Supplementary Material (Online-Only Publication) for: "Inference in Linear Dyadic Data Models with Network Spillovers"

Nathan Canen and Ko Sugiura

A Mathematical Set-up

This section lays out the mathematical set-up of our model in more detail, heavily drawing from Kojevnikov et al. (2021). We conclude with a discussion of the related statistical literature.

We first define a collection of pairs of sets of dyads. For any positive integers a, b and s, define

$$\mathcal{P}_M(a,b;s) \coloneqq \{(A,B) : A,B \subset \mathcal{M}_N, |A| = a, |B| = b, \rho_M(A,B) \geq s\},$$

where

$$\rho_M(A,B) \coloneqq \min_{m\in A} \min_{m'\in B} \rho_M(m,m'), \tag{19}$$

with $\rho_M(m,m')$ denoting the geodesic distance between dyads m and m', i.e., the smallest number of adjacent dyads between dyads m and m'. In words, the set $\mathcal{P}_M(a,b;s)$ collects all two distinct sets of active dyads whose sizes are a and b and that have no dyads in common.

Next we consider a collection of bounded Lipschitz functions. Define

$$\mathcal{L}_K \coloneqq \{\mathcal{L}_{K,c} : c \in \mathbb{N}\},$$

where

$$\mathcal{L}_{K,c} \coloneqq \{f : \mathbb{R}^{K\times c} \to \mathbb{R} : \|f\|_\infty < \infty, \mathrm{Lip}(f) < \infty\},$$

with $\|\cdot\|_\infty$ representing the supremum norm and $\mathrm{Lip}(f)$ being the Lipschitz constant.[22] In words, the set $\mathcal{L}_{K,c}$ collects all the bounded Lipschitz functions on $\mathbb{R}^{K\times c}$ and the set $\mathcal{L}_K$ moreover gathers such sets with respect to $c \in \mathbb{N}$.

[22] It is immediate to see that $\mathbb{R}$ is a normed space with respect to the Euclidean norm, while the $\mathbb{R}^{K\times c}$ can be equipped with the norm $\rho_c(x,y) \coloneqq \sum_{\ell=1}^{c} \|x_\ell - y_\ell\|$ where $x,y \in \mathbb{R}^{K\times c}$ and $\|z\| \coloneqq (z'z)^{\frac{1}{2}}$, thereby the Lipschitz constant is defined as $\mathrm{Lip}(f) \coloneqq \min\{w \in \mathbb{R} : |f(x) - f(y)| \leq w\rho_c(x,y)\ \forall x,y \in \mathbb{R}^{K\times c}\}$.

Lastly, we write

$$Y_{M,A} \coloneqq (Y_{M,m})_{m \in A},$$

and $Y_{M,B}$ is analogously defined. Let $\{\mathcal{C}_M\}_{M\geq 1}$ denote a sequence of σ-algebras and be suppressed as $\{\mathcal{C}_M\}$.

The network dependent random variables are characterized by the upper bound of their covariances, first defined in Definition 2.2 of Kojevnikov et al. (2021).

Definition A.1 (Conditional ψ-Dependence given $\{\mathcal{C}_M\}$). *A triangular array $\{Y_{M,m} \in \mathbb{R}^K : M \geq 1, m \in \{1, \ldots, M\}\}$ is called conditionally ψ-dependent given $\{\mathcal{C}_M\}$, if for each $M \in \mathbb{N}$, there exist a $\mathcal{C}_M$-measurable sequence $\theta_M \coloneqq \{\theta_{M,s}\}_{s\geq 0}$ with $\theta_{M,0} = 1$, and a collection of nonrandom function $(\psi_{a,b})_{a,b\in\mathbb{N}}$ where $\psi_{a,b} : \mathcal{L}_{K,a} \times \mathcal{L}_{K,b} \to [0.\infty)$, such that for all $(A, B) \in \mathcal{P}_M(a, b; s)$ with $s > 0$ and all $f \in \mathcal{L}_{K,a}$ and $g \in \mathcal{L}_{K,b}$,*

$$\left|Cov\big(f(Y_{M,A}), g(Y_{M,B}) \mid \mathcal{C}_M\big)\right| \leq \psi_{a,b}(f, g)\theta_{M,s} \qquad a.s.$$

Intuitively, this definition states that the upper bound must be decomposed into two components. The first part $\psi_{a,b}(f, g)$ is deterministic and depends on nonlinear Lipschitz functions f and g. The other component $\theta_{M,s}$ is stochastic and depends only on the distance of the random variables on the underlying network. The former, nonrandom component reflects the scaling of the random variables as well as that of the Lipschitz transformations, while the latter random part stands for the covariability of the two random variables. We call $\theta_{M,s}$ the dependence coefficient. We follow Kojevnikov et al. (2021) in assuming boundedness for these two components.

Assumption A.1 (Kojevnikov et al. (2021), Assumption 2.1). *The triangular array $\{Y_{M,m} \in \mathbb{R}^K : M \geq 1, m \in \{1, \ldots, M\}\}$ is conditionally ψ-dependent given $\{\mathcal{C}\}$ with the dependence coefficients $\{\theta_{M,s}\}$ satisfying the following conditions: (a) there exists a constant $C > 0$ such that $\psi_{a,b}(f, g) \leq C \times ab\big(\|f\|_\infty + Lip(f)\big)\big(\|g\|_\infty + Lip(g)\big)$; (b) $\sup_{M\geq 1} \max_{s\geq 1} \theta_{M,s} < \infty$ a.s.*

Assumption A.1 is maintained throughout the paper and employed to show asymptotic properties of our estimators such as the consistency and asymptotic normality, and the consistency of the network-robust variance estimator for dyadic data.

A.1 Additional Discussion of Assumption 3.1

Assumption 3.1 assumed that $M \to \infty$ as $N \to \infty$. This is consistent with many applications. For example, in international trade, the entry of a new country/firm to a market will most likely increase the number of trade flows in the economy; in political economy, the more members of parliaments (MEPs) there are, the more pairs of the MEPs sitting next to each other there will be (see Section 5).

This assumption is similar in spirit to Assumption 2.3 of Tabord-Meehan (2019) in which the minimum degree is assumed to grow at some constant rate relative to the number of individuals. It is milder than Assumption 2.3 of Tabord-Meehan (2019) since the latter does not allow any individual to be isolated, while Assumption 3.1 merely constrains the average degree. Similarly, this assumption is weaker than the assumption that the maximum degree in a network is bounded even when $N \to \infty$ (e.g., Penrose and Yukich (2003) and de Paula et al. (2018)).

A.2 Related Literature

As stated in the main text, our paper is related to the recent literature on inference for multiway clustering, whether OLS estimation with multiway clustering (Cameron et al. (2011)); a clustering method in high-dimensional set-ups (Chiang et al. (2021)); clustering within the time dimension (Chiang et al. (2022)); clustered inference with empirical likelihood (Chiang et al. (2022)); bootstrap methods in multiway clustering (Davezies et al., 2021; Menzel, 2021; MacKinnon et al., 2022b); clustering in the context of average treatment effects (Abadie et al. (2022)), to name but a few (see MacKinnon et al. (2022a) for review). Again, one of the common assumptions in this literature is that individual observations are divided into disjoint groups – clusters – and observations in different clusters are not correlated. To that end, MacKinnon et al. (2022c) propose measures of cluster-level influence that can be used to assess whether the underlying assumption of cluster-robust variance estimation is satisfied.

Our approach complements the existing methods similar to how inference with spatial data (e.g., Conley (1999) and Jenish and Prucha (2009)) complements one-way clustering inference. Our approach still differs from such inference with spatial data, since the latter routinely assumes the index set to be a Euclidean metric space, whose metric relies solely on the nature of the space, and uses it to define the dependence between variables. See also Ibragimov and Müller (2010) and references therein. In our context, however, the index set of dyads alone does not suffice to dictate the dependence structure because indices themselves do not inform us of the network topology. Instead, we first introduce a metric on a network among dyads and our mixing condition is based on dependence as dyads grow further apart along the network.

Our main insight in accommodating indirect spillovers is that we can rewrite the correlation structure among dyads as a dyadic network, where links denote whether they share a common member. As a result, this dyadic network describes how close/far certain dyads are from sharing members with other dyads. In doing so, the transformed problem is amenable to appropriate applications of recent developments in the statistics of random variables which are correlated along an (observable, exogenous) network. In particular, we apply asymptotic results for network-dependent random variables developed by Kojevnikov et al. (2021)[23] to an appropriately defined dyadic network, with assumptions imposed on the latter. Leung (2021) and Leung (2022) also apply the framework of Kojevnikov et al. (2021) to study, respectively, cluster-robust inference and causal inference for the case of individual-specific random variables. These papers focus on the correlation along a network over individuals, rather than over dyads. Meanwhile, Leung and Moon (2021) derive an asymptotic theory for dyadic variables in the context of networks, primarily for endogenous network formation models.

B Proofs of Main Theorems and Results

B.1 Identification of β

Assumption B.1. *For each $N \in \mathbb{N}$:*
(a) $\sup_{m\in\mathcal{M}_N} E\big[|\varepsilon_{M,m}|^2\big]$ *exists and is finite;*
(b) $\sup_{m\in\mathcal{M}_N} E\big[\|x_{M,m}\|\big]$ *exists and is finite;*
(c) $E\big[x_{M,m}x'_{M,m}\big]$ *exists with finite elements and positive definite for all* $m \in \mathcal{M}_N$*;*
(d) $E\big[\varepsilon_{M,m} \mid X_M\big] = 0$ *for all* $m \in \mathcal{M}_N$*.*

Assumption B.1 (a) and (b) are standard and jointly imply the finite existence of the second moment of $y_{M,m}$ for all $m \in \mathcal{M}_N$, which in turn implies the finite existence of the cross moment of $y_{M,m}$ and $x_{M,m}$ for all $m \in \mathcal{M}_N$. The third and fourth assumptions are also standard in the context of the linear regression models and require no multicolinearity and strict exogeneity, respectively.

