) \right| d\tilde{\gamma}(\xi_i^*, \xi_i) d\tilde{\gamma}(\xi_j^*, \xi_j) \end{aligned}$$

The fourth equality follows from the fact that the two indicator functions differ in value only if η_{ij} falls into the interval between $h_{0,n}(\xi_i, \xi_j)$ and $\hat{h}_n(\xi_i^*, \xi_j^*)$, which happens with probability $\left| \hat{h}_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i, \xi_j) \right|$. In the last line we use triangle inequality.

We now look at the last two terms:

$$\begin{aligned}
& \int \int \left| \hat{h}_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*) \right| d\tilde{\gamma}(\xi_i^*, \xi_i) d\tilde{\gamma}(\xi_j^*, \xi_j) \\
&= \int \int \left| \hat{h}_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*) \right| d\hat{F}_n(\xi_i^*) d\hat{F}_n(\xi_j^*) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left| \hat{h}_n(\xi_i^A, \xi_j^A) - h_{0,n}(\xi_i^A, \xi_j^A) \right| \\
&\leq \max_{i,j} \left| \hat{h}_n(\xi_i^A, \xi_j^A) - h_{0,n}(\xi_i^A, \xi_j^A) \right| = o_p(\rho_n)
\end{aligned}$$

by Theorem 1, where $\xi_i^A \sim F_0$ refers to the unobserved characteristics used in the formation of the matrix A which A^* is bootstrapped from.

For the other term we use Assumption 1.2, which says:

$$\inf_{\xi_j \in \text{Supp}(\xi_i)} P \left(\xi_k \in \left\{ \sup_{\xi_t \in \text{Supp}(\xi_i)} |w_0(\xi_t, \xi_k) - w_0(\xi_t, \xi_j)| < \delta \right\} \right) \geq \left(\frac{\delta}{C} \right)^{\frac{1}{\alpha}}.$$

We have:

$$\begin{aligned}
& \int \int |h_{0,n}(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i, \xi_j)| d\tilde{\gamma}(\xi_i, \xi_i^*) d\tilde{\gamma}(\xi_j, \xi_j^*) \\
&\leq \rho_n \int \int |w_0(\xi_i^*, \xi_j^*) - w_0(\xi_i, \xi_j^*)| d\tilde{\gamma}(\xi_i, \xi_i^*) d\hat{F}_n(\xi_j^*) \\
&\quad + \rho_n \int \int |w_0(\xi_i, \xi_j^*) - w_0(\xi_i, \xi_j)| dF_0(\xi_i) d\tilde{\gamma}(\xi_j, \xi_j^*) \\
&\leq \rho_n \int \sup_{\xi_j^* \in \text{Supp}(\xi_i)} |w_0(\xi_i^*, \xi_j^*) - w_0(\xi_i, \xi_j^*)| d\tilde{\gamma}(\xi_i, \xi_i^*) \\
&\quad + \rho_n \int \sup_{\xi_i \in \text{Supp}(\xi_i)} |w_0(\xi_i, \xi_j^*) - w_0(\xi_i, \xi_j)| d\tilde{\gamma}(\xi_j, \xi_j^*) \\
&= 2\rho_n \int \sup_{\xi_i \in \text{Supp}(\xi_i)} |w_0(\xi_i, \xi_j^*) - w_0(\xi_i, \xi_j)| d\tilde{\gamma}(\xi_j, \xi_j^*)
\end{aligned}$$

The first inequality is due to the definition of $h_{0,n} = \rho_n w_0$ and triangle inequality. In the second inequality we take a supremum over the repeated index and note that the support of \hat{F}_n is a subset of the support of F_0 . As the terms no longer depend on ξ_j^* and ξ_i respectively, we integrate over their distributions. The resulting two terms are equal (note that w_0 is symmetric).

Fix $\varepsilon > 0$. For every $\xi_j^* \in \text{Supp}(\xi_i)$ there exists a neighbourhood $N(\xi_j^*, \varepsilon)$ of measure at least $(\frac{\varepsilon}{C})^{\frac{1}{\alpha}}$ such that for all $\xi_j \in N(\xi_j^*, \varepsilon)$: $\sup_{\xi_i \in \text{Supp}(\xi_i)} |w_0(\xi_i, \xi_j^*) - w_0(\xi_i, \xi_j)| < \varepsilon$. Our task is to show that there exists a coupling $\tilde{\gamma}$ which aligns ξ_j^* with their corresponding neighbourhoods.

To that end, define

$$d_S(a, b) \equiv \sup_{\xi_i \in \text{Supp}(\xi_i)} |w_0(\xi_i, a) - w_0(\xi_i, b)|. \quad (36)$$

d_S is a pseudometric, i.e. it may fail positivity (the distance between two distinct points may be zero) but it satisfies all other properties of a distance (in particular the triangle inequality).

Take K points $\{a_1, \dots, a_K\} \in \text{Supp}(\xi_i)$ which are at least ε apart: $\forall 1 \leq i < j \leq K : d_S(a_i, a_j) > \varepsilon$. Form a $\frac{\varepsilon}{2}$ -neighbourhood around each a_k .

These neighbourhoods are non-overlapping: suppose there was a $b \in N(a_i, \frac{\varepsilon}{2})$ and $b \in N(a_j, \frac{\varepsilon}{2})$ for $i \neq j$. Then by triangle inequality: $d_S(a_i, a_j) \leq d_S(a_i, b) + d_S(a_j, b) \leq \varepsilon$. But we have assumed $d_S(a_i, a_j) > \varepsilon$, a contradiction.

By Assumption 1.2 we know that each of these neighbourhoods has a measure at least $(\frac{\varepsilon}{2C})^{\frac{1}{\alpha}}$.

It follows that:

$$1 \geq P_{b \sim F_0} \left(b \in \bigcup_{i=1}^K N \left(a_i, \frac{\varepsilon}{2} \right) \right) = \bigcup_{i=1}^K P_{b \sim F_0} \left(b \in N \left(a_i, \frac{\varepsilon}{2} \right) \right) \geq K \left(\frac{\varepsilon}{2C} \right)^{\frac{1}{\alpha}}$$

or $K \leq \left(\frac{\varepsilon}{2C} \right)^{-\frac{1}{\alpha}} < \infty$, so the set of $\{a_1, \dots, a_K\}$ has finite cardinality.

Take the largest K possible. Then for all $b \in \text{Supp}(\xi_i) \exists k \leq K$ such that $d_S(b, a_k) \leq \varepsilon$, or in other words $\bigcup_{i=1}^K N(a_i, \varepsilon)$ is a finite cover of $\text{Supp}(\xi_i)$. Hence we can assign each $b \in \text{Supp}(\xi_i)$ to one of the $k \in \{1, \dots, K\}$: start with $N(a_k, \frac{\varepsilon}{2})$ for all k , then for each point not yet assigned to a region add it to the region with (not necessarily unique) k which minimises the d_S distance from that point to a_k . This way we form K disjoint regions, say $\{N_k\}_{k=1}^K$, each of size at least $(\frac{\varepsilon}{2C})^{\frac{1}{\alpha}}$ and such that whenever $b_1, b_2 \in N_k \subseteq N(a_k, \varepsilon)$ we have $d_S(b_1, b_2) \leq d_S(b_1, a_k) + d_S(b_2, a_k) \leq 2\varepsilon$.

Now instead of ξ_i report $k(\xi_i)$ such that $\xi_i \in N_{k(\xi_i)}$. This means we are replacing F_0 with an empirical distribution function G_ε which takes only K values, each with probability $P_{b \sim F_0}(b \in N_k) \geq (\frac{\varepsilon}{2C})^{\frac{1}{\alpha}}$; and we replace \hat{F}_n with an empirical distribution function $\hat{G}_{\varepsilon, n}$ from G_ε . We choose $\tilde{\gamma}$ to be any coupling of F_0, \hat{F}_n consistent with the following: for $y \sim \mathcal{U}[0, 1]$ set $k(\xi_j) = G_\varepsilon^{-1}(y)$, $k(\xi_j^*) = \hat{G}_{\varepsilon, n}^{-1}(y)$.

Then:

$$\int \sup_{\xi_i \in \text{Supp}(\xi_i)} |w_0(\xi_i, \xi_j^*) - w_0(\xi_i, \xi_j)| d\tilde{\gamma}(\xi_j, \xi_j^*) \leq 2\varepsilon + M_w \int \mathbf{1}(k(\xi_j) \neq k(\xi_j^*)) d\tilde{\gamma}(\xi_j, \xi_j^*)$$

For the first inequality we note that either ξ_j, ξ_j^* fall in the same N_k and hence $d_S(\xi_j, \xi_j^*) \leq 2\varepsilon$, or they come from different subsets of the domain, in which case their maximal possible distance is

M_w . For the final term:

$$\begin{aligned}
\int \mathbf{1}(k(\xi_j) \neq k(\xi_j^*)) d\tilde{\gamma}(\xi_j, \xi_j^*) &\leq \int |k(\xi_j) - k(\xi_j^*)| d\tilde{\gamma}(\xi_j, \xi_j^*) \\
&= \int_0^1 |G_\varepsilon^{-1}(y) - \hat{G}_{\varepsilon,n}^{-1}(y)| dy \\
&= \int_1^K |G_\varepsilon(x) - \hat{G}_{\varepsilon,n}(x)| dx \\
&\leq K \sup_x |G_\varepsilon(x) - \hat{G}_{\varepsilon,n}(x)| \xrightarrow{a.s.} 0.
\end{aligned}$$

The first inequality is due to the fact that $k \in \mathbb{N}$ so if the terms are not equal their distance is at least 1. The next equality is by construction of $\tilde{\gamma}$, noting that $y \sim \mathcal{U}[0, 1]$. We then do a change of variable (we switch from integrating the horizontal distance to the vertical distance between the plots of G_ε and $\hat{G}_{\varepsilon,n}$), noting that the plots can only differ on the domain $x \in [1, K]$. We use an upper bound in terms of a supremum over x and conclude that the final expression goes to zero almost surely by Glivenko-Cantelli Theorem. Hence for all n large enough $\int \mathbf{1}(k(\xi_j) \neq k(\xi_j^*)) d\tilde{\gamma}(\xi_j, \xi_j^*) \leq \frac{\varepsilon}{M_w}$ with probability one.

Since ε was arbitrary, the overall expression is $o_p(\rho_n)$, as required. \square

For the proofs of the next section, in the appendix we split the argument into more steps and provide intermediate results which lead to the conclusions in Theorem 3, Lemma 1 and Corollary 1. The advantage of the additional steps is that they characterise moment conditions sufficient for bootstrap consistency which could be verified for other classes of functions or alternative estimators of the network-generating function. They have been left out of the main text to avoid introducing more complicated notation and improve the readability.

We begin with a general result, not specific to U-statistics. Because the many levels of randomness can get confusing very quickly, we have decided to tackle them one at a time: we firstly characterise a class of *non-random* estimators and distributions for which we get weak convergence of our statistic to the correct limit. We denote these generic non-random statistics and distribution as e.g. h_n, F_n and we can think of them as specific realisations of their random equivalents, e.g. F_n can be the empirical distribution $\hat{F}_n|\xi$ we get for a specific draw of ξ . In practice, the classes of h_n, F_n will often be wider and also contain elements which cannot be achieved as a specific realisation of our random procedure. Once we have characterised the class which ensures weak convergence to the desired limit, we show that, once we allow for randomness in ξ , the statistics based on the random \hat{h}_n, \hat{F}_n belong that class with high probability, hence they converge weakly to the same limit either almost surely or in probability.

Definition A.1. *Set $C_{w,F,\rho}$. Let \mathfrak{h} denote a set of linking functions, let \mathcal{F} denote a set of distributions, and let $(0, 1]^\mathbb{N}$ denote a set of sequences of densities $\{\rho_n\}_{n=1}^\infty$, $0 < \rho_n \leq 1$. Let $(w, F, \rho) \in \mathfrak{h} \times \mathcal{F} \times (0, 1]^\mathbb{N}$ be a triple of a function w_0 , a distribution F , and a sparsity sequence ρ . Let $\xi \sim F$ and $\xi^* \sim F_n$. For each $(w, F, \rho) \in \mathfrak{h} \times \mathcal{F} \times (0, 1]^\mathbb{N}$ let $C_{w,F,\rho}$ be the set of non-random sequences of pairs of functions and distributions $\{(h_n, F_n)\}_{n=1}^\infty$ characterised by a set of conditions on convergence of moments of the form $E_{F_n} \left(f \left(\frac{h_n}{\rho_n}(\xi^*), w(\xi^*) \right) \right) \rightarrow E_F (f(w(\xi), w(\xi)))$ as $n \rightarrow \infty$ for some class of functions $f \in \mathfrak{f}$. That is:*

$$C_{w,F,\rho} = \left\{ \{(h_n, F_n)\}_{n=1}^\infty : \forall n \in \mathbb{N}, \forall f \in \mathfrak{f} : \right. \quad (37)$$

$$\left. (h_n, F_n) \in \mathfrak{h} \times \mathcal{F} \text{ and } \lim_{n \rightarrow \infty} E_{F_n} \left(f \left(\frac{h_n}{\rho_n}(\xi^*), w(\xi^*) \right) \right) = E_F (f(w(\xi), w(\xi))) \right\}.$$

We state the general version of the result²⁷:

Theorem A.1. *Let $C_{w_0, F_0, \rho}$ be as defined in Definition A.1 and suppose that:*

(i) *the set $C_{w_0, F_0, \rho}$ contains the sequence $\{(h_{0,n}, F_0)\}_{n=1}^\infty$;*

(ii) *for any sequence $\{(h_n, F_n)\}_{n=1}^\infty$ in $C_{w_0, F_0, \rho}$, $\tilde{J}_n(t, h_n, F_n)$ converges weakly to a common*

²⁷. The structure and the proof are strongly inspired by Theorem 1.2.1 of Politis et al. (1999).

distribution $J(t, w_0, F_0)$ ²⁸;

(iii) for any sequence $\{(h_n, F_n)\}_{n=1}^\infty$ in $C_{w_0, F_0, \rho}$:

$$\lim_{n \rightarrow \infty} E_{h_n, F_n} \left[\left(f_n(A^*(h_n(\xi^*), \eta^*), \rho_n, F_n) - \tilde{f}_n(h_n(\xi^*), \rho_n, F_n) \right)^2 \right] = 0 \quad (38)$$

where $A^*(h_n(\xi^*), \eta^*)$ denotes an adjacency matrix A^* based on a vector of observations of $\xi^* \stackrel{i.i.d.}{\sim} F_n$, with Bernoulli probabilities determined by $h_n(\xi_i^*)$.

If the random sequence $\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^\infty$ belongs to $C_{w_0, F_0, \rho}$ with probability one, i.e. $\forall n \in \mathbb{N}, \forall f \in \mathcal{F}$: $(\hat{h}_n, \hat{F}_n) \in \mathcal{h} \times \mathcal{F}$ a.s. and $E_{\hat{F}_n} \left(f \left(\frac{\hat{h}_n}{\rho_n}(\xi^*), w_0(\xi^*) \right) \right) \xrightarrow{a.s.} E_{F_0} (f(w_0(\xi), w_0(\xi)))$, then:

$$1. J_n(t, \hat{h}_n, \hat{F}_n) \xrightarrow{a.s.} J(t, w_0, F_0).$$

2. If $J(t, w_0, F_0)$ is continuous in t at $t = 1 - \alpha$ and strictly increasing at $t = 1 - \alpha$:

$$J_n^{-1}(1 - \alpha, \hat{h}_n, \hat{F}_n) \xrightarrow{a.s.} J^{-1}(1 - \alpha, w_0, F_0). \quad (39)$$

3. If $J(t, w_0, F_0)$ is continuous in t at $t = 1 - \alpha$ and is strictly increasing at $t = 1 - \alpha$ and if F_0 doesn't enter the function f_n directly but only through a parameter θ ²⁹: $f_n(A(h_{0,n}(\xi), \eta), \rho_n, \theta)$, then the $(1 - \alpha)$ confidence interval for θ constructed as:

$$CI_n \left(1 - \alpha, A, \hat{h}_n, \hat{F}_n \right) = \left\{ \theta : J_n^{-1} \left(\frac{\alpha}{2}, \hat{h}_n, \hat{F}_n \right) \leq f_n(A, \rho_n, \theta) \leq J_n^{-1} \left(1 - \frac{\alpha}{2}, \hat{h}_n, \hat{F}_n \right) \right\} \quad (40)$$

is asymptotically valid:

$$P_{h_{0,n}, F_0} \left(\theta \in CI_n \left(1 - \alpha, A, \hat{h}_n, \hat{F}_n \right) \right) \xrightarrow{a.s.} 1 - \alpha. \quad (41)$$

4. If $J(t, w_0, F_0)$ is continuous in t , then

$$\sup_t \left| J_n \left(t, \hat{h}_n, \hat{F}_n \right) - \tilde{J}_n \left(t, h_{0,n}, F_0 \right) \right| \xrightarrow{a.s.} 0.$$

If the random sequence $\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^\infty$ satisfies the moment conditions for belonging to $C_{w_0, F_0, \rho}$ in probability: $E_{\hat{F}_n} \left(f \left(\frac{\hat{h}_n}{\rho_n}(\xi^*), w_0(\xi^*) \right) \right) \xrightarrow{P} E_{F_0} (f(w_0(\xi), w_0(\xi)))$, then conclusions 1.-4. above hold with \xrightarrow{P} replacing $\xrightarrow{a.s.}$ and \xrightarrow{P} replacing $\xrightarrow{a.s.}$.

²⁸. This is weaker than F_n converges weakly to F_0 .

²⁹. For example in equation (25) we have $\theta = E_{h_{0,n}, F_0}(g(A(\iota)))$.

The above results follow straight from conveniently chosen assumptions, yet they are still useful because they provides a set of sufficient conditions for the convergence of the bootstrap distribution to the correct limit, the correctness of bootstrap confidence intervals, and the consistency of bootstrap.

Proof of Theorem A.1. We start from proving 1.

For any $\{(h_n, F_n)\}_{n=1}^\infty$ in $C_{w_0, F_0, \rho}$:

$$\begin{aligned} f_n(A(h_n(\xi^*), \eta^*), \rho_n, F_n) &= \tilde{f}_n(h_n(\xi^*), \rho_n, F_n) \\ &\quad + \left(f_n(A(h_n(\xi^*), \eta^*), \rho_n, F_n) - \tilde{f}_n(h_n(\xi^*), \rho_n, F_n) \right) \end{aligned}$$

By assumption (ii), the distribution of $\tilde{f}_n(h_n(\xi^*), \rho_n, F_n)$ converges weakly to the desired limit: $\tilde{J}_n(t, h_n, F_n) \xrightarrow{weakly} J(t, w_0, F_0)$. By assumption (iii), the second term converges to 0 in second mean, hence it is $o_p(1)$ and does not affect the distribution limit³⁰. For any sequence $\{(h_n, F_n)\}_{n=1}^\infty \in C_{w_0, F_0, \rho}$ we have:

$$J_n(t, h_n, F_n) \xrightarrow{weakly} J(t, w_0, F_0) \quad \text{i.e.} \quad d(J_n(t, h_n, F_n), J(t, w_0, F_0)) \rightarrow 0.$$

The random sequence $\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^\infty$ belongs to $C_{w_0, F_0, \rho}$ with probability one, hence:

$$\begin{aligned} &P\left(\lim_{n \rightarrow \infty} d\left(J_n(t, \hat{h}_n, \hat{F}_n), J(t, w_0, F_0)\right) = 0\right) \\ &\geq P\left(\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^\infty \in C_{w_0, F_0, \rho} \text{ and } \lim_{n \rightarrow \infty} d\left(J_n(t, \hat{h}_n, \hat{F}_n), J(t, w_0, F_0)\right) = 0\right) \\ &= P\left(\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^\infty \in C_{w_0, F_0, \rho}\right) \\ &= 1, \end{aligned}$$

that is: $d\left(J_n(t, \hat{h}_n, \hat{F}_n), J(t, w_0, F_0)\right) \xrightarrow{a.s.} 0$.

For the case of convergence in probability, we use the following result:

Theorem (Billingsley (1995) Theorem 20.5 (ii)). *A necessary and sufficient condition for $X_n \xrightarrow{p} X$ is that each subsequence $\{X_{n'}\}$ has a further subsequence $\{X_{n''}\}$ such that $X_{n''} \xrightarrow{a.s.} X$.*

Given that $E_{\hat{F}_n}\left(f\left(\frac{\hat{h}_n}{\rho_n}(\xi^*), w_0(\xi^*)\right)\right) \xrightarrow{p} E_{F_0}(f(w_0(\xi), w_0(\xi)))$, for any subsequence indexed by n' there is a further subsequence indexed by n'' which satisfies $E_{\hat{F}_{n''}}\left(f\left(\frac{\hat{h}_{n''}}{\rho_{n''}}(\xi^*), w_0(\xi^*)\right)\right) \xrightarrow{a.s.}$

30. By Theorem 25.4 in Billingsley (1995): $X_n \xrightarrow{d} X$ and $X_n - Y_n \xrightarrow{p} 0$, then $Y_n \xrightarrow{d} X$. Also, if F_{X_n} and F_X denote the distribution functions of random variables X_n and X , respectively, then $X_n \xrightarrow{d} X$ means $F_{X_n} \xrightarrow{weakly} F_X$.