Identification of the linear parameter in equation (1) follows from Assumption B.1 (see Proposition B.1 in Appendix B.1).

Proposition B.1 (Identification)**.** *Under Assumption B.1, the regression parameter β in (1) is identified.*

Proof. For each $m \in \mathcal{M}_N$, premultiply the model (1) by $x_{M,m}$ to obtain

$$x_{M,m}y_{M,m} = x_{M,m}x'_{M,m}\beta + x_{M,m}\varepsilon_{M,m} \qquad \forall m \in \mathcal{M}_N.$$

[23] Vainora (2020) provides another such theoretical contribution.

Taking the expectation with respect to $\{(x_{M,m}, y_{M,m}, \varepsilon_{M,m})\}_{m\in\mathcal{M}_N}$ implies:

$$E\Big[x_{M,m}y_{M,m}\Big] = E\Big[x_{M,m}x'_{M,m}\Big]\beta + E\Big[x_{M,m}\varepsilon_{M,m}\Big].$$

The second term on the right hand side is equal to 0, due to Assumption B.1 (d). Next, Assumption B.1 (c) ensures existence of the inverse of the expectation term in the first term of the right hand side, ensuring identification. □

B.2 Consistency of $\hat{\beta}$

As usual, the Central Limit Theorem for a normalized sum requires us to have stronger conditions than what is required for consistency. Those stronger conditions were introduced in the main text as Assumptions 3.2 and 3.3. However, they are only required for Theorem 3.2. For the consistency proof (Theorem 3.1),we can replace those two assumptions by the following weaker conditions.

Assumption B.2. *There exists* $\eta > 0$ *such that* $\sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\Big[|\varepsilon_{M,m}|^{1+\eta} \mid \mathcal{C}_M\Big] < \infty$.

Assumption B.2 allows for the same interpretation as Assumption 3.2, i.e., the random error term ε_m cannot be too large, conditional on a common component. This assumption, however, is less stringent than the previous one because it now requires the finiteness of a lower moment of ε_m.

Assumption B.3. $\frac{1}{M}\sum_{s\geq 1}\delta_M^{\partial}(s;1)\theta_{M,s} \overset{a.s.}{\to} 0$ *as* $M \to \infty$.

Similar to Assumption 3.3, this assumption binds the covariance of the random variables, the dependence reflected in the dependence coefficients, and the underlying network. That is, given σ_M growing at least at the same rate of M, the composite of the density of the network and the magnitude of the correlations of the random variables must decay fast enough.

Proof of Theorem 3.1: From (8), (12) and (13), we can write

$$\begin{aligned}\hat{\beta} - \beta &= \Big(\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\Big)^{-1}\sum_{m\in\mathcal{M}_N} x_{M,m}\varepsilon_{M,m}\\ &= \Big(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\Big)^{-1}\frac{1}{M}\sum_{m\in\mathcal{M}_N} Y_{M,m}\\ &= \frac{1}{M}\sum_{m\in\mathcal{M}_N}\Big(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\Big)^{-1}Y_{M,m}.\end{aligned}$$

Define $\tilde{Y}_{M,m} := \left(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\right)^{-1} Y_{M,m}$ and let $\tilde{Y}^u_{M,m}$ be the u-th entry of $\tilde{Y}_{M,m}$. That is,

$$\begin{aligned}\tilde{Y}^u_{M,m} &= D^u Y_{M,m} \\ &= D^u x_{M,m}\varepsilon_{M,m},\end{aligned}$$

where D^u stands for the u-th row of the matrix $\left(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\right)^{-1}$. Moreover, let $\hat{\beta}^u$ and β^u, respectively, denote the u-th entry of $\hat{\beta}$ and β, so that we can write

$$\hat{\beta}^u - \beta^u = \frac{1}{M}\sum_{m\in\mathcal{M}_N}\tilde{Y}^u_{M,m},$$

In light of Assumption B.1 (d), it holds that for any $N > 0$ and for each $m \in \mathcal{M}_N$

$$\begin{aligned}E\big[\tilde{Y}^u_{M,m} \mid \mathcal{C}_M\big] &= D^u x_{M,m}\underbrace{E\big[\varepsilon_{M,m} \mid \mathcal{C}_M\big]}_{0} \\ &= 0.\end{aligned}$$

By Theorem 3.1 of Kojevnikov et al. (2021), $\left\|\frac{1}{M}\sum_{m\in\mathcal{M}_N}\left(\tilde{Y}^u_{M,m} - \underbrace{E\big[\tilde{Y}^u_{M,m} \mid \mathcal{C}_M\big]}_{0}\right)\right\|_{\mathcal{C}_M,1} \overset{a.s.}{\to} 0.$

Hence,

$$\left\|\frac{1}{M}\sum_{m\in\mathcal{M}_N}\tilde{Y}^u_{M,m}\right\|_{\mathcal{C}_M,1} \overset{a.s.}{\to} 0 \qquad M \to \infty,$$

so that

$$\begin{aligned}E\big[|\hat{\beta}^u - \beta^u|\big] &= E\Big[E\big[|\hat{\beta}^u - \beta^u| \mid \mathcal{C}_M\big]\Big] \\ &= E\Big[\|\hat{\beta}^u - \beta^u\|_{\mathcal{C}_M,1}\Big] \\ &= E\Big[\Big\|\frac{1}{M}\sum_{m\in\mathcal{M}_N}\tilde{Y}^u_{M,m}\Big\|_{\mathcal{C}_M,1}\Big] \\ &\to 0 \qquad M \to \infty,\end{aligned}$$

where the last implication is a consequence of the Dominated Convergence Theorem. In view of Assumption 3.1, this is true also with respect to N going to infinity.

Since it holds by the Markov inequality that for any $c > 0$

$$\Pr\big(|\hat{\beta}^u - \beta^u| > c\big) \leq \frac{E\big[|\hat{\beta}^u - \beta^u|\big]}{c},$$

it then follows that

$$\Pr\big(|\hat{\beta}^u - \beta^u| > c\big) \to 0,$$

as $N \to \infty$. Hence we have

$$|\hat{\beta}^u - \beta^u| \xrightarrow{p} 0 \qquad as \quad N \to \infty.$$

Finally, we can invoke the Cramér-Wold device to obtain

$$\|\hat{\beta} - \beta\|_2 \xrightarrow{p} 0 \qquad as \quad N \to \infty,$$

as desired. □

B.3 Lemma

Here we establish a lemma that is used repeatedly throughout the subsequent proofs in this paper.

Lemma B.1. *Define* $A := \lim_{N\to\infty} \frac{1}{M} \sum_{k\in\mathcal{M}} E\Big[x_{M,k}x'_{M,k}\Big]$ *and assume that Assumptions 3.1 and 3.5 hold.*

(i) $A^{-1} := \lim_{N\to\infty} \Big(\frac{1}{M} \sum_{k\in\mathcal{M}} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}$ *exists with finite elements and positive definite.*

(ii) Suppose, moreover, that Assumption 3.7 holds. Then, $\Big\|\Big(\frac{1}{M} \sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1} - \Big(\frac{1}{M} \sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}\Big\|_F \xrightarrow{p} 0.$

Proof. (i) The fact that it is positive definite follows from Assumption 3.5. The fact that the elements are finite is proved by considering element-by-element convergence. Let $x_{k,i}$ denote the i-th element of $x_{M,k}$. Then the (i,j) entry of $\frac{1}{M} \sum_{k\in\mathcal{M}} E\big[x_{M,k}x'_{M,k}\big]$ is given by: $\frac{1}{M} \sum_{k\in\mathcal{M}} E\big[x_{k,i}x_{k,j}\big]$.

We write the (i,j) entry of A as $A_{i,j}$.

From Assumption 3.5, there exists a nonnegative finite constant $C_{0,1}$ such that

$$C_{0,1} = \sup_{N\geq 1} \max_{m\in\mathcal{M}_N} E\big[x_{m,i}x_{m,j}\big],$$

so that

$$\begin{aligned}
A_{i,j} &= \lim_{N\to\infty} \frac{1}{M} \sum_{k\in\mathcal{M}_N} \underbrace{E\big[x_{k,i}x_{k,j}\big]}_{\le C_{0,1}} \\
&\le \lim_{N\to\infty} \frac{1}{M} \sum_{k\in\mathcal{M}_N} C_{0,1} \\
&= C_{0,1} \lim_{N\to\infty} \frac{1}{M} \underbrace{\sum_{k\in\mathcal{M}_N} 1}_{M} \\
&= C_{0,1}.
\end{aligned}$$

Hence $A_{i,j}$ exists with being finite. By repeating the same argument for all $i, j = 1, \ldots, K$, it holds that A exists with finite elements.

(ii) To begin with, observe that

$$\begin{aligned}
&\Big\|\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1} - \Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}\Big\|_F \\
&= \Big\|\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1} - A^{-1} + A^{-1} - \Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}\Big\|_F \\
&\le \Big\|\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1} - A^{-1}\Big\|_F + \Big\|A^{-1} - \Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}\Big\|_F.
\end{aligned}$$

Note that convergence of the second term follows from (i). Hence, we wish to prove that:

$$\Big\|\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k} - A\Big\|_F \xrightarrow{p} 0$$

To do so, we follow a strategy employed in Aronow et al. (2015) and Tabord-Meehan (2019). In light of (i), it remains to show

$$Var\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big) \to 0.$$

As in (i), we consider the element-by-element convergence, using the same notation. The variance can be expressed as a sum of covariances:

$$Var\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\Big) = \frac{1}{M^2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} Cov\big(x_{m,i}x_{m,j}, x_{m',i}x_{m',j}\big)$$
$$= \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} Cov\big(x_{m,i}x_{m,j}, x_{m',i}x_{m',j}\big).$$

Again from Assumption 3.5, there exists a nonnegative finite constant $C_{0,2}$ such that

$$C_{0,2} = \sup_{N\geq 1}\max_{m,m'\in\mathcal{M}_N} Cov\left(x_{m,i}x_{m,j}, x_{m',i}x_{m',j}\right).$$

Hence,

$$\begin{aligned}
Var\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\Big) &\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} C_{0,2} \\
&= \frac{C_{0,2}}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} 1 \\
&= \frac{C_{0,2}}{M^2}\sum_{s\geq 0} M\delta_M^{\partial}(s;1) \\
&= C_{0,2}\underbrace{\frac{1}{M}\sum_{s\geq 0}\delta_M^{\partial}(s;1)}_{\to 0} \\
&\to 0,
\end{aligned}$$

where the last implication is due to Assumption 3.7. By repeating the same argument for all $i, j = 1, \ldots, K$, we obtain

$$Var\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big) \to 0.$$

Now, by the Chebyshev's inequality, we arrive at

$$\Big\|\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k} - A\Big\|_F \xrightarrow{p} 0.$$

Furthermore, applying the Continuous Mapping Theorem yields

$$\left\| \Big(\frac{1}{M} \sum_{k \in \mathcal{M}_N} x_{M,k} x'_{M,k} \Big)^{-1} - A^{-1} \right\|_F \xrightarrow{p} 0,$$

obtaining the result. Therefore,

$$\left\| \Big(\frac{1}{M} \sum_{k \in \mathcal{M}_N} x_{M,k} x'_{M,k} \Big)^{-1} - \Big(\frac{1}{M} \sum_{k \in \mathcal{M}_N} E\big[x_{M,k} x'_{M,k} \big] \Big)^{-1} \right\|_F \xrightarrow{p} 0,$$

as desired. □

B.4 Asymptotic Normality of $\hat{\beta}$

In this subsection, we prove Theorem 3.2 under a slightly milder condition than Assumption 3.4.

Assumption B.4 (Growth Rates of Variances)**.** *There exists a sequence of (possibly random) positive numbers,* $\{\pi_{N,M}\}_{N>0}$*, such that*

$$\frac{\sigma_M^2}{\pi_{N,M} \tau_M^2} \xrightarrow{a.s.} 1 \qquad as \quad N \to \infty.$$

When $\pi_{N,M} = 1$, this assumption simplifies to Assumption 3.4, which is used for the results in the main text.

For our proof of the asymptotic distribution of $\hat{\beta}$, we require that its asymptotic variance is well-defined. The first assumption, Assumption 3.5(a)-(b), is necessary for one of the matrices in the expression to be well-defined.[24] Part (c) assures that the middle part of the asymptotic variance is non-trivial.

When Assumption 3.4 is replaced by Assumption B.4, Assumption 3.5 must also be modified accordingly.

Assumption B.5. $\lim_{N \to \infty} \frac{N \pi_{N,M}}{M^2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N} E\big[\varepsilon_{M,m} \varepsilon_{M,m'} x_{M,m} x'_{M,m'} \big]$ *exists with finite elements.*

An important comparison of Assumption B.5 can be made to the variety of assumptions used in the literature.

[24] As pointed out in Tabord-Meehan (2019), the bounded support assumption can be relaxed by imposing an alternative condition on higher-order moments (boundedness of the 16th order moment of $x_{M,m}$, in our case).

Remark B.1. *The requirement on the behavior of $AVar(\hat{\beta})$ mirrors Assumptions 2.4, 2.5 and 2.6 of Tabord-Meehan (2019): Assumption B.5 boils down to his Assumption 2.4, if it is well-defined with $\pi_{N,M} = \frac{M}{N}$; it reduces to his Assumption 2.5, if it is compatible with $\pi_{N,M} = \frac{M}{N^2}$; and it coincides with Assumption 2.6, if it is maintained with $\pi_{N,M} = \frac{M}{N^{r+1}}$ for $r \in [0,1]$. Moreover, if $AVar(\hat{\beta})$ is well-defined for $\pi_{N,M} = 1$, the expression (15) simplifies to the assumption that appears in Lemma 1 of Aronow et al. (2015).*

Proof of Theorem 3.2: From (13),

$$\begin{aligned}\sqrt{N}(\hat{\beta}-\beta) &= \Big(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\Big)^{-1}\frac{\sqrt{N}}{M}\sum_{m\in\mathcal{M}_N} Y_{M,m}\\ &= \Big(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\Big)^{-1}\frac{\sqrt{N}}{M}S_M.\end{aligned}$$

Since $\big(\frac{1}{M}\sum_{j\in\mathcal{M}_N} x_{M,j}x'_{M,j}\big)^{-1}$ converges to a well-defined limit (Lemma B.1), the asymptotic distribution of $\sqrt{N}(\hat{\beta}-\beta)$ is dictated by that of $\frac{\sqrt{N}}{M}S_M$.

First of all, we prove

$$\frac{S^u_M}{\sigma_M} \xrightarrow{d} \mathcal{N}(0,1),$$

as $N\to\infty$. Consider the scenario that $N\to\infty$, in which Assumption 3.1 implies $M\to\infty$. Denote $\tilde{S}^u_M := \frac{S^u_M}{\sigma_M}$. Let X be the $M\times K$ matrix that records the observed dyad-specific characteristics as defined in Section 2.1.1, but here the subscript M is omitted for notational simplicity. The value that X takes is denoted by x.

Under Assumptions 3.2 and 3.3, it holds by Theorem 3.2 of Kojevnikov et al. (2021) that for any $\epsilon > 0$, there exists $M_0 > 0$ such that for each $M > M_0$ and for each $x\in\mathbb{R}^{M\times K}$,

$$\sup_{t\in\mathbb{R}}\big|\Pr(\tilde{S}^u_M \le t \mid X = x) - \Phi(t)\big| < \epsilon, \tag{20}$$

where $\Phi(\cdot)$ is the CDF of a standard Normal distribution. Then, by the law of total probability, we have

$$\begin{aligned}\big|\Pr(\tilde{S}^u_M \le t) - \Phi(t)\big| &= \Big|\int \Pr(\tilde{S}^u_M \le t \mid X = x)dF_X(x) - \Phi(t)\Big|\\ &= \Big|\int \Pr(\tilde{S}^u_M \le t \mid X = x) - \Phi(t)dF_X(x)\Big|\\ &\le \int \big|\Pr(\tilde{S}^u_M \le t \mid X = x) - \Phi(t)\big|dF_X(x)\end{aligned}$$

$$\leq \int \sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}^u_M \leq t \mid X = x) - \Phi(t)\right| dF_X(x), \tag{21}$$

where $F_X(\cdot)$ denotes the probability distribution function of X. Now pick arbitrarily $\epsilon > 0$. Then there exists $M_0 > 0$ such that for each $M > M_0$

$$\begin{aligned}
\left|\Pr(\tilde{S}^u_M \leq t) - \Phi(t)\right| &\leq \int \underbrace{\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}^u_M \leq t \mid X = x) - \Phi(t)\right|}_{<\epsilon} dF_X(x) \\
&\leq \int \epsilon dF_X(x) \\
&\leq \epsilon, \qquad (22)
\end{aligned}$$

where the first and second inequalities come from (21) and (20), respectively. Since the right hand side of (22) does not depend on t, we then have that for each $M > M_0$,

$$\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}^u_M \leq t) - \Phi(t)\right| \leq \epsilon,$$

which implies

$$\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}^u_M \leq t) - \Phi(t)\right| \to 0 \qquad as \quad M \to \infty,$$

We have then shown that

$$\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}^u_M \leq t) - \Phi(t)\right| \to 0 \qquad as \quad N \to \infty,$$

from which we obtain

$$\frac{S^u_M}{\sigma_M} \xrightarrow{d} \mathcal{N}(0, 1) \qquad as \quad N \to \infty.$$

Next this can be combined with Assumption B.4 by using the Slutsky's Theorem, yielding that

$$\frac{S^u_M}{\tau_M\sqrt{\pi_{N,M}}} \xrightarrow{d} \mathcal{N}(0, 1) \qquad as \quad N \to \infty.$$

Moreover, applying the Cramér-Wold device gives

$$\frac{\tau_M^{-1}}{\sqrt{\pi_{N,M}}} S_M \xrightarrow{d} \mathcal{N}(0, I_K) \qquad as \quad N \to \infty,$$

where I_K is the $K \times K$ identity matrix and τ_M is understood as the variance-covariance matrix.[25]

Now notice that we have

$$\sqrt{N}(\hat{\beta} - \beta) = \Big(\frac{1}{M} \sum_{j \in \mathcal{M}_N} x_{M,j} x'_{M,j}\Big)^{-1} \frac{\sqrt{N}}{M} \tau_M \sqrt{\pi_{N,M}} \underbrace{\frac{\tau_M^{-1}}{\sqrt{\pi_{N,M}}} S_{M,m}}_{\xrightarrow{d} \mathcal{N}(0, I_K)} .$$