$E_{F_0}(f(w_0(\xi), w_0(\xi)))$. By what we have just shown, applied to $C_{w_0, F_0, \rho''}$, where ρ'' is the subsequence of ρ indexed by n'' , $d\left(J_{n''}\left(t, \hat{h}_{n''}, \hat{F}_{n''}\right), J\left(t, w_0, F_0\right)\right) \xrightarrow{a.s.} 0$. Applying Theorem 20.5 (ii) from Billingsley (1995) in the other direction, this means that $d\left(J_n\left(t, \hat{h}_n, \hat{F}_n\right), J\left(t, w_0, F_0\right)\right) \xrightarrow{P} 0$. Hence 1. holds.

The remaining conclusions follow by arguments identical to those in the proof of Theorem 1.2.1 in Politis et al. (1999). For 2. we use the following Lemma:

Lemma (Lemma 1.2.1 of Politis et al. (1999)). *Let $\{G_n\}$ be a sequence of distribution functions on the real line converging weakly to a distribution function G (i.e. $G_n(x) \rightarrow G(x)$ for all continuity points of G). Assume G is continuous and strictly increasing at $y = G^{-1}(1 - \alpha)$. Then,*

$$G_n^{-1}(1 - \alpha) = \inf\{x : G_n(x) \geq 1 - \alpha\} \rightarrow G^{-1}(1 - \alpha). \quad (42)$$

Proof. See Politis et al. (1999) p.10. □

Together with the conclusion from 1. that $\tilde{J}_n(t, h_n, F_n) \xrightarrow{weakly} J(t, w_0, F_0)$ for all (h_n, F_n) in $C_{w_0, F_0, \rho}$, the lemma implies that $J_n^{-1}(1 - \alpha, h_n, F_n) \rightarrow J^{-1}(1 - \alpha, w_0, F_0)$ for all (h_n, F_n) in $C_{w_0, F_0, \rho}$. Arguments identical to those in the proof of 1. show that if $\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^{\infty}$ belongs to $C_{w_0, F_0, \rho}$ with probability one, then $J_n^{-1}(1 - \alpha, \hat{h}_n, \hat{F}_n) \xrightarrow{a.s.} J^{-1}(1 - \alpha, w_0, F_0)$ and if $\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^{\infty}$ satisfies the moment conditions for belonging to $C_{w_0, F_0, \rho}$ in probability: $E_{\hat{F}_n}\left(f\left(\frac{\hat{h}_n}{\rho_n}(\xi^*), w_0(\xi^*)\right)\right) \xrightarrow{P} E_{F_0}(f(w_0(\xi), w_0(\xi)))$, then $J_n^{-1}(1 - \alpha, \hat{h}_n, \hat{F}_n) \xrightarrow{P} J^{-1}(1 - \alpha, w_0, F_0)$.

In order to show 3., we firstly prove the following Lemma:

Lemma A.6. *Let $\{G_n\}$ be a sequence of distribution functions on the real line converging weakly to a distribution function G (i.e. $G_n(x) \rightarrow G(x)$ for all continuity points of G). Let x_n be a real-valued sequence converging to x (i.e. $x_n \rightarrow x$). Assume that G is continuous and strictly increasing at x . Then,*

$$G_n(x_n) \rightarrow G(x). \quad (43)$$

Proof. Take any $\delta > 0$. Since G is continuous at x , there exists $\varepsilon > 0$ such that $x - \varepsilon$ and $x + \varepsilon$ are continuity points of G and

$$\begin{aligned} G(x - \varepsilon) - G(x) &\geq -\frac{\delta}{2} \\ G(x + \varepsilon) - G(x) &\leq \frac{\delta}{2}. \end{aligned}$$

Since $x_n \rightarrow x$, $G_n(x - \varepsilon) \rightarrow G(x - \varepsilon)$, $G_n(x) \rightarrow G(x)$ and $G_n(x + \varepsilon) \rightarrow G(x + \varepsilon)$, there exists an

$N \in \mathbb{N}$ such that for all $n \geq N$:

$$\begin{aligned} |x_n - x| &\leq \varepsilon \\ |G_n(x - \varepsilon) - G(x - \varepsilon)| &\leq \frac{\delta}{2} \\ |G_n(x) - G(x)| &\leq \frac{\delta}{2} \\ |G_n(x + \varepsilon) - G(x + \varepsilon)| &\leq \frac{\delta}{2}. \end{aligned}$$

Since G_n are weakly increasing for all n :

$$G_n(x - \varepsilon) \leq G_n(x) \leq G_n(x + \varepsilon).$$

Hence for all $n \geq N$:

$$\begin{aligned} -\delta &\leq G(x - \varepsilon) - G(x) - \frac{\delta}{2} \leq G_n(x - \varepsilon) - G(x) \leq \\ &\leq G_n(x) - G(x) \leq \\ &\leq G_n(x + \varepsilon) - G(x) \leq G(x + \varepsilon) - G(x) + \frac{\delta}{2} \leq \delta. \end{aligned}$$

i.e. $|G_n(x_n) - G(x)| \leq \delta$. □

For 3, we start with any (h_n, F_n) in $C_{w_0, F_0, \rho}$. We have:

$$\begin{aligned} &P_{h_0, n, F_0}(\theta \in CI_n(1 - \alpha, A, h_n, F_n)) \\ &= P_{h_0, n, F_0}\left(J_n^{-1}\left(\frac{\alpha}{2}, h_n, F_n\right) \leq f_n(A, \rho_n, \theta) \leq J_n^{-1}\left(1 - \frac{\alpha}{2}, h_n, F_n\right)\right) \\ &= P_{h_0, n, F_0}\left(f_n(A, \rho_n, \theta) \leq J_n^{-1}\left(1 - \frac{\alpha}{2}, h_n, F_n\right)\right) \\ &\quad - P_{h_0, n, F_0}\left(f_n(A, \rho_n, \theta) < J_n^{-1}\left(\frac{\alpha}{2}, h_n, F_n\right)\right) \\ &= J_n\left(J_n^{-1}\left(1 - \frac{\alpha}{2}, h_n, F_n\right), h_0, n, F_0\right) - J_n\left(J_n^{-1}\left(\frac{\alpha}{2}, h_n, F_n\right), h_0, n, F_0\right) \\ &\rightarrow J\left(J^{-1}\left(1 - \frac{\alpha}{2}, w_0, F_0\right), w_0, F_0\right) - J\left(J^{-1}\left(\frac{\alpha}{2}, w_0, F_0\right), w_0, F_0\right) = 1 - \alpha. \end{aligned}$$

The convergence follows from Lemma A.6 used with 2. (for the convergence of the argument) and $J_n(t, h_0, n, F_0) \xrightarrow{\text{weakly}} J(t, w_0, F_0)$ (for the convergence in distribution). Arguments identical to those in the proof of 1. show that if $\left\{\left(\hat{h}_n, \hat{F}_n\right)\right\}_{n=1}^{\infty}$ belongs to $C_{w_0, F_0, \rho}$ with probability one, then $P_{h_0, n, F_0}\left(\theta \in CI_n\left(1 - \alpha, A, \hat{h}_n, \hat{F}_n\right)\right) \xrightarrow{\text{a.s.}} 1 - \alpha$. If instead $\left\{\left(\hat{h}_n, \hat{F}_n\right)\right\}_{n=1}^{\infty}$ satisfies the moment conditions for belonging to $C_{w_0, F_0, \rho}$ only in probability, that is: $E_{\hat{F}_n}\left(f\left(\frac{\hat{h}_n}{\rho_n}(\xi^*), w_0(\xi^*)\right)\right) \xrightarrow{p}$

$E_{F_0}(f(w_0(\xi), w_0(\xi)))$, then we have $P_{h_{0,n}, F_0}(\theta \in CI_n(1 - \alpha, A, \hat{h}_n, \hat{F}_n)) \xrightarrow{P} 1 - \alpha$.

Finally, 4. follows from 1. and Polya's Theorem:

Theorem (Polya's Theorem, Satz I of Pólya (1920)). *Let X_n, X be random variables with distributions $F_n(x)$ and $F(x)$ respectively. If F is continuous*

$$X_n \xrightarrow{d} X \iff \sup_x |F_n(x) - F(x)| \rightarrow 0.$$

□

We can now provide more primitive conditions for the special class of f_n for which \tilde{f}_n is a U-statistic.

Theorem A.2 (Consistency of bootstrap for U-statistics). *Let ι be a set of m nodes and denote the adjacency matrix on the subgraph with nodes in ι and linking probabilities $h_n(\cdot, \cdot)$ by $A(h_n(\xi(\iota)), \eta(\iota))$. Let $g : \{0, 1\}^{\binom{m}{2}} \rightarrow \mathbb{R}$ be a symmetric function from a subgraph on $m < \infty$ nodes to the real line and let*

$$\begin{aligned} & f_n(A(h_n(\xi^*), \eta^*), \rho_n, F_n) \\ &= \frac{\sqrt{n}}{\binom{n}{m} \rho_n^{\tau(g)}} \sum_{1 \leq \iota_1 < \iota_2 < \dots < \iota_m \leq n} (g(A(h_n(\xi^*(\iota)), \eta^*(\iota))) - E_{h_n, F_n}(g(A(h_n(\xi^*(\iota)), \eta^*(\iota))))) \end{aligned}$$

and $\tilde{g}(h_{0,n}(\xi(\iota))) \equiv E(g(A(h_{0,n}(\xi(\iota)), \eta(\iota))) | \xi(\iota))$. There exists a normalisation $\tau(g)$ ³¹ and a function $\tilde{g} : \text{Supp}(\xi)^m \rightarrow \mathbb{R}$ such that:

- $\frac{\tilde{g}(h_{0,n}(\xi(\iota)))}{\rho_n^{\tau(g)}} = \tilde{g}(w_0(\xi(\iota))) + O(\rho_n)$
- $E_{F_0}(|\tilde{g}(w_0(\xi(j)))|) > 0$ for some $j \in \mathcal{M}_m$
- $E_{F_0}(\tilde{g}^2(w_0(\xi(j)))) < \infty \quad \forall j \in \mathcal{M}_m$
- $\text{Var}_{F_0}(E_{F_0}(\tilde{g}(w_0(\xi(\iota))) | \xi_{\iota_1})) \equiv \sigma_1^2 < \infty$

Suppose that:

$$\begin{aligned} & \sigma_1^2 > 0 \\ & \frac{n}{\binom{n}{m} \rho_n^{\tau(g)}} \rightarrow 0. \end{aligned}$$

31. For $m = 2$, if $g(0) \neq 0$ we set $\rho_n^{-\tau(g)} = 1$, $\tilde{g}(w_0(\xi_i, \xi_j)) = g(0)$ and if $g(0) = 0$ but $g(1) \neq 0$ we set $\rho_n^{-\tau(g)} = \frac{1}{\rho_n}$ and $\tilde{g}(w_0(\xi_i, \xi_j)) = g(1)w_0(\xi_i, \xi_j)$. More generally, for $m \geq 2$, $\rho_n^{-\tau(g)} = \frac{1}{\rho_n^k}$ where k is the smallest number of ones such that $g(\cdot)$ evaluated at a vector of k ones and $\binom{m}{2} - k$ zeros is non-zero.

and let $C_{w_0, F_0, \rho}$ be a set of sequences $\{(h_n, F_n)\}_{n=1}^\infty$ which satisfy:

1. $E_{F_n} \left(\left(\frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*)) \right)^2 \right) \rightarrow 0.$
2. $E_{F_n} (f(\xi^*(\iota))) \rightarrow E_{F_0} (f(\xi(\iota)))$ for all $f : \text{Supp}(\xi)^k \rightarrow \mathbb{R}$ such that $E_{F_0} (|f(\xi(\iota))|) < \infty$ for all $\iota \in \mathcal{M}_k$, for any $k \leq 2m - 1$.

Then $\{(\hat{h}_n, \hat{F}_n)\}_{n=1}^\infty$ satisfies 1.-2. in probability and we get all conclusions of Theorem A.1 in probability with $J(t, w_0, F_0) = N(0, m^2 \sigma_1^2)$.

Remark. The advantage of stating our condition as in 2. instead of directly showing that it holds when $F_n = \hat{F}_n$ because of SLLN for U-statistics is that it characterises a wider class of distributions we could resample from. For example, when we adjust the resampling distribution for the purpose of GMM by adding weights to different observations in a way that ensures the moment conditions hold in the bootstrap world.

Proof of Theorem A.2. The theorem was stated for a general $m < \infty$ but for simplicity of notation we present the proof for the case of $m = 2$. The structure of the argument remains identical if we use $m > 2$.

To show the existence of \tilde{g} and $\tau(g)$ we start by analysing the form of \tilde{g} . Since g is a function from $\{0, 1\}^{\binom{m}{2}}$ it takes at most $2^{\binom{m}{2}}$ distinct values. Each of those values is taken with probability that the input submatrix $A(\iota)$ matches a given pattern of 0s and 1s. Let $\Gamma(A(\iota))$ denote the set (of cardinality $2^{\binom{m}{2}}$) of all possible values $A(\iota)$ can take. Then:

$$\tilde{g}(h_{0,n}(\xi(\iota))) = \sum_{\gamma \in \Gamma(A(\iota))} g(\gamma) P(A(\iota) = \gamma | \xi(\iota)).$$

Conditional on $\xi(\iota)$, the elements of $A(\iota)$ are independent and $P(A_{ij} = 1 | \xi) = h_{0,n}(\xi_i, \xi_j) = \rho_n w_0(\xi_i, \xi_j) \sim \rho_n$ while $P(A_{ij} = 0 | \xi) = 1 - h_{0,n}(\xi_i, \xi_j) = 1 - \rho_n w_0(\xi_i, \xi_j) \sim 1$. The probability of the event that the upper triangle of $A(\iota)$ consists of k ones and $\binom{m}{2} - k$ zeros is proportional to ρ_n^k . The smallest k for which $g(\cdot)$ evaluated at an input γ with k ones and $\binom{m}{2} - k$ zeros in the upper triangle is non-zero is equal to the normalisation $\tau(g)$. By construction, all γ s with fewer ones have a coefficient $g(\gamma) = 0$. All γ s with more ones happen with probability proportional to ρ_n^l for $l > \tau(g)$, i.e. after a normalisation by $\rho_n^{-\tau(g)}$ are $O(\rho_n)$ and go to zero.

The terms in the sum proportional to $\rho_n^{\tau(g)}$ are of the form:

$$g(\gamma) \underbrace{h_{0,n}(\xi_{\iota_1}, \xi_{\iota_2}) \cdots h_{0,n}(\xi_{\iota_3}, \xi_{\iota_4})}_{\tau(g) \text{ terms}} (1 - h_{0,n}(\xi_{\iota_5}, \xi_{\iota_6})) \cdots (1 - h_{0,n}(\xi_{\iota_7}, \xi_{\iota_8})).$$

After a normalisation by $\rho_n^{-\tau(g)}$ we get:

$$g(\gamma) \underbrace{w_0(\xi_{\iota_1}, \xi_{\iota_2}) \dots w_0(\xi_{\iota_3}, \xi_{\iota_4})}_{\tau(g) \text{ terms}} (1 - h_{0,n}(\xi_{\iota_5}, \xi_{\iota_6})) \dots (1 - h_{0,n}(\xi_{\iota_7}, \xi_{\iota_8}))$$

we keep the $g(\gamma)w_0(\xi_{\iota_1}, \xi_{\iota_2}) \dots w_0(\xi_{\iota_3}, \xi_{\iota_4})$ part in \tilde{g} and note that the remainder of the previous term is $O(\rho_n)$.

To sum up, \tilde{g} takes the form of a finite sum of non-zero constant (value of g at a specific realisation γ) times a product of $\tau(g)$ terms of the form $w_0(\xi_{\iota_i}, \xi_{\iota_j})$.

The remaining terms in $\frac{\tilde{g}(h_{0,n}(\xi(\iota)))}{\rho_n^{\tau(g)}} - \tilde{g}(w_0(\xi(\iota)))$ vanish at the rate $O(\rho_n)$.

Since $w_0(\xi_i, \xi_j)$ is not identically equal to zero and there are non-zero coefficients $g(\gamma)$ multiplying products of $w_0(\xi_{\iota_i}, \xi_{\iota_j})$ in \tilde{g} , there exists $j \in \mathcal{M}_m$ for which $E_{F_0}(|\tilde{g}(w_0(\xi(j)))|) > 0$.

Since $w_0(\xi_i, \xi_j) < M_w$ for all ξ_i, ξ_j we have

$$E_{F_0}(\tilde{g}^2(w_0(\xi(j)))) < \binom{\binom{m}{2}}{\tau(g)}^2 \left(\max_{\gamma \in \Gamma(A(\iota))} g^2(\gamma) \right) M_w^2 < \infty$$

for all $j \in \mathcal{M}_m$ (where the first constant says that there are $\binom{m}{2}$ ones and zeros that determine the value of $A(\iota)$, there are $\binom{\binom{m}{2}}{\tau(g)}$ ways to place $\tau(g)$ ones in them, and after squaring a sum of $\binom{\binom{m}{2}}{\tau(g)}$ terms we get $\binom{\binom{m}{2}}{\tau(g)}^2$ terms, each bounded above by the remaining part of the expression). Finally, since

$$\text{Var}(E(Y|X)) = \text{Var}(Y) - E(\text{Var}(Y|X)) \leq \text{Var}(Y) = E(Y^2) - E(Y)^2 \leq E(Y^2)$$

we also get that

$$\sigma_1^2 \equiv \text{Var}_{F_0}(E_{F_0}(\tilde{g}(w_0(\xi(\iota)))|\xi_{\iota_1})) < E_{F_0}(\tilde{g}^2(w_0(\xi(\iota)))) < \infty.$$

Having established the existence of \tilde{g} and $\tau(g)$, we now check that the elements of $C_{w_0, F_0, \rho}$ satisfy condition (i)-(iii) of Theorem A.1.

The sequence $\{(h_n, F_n)\}_{n=1}^\infty = \{(h_{0,n}, F_0)\}_{n=1}^\infty$ satisfies the conditions and belongs to $C_{w_0, F_0, \rho}$ (sequences in 1. and 2. are constant and equal to the desired limit), hence (i) is satisfied.

To check condition (iii) we look at:

$$\begin{aligned} & E_{h_n, F_n} \left(\left(f_n(A^*(h_n(\xi^*), \eta^*), \rho_n, F_n) - \tilde{f}_n(h_n(\xi^*), \rho_n, F_n) \right)^2 \right) \\ &= \frac{n}{\binom{n}{2} \rho_n^{2\tau(g)}} \sum_{i^* < j^*} \sum_{k^* < l^*} E_{h_n, F_n} \left((g(A_{i^*j^*}) - E_{h_n}(g(A_{i^*j^*})|\xi^*)) (g(A_{k^*l^*}) - E_{h_n}(g(A_{k^*l^*})|\xi^*)) \right) \end{aligned}$$

To simplify the above expression notice that most terms in the summation are zero. In particular, consider different cases of overlap between the indices:

- if there is no overlap ($i^* \neq k^*$, $i^* \neq l^*$, $j^* \neq k^*$, $j^* \neq l^*$), by the independence assumption the term inside the sum is:

$$E_{h_n, F_n} \left((g(A_{i^*j^*}) - E_{h_n}(g(A_{i^*j^*})|\xi^*))^2 \right) = 0^2 = 0.$$

- If there is partial overlap (e.g. $i^* = k^*$, $j^* \neq l^*$, or any symmetric situation):

$$\begin{aligned} & E_{h_n, F_n} \left((g(A_{i^*j^*}) - E_{h_n}(g(A_{i^*j^*})|\xi^*)) (g(A_{i^*l^*}) - E_{h_n}(g(A_{i^*l^*})|\xi^*)) \right) \\ & \stackrel{LIE}{=} E_{h_n, F_n} \left(E_{h_n, F_n} \left((g(A_{i^*j^*}) - E_{h_n}(g(A_{i^*j^*})|\xi^*)) (g(A_{i^*l^*}) - E_{h_n}(g(A_{i^*l^*})|\xi^*)) \mid \xi_i^* \right) \right) \\ & \stackrel{indep}{=} E_{h_n, F_n} \left((E_{h_n, F_n} \left((g(A_{i^*j^*}) - E_{h_n}(g(A_{i^*j^*})|\xi^*)) \mid \xi_i^* \right))^2 \right) \\ & \stackrel{LIE}{=} E_{h_n, F_n} \left((E_{h_n, F_n} \left((E_{h_n}(g(A_{i^*j^*})|\xi^*) - E_{h_n}(g(A_{i^*j^*})|\xi^*)) \mid \xi_i^* \right))^2 \right) = 0. \end{aligned}$$

- If there is full overlap ($i^* = k^*$ and $j^* = l^*$, or $i^* = l^*$ and $j^* = k^*$):

$$E_{h_n, F_n} \left((g(A_{i^*j^*}) - E_{h_n}(g(A_{i^*j^*})|\xi^*))^2 \right) \leq E_{h_n, F_n} \left(g^2(A_{i^*j^*}) \right).$$

There are $\binom{n}{2}$ terms of this final form in the sum.