Hence we obtain

$$\sqrt{N}\big(\hat{\beta} - \beta\big) \xrightarrow{d} \mathcal{N}(0, AVar(\hat{\beta})) \qquad as \quad N \to \infty,$$

where

$$AVar(\hat{\beta}) := \lim_{N \to \infty} N \pi_{N,M} \Big(\sum_{k \in \mathcal{M}_N} E\big[x_{M,k} x'_{M,k}\big]\Big)^{-1} \Big(\sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N} E\big[\varepsilon_{M,m} \varepsilon_{M,m'} x_{M,m} x'_{M,m'}\big]\Big) \Big(\sum_{k \in \mathcal{M}_N} E\big[x_{M,k} x'_{M,k}\big]\Big)^{-1},$$

which is well-defined due to Lemma B.1 (i) along with Assumption B.5. When $\pi_{N,M} = 1$, this is the result in the main text. □

B.5 Lemma

In the proof of Theorem 3.3, we make use of the following lemma from Kojevnikov et al. (2021), p.903:

Lemma B.2. *Define*

$$H_M(s,r) := \{(m,j,k,l) \in \mathcal{M}_N^4 : j \in \mathcal{M}_N(m;r),\ l \in \mathcal{M}_N(k;r),\ \rho_M(\{m,j\},\{k,l\}) = s\}.$$

Then

$$|H_M(s,r)| \leq 4 M c_M(s,r;2).$$

[25] To save notation, we use the same τ_M to denote the case of one-dimensional parameter and the case of multiple-dimensional parameters.

B.6 Consistency of $\widehat{Var}(\hat{\beta})$

Proof of Theorem 3.3: Denote the variance of $\frac{S_M}{\sqrt{M}}$ as $V_{N,M} \coloneqq Var\big(\frac{S_M}{\sqrt{M}}\big)$. It can readily be shown that $V_{N,M}$ takes the form of $V_{N,M} = \sum_{s\geq 0} \Omega_{N,M}(s)$, where

$$\Omega_{N,M}(s) \coloneqq \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} E\big[Y_{M,m}Y'_{M,j}\big].$$

Following Kojevnikov et al. (2021), we define the kernel heteroskedasticity and autocorrelation consistent (HAC) estimator of $V_{N,M}$ as $\hat{V}_{N,M} \coloneqq \sum_{s\geq 0} \omega_M(s)\hat{\Omega}_{N,M}(s)$,

where $\omega_M(s) \coloneqq \omega\big(\frac{s}{b_M}\big)$ and

$$\hat{\Omega}_{N,M}(s) \coloneqq \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \hat{Y}_{M,m}\hat{Y}'_{M,j}.$$

Moreover, we define an empirical analogue of $V_{N,M}$, though infeasible, by $\tilde{V}_{N,M} \coloneqq \sum_{s\geq 0} \omega_M(s)\tilde{\Omega}_{N,M}(s)$, where

$$\tilde{\Omega}_{N,M}(s) \coloneqq \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} Y_{M,m}Y'_{M,j}.$$

Additionally, we denote a conditional version of $V_{N,M}$ by $V^c_{N,M} \coloneqq Var\big(\frac{S_M}{\sqrt{M}} \mid \mathcal{C}_M\big)$, i.e., $V^c_{N,M} = \sum_{s\geq 0} \Omega^c_{N,M}(s)$, where

$$\Omega^c_{N,M}(s) \coloneqq \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} E\big[Y_{M,m}Y'_{M,j} \mid \mathcal{C}_M\big].$$

Notice that since $E\big[Y_{M,m} \mid \mathcal{C}_M\big] = 0$ $a.s.$, it follows from the law of total variance that $V_{N,M} = E\big[V^c_{N,M}\big]$.

Notice furthermore that it holds that

$$Var(\hat{\beta}) = \frac{N}{M}\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1} V_{N,M} \Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1},$$

and

$$\widehat{Var}(\hat{\beta}) = \frac{1}{M}\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1} \hat{V}_{N,M} \Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1}.$$

Since

$$\|N\widehat{Var}(\hat{\beta}) - Var(\hat{\beta})\|_F = \frac{N}{M}\Bigg\|\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1}\hat{V}_{N,M}\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\Big)^{-1}$$
$$- \Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}V_{N,M}\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\Big)^{-1}\Bigg\|_F,$$

and $\frac{N}{M}$ is bounded due to Assumption 3.1, it thus suffices to show that

(i) $\Big\|\big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{M,k}x'_{M,k}\big)^{-1} - \big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x'_{M,k}\big]\big)^{-1}\Big\|_F \xrightarrow{p} 0$;

(ii) $\|\hat{V}_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0$.

Part (i) is already shown in Lemma B.1 (ii). Hence, it remains to prove Part (ii).

To begin with, observe that by the technique of add and subtract as well as the triangular inequality,

$$\begin{aligned}\|\hat{V}_{N,M} - V_{N,M}\|_F &= \|\hat{V}_{N,M} - \tilde{V}_{N,M} + \tilde{V}_{N,M} - V^c_{N,M} + V^c_{N,M} - V_{N,M}\|_F \\ &\le \|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F + \|\tilde{V}_{N,M} - V^c_{N,M}\|_F + \|V^c_{N,M} - V_{N,M}\|_F.\end{aligned}$$

We thus aim to prove

(1) $\|V^c_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0$;

(2) $\|\tilde{V}_{N,M} - V^c_{N,M}\|_F \xrightarrow{p} 0$;

(3) $\|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F \xrightarrow{p} 0$.

We start with:

(1) $\|V^c_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0$:
The proof proceeds in multiple steps:

(a) $E\big[\big\|V^c_{N,M} - V_{N,M}\big\|_F^2\big] \to 0$;

(b) $\big\|V^c_{N,M} - V_{N,M}\big\|_F \xrightarrow{p} 0$.

We begin with:

(a) $E\big[\big\|V^c_{N,M} - V_{N,M}\big\|_F^2\big] \to 0$:
We prove this by showing the element-wise convergence. With a slight abuse of

notation, we denote the (a,b) entry of $V^c_{N,M}$ and $V_{N,M}$ as $V^c_{a,b}$ and $V_{a,b}$, respectively. Then it is enough to verify that

$$E\big[(V^c_{a,b} - V_{a,b})^2\big] \to 0.$$

Notice that $V^c_{a,b}$ and $V_{a,b}$ are given by

$$V^c_{a,b} = \sum_{s\geq 0} \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}^\partial_N(m;s)} E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big]$$

and

$$V_{a,b} = \sum_{s\geq 0} \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}^\partial_N(m;s)} E\big[Y_{m,a}Y_{j,b}\big],$$

where $Y_{m,a}$ and $Y_{m,b}$ stand for the a-th and b-th element of $Y_{M,m}$, respectively. Note moreover that $E[V^c_{a,b}] = V_{a,b}$. Hence, we can write

$$\begin{aligned} E\big[(V^c_{a,b} - V_{a,b})^2\big] &= Var(V^c_{a,b}) \\ &= E\big[(V^c_{a,b})^2\big] - \big(V_{a,b}\big)^2 \\ &\leq E\big[(V^c_{a,b})^2\big]. \end{aligned}$$

Observe that

$$\begin{aligned} E\big[(V^c_{a,b})^2\big] &= E\Big[\Big(\sum_{s\geq 0} \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}^\partial_N(m;s)} E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big]\Big)^2\Big] \\ &= E\Big[\frac{1}{M^2} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}^\partial_N(m;s)} \sum_{t\geq 0} \sum_{k\in\mathcal{M}_N} \sum_{l\in\mathcal{M}^\partial_N(k;t)} E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big] E\big[Y_{k,a}Y_{l,b} \mid \mathcal{C}_M\big]\Big] \\ &= \frac{1}{M^2} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}^\partial_N(m;s)} \sum_{t\geq 0} \sum_{k\in\mathcal{M}_N} \sum_{l\in\mathcal{M}^\partial_N(k;t)} E\Big[E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big] E\big[Y_{k,a}Y_{l,b} \mid \mathcal{C}_M\big]\Big]. \end{aligned}$$