Combining the three cases, we get:

$$E_{h_n, F_n} \left(\left(f_n(A^*(h_n(\xi^*), \eta^*), \rho_n, F_n) - \tilde{f}_n(h_n(\xi^*), \rho_n, F_n) \right)^2 \right) \leq \frac{2E_{h_n, F_n} \left(g^2(A_{i^*j^*}) \right)}{(n-1)\rho_n^{2\tau(g)}}$$

In an analogous way to how we have defined \tilde{g} and the corresponding \tilde{g} , we let $\tilde{g}^2(h_{0,n}(\xi_i, \xi_j)) \equiv E_{h_{0,n}}(g^2(A_{ij})|\xi)^{32}$, and we can find a function $\tilde{g}^2(w_0(\xi_i, \xi_j))$ with $\frac{\tilde{g}^2(h_{0,n}(\xi_i, \xi_j))}{\rho_n^{2\tau(g)}} = \tilde{g}^2(w_0(\xi_i, \xi_j)) +$

32. Comparing to the example given earlier:

$$\begin{aligned} E(g^2(A_{1,2}, A_{2,3})|\xi) &\equiv \tilde{g}^2(h_{0,n}(\xi_1, \xi_2), h_{0,n}(\xi_2, \xi_3)) \\ &= g^2(0,0)(1-h_{0,n}(\xi_1, \xi_2))(1-h_{0,n}(\xi_2, \xi_3)) + g^2(0,1)(1-h_{0,n}(\xi_1, \xi_2))h_{0,n}(\xi_2, \xi_3) \\ &+ g^2(1,0)h_{0,n}(\xi_1, \xi_2)(1-h_{0,n}(\xi_2, \xi_3)) + g^2(1,1)h_{0,n}(\xi_1, \xi_2)h_{0,n}(\xi_2, \xi_3). \end{aligned}$$

$O(\rho_n)$, $0 < E_{F_0} \left(\left| \tilde{g}^2(w_0(\xi_i, \xi_j)) \right| \right) < \infty$ and $0 < E_{F_0} \left(\left| \tilde{g}^2(w_0(\xi_i, \xi_i)) \right| \right) < \infty$. Then:

$$\begin{aligned}
& E_{h_n, F_n} \left(\rho_n^{-\tau(g)} g^2(A_{i^* j^*}) \right) \\
&= E_{F_n} \left(E_{h_n} \left(\rho_n^{-\tau(g)} g^2(A_{i^* j^*}) \mid \xi^* \right) \right) \\
&= E_{F_n} \left(\rho_n^{-\tau(g)} \tilde{g}^2(h_n(\xi_i^*, \xi_j^*)) \right) \\
&= E_{F_n} \left(\rho_n^{-\tau(g)} \tilde{g}^2(h_{0,n}(\xi_i^*, \xi_j^*)) + \rho_n^{-\tau(g)+1} \tilde{g}'^2(\tilde{h}_n(\xi_i^*, \xi_j^*)) \frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*)) \right) \\
&\leq E_{F_n} \left(\rho_n^{-\tau(g)} \tilde{g}^2(h_{0,n}(\xi_i^*, \xi_j^*)) \right) + \underbrace{\rho_n^{-\tau(g)+1} \sup_h \left| \tilde{g}'^2(h) \right|}_{< \infty} \underbrace{E_{F_n} \left(\frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*)) \right)}_{=o(1)} \\
&= E_{F_n} \left(\tilde{g}^2(w_0(\xi_i^*, \xi_j^*)) \right) + O(\rho_n) + o(1) \\
&\xrightarrow{a.s.} E_{F_0} \left(\tilde{g}^2(w_0(\xi_i, \xi_j)) \right) < \infty.
\end{aligned}$$

Note that since the leading term of $\tilde{g}^2(h_{0,n})$ is proportional to the $\tau(g)$ th power of $h_{0,n}$, the leading term of $\tilde{g}'^2(h_{0,n})$ has a $h_{0,n}$ to the power $\tau(g) - 1$. Given the form of $\tilde{g}'^2(h_{0,n})$, which is a sum of finitely many terms of the form of a bounded constant times bounded powers of $h_{0,n}$, the whole derivative is bounded. It follows that:

$$E_{h_n, F_n} \left(\left(f_n(A^*(h_n(\xi^*), \eta^*), \rho_n, F_n) - \tilde{f}_n(h_n(\xi^*), \rho_n, F_n) \right)^2 \right) \leq O \left(\frac{1}{n \rho_n^{\tau(g)}} \right) = o(1).$$

Hence (iii) holds.

Checking (ii) is a bit more involved. We start with a Hoeffding's (martingale) decomposition³³ of $\tilde{f}_n(h_n(\xi^*), \rho_n, F_n)$ for any $\{(h_n, F_n)\}_{n=1}^\infty$ in $C_{w_0, F_0, \rho}$:

$$\begin{aligned}
\tilde{f}_n(h_n(\xi^*), \rho_n, F_n) &= \frac{\sqrt{n}}{\binom{n}{2} \rho_n^{\tau(g)}} \sum_{i < j} \tilde{g}(h_n(\xi_i^*, \xi_j^*)) - E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))) \\
&= \frac{2}{\sqrt{n} \rho_n^{\tau(g)}} \sum_{i=1}^n E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)) \mid \xi_i^*) - E_{F_n}(E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)) \mid \xi_i^*)) \\
&+ \frac{\sqrt{n}}{\binom{n}{2} \rho_n^{\tau(g)}} \sum_{i < j} (\tilde{g}(h_n(\xi_i^*, \xi_j^*)) - E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)) \mid \xi_i^*) \\
&\quad - E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)) \mid \xi_j^*) + E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)))) \\
&\equiv \tilde{U}_n(h_n, F_n) + \tilde{r}_n(h_n, F_n).
\end{aligned}$$

We firstly focus on $\tilde{U}_n(h_n, F_n)$, which is a (rescaled) average of i.i.d. terms. We add and subtract

This example illustrates why $\tilde{g}^2(h_{0,n}(\xi_1, \xi_2), h_{0,n}(\xi_2, \xi_3))$ is proportional to $\rho_n^{\tau(g)}$, not to $\rho_n^{2\tau(g)}$.

33. For more details see Chapter 5 of Serfling (2009), specifically section 5.1.5.

terms that swap h_n for $h_{0,n}$ and F_n for F_0 :

$$\begin{aligned}
\tilde{U}_n(h_n, F_n) &= \frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n \left(E_{F_n} (\tilde{g}(h_n(\xi_i^*, \xi_j^*)) | \xi_i^*) - E_{F_n} (\tilde{g}(h_n(\xi_i^*, \xi_j^*))) \right. \\
&\quad \left. - E_{F_n} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*) + E_{F_n} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))) \right) \\
&\quad + \frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n \left(E_{F_n} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*) - E_{F_n} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))) \right. \\
&\quad \left. - E_{F_0} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*) + E_{F_n} (E_{F_0} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*)) \right) \\
&\quad + \frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n (E_{F_0} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*) - E_{F_n} (E_{F_0} (\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*))) \\
&= T_1 + T_2 + T_3
\end{aligned}$$

We deal with these terms one by one.

For T_1 , we do Taylor expansion of \tilde{g} around $h_{0,n}$:

$$\tilde{g}(h_n(\xi_i^*, \xi_j^*)) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) = \tilde{g}'(\tilde{h}_n(\xi_i^*, \xi_j^*)) (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*))$$

where $\tilde{h}_n(\xi_i^*, \xi_j^*)$ is between $h_n(\xi_i^*, \xi_j^*)$ and $h_{0,n}(\xi_i^*, \xi_j^*)$. We can show that T_1 goes to zero in second mean, hence also in probability. Let:

$$\begin{aligned}
T_1 &= \frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n (E_{F_n} (\tilde{g}(h_n(\xi_i^*, \xi_j^*)) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)) | \xi_i^*) - E_{F_n} (\tilde{g}(h_n(\xi_i^*, \xi_j^*)) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)))) \\
&= \frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n b_{i^*}.
\end{aligned}$$

Note that the terms inside the sum are independent and have zero expectation:

$$\begin{aligned}
E_{F_n}(b_{i^*}) &= E_{F_n} (E_{F_n} (\tilde{g}(h_n(\xi_i^*, \xi_k^*)) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_k^*)) | \xi_i^*) - E_{F_n} (\tilde{g}(h_n(\xi_k^*, \xi_l^*)) - \tilde{g}(h_{0,n}(\xi_k^*, \xi_l^*)))) \\
&\stackrel{LIE}{=} E_{F_n} (\tilde{g}(h_n(\xi_k^*, \xi_l^*)) - \tilde{g}(h_{0,n}(\xi_k^*, \xi_l^*))) - E_{F_n} (\tilde{g}(h_n(\xi_k^*, \xi_l^*)) - \tilde{g}(h_{0,n}(\xi_k^*, \xi_l^*))) = 0
\end{aligned}$$

Hence in the expansion of the square all terms with $i \neq j$ are zero:

$$\begin{aligned}
E(T_1^2) &= E\left(\left(\frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n b_{i^*}\right)^2\right) \\
&= 4\frac{1}{n\rho_n^{2\tau(g)}} \sum_{i=1}^n E_{F_n}(b_{i^*}^2) + 8\frac{1}{n\rho_n^{2\tau(g)}} \sum_{i<j} E_{F_n}(b_{i^*}b_{j^*}) \\
&\stackrel{i.i.d.}{=} 4\rho_n^{-2\tau(g)} E_{F_n}(b_{i^*}^2) + 8\frac{1}{n\rho_n^{2\tau(g)}} \sum_{i<j} E_{F_n}(b_{i^*}) E_{F_n}(b_{j^*}) \\
&= 4\rho_n^{-2\tau(g)} E_{F_n}(b_{i^*}^2) \\
&= 4\rho_n^{-2\tau(g)} E_{F_n}\left(\left(E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))\right)|\xi_i^*\right. \\
&\quad \left.- E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))\right)^2) \\
&= 4\rho_n^{-2\tau(g)} E_{F_n}\left(\left(E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))\right)|\xi_i^*\right)^2 \\
&\quad - 4\rho_n^{-2\tau(g)} \left(E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))) - \tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))\right)^2 \\
&\leq 4\rho_n^{-2\tau(g)} E_{F_n}\left(\left(E_{F_n}(\tilde{g}'(\tilde{h}_n(\xi_i^*, \xi_j^*))) (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*))\right)|\xi_i^*\right)^2 \\
&\leq 4 \underbrace{\left(\rho_n^{-\tau(g)+1} \sup_{h \in [0, M_w \rho_n]} |\tilde{g}'(h)|\right)^2}_{< \infty} \underbrace{\left(\frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*))\right)^2}_{=o(1)} \rightarrow 0
\end{aligned}$$

In the first inequality we use the fact that the second term is negative and smaller in magnitude than the first. We then pull the supremum over derivatives of \tilde{g} out of the expectation, use Jensen's inequality to put the square inside the inner expectation, apply the law of iterated expectations, and use the assumption 1. to get the conclusion.

As mentioned before, the derivative of \tilde{g} is bounded for any choice of g , and as we take a derivative with respect to h the leading term of the \tilde{g}' becomes proportional to power one lower than \tilde{g} , i.e. $\rho_n^{-\tau(g)+1} \tilde{g}' = O_p(1)$.

For the middle term, T_2 , we show that it goes to zero in mean squared. To simplify notation, let $T_2 = \frac{2}{\sqrt{n}\rho_n^{\tau(g)}} \sum_{i=1}^n a_{i^*}$ and notice that:

$$\begin{aligned}
E_{F_n}(a_{i^*}) &= \underbrace{E_{F_n}(E_{F_n}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))|\xi_i^*)) - E_{F_n}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*)))}_{=0} \\
&\quad - \underbrace{E_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))|\xi_i^*)) + E_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j^*))|\xi_i^*))}_{=0} = 0.
\end{aligned}$$

Then we have:

$$\begin{aligned}
E(T_2^2) &= E\left(\left(\frac{2}{\sqrt{n}\rho_n^{\tau(g)}}\sum_{i=1}^n a_{i^*}\right)^2\right) \\
&= 4\frac{1}{n\rho_n^{2\tau(g)}}\sum_{i=1}^n E_{F_n}(a_{i^*}^2) + 8\frac{1}{n\rho_n^{2\tau(g)}}\sum_{i<j} E_{F_n}(a_{i^*}a_{j^*}) \\
&\stackrel{i.i.d.}{=} 4\rho_n^{-2\tau(g)}E_{F_n}(a_{i^*}^2) + 8\frac{1}{n\rho_n^{2\tau(g)}}\sum_{i<j} E_{F_n}(a_{i^*})E_{F_n}(a_{j^*}) \\
&= 4\rho_n^{-2\tau(g)}E_{F_n}(a_{i^*}^2) \\
&= 4E_{F_n}\left(\left(E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))|\xi_i^*) + E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))) - E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*)\right.\right. \\
&\quad \left.\left.+ E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))) - E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))) + E_{F_n}(E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*))\right)^2\right) + O(\rho_n) \\
&\leq 8E_{F_n}\left(\underbrace{\left(E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))|\xi_i^*)\right)^2}_{\rightarrow E_{F_0}\left(\left(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))|\xi_i)\right)^2\right)}\right) + 8E_{F_n}\left(\underbrace{\left(E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*)\right)^2}_{\rightarrow E_{F_0}\left(\left(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))|\xi_i)\right)^2\right)}\right) \\
&\quad - 16E_{F_n}\left(\underbrace{E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))|\xi_i^*)E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*)}_{\rightarrow E_{F_0}\left(\left(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))|\xi_i)\right)^2\right)}\right) \\
&\quad + 8\underbrace{\left(E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))) - E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j)))\right)^2}_{\rightarrow 0} \\
&\quad + 8\underbrace{\left(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))) - E_{F_n}(E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*))\right)^2}_{\rightarrow 0} + O(\rho_n) \\
&\rightarrow 0
\end{aligned}$$

In the 5th equality we plug in the definition of a_{i^*} , we add and subtract the term $E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j)))$, we bring the normalisation by $\rho_n^{-2\tau(g)}$ inside the expectation and use $\frac{\tilde{g}(\xi_i, \xi_j)}{\rho_n^{\tau(g)}} = \tilde{g}(w_0(\xi_i, \xi_j)) + O(\rho_n)$. In the next step, we apply $(a+b)^2 \leq 2a^2 + 2b^2$, where a corresponds to the first four terms in the previous summation, for which we expand the square, and b corresponds to the last two terms. We now verify that we can apply property 2. to all resulting terms:

- By the independence between ξ_j^* and ξ_k^* when $j \neq k$ and the law of iterated expectations we can rewrite the first term as:

$$\begin{aligned}
E_{F_n}\left(\left(E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))|\xi_i^*)\right)^2\right) &= E_{F_n}\left(E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))|\xi_i^*)E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_k^*))|\xi_i^*)\right) \\
&= E_{F_n}(\tilde{g}(w_0(\xi_i^*, \xi_j^*))\tilde{g}(w_0(\xi_i^*, \xi_k^*)))
\end{aligned}$$

We now check the conditions for 2. when all indices are unique:

$$\begin{aligned}
E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))\tilde{g}(w_0(\xi_i, \xi_k))|) &\stackrel{LIE}{=} E_{F_0} (E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))\tilde{g}(w_0(\xi_i, \xi_k))| | \xi_i)) \\
&\leq E_{F_0} (E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))| | \xi_i) E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_k))| | \xi_i)) \\
&= E_{F_0} \left(E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))| | \xi_i)^2 \right) \\
&\leq E_{F_0} (E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j)) | \xi_i)) \\
&\stackrel{LIE}{=} E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j))) < \infty.
\end{aligned}$$

When two indices are repeated we use Cauchy-Schwarz inequality:

$$E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))\tilde{g}(w_0(\xi_i, \xi_i))|) \leq \sqrt{E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j))) E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_i)))} < \infty.$$

And when all indices are equal the condition $E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_i))) < \infty$ follows straight from the assumptions. Hence we have

$$\begin{aligned}
E_{F_n} \left((E_{F_n} (\tilde{g}(w_0(\xi_i^*, \xi_j^*)) | \xi_i^*))^2 \right) &\rightarrow E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))\tilde{g}(w_0(\xi_i, \xi_k))|) \\
&= E_{F_0} \left((E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i))^2 \right).
\end{aligned}$$

- For the second term, we can verify the condition for 2. when the indices are unique:

$$\begin{aligned}
E_{F_0} \left(\left| E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i)^2 \right| \right) &= E_{F_0} \left(E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i)^2 \right) \\
&\leq E_{F_0} (E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j)) | \xi_i)) \\
&\stackrel{LIE}{=} E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j))) < \infty,
\end{aligned}$$

where the inequality follows from Jensen's inequality. When the indices are repeated:

$$E_{F_0} \left(\left| E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_i)) | \xi_i)^2 \right| \right) = E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_i))) < \infty.$$

$$\text{hence } E_{F_n} \left((E_{F_n} (\tilde{g}(w_0(\xi_i^*, \xi_j^*)) | \xi_i^*))^2 \right) \rightarrow E_{F_0} \left((E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i))^2 \right).$$

- The third term can be rewritten as:

$$\begin{aligned}
E_{F_n} (E_{F_n} (\tilde{g}(w_0(\xi_i^*, \xi_j^*)) | \xi_i^*) E_{F_0} (\tilde{g}(w_0(\xi_i^*, \xi_j)) | \xi_i^*)) \\
= E_{F_n} (\tilde{g}(w_0(\xi_i^*, \xi_j^*)) E_{F_0} (\tilde{g}(w_0(\xi_i^*, \xi_j)) | \xi_i^*)).
\end{aligned}$$

Using Jensen's inequality, we verify the condition for 2. when the indices are unique:

$$\begin{aligned}
& E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j)) E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_k)) | \xi_i)|) \\
& \leq E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))| E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_k))| | \xi_i)) \\
& \stackrel{LIE}{=} E_{F_0} (E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))| | \xi_i) E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_k))| | \xi_i)) \\
& = E_{F_0} \left(E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_j))| | \xi_i)^2 \right) \\
& \leq E_{F_0} (E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j)) | \xi_i)) \\
& \stackrel{LIE}{=} E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j))) < \infty
\end{aligned}$$

and using Jensen's and Cauchy-Schwarz inequalities we verify it when the indices are equal:

$$\begin{aligned}
E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_i)) E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i)|) & \leq E_{F_0} (E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_i)) \tilde{g}(w_0(\xi_i, \xi_j))| | \xi_i)) \\
& \stackrel{LIE}{=} E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_i)) \tilde{g}(w_0(\xi_i, \xi_j))|) \\
& \leq \sqrt{E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j))) E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_i)))} < \infty.
\end{aligned}$$

hence

$$\begin{aligned}
& E_{F_n} (E_{F_n} (\tilde{g}(w_0(\xi_i^*, \xi_j^*)) | \xi_i^*) E_{F_0} (\tilde{g}(w_0(\xi_i^*, \xi_j)) | \xi_i^*)) \\
& \rightarrow E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i)) = E_{F_0} \left((E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i))^2 \right)
\end{aligned}$$

- For the fourth term we can verify that

$$\begin{aligned}
E_{F_0} (|E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)) | \xi_i)|) & \leq \sqrt{E_{F_0} (E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j)) | \xi_i))} \\
& \stackrel{LIE}{=} \sqrt{E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_j)))} < \infty, \\
E_{F_0} (|E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_i)) | \xi_i)|) & = E_{F_0} (|\tilde{g}(w_0(\xi_i, \xi_i))|) \leq \sqrt{E_{F_0} (\tilde{g}^2(w_0(\xi_i, \xi_i)))} < \infty.
\end{aligned}$$

hence $E_{F_n} (\tilde{g}(w_0(\xi_i^*, \xi_j^*))) \rightarrow E_{F_0} (\tilde{g}(w_0(\xi_i, \xi_j)))$.

We combine all terms using continuous mapping theorem and see that they all cancel out and the limit is zero.

For T_3 , we can write:

$$T_3 = 2\rho_n^{-\tau(g)} \sqrt{\text{Var}_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*))} \\ \times \sum_{i=1}^n \frac{E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*) - E_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*))}{\sqrt{n}\sqrt{\text{Var}_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*))}}.$$

Denote the terms inside the sum by X_{in} . They have zero expectation:

$$E_{F_n}(X_{in}) = \frac{E_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*) - E_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*)))}{\sqrt{n}\sqrt{\text{Var}_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*))}} = 0.$$

Their variances sum to 1 for each n :

$$\sum_{i=1}^n \text{Var}_{F_n}(X_{in}) = n \frac{\text{Var}_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*))}{n\text{Var}_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*))} = 1.$$

And for all n when $i \neq j$ the terms X_{in} and X_{jn} are independent and identically distributed. Hence by Lindeberg-Levy CLT for triangular arrays their sum converges in distribution to a standard normal random variable.

For the multiplier term we have:

$$\rho_n^{-2\tau(g)} \text{Var}_{F_n}(E_{F_0}(\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*)) = \text{Var}_{F_n}\left(E_{F_0}\left(\rho_n^{-\tau(g)}\tilde{g}(h_{0,n}(\xi_i^*, \xi_j))|\xi_i^*\right)\right) \\ = \text{Var}_{F_n}(E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*)) + O(\rho_n) \\ = E_{F_n}\left(E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*)^2\right) - (E_{F_n}(E_{F_0}(\tilde{g}(w_0(\xi_i^*, \xi_j))|\xi_i^*)))^2 + O(\rho_n) \\ \rightarrow E_{F_0}\left(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))|\xi_i)^2\right) - (E_{F_0}(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))|\xi_i)))^2 \\ = \text{Var}_{F_0}(E_{F_0}(\tilde{g}(w_0(\xi_i, \xi_j))|\xi_i)) \equiv \sigma_1^2 < \infty.$$

The first equality is pulling the normalisation inside the variance. The second equality applies the definition of \tilde{g} . The third equality is rewriting variance in terms of expectations. The limit follows from 2. (we have already checked that the relevant absolute moments are finite when we were checking conditions for convergence of T_2 , terms two and four) and the continuous mapping theorem. The final line is by definition. Hence $T_3 \xrightarrow{d} N(0, 4\sigma_1^2)$.

It remains to show that $\tilde{r}_n(h_n, F_n) = o_p(1)$. We can check that the expression is \sqrt{n} times a U-statistic with a kernel function $G(\xi_i^*, \xi_j^*) = \rho_n^{-\tau(g)}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)) - E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))|\xi_i^*) - E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*))|\xi_j^*) + E_{F_n}(\tilde{g}(h_n(\xi_i^*, \xi_j^*)))$). Note that $E(G(\xi_i^*, \xi_j^*)) = E(G(\xi_i^*, \xi_j^*)|\xi_i^*) = 0$, i.e. it is a degenerate U-statistic with $\text{Var}(E(G(\xi_i^*, \xi_j^*)|\xi_i^*)) = 0$. We could show that the whole term is negligible by convergence in second mean from definition, or rely on a Theorem from section 5.3.2

in Serfling (2009) which, in the present setting, can be stated as:

Lemma (Theorem 5.3.2 in Serfling (2009)). *If $E_{F_n} \left((G(\xi_i^*, \xi_j^*))^2 \right) < \infty$ then*

$$E_{F_n} \left((\tilde{r}_n(h_n, F_n))^2 \right) = O \left(\frac{1}{n} \right) = o(1).$$

By Jensen's inequality and the law of large numbers, $\rho_n^{-2\tau(g)} E_{F_n} (\tilde{g}^2 (h_n (\xi_i^*, \xi_j^*)))$ is an upper bound for all terms in the expansion of $E_{F_n} \left((G(\xi_i^*, \xi_j^*))^2 \right)$. Hence the sufficient condition is implied by:

$$\begin{aligned} \rho_n^{-2\tau(g)} E_{F_n} (\tilde{g}^2 (h_n (\xi_i^*, \xi_j^*))) &\leq \underbrace{2 E_{F_n} (\tilde{g}^2 (w_0 (\xi_i^*, \xi_j^*)))}_{\rightarrow E_{F_0} (\tilde{g}^2 (w_0 (\xi_i, \xi_j))) < \infty} + O(\rho_n) \\ &+ 2 \underbrace{\left(\sup_{h \in [0, M_w \rho_n]} \left| \frac{\tilde{g}'(h)}{\rho_n^{\tau(g)-1}} \right| \right)^2}_{< \infty} \underbrace{E_{F_n} \left(\left(\frac{1}{\rho_n} (h_n (\xi_i^*, \xi_j^*) - \hat{h}_n (\xi_i^*, \xi_j^*)) \right)^2 \right)}_{\rightarrow 0} \\ &\leq 2E_{F_0} (\tilde{g}^2 (w_0 (\xi_i, \xi_j))) + o(1). \end{aligned}$$

Hence for any $\varepsilon > 0$ we can find an N sufficiently large so that the condition is satisfied:

$$E_{F_n} \left((G(\xi_i^*, \xi_j^*))^2 \right) < 8E_{F_0} (\tilde{g}^2 (w_0 (\xi_i, \xi_j))) + \varepsilon < \infty \text{ for all } n > N.$$

Moving on to the second part of the proof, we check that the sequence $\{\hat{h}_n, \hat{F}_n\}_{n=1}^\infty$ satisfies assumptions 1. and 2. in probability:

1. Follows from Theorem 1:

$$\begin{aligned} E_{\hat{F}_n} \left(\left(\frac{1}{\rho_n} (\hat{h}_n (\xi_i^*, \xi_j^*) - h_{0,n} (\xi_i^*, \xi_j^*)) \right)^2 \right) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{\rho_n} (\hat{h}_n (\xi_i, \xi_j) - h_{0,n} (\xi_i, \xi_j)) \right)^2 \\ &\leq \left(\max_{i,j} \left| \frac{\hat{h}_n (\xi_i, \xi_j) - h_{0,n} (\xi_i, \xi_j)}{\rho_n} \right| \right)^2 \\ &= o_p(1)^2 = o_p(1). \end{aligned}$$

2. Let $f : \text{Supp}(\xi)^3 \rightarrow \mathbb{R}$ be any symmetric function for which $E_{F_0} (|f(\xi_i, \xi_j, \xi_k)|) < \infty$,

$E_{F_0}(|f(\xi_i, \xi_i, \xi_j)|) < \infty$ and $E_{F_0}(|f(\xi_i, \xi_i, \xi_i)|) < \infty$. We have:

$$\begin{aligned}
E_{\hat{F}_n}(f(\xi_i^*, \xi_j^*, \xi_k^*)) &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n f(\xi_i, \xi_j, \xi_k) \\
&= \underbrace{\frac{(n-1)(n-2)}{n^2}}_{\rightarrow 1} \underbrace{\frac{1}{\binom{n}{3}} \sum_{i < j < k} f(\xi_i, \xi_j, \xi_k)}_{\xrightarrow{a.s.} E_{F_0}(f(\xi_i, \xi_j, \xi_k)) < \infty} + \underbrace{\frac{3(n-1)}{n^2}}_{\rightarrow 0} \underbrace{\frac{1}{\binom{n}{2}} \sum_{i < j} f(\xi_i, \xi_i, \xi_j)}_{\xrightarrow{a.s.} E_{F_0}(f(\xi_i, \xi_i, \xi_j)) < \infty} \\
&\quad + \underbrace{\frac{1}{n^2}}_{\rightarrow 0} \underbrace{\frac{1}{n} \sum_{i=1}^n f(\xi_i, \xi_i, \xi_i)}_{\xrightarrow{a.s.} E_{F_0}(f(\xi_i, \xi_i, \xi_i)) < \infty} \\
&\xrightarrow{a.s.} E_{F_0}(f(\xi_i, \xi_j, \xi_k))
\end{aligned}$$

The first equality follows from the definition of the empirical distribution function \hat{F}_n . The convergence of the two terms in the second line follows from the SLLN for U-statistics (see e.g. Theorem A. in section 5.4 of Serfling (2009), p.190) given that $E_{F_0}(|f(\xi_i, \xi_j, \xi_k)|) < \infty$ and $E_{F_0}(|f(\xi_i, \xi_i, \xi_j)|) < \infty$. The convergence of the term in the third line follows from Kolmogorov's SLLN for i.i.d. random variables which applies under the assumption that $E_{F_0}(|f(\xi_i, \xi_i, \xi_i)|) < \infty$. The final line is by continuous mapping theorem for almost sure convergence. Condition 2. holds almost surely (hence also in probability), but because we only get condition 1. in probability the overall result is for convergence weakly in probability. \square

The above result was stated for a normalisation using the unknown ρ_n . We now show that the conclusions remain true when we replace it with an estimate.

Corollary A.1. *Under the assumptions of Theorem A.2*

$$\hat{\rho}_n - \rho_n = o_p(\rho_n),$$

hence

$$f_n\left(A\left(\hat{h}_n(\xi^*), \eta^*\right), \hat{\rho}_n, \hat{F}_n\right) = f_n\left(A\left(\hat{h}_n(\xi^*), \eta^*\right), \rho_n, \hat{F}_n\right) + o_p(1)$$

and we get all conclusions of Theorem A.1 in probability with $J(t, w_0, F_0) = N(0, m^2 \sigma_1^2)$ for

$$\begin{aligned} & f_n \left(A \left(\hat{h}_n(\xi^*), \eta^* \right), \hat{\rho}_n, \hat{F}_n \right) \\ &= \frac{\sqrt{n}}{\binom{n}{m} \hat{\rho}_n^{\tau(g)}} \sum_{1 \leq \iota_1 < \dots < \iota_m \leq n} \left(g \left(A \left(\hat{h}_n(\xi^*(\iota)), \eta^*(\iota) \right) \right) - E_{\hat{h}_n, \hat{F}_n} \left(g \left(A \left(\hat{h}_n(\xi^*(\iota)), \eta^*(\iota) \right) \right) \right) \right). \end{aligned}$$

Proof of Corollary A.1. Let ρ_n^* denote the density of a bootstrap adjacency matrix formed by $\xi^* \sim F_n$ with linking probabilities h_n . We can write

$$f_n \left(A(h_n(\xi^*), \eta^*), \rho_n^*, F_n \right) = \left(\frac{\rho_n}{\hat{\rho}_n^*} \right)^{\tau(g)} f_n \left(A(h_n(\xi^*), \eta^*), \rho_n, F_n \right)$$

hence it is sufficient to show $\rho_n^* - \rho_n = o_p(\rho_n)$ which, by Slutsky's theorem, implies $\left(\frac{\rho_n}{\rho_n^*} \right)^{\tau(g)} \xrightarrow{p} 1$.

Applying Theorem 3 to $g(A_{ij}) = A_{ij}$, for which $m = 2$ and $\tau(g) = 1$, we have:

$$\frac{\sqrt{n}}{\binom{n}{2} \rho_n} \sum_{1 \leq i^* < j^* \leq n} (A_{i^* j^*} - E_{h_n, F_n}(A_{i^* j^*})) = O_p(1)$$

where the expression is bounded in probability because it has a well-defined limiting distribution.

Hence

$$\begin{aligned} \rho_n^* &= \frac{1}{\binom{n}{2}} \sum_{1 \leq i^* < j^* \leq n} A_{i^* j^*} \\ &= E_{h_n, F_n}(A_{i^* j^*}) + \underbrace{\frac{\rho_n}{\sqrt{n}} O_p(1)}_{= \frac{1}{\sqrt{n}} O_p(\rho_n) = o_p(\rho_n)} \\ &= E_{F_n}(E_{h_n}(A_{i^* j^*} | \xi^*)) + o_p(\rho_n) \\ &= E_{F_n}(E_{h_n}(h_n(\xi_i^*, \xi_j^*) | \xi^*)) + o_p(\rho_n) \\ &= \rho_n \left(E_{F_n} \left(\frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*)) \right) + E_{F_n} \left(\frac{1}{\rho_n} h_{0,n}(\xi_i^*, \xi_j^*) \right) \right) + o_p(\rho_n) \\ &\leq \rho_n \left(\sqrt{E_{F_n} \left(\left(\frac{1}{\rho_n} (h_n(\xi_i^*, \xi_j^*) - h_{0,n}(\xi_i^*, \xi_j^*)) \right)^2 \right)} + E_{F_n}(w_0(\xi_i^*, \xi_j^*)) \right) + o_p(\rho_n) \\ &= \rho_n (o(1) + E_{F_0}(w_0(\xi_i, \xi_j) + o(1))) + o_p(\rho_n) \\ &= \rho_n (o(1) + 1 + o(1)) + o_p(\rho_n) \\ &= \rho_n + o_p(\rho_n) \end{aligned}$$

The first equality is by definition, the second follows from the above expression and result from Theorem 3. The third equality uses the law of iterated expectations. The fourth equality is by

definition of A . For the fifth equality we add and subtract $E_{F_n}(h_{0,n}(\xi_i^*, \xi_j^*))$ and pull ρ_n out of the bracket. The inequality is due to Jensen's inequality where the final term is transformed according to the definition of w_0 . The sixth equality uses assumptions 1. and 2.. The seventh equality is due to the definition of w_0 which is assumed to integrate to 1.

The above derivation applies to the case of $\rho_n^* = \hat{\rho}_n$ (for $h_n = h_{0,n}$ and $f_n = F_0$), proving that $f_n(A(h_{0,n}(\xi), \eta), \hat{\rho}_n, F_0)$ and $f_n(A(h_{0,n}(\xi), \eta), \rho_n, F_0)$ have the same asymptotic limit.

If we replaces ρ_n^* with $\hat{\rho}_n^*$ the $o(1)$ terms in the derivation are replaced by $o_p(1)$, which does not affect the overall result. Hence we also get the same limit of $f_n(A(\hat{h}_n(\xi^*), \eta^*), \hat{\rho}_n^*, \hat{F}_n)$. \square

We note that all conclusions of Theorem 3, Lemma 1 and Corollary 1 follow from Theorem A.2 and Corollary A.1, hence they have also been proven.

A.3 Useful results

For reference, we list some results which we use in our proofs:

Theorem (Bernstein's inequality for bounded random variables³⁴). *Let Z_1, \dots, Z_n be independent random variables. Assume that there exist some positive constant M such that $|Z_i| \leq M$ with probability one for each i . Let also $\sigma^2 = \frac{1}{n} \sum_{i=1}^n V(Z_i)$. Then, for all $\varepsilon > 0$:*

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n (Z_i - E(Z_i))\right| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{n\varepsilon^2}{2(\sigma^2 + \frac{1}{3}M\varepsilon)}\right). \quad (44)$$

B Appendix: additional tables, plots, codes

B.1 Codes

In this section we present some of the codes used for simulations. A full package should eventually become available online.

We start with the definitions of different distances and estimators of the linking functions:

```
def D2(A):
    #a function which maps A into a matrix of distances D2
    n = len(A)
    V = (1/n)*np.matmul(A, A)
    C = (1/n)*np.matmul(V, V)
    B = np.matmul(np.diag(np.diag(C)), np.ones((n, n)))
    D = B+B.T-2*C
    return D
```

³⁴. Copied after Zeleneev (2020).

```

def Dmax(A):
    #a function which maps A into a matrix of distances Dmax
    n = len(A)
    V = (1/n)*np.matmul(A,A)
    F = torch.tensor(np.tensordot(np.ones(n),V,0))
    G = torch.transpose(F, 1,0)
    H = F-G
    J=1-torch.transpose(np.fmax(torch.eye(n).repeat(n, 1, 1),np.tensordot(np.eye(n)
        ,np.ones(n),0)),2,1)

    D = np.array(torch.amax(np.fmin(abs(H),J), dim=2))
    return D

```

```

def HK1h(D,A,h):
    # gives a kernel approximation to the linking function based on a one-way
        normal kernel, with bandwidth h,
        based on distance D

    n= len(A)
    K = np.exp(-0.5*(D/h)**2)
    K[np.isnan(K)] = 0
    T=np.matmul(K,A)
    B=np.matmul(K,np.ones((n,n)))-K
    H = T/B
    H=(H+np.transpose(H))/2
    return H

```

```

def HK2h(D,A,h):
    # gives a kernel approximation to the linking function based on a two-way
        normal kernel, with bandwidth h,
        based on distance D

    n= len(A)
    K = np.exp(-0.5*(D/h)**2)
    K[np.isnan(K)] = 0
    H=np.matmul(K,np.matmul(A,K))/np.matmul(K,np.matmul(1-np.eye(n),K))
    return H

```

```

def HNN1(D,A):
    #gives a kernel approximation to h based on one-way nearest neighbours, with
        bandwidth h, based on distance D
    #uses the optimal neighbourhood size  $(n \log(n))^{1/2}$ 

    n= len(A)
    N_size = round(np.sqrt(n*np.log(n)))
    N=np.argpartition(D, N_size+1)[:,:N_size+1]

```

```

mask = np.ones((n,N_size+1), dtype=bool)
mask[range(n), np.argmax(N==np.array(range(n)).reshape(n,1), axis=1)] = False
N = N[mask].reshape(n, N_size)
mask2 = np.zeros((n,n), dtype=bool)
mask2[np.tile(np.array(range(n)).reshape(n,1),N_size), N] = True
mask_long=np.tile(mask2,(n,1))
A_long=np.tile(A,(1,n)).reshape(n*n,n)
Amlong = A_long[mask_long].reshape(n,n, N_size)
H=np.sum(Amlong,2)/N_size
H=(H+np.transpose(H))/2
return H

```

In simulations we generate the true matrices using one of the following functions:

```

def high_rho_generate(n, r, rep):
    #generates a rep number of true n by n matrices from the high rho function with
    #density r/1.35, outputs only the
    #adjacency matrices

    A_true = []
    for s in range(rep):
        w = np.random.uniform(0,1,(n))
        u = np.random.uniform(0,1,(n,n))
        eta = np.tril(u) + np.tril(u, -1).T
        Wi = np.tensordot(w,np.ones(n),0)
        Wj = np.tensordot(np.ones(n),w,0)
        A = (eta < r*(1-((abs(0.5-Wi)<0.05) & (abs(0.5-Wj)<0.05))))*(1-0.5*(abs(0.5-
        Wi)+abs(0.5-Wj)))/(0.975))*1
        np.fill_diagonal(A, 0)
        A_true.append(A)
    return A_true

```

```

def horse_generate(n, r, rep):
    #generates a rep number of true n by n matrices from the horseshoe function
    #with density r/4.44, outputs only the
    #adjacency matrices

    A_true = []
    for s in range(rep):
        w = np.random.uniform(0,1,(n))
        u = np.random.uniform(0,1,(n,n))
        eta = np.tril(u) + np.tril(u, -1).T
        Wi = np.tensordot(w,np.ones(n),0)
        Wj = np.tensordot(np.ones(n),w,0)
        A = (eta < r*((np.exp(-200*(Wi-Wj**2)**2)+np.exp(-200*(Wj-Wi**2)**2))/2))*1
        np.fill_diagonal(A, 0)

```

```

    A_true.append(A)
return A_true

```

```

def product_generate_A_h_xi(n, r, rep):
    #generates a rep number of true n by n matrices from the product function with
    #density r/4, outputs the adjacency
    #matrices, the true linking function,
    #and the true values of the underlying
    #characteristics  $\xi_i$ 

    A_true = []
    xi_true = []
    h_true = []
    for s in range(rep):
        w = np.random.uniform(0,1,(n))
        u = np.random.uniform(0,1,(n,n))
        eta = np.tril(u) + np.tril(u, -1).T
        Wi = np.tensordot(w,np.ones(n),0)
        Wj = np.tensordot(np.ones(n),w,0)
        h = r*Wi*Wj
        A = (eta < h)*1
        np.fill_diagonal(A, 0)
        np.fill_diagonal(h, 0)
        A_true.append(A)
        h_true.append(h)
        xi_true.append(list(w))
    return (A_true, h_true, np.array(xi_true))

```

Code for running bootstrap and finding the optimal bandwidth:

```

def boot_HK1h(A,h,B):
    #outputs B bootstrapped adjacency matrices based on matrix A with bandwidth h
    #using linking function estimate HK1

    n=len(A)
    H_true = HK1h(D2(A),A,h)
    #choose nodes for bootstrap villages:
    v = np.random.randint(0, n, size=(B,n))
    #generate new adjacency matrices
    row = np.tensordot(v,np.ones(n),0).astype(int)
    column = np.tile(np.array(v),n).reshape(B,n,n)
    G = H_true[row,column]
    u = np.random.rand(B, n, n)
    m = np.tril(u) + np.transpose(np.tril(u, -1),[0,2,1])
    A_boot = (m < G)*1
    [np.fill_diagonal(A_boot[i], 0) for i in range(B)]

```

```
return A_boot
```

```
def HK1h_loo(D,A,h):  
    #gives a leave-one-out kernel approximation to h based on a one-way normal  
    #kernel, with bandwidth h, based on  
    #distance D  
  
    n= len(A)  
    K = np.exp(-0.5*(D/h)**2)  
    K[np.isnan(K)] = 0  
    T = np.matmul(K,A) - np.matmul(np.diag(np.diag(K)),np.ones((n,n)))*A  
    B = np.matmul(K,np.ones((n,n)))-(K-np.diag(np.diag(K))-np.matmul(np.diag(np.  
    #diag(K)),np.ones((n,n)))  
  
    H = T/B  
    H=(H+np.transpose(H))/2  
    return H
```

```
def log_likelihood(A,H):  
    #the log-likelihood estimation for an adjacency matrix A under the assumption  
    #it comes from a distribution with  
    #linking probabilities in H  
  
    log_likelihood = np.sum(A*np.log(H)+(1-A)*np.log(1-H))  
    return log_likelihood
```

```
def ll(h, A):  
    #the leave-one-out log-likelihood objective function for use in minimising  
    #procedures  
  
    return -log_likelihood(A,HK1h_loo(D2(A),A,h))
```