By the Cauchy-Schwartz inequality,

$$E\big[\varepsilon_m\varepsilon_j \mid \mathcal{C}_M\big] \leq \big(E\big[\varepsilon^2_m \mid \mathcal{C}_M\big]\big)^{\frac{1}{2}} \big(E\big[\varepsilon^2_m \mid \mathcal{C}_M\big]\big)^{\frac{1}{2}},$$

it then follows from from Assumption 3.6 (a) that there exists an a.s.-bounded function $\bar{C}_1$ such that $E\big[\varepsilon_m\varepsilon_j \mid \mathcal{C}_M\big] \leq \bar{C}_1$ *a.s.* Similarly, we have an a.s.-bounded function $\bar{C}_2$ such that $E\big[\varepsilon_k\varepsilon_l \mid \mathcal{C}_M\big] \leq \bar{C}_2$ $a.s.$ Then,

$$\begin{aligned} E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big] &= E\big[\varepsilon_m\varepsilon_j x_{m,a}x_{j,b} \mid \mathcal{C}_M\big] \\ &= x_{m,a}x_{j,b} \underbrace{E\big[\varepsilon_m\varepsilon_j \mid \mathcal{C}_M\big]}_{\leq \bar{C}_1} \\ &\leq x_{m,a}x_{j,b}\bar{C}_1 \quad a.s., \end{aligned}$$

where $x_{m,a}$ represents the a-th element of $x_{M,m}$ and $x_{j,b}$ the b-th element of $x_{M,j}$. Analogously, one obtains $E\big[Y_{k,a}Y_{l,b} \mid \mathcal{C}_M\big] \leq x_{k,a}x_{l,b}\bar{C}_2$ $a.s.$ Once again, through the multiple application of the Cauchy-Schwartz inequality, it follows that

$$\begin{aligned} E\Big[E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big]E\big[Y_{k,a}Y_{l,b} \mid \mathcal{C}_M\big]\Big] &\leq E\big[\bar{C}_1\bar{C}_2 x_{m,a}x_{j,b}x_{k,a}x_{l,b}\big] \\ &\leq \Big(E\big[(\bar{C}_1\bar{C}_2)^2\big]\Big)^{\frac{1}{2}}\Big(E\big[(x_{m,a}x_{j,b}x_{k,a}x_{l,b})^2\big]\Big)^{\frac{1}{2}} \\ &\leq \Big(E\big[(\bar{C}_1)^4\big]\Big)^{\frac{1}{4}}\Big(E\big[(\bar{C}_2)^4\big]\Big)^{\frac{1}{4}} \\ &\quad \times \Big(E\big[x_{m,a}^8\big]\Big)^{\frac{1}{8}}\Big(E\big[x_{j,b}^8\big]\Big)^{\frac{1}{8}}\Big(E\big[x_{l,a}^8\big]\Big)^{\frac{1}{8}}\Big(E\big[x_{l,b}^8\big]\Big)^{\frac{1}{8}}. \end{aligned}$$

We note here that Assumption 3.5 ensures that there exists a nonnegative finite constant $C_{m,a}$ such that $E\big[x_{l,a}^8\big] < C_{m,a}$, with the same argument holding true for $x_{j,b}$, $x_{k,a}$ and $x_{l,b}$ as well. Hence,

$$E\Big[E\big[Y_{m,a}Y_{j,b} \mid \mathcal{C}_M\big]E\big[Y_{k,a}Y_{l,b} \mid \mathcal{C}_M\big]\Big] \leq \bar{C},$$

where $\bar{C}$ is a nonnegative finite constant that is appropriately defined.

Substituting this into the inequality above,

$$\begin{aligned} E\big[(V_{a,b}^c)^2\big] &\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(k;t)}\bar{C} \\ &= \frac{\bar{C}}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(k;t)} 1 \\ &= \frac{\bar{C}}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)} 1 \\ &= \frac{\bar{C}}{M^2}\sum_{s\geq 0}\underbrace{|H_M(s;b_M)|}_{\leq 4Mc_M(s,b_M;2)} \\ &\leq \frac{\bar{C}}{M^2}\sum_{s\geq 0} 4Mc_M(s,b_M;2) \end{aligned}$$

$$= 4\bar{C} \underbrace{\frac{1}{M}\sum_{s\geq 0} c_M(s, b_M; 2)}_{\to 0}$$

$$\to 0,$$

where the second inequality comes from Lemma B.2, and the last implication is due to Assumption 3.7. Therefore we have shown that

$$E\left[(V^c_{a,b} - V_{a,b})^2\right] \to 0.$$

By repeating the same argument for each $a, b = 1, \ldots, K$, it follows that

$$E\left[\|V^c_{N,M} - V_{M,M}\|^2_F\right] \to 0.$$

(b) $\left\|V^c_{N,M} - V_{N,M}\right\|_F \xrightarrow{p} 0$:
By the Chebyshev's inequality and the result of part (a), we complete part (1) as it follows that for any $\eta > 0$,

$$\Pr(\left\|V^c_{N,M} - V_{N,M}\right\|_F > \eta) < \frac{1}{\eta^2} \underbrace{E\left[\left\|V^c_{N,M} - V_{N,M}\right\|^2_F\right]}_{\to 0}$$

$$\to 0.$$

(2) $\|\tilde{V}_{N,M} - V^c_{N,M}\|_F \xrightarrow{p} 0$:
This immediately follows from applying Proposition 4.1 of Kojevnikov et al. (2021)[26] and the Dominated Convergence Theorem in the Markov inequality: i.e.,

$$\begin{aligned}
\Pr\big(\|\tilde{V}_{N,M} - V^c_{N,M}\|_F \geq \eta\big) &\leq \frac{1}{\eta} E\left[\|\tilde{V}_{N,M} - V^c_{N,M}\|_F\right] \\
&= \frac{1}{\eta} E\Big[\underbrace{E\left[\|\tilde{V}_{N,M} - V^c_{N,M}\|_F \mid \mathcal{C}_M\right]}_{\xrightarrow{a.s.} 0}\Big] \\
&\to 0,
\end{aligned}$$

for any $\eta > 0$.

(3) $\|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F \xrightarrow{p} 0$:

[26] Notice that the definitions of $V_{N,M}$, $\hat{V}_{N,M}$, $\tilde{V}_{N,M}$ and $V^c_{N,M}$ are slightly different from those used in Proposition 4.1 of Kojevnikov et al. (2021).

First, we have[27]

$$
\begin{aligned}
\|\hat{V}_{N,M}-\tilde{V}_{N,M}\|_F &= \left\| \sum_{s\geq 0}\omega_M(s)\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\hat{\varepsilon}_m\hat{\varepsilon}_j x_m x_j' - \sum_{s\geq 0}\omega_M(s)\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\varepsilon_m\varepsilon_j x_m x_j' \right\|_F \\
&= \left\| \sum_{s\geq 0}\omega_M(s)\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\big(\hat{\varepsilon}_m\hat{\varepsilon}_j-\varepsilon_m\varepsilon_j\big)x_m x_j' \right\|_F \\
&\leq \left\| \sum_{s\geq 0}\underbrace{|\omega_M(s)|}_{\leq 1}\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\big(\hat{\varepsilon}_m\hat{\varepsilon}_j-\varepsilon_m\varepsilon_j\big)x_m x_j' \right\|_F \\
&\leq \left\| \sum_{s\geq 0}\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\big(\hat{\varepsilon}_m\hat{\varepsilon}_j-\varepsilon_m\varepsilon_j\big)x_m x_j' \right\|_F \\
&\leq \sum_{s\geq 0}\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\big|\hat{\varepsilon}_m\hat{\varepsilon}_j-\varepsilon_m\varepsilon_j\big|\|x_m x_j'\|_F.
\end{aligned}
$$

Observe that, by definition, $\hat{\varepsilon}_m$ can be written as $\hat{\varepsilon}_m = \varepsilon_m - x_m'(\hat{\beta}-\beta)$. Hence

$$
\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j = -\varepsilon_m(\hat{\beta}-\beta)'x_j - x_m'(\hat{\beta}-\beta)\varepsilon_j + x_m'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_j,
$$

so that by the triangular inequality,

$$
\big|\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j\big| \leq \|\hat{\beta}-\beta\|_2\|x_j\|_2|\varepsilon_m| + \|\hat{\beta}-\beta\|_2\|x_m\|_2|\varepsilon_j| + \|\hat{\beta}-\beta\|_2^2\|x_m\|_2\|x_j\|_2,
$$

for each $m, j \in \mathcal{M}_N$. Hence $\|\hat{V}_{N,M}-\tilde{V}_{N,M}\|_F$ can be bounded as

$$
\begin{aligned}
&\|\hat{V}_{N,M}-\tilde{V}_{N,M}\|_F \\
&\leq \|\hat{\beta}-\beta\|_2\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\|x_m\|_2\|x_j\|_2^2|\varepsilon_m| + \|\hat{\beta}-\beta\|_2\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\|x_m\|_2^2\|x_j\|_2|\varepsilon_j| \\
&\quad + \|\hat{\beta}-\beta\|_2^2\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\|x_m\|_2^2\|x_j\|_2^2.
\end{aligned}
$$

Denote

$$
\begin{aligned}
R_{N,1} &:= \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\|x_m\|_2\|x_j\|_2^2|\varepsilon_m|; \\
R_{N,2} &:= \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\|x_m\|_2^2\|x_j\|_2|\varepsilon_j|;
\end{aligned}
$$

[27] To lighten the notational burden, we drop the M subscript from $\{x_{M,m}\}_{m\in\mathcal{M}_N}$ and $\{\varepsilon_{M,m}\}_{m\in\mathcal{M}_N}$ in the rest of the proof.

$$R_{N,3} \coloneqq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2^2.$$

Now, since by Theorem 3.1, $\|\hat{\beta}-\beta\|_2 \xrightarrow{p} 0$, and the application of the Continuous Mapping Theorem yields $\|\hat{\beta}-\beta\|_2^2 \xrightarrow{p} 0$, it thus suffices to prove that each of $R_{N,1}$, $R_{N,2}$ and $R_{N,3}$ converges in probability to a finite number. In proving this, we follow a strategy employed in Aronow et al. (2015) and Tabord-Meehan (2019).