A sample Monte Carlo simulation code using the above definitions and the WARP procedure from Giacomini, Politis, and White (2013) to obtain confidence interval coverage:

```
#define the output data frame:  
df_loo = pd.DataFrame(columns=['S', 'B', 'n', 'rho','average for true graphs', '  
    #true value', 'alpha', 'proportion of  
    #bootstrap CI that cover truth', 'average  
    #length of bootstrap CI', 'statistic','h']  
    )  
  
#run the simulations:  
ALPHA = [0.01, 0.05, 0.1, 0.15, 0.2, 0.3] #sizes of confidence intervals  
NN = [25, 50, 100, 150, 200, 300] #sample sizes  
SS = [1000] #number of true graphs  
RR = [1, 0.75,0.5,0.25,0.1] #sparsity level
```



```

for n in NN:
    for r in RR:
        S_max = max(SS)
        (A_true, h_true, xi_true) = product_generate_A_h_xi(n, r, S_max)
        A_boot = []
        h_list = []
        B=1 #because of WARP we only need one bootstrap replication

        true_density = r*0.25

        for A in A_true[0:min(len(A_true),S_max)]:
            #find the optimal bandwidth by minimising the leave-one-out log-likelihood
            h_guess = 0.2090189845643738*true_density**1.38258532*n**(-1.55268817)*
                np.log(n)**1.82661653
            res = minimize(ll, h_guess, args=A, method = 'Nelder-Mead', tol=1e-7,
                bounds=((0,1.1),))

            h = res.x[0]
            #do bootstrap for matrix A using the optimal bandwidth
            Ab = boot_HK1h(A,h,B)
            #save the bootstrapped adjacency matrices and the bandwidth
            A_boot.append(Ab)
            h_list.append(h)

        #estimate the statistic of interest for true and bootstrapped graphs
        true_density_all = [nx.density(nx.from_numpy_array(A_true[s])) for s in
            range(S_max)]

        true_density_mean = np.mean(true_density_all)
        boot_density_all = [nx.density(nx.from_numpy_array(A_boot[s][0])) for s in
            range(S_max)]

        #find the confidence interval coverage using WARP
        for S in SS:
            if (S<=len(A_true)):
                true_density_vec = true_density_all[0:S]
                boot_density = boot_density_all[0:S]
                stat = 'density'
                density_minus_true = np.array(boot_density) - np.array(
                    true_density_vec)

                for alpha in ALPHA:
                    qu= np.percentile(density_minus_true, 100*(1-alpha/2))

```

```

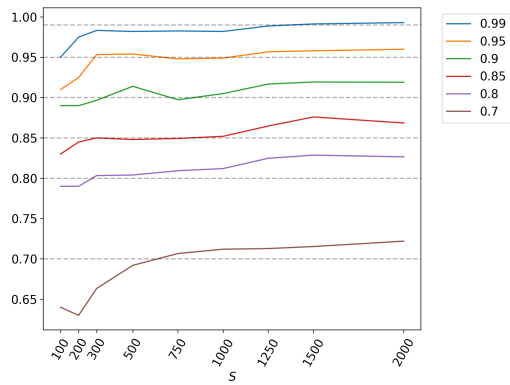
ql= np.percentile(density_minus_true, 100*(alpha/2))
bl = true_density_vec - qu
bu = true_density_vec + ql
co= np.mean([bl[i] <= true_density <= bu[i] for i in range(S)])
me = np.mean(bu-bl)
df_loo = df_loo._append({'S': S, "B": B, 'n': n, "rho": r, '
                        average for true
                        graphs':
                        true_density_mean, '
                        true value':
                        true_density, 'alpha'
                        : alpha, 'proportion
                        of bootstrap CI that
                        cover truth': co, '
                        average length of
                        bootstrap CI': me, '
                        statistic': stat, 'h'
                        : np.mean(h_list)},
                        ignore_index = True)

#save the output
df_loo.to_csv('df_loo_product_n_25_300_true_dens.csv')

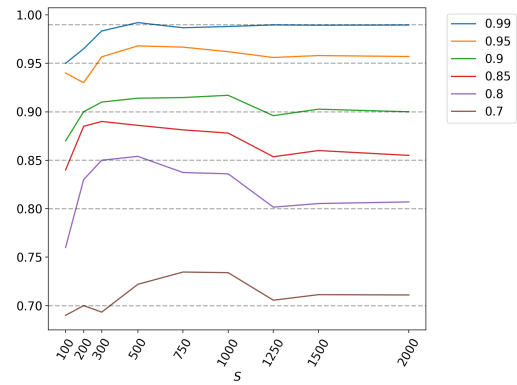
```

B.2 Monte Carlo simulations: tables and a sensitivity check

Since we are using the WARP procedure instead of traditional Monte Carlo simulations, we test its sensitivity by checking the effect of varying the number of simulated true graphs S rather than the number of bootstrap replications B , which is always kept at $B = 1$. Fig. 11a shows that the predictions for different statistics stabilise above S around 750 or higher. This is true in most simulations (see Table 4, Table 5), with the exception of networks with high density such as that in Fig. 11b (and Table 6) in which the predictions don't stabilise until $S = 1250$ or even $S = 1500$. In all other sections we use $S = 1000$. Running more repetitions is computationally expensive and provides little advantage in terms of accuracy in the majority of cases.



(a) Confidence interval coverage for density using the product generating function at $n = 500$ and $\rho_n = 0.1875$.



(b) Confidence interval coverage for transitivity using the high density generating function at $n = 500$ and $\rho_n = 0.759$.

Figure 11: Confidence interval coverage for different number of simulated true graphs S based on Monte Carlo simulations.

n	statistic	ρ_n	average for true graphs	Proportion of bootstrap CI that cover truth for						
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$	
500	λ_1	0.024	17.116	1.000	0.971	0.925	0.858	0.804	0.660	
		0.078	52.502	0.998	0.961	0.900	0.854	0.811	0.706	
		0.139	93.282	0.997	0.956	0.884	0.844	0.799	0.685	
		0.250	166.453	0.995	0.954	0.887	0.825	0.778	0.692	
	λ_3	0.024	7.823	0.005	0.000	0.000	0.000	0.000	0.000	
		0.078	12.825	0.000	0.000	0.000	0.000	0.000	0.000	
		0.139	15.582	0.000	0.000	0.000	0.000	0.000	0.000	
		0.250	17.042	0.000	0.000	0.000	0.000	0.000	0.000	
	λ_{10}	0.024	7.022	0.997	0.985	0.943	0.876	0.803	0.596	
		0.078	11.627	0.391	0.058	0.018	0.008	0.004	0.002	
		0.139	14.197	0.082	0.000	0.000	0.000	0.000	0.000	
		0.250	15.667	0.000	0.000	0.000	0.000	0.000	0.000	
	Louvain CDA modularity	0.024	0.243	0.000	0.000	0.000	0.000	0.000	0.000	
		0.078	0.119	0.001	0.000	0.000	0.000	0.000	0.000	
		0.139	0.082	0.015	0.001	0.001	0.001	0.000	0.000	
		0.250	0.052	0.198	0.061	0.021	0.008	0.006	0.002	
	density	0.024	0.024	0.405	0.177	0.096	0.059	0.044	0.024	
		0.078	0.078	0.989	0.909	0.811	0.752	0.699	0.620	
		0.139	0.139	0.993	0.955	0.889	0.827	0.771	0.672	
		0.250	0.250	0.992	0.953	0.897	0.839	0.788	0.659	
	max betweenness centrality	0.024	0.019	0.304	0.058	0.033	0.023	0.013	0.010	
		0.078	0.011	1.000	0.965	0.809	0.692	0.591	0.439	
		0.139	0.009	1.000	0.995	0.981	0.923	0.847	0.668	
		0.250	0.008	1.000	0.997	0.969	0.901	0.835	0.713	
	transitivity	0.024	0.043	1.000	1.000	1.000	0.997	0.993	0.952	
		0.078	0.138	1.000	0.967	0.928	0.899	0.864	0.774	
		0.139	0.247	0.985	0.954	0.903	0.859	0.818	0.719	
		0.250	0.443	0.991	0.947	0.890	0.835	0.779	0.672	
	triangle density	0.024	0.000	0.990	0.852	0.706	0.563	0.456	0.309	
		0.078	0.001	0.998	0.965	0.905	0.845	0.798	0.701	
		0.139	0.006	0.991	0.959	0.902	0.840	0.794	0.692	
		0.250	0.037	0.992	0.956	0.882	0.828	0.779	0.678	
	1000	λ_1	0.013	18.974	0.894	0.677	0.470	0.379	0.270	0.157
			0.058	77.998	0.991	0.951	0.886	0.847	0.812	0.715
			0.120	160.817	0.988	0.959	0.911	0.859	0.814	0.749
			0.250	332.967	0.987	0.959	0.904	0.852	0.810	0.747
		λ_3	0.013	8.573	0.000	0.000	0.000	0.000	0.000	0.000
			0.058	16.482	0.000	0.000	0.000	0.000	0.000	0.000
			0.120	21.588	0.000	0.000	0.000	0.000	0.000	0.000
			0.250	24.642	0.000	0.000	0.000	0.000	0.000	0.000
		λ_{10}	0.013	8.002	0.044	0.000	0.000	0.000	0.000	0.000
			0.058	15.493	0.000	0.000	0.000	0.000	0.000	0.000
			0.120	20.377	0.000	0.000	0.000	0.000	0.000	0.000
			0.250	23.414	0.000	0.000	0.000	0.000	0.000	0.000
		Louvain CDA modularity	0.013	0.234	0.000	0.000	0.000	0.000	0.000	0.000
			0.058	0.100	0.000	0.000	0.000	0.000	0.000	0.000
			0.120	0.064	0.000	0.000	0.000	0.000	0.000	0.000
			0.250	0.037	0.000	0.000	0.000	0.000	0.000	0.000
density		0.013	0.013	0.000	0.000	0.000	0.000	0.000	0.000	
		0.058	0.058	0.937	0.849	0.777	0.705	0.669	0.576	
		0.120	0.120	0.984	0.943	0.892	0.836	0.800	0.725	
		0.250	0.250	0.992	0.953	0.896	0.846	0.800	0.732	
max betweenness centrality		0.013	0.011	0.101	0.010	0.006	0.004	0.003	0.001	
		0.058	0.006	1.000	0.960	0.872	0.776	0.690	0.531	
		0.120	0.005	0.998	0.994	0.975	0.957	0.916	0.775	
		0.250	0.004	1.000	0.993	0.982	0.947	0.899	0.791	
transitivity		0.013	0.024	1.000	1.000	0.998	0.982	0.955	0.880	
		0.058	0.103	0.997	0.980	0.942	0.906	0.864	0.778	
		0.120	0.214	0.983	0.952	0.913	0.868	0.835	0.733	
		0.250	0.444	0.988	0.959	0.922	0.880	0.827	0.716	
triangle density		0.013	0.000	0.385	0.144	0.064	0.040	0.027	0.011	
		0.058	0.000	0.991	0.946	0.884	0.846	0.804	0.714	
		0.120	0.004	0.988	0.967	0.921	0.857	0.809	0.745	
		0.250	0.037	0.984	0.966	0.910	0.856	0.807	0.746	

Table 1: Confidence interval coverage for different densities based on Monte Carlo simulations using the product generating function when $S = 1000$.

n	statistic	ρ_n	average for true graphs	Proportion of bootstrap CI that cover truth for					
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$
250	λ_1	0.127	33.515	0.993	0.961	0.885	0.805	0.744	0.614
		0.307	79.477	0.995	0.965	0.932	0.892	0.851	0.754
		0.476	122.926	0.992	0.976	0.938	0.901	0.845	0.748
		0.740	190.475	0.989	0.970	0.932	0.886	0.838	0.732
	λ_3	0.127	10.044	0.000	0.000	0.000	0.000	0.000	0.000
		0.307	13.573	0.000	0.000	0.000	0.000	0.000	0.000
		0.476	14.350	0.000	0.000	0.000	0.000	0.000	0.000
		0.740	12.038	0.608	0.107	0.020	0.004	0.000	0.000
	λ_{10}	0.127	8.891	0.000	0.000	0.000	0.000	0.000	0.000
		0.307	12.043	0.000	0.000	0.000	0.000	0.000	0.000
		0.476	12.728	0.000	0.000	0.000	0.000	0.000	0.000
		0.740	10.505	0.996	0.972	0.947	0.886	0.839	0.722
	Louvain CDA modularity	0.127	0.137	0.000	0.000	0.000	0.000	0.000	0.000
		0.307	0.074	0.000	0.000	0.000	0.000	0.000	0.000
		0.476	0.049	0.000	0.000	0.000	0.000	0.000	0.000
		0.740	0.024	0.981	0.900	0.787	0.716	0.610	0.440
	density	0.130	0.130	1.000	1.000	0.991	0.977	0.959	0.885
		0.314	0.314	0.998	0.986	0.964	0.930	0.886	0.802
		0.488	0.488	0.996	0.980	0.946	0.902	0.858	0.744
		0.759	0.759	0.993	0.968	0.927	0.887	0.840	0.741
	max betweenness centrality	0.127	0.009	0.750	0.271	0.119	0.079	0.067	0.040
		0.307	0.005	1.000	0.813	0.605	0.466	0.390	0.249
		0.476	0.003	0.997	0.852	0.761	0.639	0.533	0.379
		0.740	0.001	0.994	0.930	0.882	0.821	0.766	0.651
transitivity	0.127	0.132	0.999	0.990	0.971	0.923	0.886	0.771	
	0.307	0.318	0.995	0.970	0.931	0.893	0.856	0.749	
	0.476	0.494	0.992	0.976	0.944	0.907	0.857	0.745	
	0.740	0.768	0.993	0.970	0.943	0.884	0.824	0.728	
triangle density	0.127	0.002	0.999	0.991	0.968	0.910	0.844	0.769	
	0.307	0.032	0.995	0.975	0.948	0.917	0.871	0.783	
	0.476	0.119	0.994	0.981	0.944	0.903	0.854	0.755	
	0.740	0.446	0.989	0.971	0.934	0.893	0.836	0.731	
500	λ_1	0.071	37.783	0.966	0.846	0.712	0.602	0.504	0.325
		0.230	119.330	0.998	0.977	0.933	0.907	0.859	0.787
		0.413	213.258	0.999	0.980	0.947	0.904	0.864	0.781
		0.740	381.727	0.999	0.960	0.926	0.879	0.833	0.752
	λ_3	0.071	11.362	0.000	0.000	0.000	0.000	0.000	0.000
		0.230	18.167	0.000	0.000	0.000	0.000	0.000	0.000
		0.413	20.806	0.000	0.000	0.000	0.000	0.000	0.000
		0.740	17.709	0.000	0.000	0.000	0.000	0.000	0.000
	λ_{10}	0.071	10.551	0.000	0.000	0.000	0.000	0.000	0.000
		0.230	16.902	0.000	0.000	0.000	0.000	0.000	0.000
		0.413	19.377	0.000	0.000	0.000	0.000	0.000	0.000
		0.740	16.336	0.000	0.000	0.000	0.000	0.000	0.000
	Louvain CDA modularity	0.071	0.135	0.000	0.000	0.000	0.000	0.000	0.000
		0.230	0.066	0.000	0.000	0.000	0.000	0.000	0.000
		0.413	0.041	0.000	0.000	0.000	0.000	0.000	0.000
		0.740	0.018	0.671	0.153	0.038	0.019	0.005	0.002
	density	0.073	0.073	1.000	0.998	0.990	0.975	0.946	0.882
		0.236	0.236	0.999	0.992	0.977	0.938	0.905	0.839
		0.423	0.423	0.999	0.984	0.953	0.915	0.885	0.791
		0.759	0.759	0.998	0.965	0.921	0.883	0.845	0.755
	max betweenness centrality	0.071	0.005	0.097	0.031	0.012	0.009	0.006	0.002
		0.230	0.003	0.959	0.643	0.462	0.352	0.269	0.190
		0.413	0.002	0.995	0.902	0.764	0.630	0.517	0.344
		0.740	0.001	0.988	0.943	0.891	0.846	0.791	0.671
transitivity	0.071	0.074	1.000	0.999	0.979	0.953	0.887	0.772	
	0.230	0.239	0.998	0.987	0.952	0.906	0.872	0.792	
	0.413	0.428	0.999	0.986	0.959	0.921	0.882	0.808	
	0.740	0.769	0.997	0.962	0.930	0.897	0.835	0.754	
triangle density	0.071	0.000	0.998	0.977	0.911	0.849	0.809	0.668	
	0.230	0.013	0.998	0.988	0.957	0.927	0.890	0.811	
	0.413	0.077	0.999	0.983	0.956	0.920	0.862	0.797	
	0.740	0.446	0.999	0.957	0.928	0.885	0.833	0.754	

Table 2: Confidence interval coverage for different densities based on Monte Carlo simulations using the high density generating function when $S = 1000$.

n	statistic	ρ_n	average for true graphs	Proportion of bootstrap CI that cover truth for						
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$	
750	λ_1	0.008	7.696	0.981	0.899	0.700	0.563	0.446	0.311	
		0.029	25.866	0.991	0.936	0.890	0.841	0.786	0.673	
		0.058	49.481	0.985	0.944	0.880	0.828	0.794	0.696	
		0.113	95.648	0.983	0.946	0.897	0.829	0.779	0.681	
	λ_3	0.008	5.338	0.000	0.000	0.000	0.000	0.000	0.000	
		0.029	12.664	0.758	0.567	0.431	0.333	0.279	0.209	
		0.058	22.937	0.979	0.898	0.825	0.764	0.709	0.619	
		0.113	43.090	0.984	0.902	0.833	0.759	0.705	0.610	
	λ_{10}	0.008	4.773	0.698	0.171	0.052	0.017	0.012	0.004	
		0.029	8.859	0.000	0.000	0.000	0.000	0.000	0.000	
		0.058	11.656	0.000	0.000	0.000	0.000	0.000	0.000	
		0.113	14.127	0.989	0.937	0.876	0.831	0.793	0.701	
	Louvain CDA modularity	0.008	0.436	0.000	0.000	0.000	0.000	0.000	0.000	
		0.029	0.323	0.891	0.806	0.712	0.646	0.585	0.494	
		0.058	0.307	0.974	0.915	0.836	0.756	0.689	0.589	
		0.113	0.294	0.988	0.948	0.903	0.842	0.754	0.673	
	density	0.008	0.008	0.007	0.000	0.000	0.000	0.000	0.000	
		0.029	0.029	0.952	0.719	0.569	0.413	0.317	0.203	
		0.058	0.058	0.980	0.932	0.876	0.821	0.758	0.678	
		0.113	0.113	0.993	0.948	0.905	0.855	0.802	0.710	
	max betweenness centrality	0.008	0.023	0.806	0.334	0.176	0.125	0.094	0.062	
		0.029	0.008	1.000	0.990	0.885	0.797	0.692	0.508	
		0.058	0.006	1.000	0.969	0.869	0.713	0.586	0.464	
		0.113	0.005	1.000	0.996	0.960	0.890	0.811	0.665	
	transitivity	0.008	0.010	1.000	1.000	1.000	0.998	0.995	0.973	
		0.029	0.039	0.987	0.946	0.901	0.848	0.794	0.714	
		0.058	0.076	0.983	0.945	0.884	0.840	0.809	0.699	
		0.113	0.148	0.976	0.926	0.870	0.824	0.782	0.668	
	triangle density	0.008	0.000	1.000	0.998	0.992	0.972	0.935	0.869	
		0.029	0.000	0.998	0.961	0.920	0.883	0.835	0.743	
		0.058	0.000	0.992	0.930	0.894	0.846	0.800	0.709	
		0.113	0.002	0.976	0.930	0.876	0.824	0.771	0.682	
	1000	λ_1	0.006	8.002	0.977	0.788	0.611	0.472	0.332	0.204
			0.026	30.381	0.998	0.941	0.894	0.844	0.787	0.689
			0.054	61.907	0.998	0.942	0.886	0.831	0.792	0.694
			0.113	127.414	0.996	0.950	0.900	0.844	0.781	0.672
		λ_3	0.006	5.528	0.000	0.000	0.000	0.000	0.000	0.000
			0.026	14.699	0.903	0.682	0.563	0.461	0.416	0.333
			0.054	28.486	0.986	0.947	0.891	0.846	0.789	0.678
			0.113	57.194	0.989	0.934	0.887	0.836	0.776	0.694
		λ_{10}	0.006	5.000	0.141	0.001	0.000	0.000	0.000	0.000
			0.026	9.885	0.000	0.000	0.000	0.000	0.000	0.000
			0.054	13.380	0.000	0.000	0.000	0.000	0.000	0.000
			0.113	16.554	0.995	0.972	0.914	0.848	0.801	0.697
		Louvain CDA modularity	0.006	0.425	0.000	0.000	0.000	0.000	0.000	0.000
			0.026	0.319	0.958	0.898	0.834	0.786	0.731	0.604
			0.054	0.302	0.971	0.881	0.804	0.747	0.643	0.557
			0.113	0.291	0.998	0.969	0.924	0.875	0.834	0.707
density		0.006	0.006	0.000	0.000	0.000	0.000	0.000	0.000	
		0.026	0.026	0.810	0.537	0.366	0.285	0.173	0.101	
		0.054	0.054	0.980	0.895	0.828	0.776	0.735	0.605	
		0.113	0.113	0.988	0.966	0.909	0.869	0.793	0.690	
max betweenness centrality		0.006	0.018	0.658	0.210	0.104	0.066	0.045	0.028	
		0.026	0.006	1.000	0.995	0.897	0.776	0.697	0.507	
		0.054	0.004	1.000	0.936	0.765	0.643	0.522	0.370	
		0.113	0.003	1.000	0.997	0.918	0.874	0.843	0.706	
transitivity		0.006	0.008	1.000	1.000	1.000	0.998	0.993	0.968	
		0.026	0.034	0.993	0.957	0.898	0.846	0.794	0.685	
		0.054	0.071	0.991	0.945	0.888	0.815	0.763	0.654	
		0.113	0.147	0.992	0.944	0.868	0.800	0.747	0.625	
triangle density		0.006	0.000	1.000	0.996	0.974	0.946	0.906	0.800	
		0.026	0.000	0.998	0.980	0.937	0.872	0.818	0.716	
		0.054	0.000	0.993	0.957	0.875	0.832	0.765	0.649	
		0.113	0.002	0.994	0.943	0.886	0.805	0.737	0.606	