First let us study the expectation of $R_{N,1}$. By applying the Cauchy-Schwartz inequality repeatedly, we have that

$$E\big[R_{N,1}\big] \leq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \Big(\big(E\big[\|x_m\|_2^2\big]\big)^{\frac{1}{2}}\big(E\big[\|x_j\|_2^8\big]\big)^{\frac{1}{2}}\Big)^{\frac{1}{2}}\big(E\big[E\big[|\varepsilon_m|^2 \mid \mathcal{C}_M\big]\big]\big)^{\frac{1}{2}}.$$

Here, in light of Assumption 3.6, there exists an a.s.-bounded function C_1 such that $C_1 = \sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\big[|\varepsilon_m|^2 \mid \mathcal{C}_M\big]$, and moreover by Assumption 3.5, there exists a nonnegative finite number $C_2 > 0$ such that $C_2 = \sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\big[\|x_m\|_2^8\big]$.

With a slight abuse of notation, we have for every $N > 0$

$$\begin{aligned}
E\left[R_{N,1}\right] &\leq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} C_1C_2 \\
&= \frac{C_1C_2}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} 1 \\
&= C_1C_2\sum_{s\geq 0}\underbrace{\frac{1}{M}\sum_{m\in\mathcal{M}_N}|\mathcal{M}_N^{\partial}(m;s)|}_{\delta_M^{\partial}(s;1)} \\
&= C_1C_2\underbrace{\sum_{s\geq 0}\delta_M^{\partial}(s;1)}_{<\infty} \\
&< C,
\end{aligned}$$

for some constant $C \in (0,\infty)$, where the last inequality is because of Assumption 3.7.

Next let us study the variance of $R_{N,1}$. It suffices to show that $E\big[R_{N,1}^2\big] \to 0$, By the Cauchy-Schwartz inequality, it holds that

$$E\big[R_{N,1}^2\big] = E\Big[\frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(k;t)} \|x_m\|_2\|x_j\|_2^2\|x_k\|_2\|x_l\|_2^2|\varepsilon_m||\varepsilon_k|\Big]$$

$$\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^\partial(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^\partial(k;t)}\left(E\left[\|x_m\|_2^2\|x_j\|_2^4\|x_k\|_2^2\|x_l\|_2^4\right]\right)^{\frac{1}{2}}\left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}}.$$

Here, by Assumption 3.5 and the Cauchy-Schwartz inequality, there exists a nonnegative finite constant $C_3 > 0$ such that $C_3 = \sup_{N\leq 1}\max_{m,j,k,l\in\mathcal{M}_N} E\left[\|x_m\|_2^2\|x_j\|_2^4\|x_k\|_2^2\|x_l\|_2^4\right]$. Then, with a slight abuse of notation in writing $C_3^{\frac{1}{2}}$ as C_3, we have

$$\begin{aligned}
E\left[R_{N,1}^2\right] &= \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^\partial(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^\partial(k;t)}\left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}} \\
&= \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)}\left(E\left[E\left[|\varepsilon_m|^2|\varepsilon_k|^2 \mid \mathcal{C}_M\right]\right]\right)^{\frac{1}{2}}.
\end{aligned}$$

Corollary A.2 of Kojevnikov et al. (2021) shows that there exists a nonnegative finite constant C_4 such that $E\left[|\varepsilon_m|^2|\varepsilon_k|^2 \mid \mathcal{C}_M\right] \leq C_4\bar{\theta}\theta_{M,s}^{1-\frac{4}{p}}$, where $\bar{\theta} := \sup_{M\geq 1}\max_{s\geq 1}\theta_{M,s}$. Upon applying Lemma B.2 from the Appendix, we obtain

$$E\left[R_{N,1}^2\right] \leq \frac{C_3C_4'}{M^2}\sum_{s\geq 0}\left(E\left[\theta_{M,s}^{1-\frac{4}{p}}\right]\right)^{\frac{1}{2}}4Mc_M(s,b_M;2) = \frac{4C_3C_4''}{M}\sum_{s\geq 0}c_M(s,b_M;2) \to 0,$$

where we apply Assumption 3.7 for the last implication, and C_4' and C_4'' are nonnegative finite constants defined appropriately. Hence we have shown that $R_{N,1}$ converges to a finite constant.

The proof of $R_{N,2}$ is analogous.

It remains to show that $R_{N,3}$ converges in probability to a finite constant. Let us first study the expectation of $R_{N,3}$. Observe that

$$E\left[R_{N,3}\right] = \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^\partial(m;s)} E\left[\|x_m\|_2^2\|x_j\|_2^2\right].$$

By Assumption 3.5, there exists a nonnegative finite number $C_5 > 0$ such that $C_5 = \sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\left[\|x_m\|_2^2\|x_j\|_2^2\right]$. Hence for every $N > 0$,

$$\begin{aligned}
E\left[R_{N,3}\right] &= \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^\partial(m;s)}\underbrace{E\left[\|x_m\|_2^2\|x_j\|_2^2\right]}_{\leq C_5} \\
&\leq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^\partial(m;s)} C_5 \\
&= \frac{C_5}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^\partial(m;s)} 1
\end{aligned}$$

$$= C_5 \sum_{s \geq 0} \underbrace{\frac{1}{M} \sum_{m \in \mathcal{M}_N} |\mathcal{M}_N^{\partial}(m; s)|}_{\delta_M^{\partial}(s;1)}$$

$$= C_5 \underbrace{\sum_{s \geq 0} \delta_M^{\partial}(s; 1)}_{<\infty}$$

$$< C,$$

where we apply Assumption 3.7 in the last implication and a constant $C \in (0, \infty)$ is appropriately defined.

Next let us consider the variance of $R_{M,3}$:

$$E\big[R_{N,3}^2\big] = \frac{1}{M^2} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{j \in \mathcal{M}_N^{\partial}(m;s)} \sum_{t \geq 0} \sum_{k \in \mathcal{M}_N} \sum_{l \in \mathcal{M}_N^{\partial}(k;t)} E\big[\|x_m\|_2^2 \|x_j\|_2^2 \|x_k\|_2^2 \|x_l\|_2^2\big].$$

Once again, Assumption 3.5 and the Cauchy-Schwartz inequality imply that there exists a nonnegative finite number $C_6 > 0$ such that $C_6 = \sup_{N \geq 1} \max_{m,j,k,l \in \mathcal{M}_N} E\big[\|x_m\|_2^2 \|x_j\|_2^2 \|x_k\|_2^2 \|x_l\|_2^2\big]$. Then by Lemma B.2,

$$E\big[R_{N,3}^2\big] \leq \frac{1}{M^2} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{j \in \mathcal{M}_N^{\partial}(m;s)} \sum_{t \geq 0} \sum_{k \in \mathcal{M}_N} \sum_{l \in \mathcal{M}_N^{\partial}(k;t)} C_6 = \frac{4C_6}{M} \sum_{s \geq 0} c_M(s, b_M; 2) \to 0,$$

where the last implication is a consequence of Assumption 3.7 (ii).

Therefore we have shown that

$$\begin{aligned}
&\big\|\hat{V}_{N,M} - \tilde{V}_{N,M}\big\|_F \\
&\leq \big\|\hat{\beta} - \beta\big\|_2 \underbrace{\frac{1}{M} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{j \in \mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2 \|x_j\|_2^2 |\varepsilon_m|}_{R_{M,1}} + \big\|\hat{\beta} - \beta\big\|_2 \underbrace{\frac{1}{M} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{j \in \mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2 \|x_j\|_2 |\varepsilon_j|}_{R_{M,2}} \\
&\quad + \big\|\hat{\beta} - \beta\big\|_2^2 \underbrace{\frac{1}{M} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{j \in \mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2 \|x_j\|_2^2}_{R_{M,3}} \\
&= \underbrace{\big\|\hat{\beta} - \beta\big\|_2}_{\xrightarrow{p} 0} \underbrace{R_{M,1}}_{<\infty} + \underbrace{\big\|\hat{\beta} - \beta\big\|_2}_{\xrightarrow{p} 0} \underbrace{R_{M,2}}_{<\infty} + \underbrace{\big\|\hat{\beta} - \beta\big\|_2^2}_{\xrightarrow{p} 0} \underbrace{R_{M,3}}_{<\infty} \xrightarrow{p} 0,
\end{aligned}$$

which proves $\|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F \xrightarrow{p} 0$.

To sum up, combining parts (1), (2) and (3), we have

$$\begin{aligned}\|\hat{V}_{N,M} - V_{N,M}\|_F &\leq \underbrace{\|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F}_{\xrightarrow{p}0} + \underbrace{\|\tilde{V}_{N,M} - V^c_{N,M}\|_F}_{\xrightarrow{p}0} + \underbrace{\|V^c_{N,M} - V_{N,M}\|_F}_{\xrightarrow{p}0} \\ &\xrightarrow{p} 0,\end{aligned}$$

which completes the proof. □

B.7 Corollary 3.1

Proof of Corollary 3.1: For simplicity we denote

$$\hat{V}^{Dyad}_{N,M} := \Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} \mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1},$$

and

$$\hat{V}^{Network}_{N,M} := \Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1},$$ [28]

where we choose the kernel function and the lag truncation parameter so that the weights become equal one for all active dyads: namely, we use the mean-shifted rectangular kernel with the lag truncation being the length of the longest path in the network. Define moreover $\widetilde{Var}(\hat{\beta})$ to be the same variance as in the main text —
$N\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x_{M,k}'\big]\Big)^{-1}\Big(\frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^\partial_N(m;s)} E\big[Y_{M,m}Y_{M,m'}'\big]\Big)\Big(\frac{1}{M}\sum_{k\in\mathcal{M}_N} E\big[x_{M,k}x_{M,k}'\big]\Big)$
but now applied to the network-regression model (1) and (2). By the triangular inequality,

$$\begin{aligned}\big\|N\hat{V}^{Network}_{N,M} - N\hat{V}^{Dyad}_{N,M}\big\|_F &= \big\|N\hat{V}^{Network}_{N,M} - \widetilde{Var}(\hat{\beta}) + \widetilde{Var}(\hat{\beta}) - N\hat{V}^{Dyad}_{N,M}\big\|_F \\ &\leq \big\|N\hat{V}^{Network}_{N,M} - \widetilde{Var}(\hat{\beta})\big\|_F + \big\|\widetilde{Var}(\hat{\beta}) - N\hat{V}^{Dyad}_{N,M}\big\|_F.\end{aligned}$$