Table 3: Confidence interval coverage for different densities based on Monte Carlo simulations using the horseshoe generating function when $S = 1000$.

n	average $a^{(opt)}$	statistic	average for true graphs	S	Proportion of bootstrap CI that cover truth for							
					$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$		
300	0.000105	density	0.187	100	1.000	0.990	0.930	0.880	0.880	0.760		
				200	0.995	0.950	0.910	0.870	0.840	0.770		
				300	0.993	0.947	0.910	0.850	0.803	0.740		
				500	0.996	0.950	0.916	0.854	0.806	0.726		
				750	0.996	0.943	0.897	0.835	0.795	0.709		
				1000	0.996	0.939	0.897	0.847	0.812	0.712		
				1250	0.993	0.941	0.902	0.845	0.811	0.714		
				1500	0.995	0.945	0.901	0.846	0.807	0.713		
				2000	0.995	0.949	0.900	0.849	0.809	0.719		
				100	1.000	0.990	0.920	0.890	0.840	0.700		
		λ_1	75.206			200	0.995	0.955	0.905	0.870	0.815	0.730
						300	0.993	0.950	0.890	0.853	0.820	0.733
						500	0.996	0.944	0.894	0.860	0.824	0.722
						750	0.996	0.941	0.892	0.843	0.819	0.719
						1000	0.994	0.942	0.898	0.843	0.813	0.725
						1250	0.993	0.941	0.897	0.850	0.818	0.721
						1500	0.993	0.949	0.899	0.853	0.815	0.724
						2000	0.995	0.952	0.899	0.851	0.815	0.719
						100	1.000	0.980	0.940	0.890	0.810	0.700
						transitivity	0.332			200	1.000	0.965
300	0.997	0.947	0.917	0.870	0.817					0.700		
500	0.992	0.946	0.896	0.836	0.796					0.702		
750	0.991	0.949	0.904	0.841	0.815					0.715		
1000	0.990	0.946	0.888	0.847	0.812					0.712		
1250	0.990	0.946	0.890	0.839	0.806					0.712		
1500	0.989	0.946	0.893	0.845	0.809					0.718		
2000	0.990	0.949	0.905	0.848	0.805					0.712		
100	0.950	0.910	0.890	0.830	0.790					0.640		
500	0.000059	density	0.188	200	0.975					0.925	0.890	0.845
				300	0.983	0.953	0.897	0.850	0.803	0.663		
				500	0.982	0.954	0.914	0.848	0.804	0.692		
				750	0.983	0.948	0.897	0.849	0.809	0.707		
				1000	0.982	0.949	0.905	0.852	0.812	0.712		
				1250	0.989	0.957	0.917	0.865	0.825	0.713		
				1500	0.991	0.958	0.919	0.876	0.829	0.715		
				2000	0.993	0.960	0.919	0.869	0.827	0.722		
				100	0.940	0.910	0.860	0.790	0.780	0.660		
				λ_1	125.416			200	0.990	0.920	0.890	0.840
		300	0.990					0.947	0.910	0.867	0.817	0.670
		500	0.986					0.954	0.928	0.858	0.828	0.668
		750	0.991					0.951	0.913	0.863	0.815	0.696
		1000	0.988					0.951	0.915	0.866	0.811	0.709
		1250	0.990					0.954	0.914	0.872	0.820	0.716
		1500	0.991					0.955	0.919	0.875	0.817	0.719
		2000	0.995					0.959	0.920	0.881	0.816	0.716
		100	0.990					0.960	0.800	0.770	0.710	0.630
		transitivity	0.333							200	0.990	0.950
				300	0.987	0.953	0.930			0.863	0.803	0.717
500	0.986			0.964	0.926	0.874	0.814			0.706		
750	0.988			0.969	0.931	0.875	0.825			0.727		
1000	0.982			0.957	0.922	0.865	0.812			0.722		
1250	0.993			0.962	0.921	0.862	0.815			0.720		
1500	0.990			0.961	0.917	0.862	0.815			0.713		
2000	0.997			0.963	0.923	0.874	0.824			0.713		

Table 4: Confidence interval coverage for different number of simulated true graphs S based on Monte Carlo simulations using the product generating function when $n = 300$ or $n = 500$ and $\rho_n = 0.1874$.

n	average $a^{(opt)}$	statistic	average for true graphs	S	Proportion of bootstrap CI that cover truth for							
					$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$		
300	0.000046	density	0.112	100	0.990	0.990	0.960	0.960	0.950	0.770		
				200	0.990	0.980	0.955	0.945	0.925	0.850		
				300	0.990	0.983	0.963	0.940	0.903	0.827		
				500	0.992	0.976	0.952	0.922	0.880	0.802		
				750	0.992	0.972	0.948	0.905	0.860	0.780		
				1000	0.994	0.974	0.945	0.901	0.866	0.781		
				1250	0.995	0.975	0.943	0.898	0.858	0.775		
				1500	0.995	0.975	0.937	0.899	0.859	0.774		
				2000	0.998	0.973	0.940	0.901	0.855	0.761		
				100	0.970	0.960	0.810	0.720	0.700	0.630		
		λ_2	25.306			200	0.980	0.965	0.910	0.875	0.800	0.665
						300	0.990	0.963	0.930	0.900	0.830	0.680
						500	0.994	0.954	0.910	0.876	0.818	0.662
						750	0.991	0.953	0.905	0.861	0.800	0.669
						1000	0.995	0.959	0.911	0.859	0.805	0.669
						1250	0.995	0.957	0.912	0.866	0.819	0.674
						1500	0.995	0.965	0.917	0.871	0.821	0.682
						2000	0.997	0.963	0.919	0.869	0.819	0.693
						100	0.980	0.850	0.810	0.780	0.750	0.670
						transitivity	0.146			200	1.000	0.910
300	0.987	0.873	0.820	0.793	0.743					0.663		
500	0.986	0.912	0.846	0.802	0.772					0.690		
750	0.983	0.920	0.856	0.809	0.772					0.667		
1000	0.995	0.932	0.881	0.813	0.777					0.674		
1250	0.994	0.934	0.885	0.818	0.779					0.674		
1500	0.995	0.943	0.889	0.830	0.791					0.685		
2000	0.994	0.945	0.885	0.822	0.776					0.681		
100	0.970	0.910	0.860	0.830	0.810					0.750		
500	0.000031	density	0.112	200	0.985					0.955	0.890	0.855
				300	0.990	0.953	0.937	0.873	0.820	0.773		
				500	1.000	0.962	0.946	0.892	0.836	0.746		
				750	0.999	0.961	0.949	0.896	0.848	0.761		
				1000	0.998	0.961	0.936	0.878	0.832	0.748		
				1250	0.997	0.966	0.940	0.890	0.843	0.746		
				1500	0.996	0.966	0.937	0.886	0.831	0.731		
				2000	0.998	0.964	0.931	0.883	0.834	0.743		
				100	0.990	0.940	0.900	0.830	0.770	0.710		
				λ_2	41.732			200	1.000	0.970	0.905	0.795
		300	0.997					0.943	0.887	0.817	0.783	0.717
		500	0.998					0.948	0.910	0.844	0.802	0.734
		750	0.999					0.955	0.897	0.829	0.797	0.727
		1000	0.998					0.960	0.905	0.834	0.801	0.725
		1250	0.998					0.962	0.900	0.842	0.806	0.731
		1500	0.998					0.958	0.896	0.834	0.797	0.715
		2000	0.998					0.956	0.905	0.848	0.804	0.726
		100	0.990					0.950	0.890	0.860	0.750	0.690
		transitivity	0.147							200	0.985	0.960
				300	0.983	0.947	0.930			0.850	0.807	0.700
500	0.988			0.958	0.908	0.838	0.790			0.670		
750	0.989			0.953	0.896	0.843	0.795			0.676		
1000	0.989			0.955	0.898	0.855	0.812			0.697		
1250	0.990			0.958	0.904	0.859	0.813			0.701		
1500	0.991			0.951	0.901	0.859	0.806			0.703		
2000	0.991			0.956	0.890	0.846	0.792			0.684		

Table 5: Confidence interval coverage for different number of simulated true graphs S based on Monte Carlo simulations using the horseshoe generating function when $\rho_n = 0.1125$.

ρ_n	statistic	average for true graphs	S	Proportion of bootstrap CI that cover truth for					
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$
0.569	density	0.570	100	1.000	0.990	0.970	0.950	0.910	0.810
			200	1.000	0.980	0.955	0.930	0.855	0.760
			300	1.000	0.980	0.947	0.913	0.843	0.757
			500	1.000	0.980	0.946	0.890	0.832	0.754
			750	1.000	0.975	0.925	0.855	0.815	0.715
			1000	0.996	0.975	0.928	0.858	0.821	0.733
			1250	0.998	0.970	0.918	0.851	0.815	0.718
			1500	0.997	0.969	0.924	0.863	0.820	0.723
			2000	0.994	0.970	0.924	0.862	0.820	0.722
			λ_1	286.543	100	1.000	0.940	0.800	0.740
	200	0.985			0.960	0.905	0.840	0.770	0.650
	300	0.983			0.940	0.910	0.840	0.750	0.623
	500	0.984			0.940	0.898	0.822	0.742	0.638
	750	0.992			0.948	0.907	0.841	0.792	0.687
	1000	0.993			0.944	0.903	0.842	0.784	0.701
	1250	0.996			0.962	0.908	0.848	0.796	0.700
	1500	0.997			0.963	0.907	0.862	0.797	0.699
	2000	0.998			0.966	0.910	0.868	0.814	0.708
	transitivity	0.576			100	1.000	0.920	0.870	0.730
			200	0.985	0.955	0.900	0.840	0.795	0.670
300			0.983	0.927	0.903	0.803	0.753	0.630	
500			0.986	0.940	0.886	0.834	0.760	0.652	
750			0.999	0.949	0.908	0.856	0.795	0.688	
1000			0.998	0.948	0.901	0.855	0.800	0.699	
1250			0.998	0.967	0.909	0.866	0.806	0.698	
1500			0.997	0.967	0.911	0.869	0.807	0.691	
2000			0.998	0.969	0.912	0.869	0.821	0.703	
0.759			density	0.759	100	0.990	0.990	0.920	0.920
	200	0.995			0.970	0.925	0.900	0.840	0.690
	300	0.990			0.977	0.917	0.870	0.840	0.723
	500	0.988			0.974	0.918	0.882	0.842	0.712
	750	0.996			0.977	0.928	0.893	0.849	0.747
	1000	0.990			0.970	0.926	0.895	0.861	0.748
	1250	0.992			0.965	0.928	0.890	0.855	0.739
	1500	0.990			0.958	0.915	0.873	0.827	0.707
	2000	0.990			0.957	0.906	0.861	0.818	0.700
	λ_1	381.690			100	0.960	0.940	0.860	0.850
			200	0.970	0.945	0.900	0.895	0.870	0.675
			300	0.987	0.957	0.910	0.897	0.813	0.710
			500	0.992	0.964	0.918	0.882	0.834	0.732
			750	0.987	0.967	0.907	0.863	0.832	0.737
			1000	0.989	0.962	0.912	0.874	0.833	0.739
			1250	0.990	0.958	0.903	0.845	0.809	0.710
			1500	0.990	0.961	0.908	0.850	0.813	0.711
			2000	0.989	0.957	0.902	0.849	0.814	0.708
			transitivity	0.768	100	0.950	0.940	0.870	0.840
	200	0.965			0.930	0.900	0.885	0.830	0.700
300	0.983	0.957			0.910	0.890	0.850	0.693	
500	0.992	0.968			0.914	0.886	0.854	0.722	
750	0.987	0.967			0.915	0.881	0.837	0.735	
1000	0.988	0.962			0.917	0.878	0.836	0.734	
1250	0.990	0.956			0.896	0.854	0.802	0.706	
1500	0.989	0.958			0.903	0.860	0.805	0.711	
2000	0.990	0.957			0.900	0.855	0.807	0.711	

Table 6: Confidence interval coverage for different number of simulated true graphs S based on Monte Carlo simulations using the high density generating function when $n = 500$.

n	statistic	average for true graphs	method	Proportion of bootstrap CI that cover truth for					
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$
100	λ_1	17.222	HK1	0.998	0.982	0.940	0.895	0.854	0.775
			HK2	0.998	0.986	0.939	0.894	0.861	0.769
			HNN1	0.947	0.777	0.648	0.590	0.507	0.397
			emp	0.999	0.953	0.912	0.877	0.842	0.742
	λ_2	6.418	HK1	0.974	0.858	0.704	0.599	0.504	0.341
			HK2	1.000	0.947	0.879	0.796	0.679	0.514
			HNN1	0.966	0.896	0.771	0.675	0.612	0.537
			emp	0.343	0.057	0.012	0.004	0.002	0.000
	transitivity	0.220	HK1	0.999	0.992	0.971	0.938	0.907	0.847
			HK2	0.998	0.984	0.963	0.933	0.901	0.831
			HNN1	0.986	0.937	0.838	0.785	0.739	0.665
			emp	0.992	0.966	0.897	0.839	0.800	0.703
300	λ_1	50.535	HK1	0.995	0.959	0.918	0.876	0.836	0.739
			HK2	0.994	0.958	0.923	0.889	0.859	0.765
			HNN1	0.936	0.847	0.724	0.647	0.597	0.509
			emp	0.990	0.953	0.900	0.859	0.824	0.733
	λ_2	11.778	HK1	0.000	0.000	0.000	0.000	0.000	0.000
			HK2	0.001	0.000	0.000	0.000	0.000	0.000
			HNN1	0.978	0.915	0.857	0.814	0.754	0.651
			emp	0.000	0.000	0.000	0.000	0.000	0.000
	transitivity	0.221	HK1	0.997	0.982	0.940	0.897	0.859	0.759
			HK2	0.996	0.984	0.954	0.904	0.860	0.782
			HNN1	0.981	0.919	0.872	0.823	0.747	0.629
			emp	0.988	0.947	0.897	0.847	0.781	0.678
500	λ_1	83.925	HK1	0.983	0.955	0.904	0.849	0.782	0.680
			HK2	0.986	0.944	0.902	0.858	0.791	0.701
			HNN1	0.950	0.860	0.773	0.682	0.633	0.534
			emp	0.981	0.937	0.888	0.851	0.817	0.708
	λ_2	15.448	HK1	0.000	0.000	0.000	0.000	0.000	0.000
			HK2	0.000	0.000	0.000	0.000	0.000	0.000
			HNN1	0.982	0.915	0.854	0.814	0.753	0.658
			emp	0.000	0.000	0.000	0.000	0.000	0.000
	transitivity	0.222	HK1	0.986	0.945	0.899	0.867	0.814	0.708
			HK2	0.987	0.956	0.911	0.851	0.812	0.734
			HNN1	0.982	0.920	0.838	0.780	0.734	0.630
			emp	0.982	0.968	0.905	0.837	0.772	0.692

Table 7: Confidence interval coverage based on Monte Carlo simulations for different bootstrap methods for the true graphs from the product generating function with density $\rho_n = 0.125$, $S = 1000$ true graphs, with sample size n ranging from 100 to 500. The methods are: HK1 (our main method based on $\hat{h} \equiv \hat{h}^{(K1)}$ with $a^{(opt)}$), HK2 (our bootstrap method but using the linking function estimator $\hat{h}^{(K2)}$ with $a^{(optK2)}$ based on $\hat{h}^{(K2)}$), HNN1 (our bootstrap method but but using the linking function estimator $\hat{h}^{(N1)}$ from Zhang, Levina, and Zhu (2017) with their optimal choice of neighbourhood size), emp (empirical bootstrap from Green and Shalizi (2022)), dot_prod_k (the bootstrap method from Levin and Levina (2019) based on assuming a k -dimensional ξ_i).

statistic	n	average for true graphs	method	Proportion of bootstrap CI that cover truth for							
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$		
density	250	0.160767	HK1	0.991	0.959	0.914	0.874	0.823	0.736		
			HK2	0.993	0.956	0.908	0.868	0.826	0.721		
			HNN1	0.981	0.914	0.873	0.818	0.762	0.636		
			LLS_L	1.000	1.000	1.000	0.999	0.999	0.990		
			asyp estimated var	0.993	0.973	0.948	0.909	0.864	0.781		
			asyp infeasible var	0.990	0.943	0.895	0.846	0.795	0.701		
			dot_prod_1	1.000	0.999	0.998	0.990	0.973	0.913		
			dot_prod_3	1.000	0.999	0.998	0.991	0.967	0.906		
			emp	0.990	0.961	0.910	0.871	0.828	0.742		
	500	0.139211	HK1	0.993	0.955	0.889	0.827	0.771	0.672		
			HK2	0.993	0.954	0.872	0.816	0.753	0.667		
			HNN1	0.987	0.896	0.795	0.730	0.675	0.590		
			asyp estimated var	0.998	0.978	0.948	0.907	0.868	0.770		
			asyp infeasible var	0.993	0.952	0.900	0.855	0.797	0.676		
			dot_prod_1	1.000	1.000	0.996	0.988	0.979	0.920		
			dot_prod_3	1.000	1.000	0.997	0.991	0.985	0.936		
			emp	0.991	0.942	0.891	0.841	0.784	0.677		
			750	0.127979	HK1	0.979	0.937	0.877	0.842	0.788	0.684
	HK2	0.973			0.928	0.873	0.824	0.774	0.665		
	HNN1	0.944			0.866	0.775	0.721	0.659	0.556		
	asyp estimated var	0.994			0.980	0.953	0.914	0.883	0.799		
	asyp infeasible var	0.989			0.948	0.901	0.852	0.803	0.704		
	dot_prod_1	1.000			0.999	0.993	0.982	0.964	0.920		
	dot_prod_3	1.000			0.999	0.993	0.982	0.965	0.923		
	emp	0.978			0.933	0.887	0.835	0.795	0.691		
	triangle density	250			0.009854	HK1	0.990	0.969	0.942	0.886	0.852
			HK2	0.990		0.974	0.941	0.902	0.869	0.749	
HNN1			0.990	0.946		0.902	0.839	0.788	0.675		
LLS_L			1.000	0.999		0.995	0.983	0.968	0.925		
LLS_Q			1.000	0.999		0.995	0.983	0.966	0.923		
asyp infeasible var			0.987	0.953		0.905	0.852	0.794	0.697		
dot_prod_1			1.000	0.999		0.997	0.983	0.975	0.930		
dot_prod_3			1.000	0.999		0.999	0.983	0.976	0.933		
emp			0.994	0.966		0.935	0.871	0.821	0.732		
500		0.006402	HK1	0.991	0.959	0.902	0.840	0.794	0.692		
			HK2	0.997	0.960	0.896	0.847	0.790	0.700		
			HNN1	0.986	0.929	0.826	0.760	0.704	0.606		
			asyp infeasible var	0.993	0.951	0.889	0.852	0.797	0.695		
			dot_prod_1	1.000	1.000	0.997	0.987	0.970	0.911		
			dot_prod_3	1.000	1.000	0.998	0.993	0.973	0.921		
			emp	0.996	0.951	0.882	0.839	0.789	0.692		
			750	0.004970	HK1	0.987	0.946	0.898	0.854	0.805	0.695
					HK2	0.987	0.943	0.892	0.852	0.793	0.697
HNN1		0.962			0.904	0.820	0.752	0.703	0.598		
asyp infeasible var		0.987			0.940	0.900	0.852	0.803	0.703		
dot_prod_1		1.000			1.000	0.995	0.985	0.965	0.919		
dot_prod_3		1.000			1.000	0.995	0.985	0.967	0.922		
emp		0.977			0.938	0.900	0.841	0.804	0.709		

Table 8: Confidence interval coverage based on Monte Carlo simulations for different bootstrap methods for the true graphs from the product generating function with density $\rho_n \sim \sqrt[4]{\frac{\log(n)}{n}}$, $S = 1000$ true graphs. The methods are: HK1 (our main method based on $\hat{h} \equiv \hat{h}^{(K1)}$ with $a^{(opt)}$), HK2 (our bootstrap method but using the linking function estimator $\hat{h}^{(K2)}$ with $a^{(optK2)}$ based on $\hat{h}^{(K2)}$), HNN1 (our bootstrap method but but using the linking function estimator $\hat{h}^{(NN1)}$ from Zhang, Levina, and Zhu (2017) with their optimal choice of neighbourhood size), emp (empirical bootstrap from Green and Shalizi (2022)), dot_prod_ k (the bootstrap method from Levin and Levina (2019) based on assuming a k -dimensional ξ_i), asyp estimated var (the asymptotic distribution from Bickel, Chen, and Levina (2011) with variance estimated according to the formula in Green and Shalizi (2022)), asyp infeasible var (the asymptotic distribution from Bickel, Chen, and Levina (2011) with the true theoretical variance), LLS_L and LLS_Q (the linear and quadratic methods from Lin, Lunde, and Sarkar (2020)).