Since Theorem 3.3 implies $\big\|N\hat{V}^{Network}_{N,M} - \widetilde{Var}(\hat{\beta})\big\|_F \xrightarrow{p} 0$, then in the limit we are left with

$$\big\|N\hat{V}^{Network}_{N,M} - N\hat{V}^{Dyad}_{N,M}\big\|_F \leq \big\|\widetilde{Var}(\hat{\beta}) - N\hat{V}^{Dyad}_{N,M}\big\|_F. \tag{23}$$

Now we prove the statement by way of contradiction. Assume for the sake of contradiction that the dyadic-robust variance estimator $\hat{V}^{Dyad}_{N,M}$ is consistent, i.e., $\big\|\widetilde{Var}(\hat{\beta}) - N\hat{V}^{Dyad}_{N,M}\big\|_F \xrightarrow{p} 0$. This, combined with the inequality (23), implies $\big\|N\hat{V}^{Network}_{N,M} - N\hat{V}^{Dyad}_{N,M}\big\|_F \xrightarrow{p} 0$. Now, observe

[28] For the sake of brevity, we suppress the M from subscript throughout this proof.

that

$$
\begin{aligned}
&\big\|N\hat{V}^{Network}_{N,M} - N\hat{V}^{Dyad}_{N,M}\big\|_F \\
&= \Big\|N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1} \\
&\qquad - N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} \mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big\|_F \\
&= \Big\|N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1} \\
&\qquad - N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big\|_F \\
&= \Big\|N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{s=\{0,1\}}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \underbrace{\Big(h_{m,m'} - \mathbb{1}_{m,m'}\Big)}_{0}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1} \\
&\qquad + N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \Big(\underbrace{h_{m,m'}}_{1} - \underbrace{\mathbb{1}_{m,m'}}_{0}\Big)\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big\|_F \\
&= \Big\|N\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big\|_F \\
&= \frac{N}{M}\Big\|\Big(\frac{1}{M}\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big(\frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\Big)\Big(\frac{1}{M}\sum_{m\in\mathcal{M}_N} x_m x_m'\Big)^{-1}\Big\|_F.
\end{aligned}
$$

We prove that the inside the Frobenius norm does not converge in probability to zero.

First it can immediately be shown, by Lemma B.1 (ii), that the "bread" part $\big(\frac{1}{M}\sum_{m\in\mathcal{M}_N} x_m x_m'\big)^{-1}$ converges to $\big(\frac{1}{M}\sum_{m\in\mathcal{M}_N} E\big[x_m x_m'\big]\big)^{-1}$.

Next plugging the definition of $\hat{\varepsilon}$ into the middle part, we have

$$
\begin{aligned}
&\frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}' \\
&= \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \{\varepsilon_m + x_m'(\beta - \hat{\beta})\}\{\varepsilon_{m'} + x_{m'}'(\beta - \hat{\beta})\}x_m x_{m'}' \\
&= \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \varepsilon_m\varepsilon_{m'}x_m x_{m'}' + \varepsilon_m(\beta - \hat{\beta})x_{m'}x_m x_{m'}' + x_m'(\beta - \hat{\beta})\varepsilon_{m'}x_m x_{m'}' + x_m'(\beta - \hat{\beta})(\beta - \hat{\beta})'x_{m'}x_m x_{m'}' \\
&= \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \varepsilon_m\varepsilon_{m'}x_m x_{m'}' + \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \varepsilon_m(\beta - \hat{\beta})x_{m'}x_m x_{m'}' \\
&\quad + \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} x_m'(\beta - \hat{\beta})\varepsilon_{m'}x_m x_{m'}' + \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} x_m'(\beta - \hat{\beta})(\beta - \hat{\beta})'x_{m'}x_m x_{m'}'.
\end{aligned}
$$

Denote

$$
Q_{M,1} := \frac{1}{M}\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}^{\partial}_N(m;s)} \varepsilon_m\varepsilon_{m'}x_m x_{m'}'
$$

$$Q_{M,2} \coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m (\beta - \hat{\beta}) x_{m'} x_m x'_{m'}$$

$$Q_{M,3} \coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} x'_m (\beta - \hat{\beta}) \varepsilon_{m'} x_m x'_{m'}$$

$$Q_{M,4} \coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} x'_m (\beta - \hat{\beta})(\beta - \hat{\beta})' x_{m'} x_m x'_{m'}.$$

From Theorem 3.1, it can be seen that $Q_{M,2}$, $Q_{M,3}$ and $Q_{M,4}$ either converge to zero or diverge as N goes to infinity. When it comes to $Q_{M,1}$, observe that

$$\begin{aligned} E\Big[Q_{M,1}\Big] &= E\Big[\frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m \varepsilon_{m'} x_m x'_{m'}\Big] \\ &= \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} E\Big[\varepsilon_m \varepsilon_{m'} x_m x'_{m'}\Big], \end{aligned}$$

which never equals to zero due to the hypothesis (17) of this corollary. In either case, the middle part does not converge in probability to zero, meaning that $\left\| N\hat{V}_{N,M}^{Network} - N\hat{V}_{N,M}^{Dyad} \right\|_F \xrightarrow{p} 0$ is not true. This, however, contradicts the implication of the assumption that the dyadic-robust variance estimator is consistent. Hence, by means of contradiction, we conclude that the dyadic-robust variance estimator is not consistent, which completes the proof. □

B.8 Example 3.1

Proof of Example 3.1

By the inequality of arithmetic and geometric means, the left-hand side of (17) can be bounded by

$$\begin{aligned} \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} E\big[\varepsilon_{M,m} \varepsilon_{M,m'} x_{M,m} x_{M,m'}\big] &= \sum_{s \geq 2} \gamma^s \delta_M^{\partial}(s) \\ &\geq (S-1) \Big(\prod_{s \geq 2} \gamma^s \Big)^{1/(S-1)} \Big(\prod_{s \geq 2} \delta_M^{\partial}(s) \Big)^{1/(S-1)}, \end{aligned}$$

where $S \geq 2$ denotes the length of the longest path in the network. As the first and third terms in both estimators are the same, then using the proposed network-robust estimator will be desirable if the middle term (i.e., the left-hand side above) is larger than the tolerated threshold, B:

$$(S-1)\left(\prod_{s\geq 2}\gamma^s\right)^{1/(S-1)}\left(\prod_{s\geq 2}\delta_M^{\partial}(s)\right)^{1/(S-1)} > B.$$

Passing logs on both sides yields the results. The lower bound is attained if $\gamma^s\delta_M^{\partial}(s) = \gamma^{s'}\delta_M^{\partial}(s')$ for all $s, s' = 2, \ldots, S$. Note that S and the network densities $\{\delta_M^{\partial}(s)\}_{s\geq 2}$ can be estimated following the definition (14), as a (sample) network is observable.

C Additional Monte Carlo Simulation Results

C.1 Summary Statistics

Table 3 shows summary statistics (i.e., the average and maximum degrees) of the networks across nodes that are used in our simulation study.

The maximum degree and the average degree increase monotonically as we increase the parameters in both specifications. The number of active edges (i.e., dyads) also increases with the sample size regardless of the specification. This reflects that each node tends to have more direct links as the network becomes denser. In our exercises, the number of indirectly linked dyads also increases with network denseness. However, this is due to our simulated networks being relatively sparse. In other settings, the number of indirect connections may decrease with network density.

Table 3: Summary Statistics of Networks among Nodes in the Simulations

		Specification 1			Specification 2		
N		$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
500	d_{max}	23	40	41	5	7	8
	d_{ave}	0.8020	1.5780	2.3540	0.4760	0.9680	1.4800
1000	d_{max}	26	36	47	4	7	8
	d_{ave}	0.8590	1.7000	2.5410	0.4980	0.9810	1.5010
5000	d_{max}	53	125	130	6	9	10
	d_{ave}	0.9326	1.8618	2.7910	0.4952	1.0016	1.5114

Notes: Observation units in this table are nodes (individuals) as usual in the literature. The maximum degree, d_{max}, means the maximum number of nodes that are adjacent to a node, and the average degree, d_{ave}, is the average number of nodes adjacent to each node of the network.

Table 4 reports the degree characteristics of the networks when viewed as networks over the active edges. The table provides the average degree, the maximum degree, and the number of active edges (i.e., dyads).

Table 4: Summary Statistics of Networks among Dyads in the Simulations

		Specification 1			Specification 2		
N		$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
500	d_{act}	401	788	1175	238	484	740
	d_{max}	32	45	70	4	9	14
	d_{ave}	3.6858	6.0063	9.1881	0.9580	2.0248	3.0027
1000	d_{act}	859	1699	2540	498	981	1501
	d_{max}	35	55	76	5	8	10
	d_{ave}	3.9581	6.7810	9.2047	1.0341	1.9888	2.9594
5000	d_{act}	4663	9305	13952	2476	5008	7557
	d_{max}	74	161	210	7	12	15
	d_{ave}	5.2989	9.5159	12.8521	1.0137	2.0228	3.0341

Note: Observation units in this table are active edges (dyads), which departs from the convention. Active edges are edges that are at work in the original network over the nodes. The number of active edges is denoted by d_{act}. The maximum degree, d_{max}, expresses the maximum number of edges that are adjacent to an edge, and the average degree, d_{ave}, is the average number of edges adjacent to each edge of the network.