ρ_n	statistic	average for true graphs	method	Proportion of bootstrap CI that cover truth for							
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$		
0.056	density	0.056	HK1	0.989	0.927	0.864	0.786	0.721	0.634		
			HK2	0.968	0.825	0.654	0.514	0.419	0.346		
			HNN1	0.253	0.066	0.034	0.017	0.005	0.001		
			dot_prod_1	0.005	0.000	0.000	0.000	0.000	0.000		
			dot_prod_3	1.000	0.998	0.985	0.973	0.955	0.901		
			emp	0.999	0.985	0.960	0.920	0.875	0.800		
	λ_1	32.530	HK1	0.988	0.968	0.931	0.866	0.820	0.722		
			HK2	0.996	0.977	0.945	0.898	0.849	0.767		
			HNN1	0.925	0.751	0.631	0.522	0.440	0.332		
			emp	0.979	0.925	0.871	0.813	0.763	0.669		
			λ_2	21.481	HK1	0.997	0.957	0.903	0.860	0.827	0.729
					HK2	0.991	0.911	0.807	0.748	0.686	0.589
	transitivity	0.073	HNN1	0.970	0.845	0.696	0.585	0.507	0.409		
			emp	0.935	0.811	0.694	0.631	0.560	0.480		
			HK1	0.997	0.938	0.890	0.848	0.800	0.672		
			HK2	0.970	0.930	0.867	0.826	0.781	0.688		
			HNN1	0.966	0.901	0.818	0.767	0.702	0.590		
			emp	0.993	0.956	0.904	0.851	0.786	0.692		
0.084	density	0.084	HK1	0.990	0.953	0.914	0.874	0.833	0.722		
			HK2	0.990	0.950	0.909	0.861	0.794	0.650		
			HNN1	0.910	0.854	0.732	0.619	0.563	0.436		
			dot_prod_1	0.000	0.000	0.000	0.000	0.000	0.000		
			dot_prod_3	0.999	0.996	0.987	0.970	0.948	0.903		
			emp	0.996	0.986	0.957	0.914	0.875	0.797		
	λ_1	48.240	HK1	0.997	0.971	0.912	0.863	0.817	0.705		
			HK2	0.998	0.976	0.934	0.875	0.821	0.721		
			HNN1	0.973	0.919	0.843	0.809	0.760	0.638		
			emp	0.997	0.967	0.900	0.837	0.786	0.681		
			λ_2	31.542	HK1	0.996	0.971	0.931	0.876	0.827	0.730
					HK2	0.993	0.951	0.896	0.854	0.797	0.672
	transitivity	0.110	HNN1	0.989	0.959	0.902	0.834	0.805	0.694		
			emp	0.991	0.941	0.876	0.824	0.780	0.663		
			HK1	0.997	0.949	0.886	0.825	0.779	0.669		
			HK2	0.996	0.950	0.896	0.845	0.796	0.702		
			HNN1	0.955	0.885	0.814	0.761	0.688	0.578		
			emp	0.998	0.965	0.922	0.874	0.819	0.738		
0.113	density	0.113	HK1	0.993	0.975	0.922	0.866	0.823	0.732		
			HK2	0.989	0.970	0.914	0.869	0.829	0.718		
			HNN1	0.990	0.937	0.875	0.827	0.777	0.661		
			dot_prod_1	0.000	0.000	0.000	0.000	0.000	0.000		
			dot_prod_3	1.000	0.993	0.985	0.975	0.958	0.898		
			emp	0.993	0.982	0.955	0.918	0.870	0.792		
	λ_1	63.879	HK1	0.997	0.968	0.908	0.856	0.781	0.662		
			HK2	1.000	0.973	0.916	0.874	0.807	0.690		
			HNN1	0.989	0.937	0.879	0.789	0.738	0.654		
			emp	0.993	0.958	0.899	0.839	0.798	0.693		
			λ_2	41.713	HK1	0.994	0.961	0.909	0.855	0.810	0.694
					HK2	0.994	0.941	0.898	0.845	0.793	0.697
	transitivity	0.147	HNN1	0.988	0.955	0.901	0.857	0.799	0.686		
			emp	0.994	0.945	0.883	0.814	0.749	0.643		
			HK1	0.996	0.956	0.895	0.820	0.766	0.672		
			HK2	1.000	0.957	0.902	0.837	0.784	0.661		
			HNN1	0.956	0.846	0.771	0.694	0.655	0.550		
			emp	0.989	0.944	0.891	0.846	0.792	0.682		

Table 9: Confidence interval coverage based on Monte Carlo simulations for different bootstrap methods for the true graphs from the horseshoe generating function with sample size $n = 500$, $S = 1000$ true graphs, and different values of density ρ_n ranging from 0.056 to 0.1125. The methods are: HK1 (our main method based on $\hat{h} \equiv \hat{h}^{(K1)}$ with $a^{(opt)}$), HK2 (our bootstrap method but using the linking function estimator $\hat{h}^{(K2)}$ with $a^{(optK2)}$ based on $\hat{h}^{(K2)}$), HNN1 (our bootstrap method but using the linking function estimator $\hat{h}^{(NN1)}$ from Zhang, Levina, and Zhu (2017) with their optimal choice of neighbourhood size), emp (empirical bootstrap from Green and Shalizi (2022)), dot_prod_ k (the bootstrap method from Levin and Levina (2019) based on assuming a k -dimensional ξ_i).

ρ_n	statistic	average for true graphs	method	Proportion of bootstrap CI that cover truth for							
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.3$		
0.380	density	0.380	HK1	0.998	0.984	0.964	0.928	0.879	0.800		
			HK2	0.995	0.978	0.954	0.931	0.888	0.795		
			HNN1	0.982	0.921	0.859	0.783	0.735	0.643		
			dot_prod_1	1.000	0.996	0.987	0.975	0.948	0.873		
			dot_prod_3	1.000	0.996	0.986	0.973	0.944	0.873		
			emp	0.998	0.983	0.948	0.932	0.885	0.772		
	λ_1	114.861	HK1	0.994	0.964	0.937	0.892	0.851	0.777		
			HK2	0.999	0.970	0.939	0.907	0.869	0.793		
			HNN1	0.990	0.973	0.926	0.878	0.834	0.728		
			emp	1.000	0.984	0.957	0.935	0.902	0.817		
			λ_2	15.962	HK1	0.000	0.000	0.000	0.000	0.000	0.000
					HK2	0.000	0.000	0.000	0.000	0.000	0.000
	transitivity	0.384	HNN1	1.000	0.992	0.980	0.952	0.929	0.847		
			emp	0.000	0.000	0.000	0.000	0.000	0.000		
			HK1	0.994	0.973	0.937	0.902	0.868	0.786		
			HK2	1.000	0.981	0.937	0.897	0.859	0.775		
			HNN1	0.997	0.973	0.948	0.915	0.880	0.774		
			emp	0.998	0.973	0.888	0.839	0.778	0.694		
0.570	density	0.570	HK1	0.994	0.969	0.945	0.904	0.849	0.756		
			HK2	0.993	0.973	0.923	0.882	0.827	0.713		
			HNN1	0.994	0.964	0.926	0.889	0.836	0.731		
			dot_prod_1	1.000	0.996	0.991	0.973	0.947	0.886		
			dot_prod_3	1.000	0.996	0.987	0.963	0.939	0.857		
			emp	0.991	0.945	0.915	0.863	0.814	0.723		
	λ_1	171.787	HK1	0.999	0.978	0.941	0.905	0.852	0.767		
			HK2	0.999	0.978	0.941	0.900	0.857	0.779		
			HNN1	0.990	0.972	0.939	0.884	0.838	0.740		
			emp	0.990	0.971	0.937	0.887	0.850	0.757		
			λ_2	15.963	HK1	0.000	0.000	0.000	0.000	0.000	0.000
					HK2	0.000	0.000	0.000	0.000	0.000	0.000
	transitivity	0.576	HNN1	1.000	0.993	0.977	0.949	0.895	0.832		
			emp	0.000	0.000	0.000	0.000	0.000	0.000		
			HK1	0.998	0.979	0.934	0.908	0.864	0.770		
			HK2	0.997	0.978	0.940	0.902	0.867	0.776		
			HNN1	0.987	0.969	0.920	0.877	0.830	0.730		
			emp	0.985	0.938	0.891	0.819	0.787	0.652		
0.759	density	0.759	HK1	0.994	0.963	0.911	0.872	0.841	0.737		
			HK2	0.992	0.961	0.919	0.870	0.825	0.747		
			HNN1	0.986	0.942	0.902	0.859	0.809	0.694		
			dot_prod_1	1.000	0.998	0.990	0.975	0.945	0.881		
			dot_prod_3	1.000	0.997	0.986	0.959	0.922	0.850		
			emp	0.985	0.947	0.890	0.835	0.810	0.706		
	λ_1	228.714	HK1	0.996	0.965	0.926	0.884	0.842	0.731		
			HK2	0.993	0.959	0.926	0.877	0.841	0.734		
			HNN1	0.991	0.962	0.931	0.865	0.819	0.702		
			emp	0.983	0.953	0.914	0.864	0.813	0.702		
			λ_2	13.730	HK1	0.020	0.000	0.000	0.000	0.000	0.000
					HK2	0.978	0.817	0.615	0.498	0.379	0.217
	transitivity	0.768	HNN1	0.999	0.988	0.965	0.932	0.892	0.784		
			emp	0.000	0.000	0.000	0.000	0.000	0.000		
			HK1	0.994	0.961	0.916	0.885	0.833	0.720		
			HK2	0.993	0.961	0.916	0.886	0.829	0.732		
			HNN1	0.992	0.965	0.932	0.864	0.815	0.712		
			emp	0.970	0.920	0.872	0.814	0.776	0.643		

Table 10: Confidence interval coverage based on Monte Carlo simulations for different bootstrap methods for the true graphs from the high density generating function with sample size $n = 300$, $S = 1000$ true graphs, and different values of density ρ_n ranging from 0.38 to 0.79. The methods are: HK1 (our main method based on $\hat{h} \equiv \hat{h}^{(K1)}$ with $a^{(opt)}$), HK2 (our bootstrap method but using the linking function estimator $\hat{h}^{(K2)}$ with $a^{(optK2)}$ based on $\hat{h}^{(K2)}$), HNN1 (our bootstrap method but using the linking function estimator $\hat{h}^{(NN1)}$ from Zhang, Levina, and Zhu (2017) with their optimal choice of neighbourhood size), emp (empirical bootstrap from Green and Shalizi (2022)), dot_prod_ k (the bootstrap method from Levin and Levina (2019) based on assuming a k -dimensional ξ_i).

true value	c	average $a^{(opt)}$	Proportion of bootstrap CI that cover truth for					
			$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.15$	$\alpha=0.2$	$\alpha=0.3$
0.1109	0.10	1.045577e-06	0.990	0.962	0.936	0.857	0.818	0.736
	0.25	2.613943e-06	0.997	0.987	0.949	0.913	0.862	0.783
	0.50	5.227885e-06	0.999	0.984	0.956	0.927	0.887	0.790
	0.75	7.841828e-06	0.995	0.976	0.950	0.912	0.865	0.767
	0.90	9.410194e-06	0.999	0.987	0.961	0.926	0.880	0.778
	1.00	1.045577e-05	0.996	0.989	0.961	0.947	0.912	0.834
	1.10	1.150135e-05	0.999	0.985	0.971	0.948	0.912	0.829
	1.25	1.306971e-05	0.998	0.979	0.947	0.913	0.842	0.743
	2	2.091154e-05	0.477	0.223	0.117	0.071	0.051	0.032
	4	4.182308e-05	0.000	0.000	0.000	0.000	0.000	0.000
0.2218	10	1.045577e-04	0.000	0.000	0.000	0.000	0.000	0.000
	0.10	3.524835e-06	0.987	0.941	0.886	0.849	0.791	0.704
	0.25	8.812088e-06	0.997	0.956	0.908	0.862	0.826	0.715
	0.50	1.762418e-05	0.996	0.973	0.926	0.903	0.853	0.771
	0.75	2.643626e-05	0.993	0.953	0.913	0.878	0.806	0.713
	0.90	3.172352e-05	0.984	0.955	0.930	0.903	0.869	0.769
	1.00	3.524835e-05	0.995	0.969	0.947	0.905	0.856	0.745
	1.10	3.877319e-05	0.991	0.959	0.941	0.891	0.850	0.743
	1.25	4.406044e-05	0.995	0.968	0.940	0.900	0.860	0.763
	2	7.049670e-05	0.992	0.939	0.883	0.819	0.768	0.691
0.3327	4	1.409934e-04	0.785	0.473	0.390	0.317	0.258	0.171
	10	3.524835e-04	0.012	0.003	0.002	0.000	0.000	0.000
	0.10	5.842118e-06	0.990	0.935	0.883	0.837	0.784	0.676
	0.25	1.460529e-05	0.974	0.935	0.881	0.816	0.769	0.667
	0.50	2.921059e-05	0.991	0.968	0.909	0.865	0.802	0.674
	0.75	4.381588e-05	0.992	0.966	0.927	0.893	0.834	0.722
	0.90	5.257906e-05	0.990	0.956	0.907	0.873	0.830	0.700
	1.00	5.842118e-05	0.991	0.958	0.921	0.874	0.829	0.716
	1.10	6.426330e-05	0.993	0.965	0.929	0.881	0.835	0.720
	1.25	7.302647e-05	0.993	0.955	0.899	0.868	0.818	0.690
0.4436	2	1.168424e-04	0.990	0.955	0.923	0.876	0.824	0.710
	4	2.336847e-04	0.969	0.917	0.854	0.808	0.739	0.647
	10	5.842118e-04	0.732	0.509	0.384	0.312	0.277	0.195
	0.10	7.561251e-06	0.995	0.955	0.903	0.852	0.788	0.708
	0.25	1.890313e-05	0.994	0.954	0.903	0.852	0.805	0.719
	0.50	3.780625e-05	0.990	0.949	0.894	0.857	0.823	0.722
	0.75	5.670938e-05	0.994	0.969	0.921	0.853	0.811	0.712
	0.90	6.805126e-05	0.993	0.962	0.900	0.862	0.839	0.763
	1.00	7.561251e-05	0.991	0.954	0.887	0.848	0.803	0.691
	1.10	8.317376e-05	0.986	0.948	0.899	0.861	0.820	0.715
	1.25	9.451564e-05	0.990	0.952	0.906	0.868	0.830	0.749
	2	1.512250e-04	0.989	0.937	0.884	0.849	0.806	0.715
	4	3.024500e-04	0.989	0.955	0.911	0.851	0.796	0.707
	10	7.561251e-04	0.951	0.875	0.792	0.731	0.673	0.573

Table 11: Confidence interval coverage for transitivity at different bandwidths $c \times a^{(opt)}$ and at different densities ρ_n , based on Monte Carlo simulations using the product generating function when $n = 500$ and $S = 1000$.

ρ_n	c	average $a^{(opt)}$	Proportion of bootstrap CI that cover truth for					
			$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.15$	$\alpha=0.2$	$\alpha=0.3$
0.028125	0.01	1.402691e-08	1.000	0.991	0.979	0.962	0.935	0.854
	0.10	1.402691e-07	1.000	0.988	0.976	0.960	0.931	0.843
	0.25	3.506728e-07	1.000	0.992	0.978	0.965	0.944	0.870
	0.50	7.013456e-07	0.999	0.989	0.970	0.949	0.918	0.842
	0.75	1.052018e-06	0.999	0.945	0.884	0.831	0.778	0.666
	0.90	1.262422e-06	0.986	0.831	0.733	0.570	0.450	0.317
	1.00	1.402691e-06	0.853	0.538	0.370	0.256	0.194	0.102
	1.10	1.542960e-06	0.687	0.325	0.174	0.105	0.064	0.025
	1.25	1.753364e-06	0.355	0.092	0.026	0.008	0.005	0.001
	2	2.805383e-06	0.000	0.000	0.000	0.000	0.000	0.000
	4	5.610765e-06	0.020	0.005	0.001	0.000	0.000	0.000
	10	1.402691e-05	0.853	0.706	0.593	0.519	0.457	0.379
	20	2.805383e-05	0.979	0.931	0.848	0.796	0.746	0.634
	100	1.402691e-04	0.981	0.949	0.887	0.829	0.755	0.658
	0.056250	0.01	6.201768e-08	0.997	0.985	0.964	0.921	0.887
0.10		6.201768e-07	0.997	0.984	0.957	0.929	0.877	0.799
0.25		1.550442e-06	0.998	0.979	0.956	0.926	0.883	0.789
0.50		3.100884e-06	0.997	0.988	0.966	0.931	0.897	0.808
0.75		4.651326e-06	0.997	0.983	0.958	0.916	0.880	0.800
0.90		5.581592e-06	0.996	0.969	0.930	0.856	0.810	0.704
1.00		6.201768e-06	0.985	0.930	0.864	0.806	0.751	0.647
1.10		6.821945e-06	0.986	0.876	0.799	0.720	0.654	0.543
1.25		7.752210e-06	0.939	0.814	0.720	0.610	0.502	0.313
2		1.240354e-05	0.835	0.465	0.268	0.185	0.126	0.067
4		2.480707e-05	0.641	0.363	0.202	0.139	0.111	0.067
10		6.201768e-05	0.767	0.638	0.525	0.466	0.428	0.343
20		1.240354e-04	0.930	0.825	0.771	0.702	0.657	0.577
100		6.201768e-04	0.969	0.871	0.802	0.745	0.698	0.599
0.084375		0.01	1.469518e-07	0.997	0.982	0.947	0.917	0.867
	0.10	1.469518e-06	0.992	0.977	0.947	0.910	0.867	0.764
	0.25	3.673796e-06	0.992	0.977	0.950	0.922	0.886	0.792
	0.50	7.347591e-06	0.995	0.984	0.945	0.909	0.861	0.787
	0.75	1.102139e-05	0.992	0.980	0.953	0.918	0.882	0.774
	0.90	1.322566e-05	0.988	0.959	0.933	0.888	0.835	0.723
	1.00	1.469518e-05	0.992	0.965	0.933	0.900	0.849	0.750
	1.10	1.616470e-05	0.982	0.943	0.906	0.857	0.786	0.695
	1.25	1.836898e-05	0.990	0.953	0.903	0.813	0.772	0.653
	2	2.939036e-05	0.982	0.920	0.819	0.766	0.705	0.593
	4	5.878073e-05	0.965	0.862	0.741	0.666	0.610	0.439
	10	1.469518e-04	0.829	0.668	0.581	0.485	0.412	0.323
	20	2.939036e-04	0.902	0.783	0.722	0.655	0.607	0.520
	100	1.469518e-03	0.944	0.852	0.784	0.718	0.666	0.550
	0.112500	0.01	3.076687e-07	0.997	0.978	0.926	0.885	0.850
0.10		3.076687e-06	0.997	0.961	0.926	0.885	0.845	0.732
0.25		7.691717e-06	1.000	0.967	0.923	0.882	0.847	0.754
0.50		1.538343e-05	0.997	0.971	0.927	0.891	0.849	0.731
0.75		2.307515e-05	0.997	0.964	0.925	0.884	0.844	0.740
0.90		2.769018e-05	0.994	0.966	0.921	0.884	0.849	0.741
1.00		3.076687e-05	0.996	0.954	0.911	0.871	0.823	0.738
1.10		3.384355e-05	0.997	0.960	0.911	0.878	0.809	0.725
1.25		3.845858e-05	0.996	0.949	0.904	0.861	0.792	0.715
2		6.153373e-05	0.992	0.951	0.894	0.850	0.806	0.714
4		1.230675e-04	0.987	0.933	0.866	0.805	0.748	0.657
10		3.076687e-04	0.884	0.729	0.634	0.555	0.499	0.418
20		6.153373e-04	0.843	0.773	0.679	0.604	0.529	0.462
100		3.076687e-03	0.909	0.823	0.760	0.673	0.626	0.525

Table 12: Confidence interval coverage for density at different bandwidths $c \times a^{(opt)}$ and at different densities ρ_n , based on Monte Carlo simulations using the horseshoe generating function when $n = 500$ and $S = 1000$.

ρ_n	c	average $a^{(opt)}$	Proportion of bootstrap CI that cover truth for					
			$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.15$	$\alpha=0.2$	$\alpha=0.3$
0.37950	0.001	9.252546e-08	0.000	0.000	0.000	0.000	0.000	0.000
	0.01	9.252546e-07	0.000	0.000	0.000	0.000	0.000	0.000
	0.1	9.257237e-06	0.000	0.000	0.000	0.000	0.000	0.000
	0.25	2.314309e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	4.628619e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.75	6.942928e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.9	8.331514e-05	0.000	0.000	0.000	0.000	0.000	0.000
	1	9.257237e-05	0.000	0.000	0.000	0.000	0.000	0.000
	1.1	1.018296e-04	0.000	0.000	0.000	0.000	0.000	0.000
	1.25	1.157155e-04	0.000	0.000	0.000	0.000	0.000	0.000
	2	1.851447e-04	1.000	0.990	0.983	0.956	0.920	0.852
	4	3.702895e-04	0.999	0.993	0.981	0.953	0.897	0.820
	10	9.257237e-04	0.999	0.987	0.941	0.896	0.845	0.765
	100	9.252546e-03	0.997	0.959	0.899	0.828	0.773	0.661
1000	9.252546e-02	0.997	0.955	0.899	0.833	0.791	0.668	
0.56925	0.001	1.108253e-07	0.000	0.000	0.000	0.000	0.000	0.000
	0.01	1.108253e-06	0.000	0.000	0.000	0.000	0.000	0.000
	0.1	1.107591e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.25	2.768978e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	5.537956e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.75	8.306935e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.9	9.968322e-05	0.000	0.000	0.000	0.000	0.000	0.000
	1	1.107591e-04	0.000	0.000	0.000	0.000	0.000	0.000
	1.1	1.218350e-04	0.000	0.000	0.000	0.000	0.000	0.000
	1.25	1.384489e-04	0.000	0.000	0.000	0.000	0.000	0.000
	2	2.215183e-04	1.000	0.990	0.974	0.956	0.936	0.853
	4	4.430365e-04	0.998	0.976	0.938	0.889	0.837	0.744
	10	1.107591e-03	0.983	0.931	0.842	0.771	0.705	0.585
	100	1.108253e-02	0.907	0.738	0.505	0.435	0.346	0.259
1000	1.108253e-01	0.885	0.640	0.495	0.437	0.380	0.259	
0.75900	0.001	9.286943e-08	0.000	0.000	0.000	0.000	0.000	0.000
	0.01	9.286943e-07	0.000	0.000	0.000	0.000	0.000	0.000
	0.1	9.277168e-06	0.000	0.000	0.000	0.000	0.000	0.000
	0.25	2.319292e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	4.638584e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.75	6.957876e-05	0.000	0.000	0.000	0.000	0.000	0.000
	0.9	8.349452e-05	0.000	0.000	0.000	0.000	0.000	0.000
	1	9.277168e-05	0.000	0.000	0.000	0.000	0.000	0.000
	1.1	1.020489e-04	0.030	0.001	0.000	0.000	0.000	0.000
	1.25	1.159646e-04	0.860	0.650	0.449	0.303	0.221	0.133
	2	1.855434e-04	0.999	0.982	0.954	0.926	0.879	0.787
	4	3.710867e-04	1.000	0.981	0.952	0.910	0.856	0.755
	10	9.277168e-04	0.982	0.914	0.867	0.793	0.735	0.650
	100	9.286943e-03	0.929	0.865	0.738	0.648	0.595	0.479
1000	9.286943e-02	0.906	0.772	0.659	0.605	0.535	0.444	

Table 13: Confidence interval coverage for λ_{10} at different bandwidths $c \times a^{(opt)}$ and at different densities ρ_n , based on Monte Carlo simulations using the high density generating function when $n = 500$ and $S = 1000$.

B.3 Plots

The linking functions we consider are:

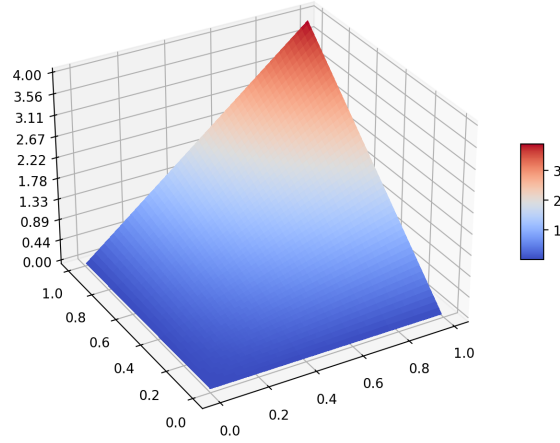


Figure 12: Dot product linking function: $h(\xi_i, \xi_j) = \rho_n \times 4\xi_i\xi_j$.

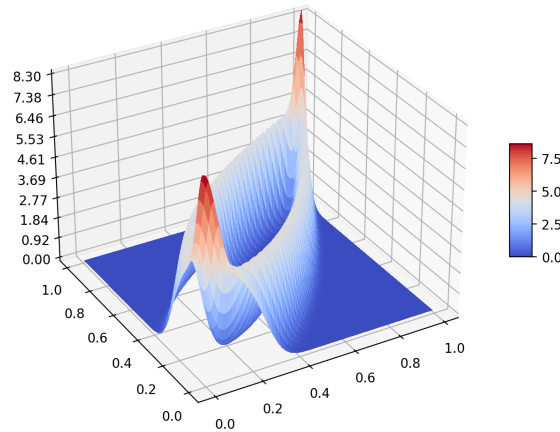


Figure 13: Horseshoe linking function: $h(\xi_i, \xi_j) = \rho_n \times 4.44286 \left(e^{-200(\xi_i - \xi_j^2)^2} + e^{-200(\xi_j - \xi_i^2)^2} \right)$.

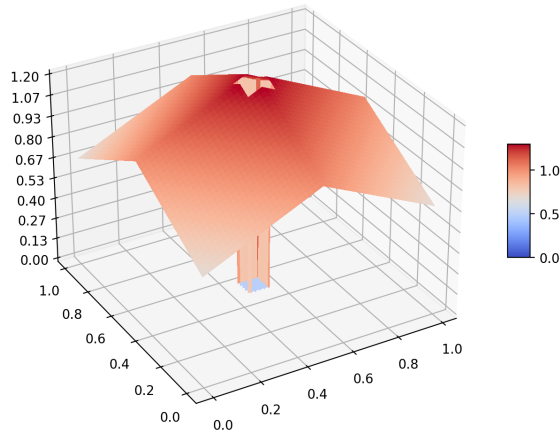


Figure 14: High-density linking function:

$$h(\xi_i, \xi_j) = \rho_n \times 1.35 (1 - \mathbb{1}(|\frac{1}{2} - \xi_i| \leq 0.05)) \mathbb{1}(|\frac{1}{2} - \xi_j| \leq 0.05)) (1 - \frac{1}{2} (|\frac{1}{2} - \xi_i| + |\frac{1}{2} - \xi_j|)).$$

C Extensions and alternative specifications

C.1 Possible extensions of application

On a technical level, one possible extension would be analysing different ways of aggregating the twelve observed types of household interactions into the adjacency matrix. Like the original paper, we have used the union of the twelve characteristic-specific adjacency matrices, but there are many other possible choices, e.g. taking an intersection (this may be less desirable as it leads to a significantly sparser network) or an average (which gives a weighted adjacency matrix). Our method allows for comparison of the adjacency matrices achieved through different aggregating functions and checking if they lead to different structures. For example, we can compare the largest eigenvalues λ_1 obtained for villages using different aggregating functions and check if they have overlapping confidence intervals. If they don't, this shows that the choice of the aggregating function is not without loss of generality.

Another possible modification of our procedure would be estimating the linking function under the assumption that each of the twelve characteristics is a separate draw from the Bernoulli distribution. We could redefine the distance function to depend directly on the twelve characteristics instead of a single aggregate adjacency matrix. Let \mathbf{A}_{ij} denote a 12×1 vector of indicators whether households i and j are related according to the twelve characteristics. Let $\|\cdot\|$ be some vector norm

(e.g. max norm, min norm, Euclidean norm, or a weighted norm³⁵). Then we can define

$$d_{ij}^{(\|\cdot\|,2)} = \left(\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{s=1}^n \|\text{diag}(\mathbf{A}_{ts})(\mathbf{A}_{is} - \mathbf{A}_{js})\| \right)^2 \right)^{\frac{1}{2}}$$

where $\text{diag}(v)$ is a diagonal matrix with diagonal entries from a vector v . To obtain the bootstrap version of adjacency matrices \mathbf{A} we could draw from a joint Bernoulli distribution with probabilities estimated using \hat{h}_n based on the distance $d_{ij}^{(\|\cdot\|,2)}$ and the adjacency matrices for individual characteristics, and a covariance matrix equal to the sample covariance between different characteristics.

C.2 Sensitivity checks

We rerun the estimation for a subset of villages using 300 repetitions of the simulated information spreading through the network to estimate the simulated moments instead of the original 75. The outcomes (middle panel in Fig. 15) for that subsample were very similar to the outcomes based on the original specification (left panel in Fig. 15). We conclude that 75 simulations are sufficient for the estimation of simulated moments.

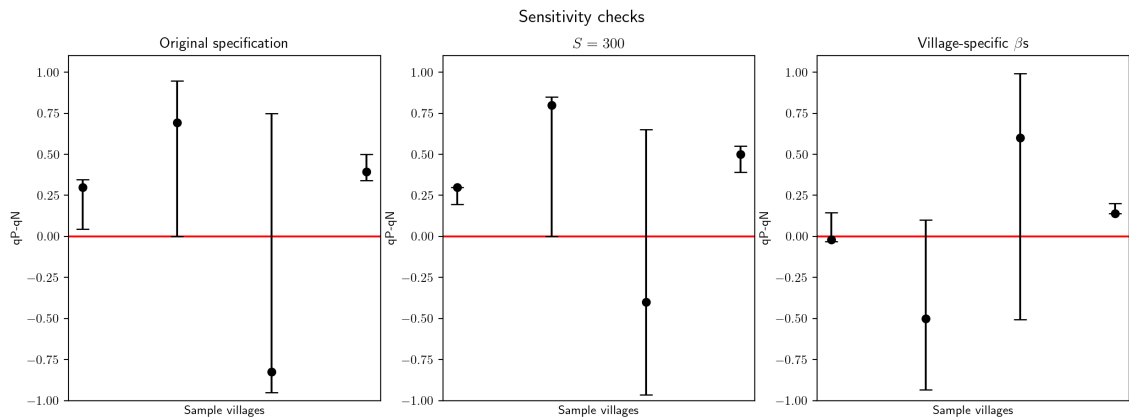


Figure 15: A comparison of the estimates of $q^P - q^N$ for a subset of villages with 95% confidence intervals based simulated moments estimated using the original specification with 75 simulations and β estimated using all villages (left), 300 simulations and β estimated using all villages (middle), and 75 simulations and β estimated using village-specific data only (right).

We also tried to estimate the β coefficients using village-specific data (rather than aggregating over all villages). This did make a difference for the estimates and confidence intervals (right panel of Fig. 15), though the conclusions remain similar. However, since the regression used to identify β is run using only the information about the leaders, we found the sample sizes for individual villages

³⁵. A weighted norm could be of the form $\|x\| = \sqrt{x'W_x x}$, with a weight matrix W_x that may depend on the adjacency matrices, e.g. $W_x = \left(\frac{1}{n} \sum_{v=1}^n x_v x_v'\right)^{-1}$.

too small to give reliable estimates. Hence we chose to use aggregate β in our main simulations.

C.3 Alternative bootstrap procedure: links only

Instead of the procedure we use in the main paper, we could skip the step 2. of resampling nodes from the original graph and go straight into resampling links according to \hat{h}_n for the original set of nodes. The motivation for this procedure is similar to the one we use: the original sample comes from the true data generating distribution, we have a good estimate for the distribution of the adjacency matrix, hence the networks simulated this way should preserve the structure of the original network. Skipping one step in the simulation simplifies the procedure, improves computational time and would also simplify the proofs. We run some simulations using this method and found that, unfortunately, it does not perform as well as our main approach. Table Table 14 shows the results of some of our simulations. We see that the confidence interval coverage is very poor, other than for a few special cases where the sample size is small ($n = 25$, in which case the bias is small relative to variance and the true value may still be included in the confidence interval) or we are estimating a statistic which is relatively tricky to estimate (e.g. λ_2) and hence measured with more variation than e.g. λ_1 or density.

Our hypothesis for why the performance is so poor is that if the bandwidth a_n is small, or if for some individuals there are no close neighbours, our procedure becomes similar to the empirical bootstrap of Green and Shalizi (2022): we draw a link between two individuals if and only if they were linked in the original graph. Hence if we don't resample individuals, the bootstrapped graph may become too similar to the original graph, at least on the subgraphs consisting of individuals with few neighbours. This can lead to insufficient variation in the bootstrapped graphs and worse performance of the bootstrap procedure. Fig. 16 and Fig. 17 shows a comparison of our original method (in blue) and the version which only resamples links while keeping the original nodes (in orange). While both do a good job of replicating the statistic values in the bootstrapped graph (dashed black line), the version which only resamples links is too concentrated around the value in the bootstrapped graph and often misses the population value of the statistic (red solid line), leading to poor confidence interval coverage of the links-only procedure.

This indicates that if one is interested in uncovering the population values, our main procedure is more reliable. However, in applications where we are only interested in confidence intervals for a specific sample, bootstrapping links only does provide narrower confidence intervals and would be preferred.

generating function	n	ρ_n	95% CI coverage for				
			density	transitivity	λ_1	λ_2	
high density	25	0.379500	0.803	0.848	0.791	0.978	
		0.569250	0.803	0.788	0.776	0.979	
		0.759000	0.755	0.730	0.752	0.973	
	100	0.379500	0.541	0.537	0.495	0.824	
		0.569250	0.509	0.554	0.528	0.925	
		0.759000	0.532	0.518	0.529	0.963	
	300	0.379500	0.168	0.144	0.156	0.662	
		0.569250	0.211	0.222	0.227	0.824	
		0.759000	0.340	0.337	0.339	0.947	
	500	0.379500	0.068	0.059	0.058	0.410	
		0.569250	0.087	0.110	0.092	0.673	
		0.759000	0.251	0.282	0.260	0.950	
	horseshoe	25	0.056250	0.727	0.920	0.838	0.956
			0.084375	0.776	0.910	0.770	0.936
			0.112500	0.745	0.757	0.715	0.916
100		0.056250	0.530	0.520	0.406	0.819	
		0.084375	0.616	0.419	0.387	0.717	
		0.112500	0.686	0.365	0.359	0.679	
200		0.056250	0.430	0.344	0.243	0.659	
		0.084375	0.585	0.308	0.302	0.606	
		0.112500	0.592	0.270	0.287	0.501	
300		0.056250	0.433	0.285	0.215	0.575	
		0.084375	0.589	0.266	0.227	0.504	
		0.112500	0.568	0.237	0.240	0.460	
500		0.056250	0.407	0.213	0.158	0.498	
		0.084375	0.530	0.215	0.221	0.413	
		0.112500	0.490	0.217	0.218	0.326	
product	25	0.125000	0.562	0.907	0.652	0.911	
		0.250000	0.484	0.785	0.513	0.963	
	100	0.125000	0.331	0.567	0.366	0.887	
		0.250000	0.263	0.423	0.298	0.957	
	300	0.125000	0.203	0.253	0.164	0.867	
		0.250000	0.124	0.177	0.136	0.914	
	500	0.125000	0.156	0.165	0.140	0.836	
		0.250000	0.089	0.126	0.090	0.865	

Table 14: 95% confidence interval coverage for density, transitivity, λ_1 and λ_2 for different generating functions, different sample sizes n from 25 to 500 and at different densities ρ_n , based on Monte Carlo simulations using a version of the algorithm which keeps the original set of nodes and only resamples the links between them.

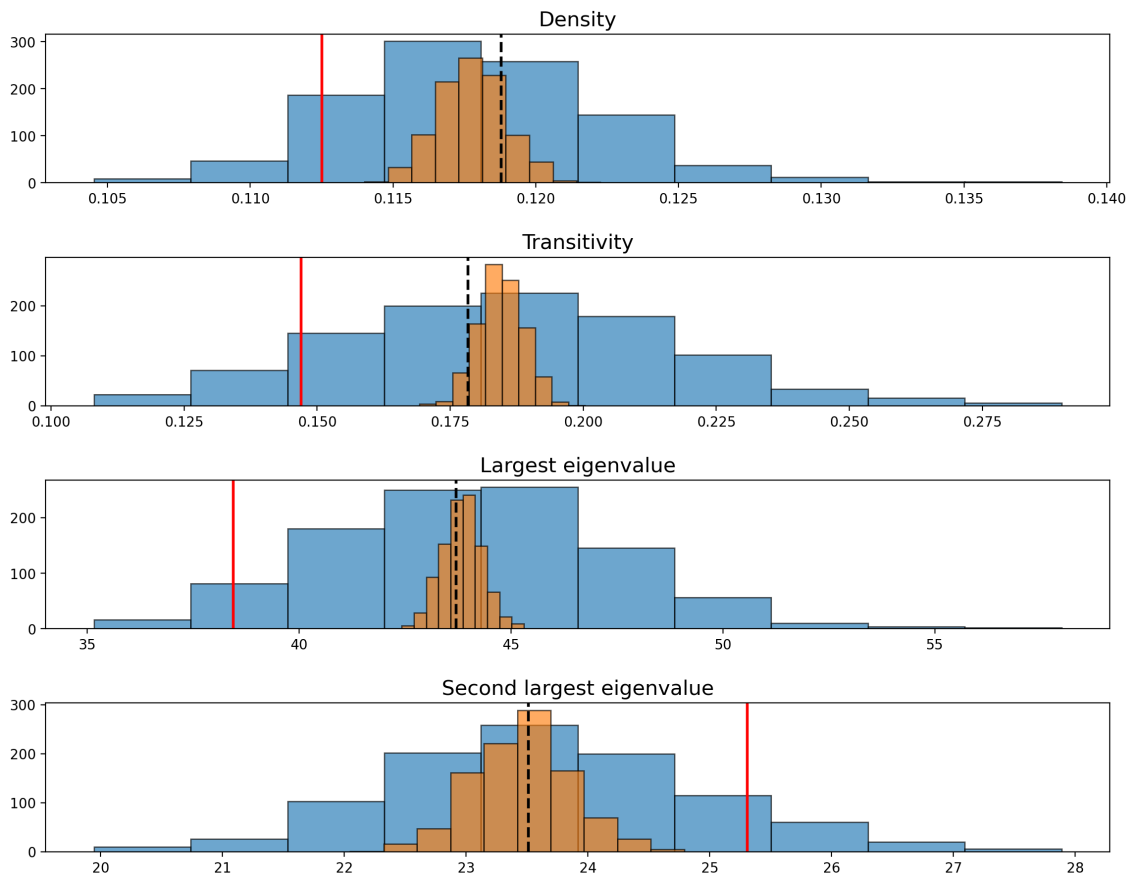


Figure 16: Comparison of histograms for our main bootstrap approach (in blue) and for the links-only version (in orange) from estimation of one specific network from the horseshoe generating function with $n = 300$ and $\rho_n = 0.1125$. The estimated statistics are, from top to bottom: density, transitivity, λ_1 and λ_2 . The red solid line denotes the population true value of the statistic while the black dashed line denotes the value of the statistic in the bootstrapped graph.

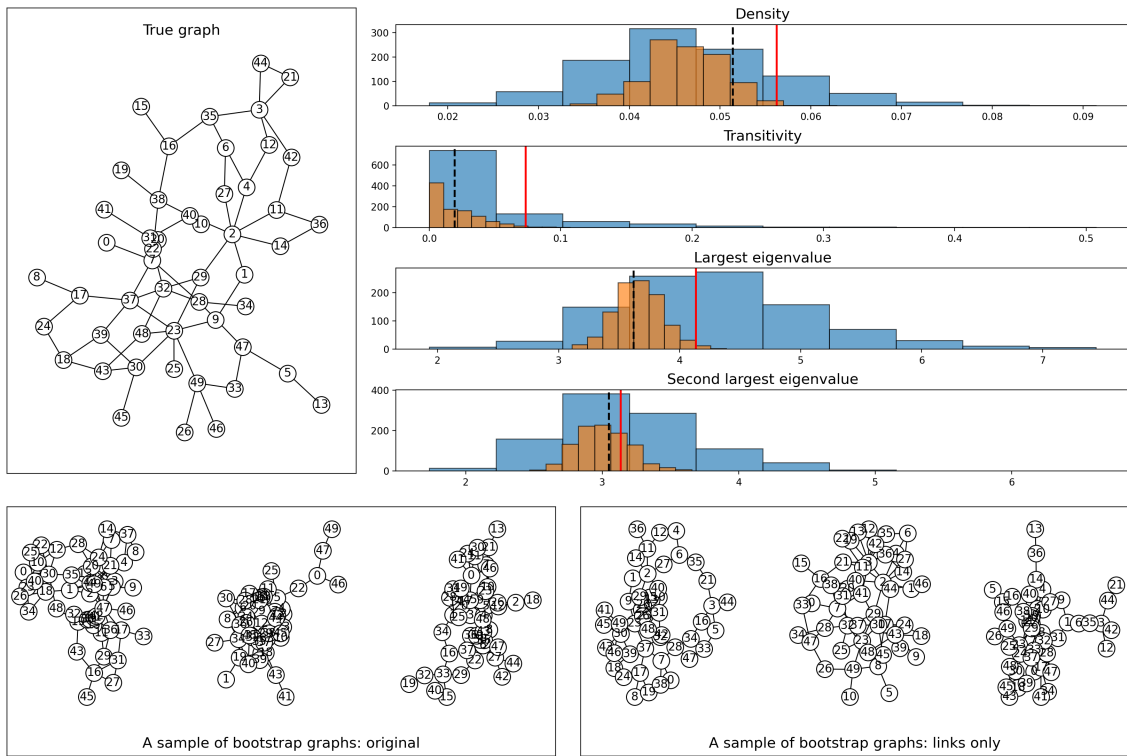


Figure 17: Comparison of histograms for our main bootstrap approach (in blue) and for the links-only version (in orange) from estimation of one specific network from the horseshoe generating function with $n = 50$ and $\rho_n = 0.05625$. The estimated statistics are, from top to bottom: density, transitivity, λ_1 and λ_2 . The red solid line denotes the population true value of the statistic while the black dashed line denotes the value of the statistic in the bootstrapped graph.