C.2 Additional Details on the Design

We draw $\varepsilon_m := \sum_{m'} \gamma_{m,m'}\eta_{m,m'}$, where $\gamma_{m,m'}$ equals γ^s if the distance between m and m' is s, and 0 otherwise, for $\gamma \in [0,1]$[29] and $s \in \{1, \ldots, S\}$ with S being the maximum geodesic distance that the spillover propagates to. Each $\eta_{m,m'}$ is drawn *i.i.d.* from $\mathcal{N}(0,1)$. Hence, γ controls the strength of spillover effects, representing their decay rate.

C.3 $S = 2$ and $\gamma = 0.8$

In this section, we further discuss the results of the Monte Carlo simulations presented in the main text. The asymptotic behaviors of the three variance estimators are illustrated in Figure 2, where the horizontal axes represent the sample size and the vertical axes indicate the standard error of the regression coefficient. The boxplots show the 25th and 75th percentiles across simulations, as well as the median, with the whiskers indicating the bounds that are not

[29] In this simulation, we focus on cases of positive spillovers, as negative spillovers can be analyzed analogously.

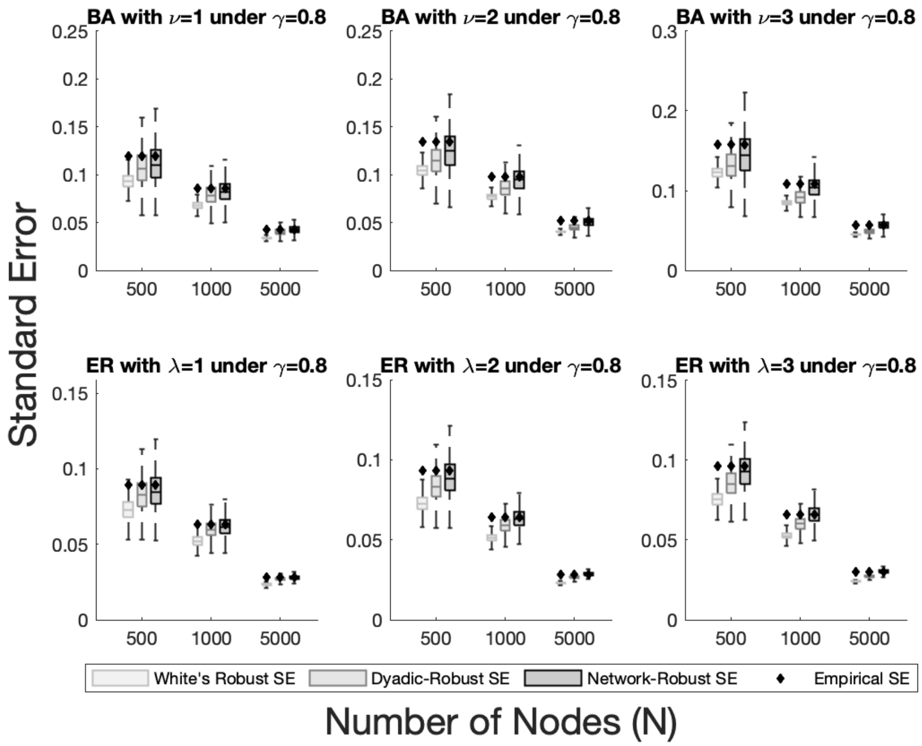
considered as outliers. The whisker length is set to cover ±2.7 times the standard deviation of the standard-error estimates. The light-, medium- and dark-gray boxplots describe the distribution of the Eicker-Huber-White, the dyadic-robust and our proposed network-robust variance estimates across simulations, respectively. The diamonds indicate the empirical standard errors of the estimates of the regression coefficients, what Aronow et al. (2015) call the true standard error. It is unsurprising that the empirical standard errors are the same across different variance estimators, as we use the same $\hat{\beta}$. The boxplots show that as the sample size increases, the variation of the network-robust variance estimator shrinks, reaching the empirical standard error (the diamonds). This is as expected since this estimator is consistent for the true variance (Theorem 3.3). The estimates appear to vary little for moderate sample sizes (e.g., $N = 1000$). However, the other variance estimators (the light- and medium-gray boxplots) converge to lower values than the empirical standard errors (the diamonds), verifying their inconsistency in this environment with network spillovers, as shown by Corollary 3.1. As we make such spillovers very small (e.g., $\gamma = 0.2$ in Appendix C.5), all estimators have similar performance. This highlights the role of condition (17): namely, the dyadic-robust variance estimator might perform satisfactorily well as long as higher-order correlations beyond immediate neighbors are negligible.

Table 5 describes the standard deviations of the estimated regression coefficients (what Aronow et al. (2015) calls the true standard errors) and the means of the estimated standard errors for each variance estimator. The round brackets indicate the biases of each estimate relative to the true standard error in percentage (%). For instance, the Eicker-Huber-White variance estimator and the dyadic-robust variance estimator, when applied to Specification 1 with $\nu = 3$, underestimate the true standard error by 21.45% and 14.14%, respectively.

C.4 $S = 2$ and $\gamma = 0.8$ with Higher Density Parameters

This subsection examines how an increase in the number of connected dyads affects the performance of the dyadic-robust variance estimator. Table 6 reports the results for the case of $S = 2$ with $\gamma = 0.8$, i.e., the same combination as the main text (Table 1), but for denser networks which set $\nu = 4, 5$ for Specification 1 and $\lambda = 4, 5$ for Specification 2. We find that, while our estimator performs well (with coverage close to the nominal level), the bias in the Eicker-Huber-White and dyadic-robust estimators variance estimators are present and increase as the network becomes denser.

Figure 2: Boxplots of Standard Errors for Specifications 1 and 2 ($S = 2$, $\gamma = 0.8$)



Note: This figure shows boxplots describing the estimated standard errors and the empirical standard errors for various combinations of parameters under Specification 1 (Barabási-Albert networks) and Specification 2 (Erdös-Renyi networks). The horizontal axis shows the number of nodes and the vertical axis represents the the standard error of the coefficient. The shaded boxes represent the 25th, 50th and 75th percentiles of estimated standard errors with the whiskers indicating the most extreme values that are not considered as outliers. The light-gray box illustrates the Eicker-Huber-White standard error, the medium-gray one the dyadic-robust standard error and the dark-gray one the network-robust standard error. The diamonds stand for the empirical standard error, defined as the standard deviation of the estimates of the regression coefficient. The estimator is considered as not covering the true standard error when the diamond is outside of the shaded area.

Table 5: Means and Biases of the Standard Errors: $N = 5000$, $S = 2$, $\gamma = 0.8$.

	Specification 1			Specification 2		
	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
True	0.0430	0.0518	0.0570	0.0285	0.0283	0.0302
Eicker-Huber-White	0.0337	0.0404	0.0448	0.0234	0.0229	0.0239
(Bias %)	(-21.61)	(-21.92)	(-21.45)	(-17.91)	(-19.14)	(-20.68)
Dyadic-robust	0.0403	0.0453	0.0490	0.0270	0.0266	0.0275
(Bias %)	(-6.19)	(-12.54)	(-14.14)	(-5.01)	(-6.12)	(-8.93)
Network-robust	0.0425	0.0509	0.0565	0.0280	0.0285	0.0302
(Bias %)	(-1.09)	(-1.78)	(-0.92)	(-1.70)	(0.58)	(0.09)

Note: This table shows the standard deviations of the estimated regression coefficients (the true standard error) and the means of the estimated standard errors for each variance estimator with the round brackets indicating the biases relative to the true standard error in percentage (%). To facilitate the comparison, the biases are rounded off to the second decimal places.

C.5 $S = 2$ and $\gamma = 0.2$

Table 7 presents the empirical coverage probability and average length of confidence intervals for β at 5% nominal size when $S = 2$ and $\gamma = 0.2$. The associated boxplots are given in Figure 3. Since the magnitude of spillovers is now much smaller than the case of $\gamma = 0.8$, there are only minor differences in performance between the network-robust variance estimator and the other two existing methods (namely, the Eicker-Huber-White and dyadic-robust variance estimators). In terms of convergence, the comparable performance of the dyadic-robust-variance estimator is evident in Figure 3.

Comparing Table 7 to Table 1 highlights the impact of spillovers on the variance estimators. When the spillovers are substantially weak (e.g., $\gamma = 0.2$), the dyadic-robust variance estimator can serve as a good substitute for the network-robust one. In the case of relatively high spillovers (e.g., $\gamma = 0.8$), on the other hand, there are evident biases (around 4 percentage points for Specification 1 and 3 percentage points for Specification 2 when $N = 5000$). Based on this comparison, we suggest that the network-robust variance estimator be used when the correlations are expected to be relatively strong.

Table 6: The empirical coverage probability and average length of confidence intervals for β at 95% nominal level: $S = 2$, $\gamma = 0.8$, higher denseness parameters.

		Specification 1		Specification 2	
	N	$\nu = 4$	$\nu = 5$	$\lambda = 4$	$\lambda = 5$
		Coverage Probability			
Eicker-Huber-White	500	0.8758	0.8688	0.8744	0.8706
	1000	0.8752	0.8688	0.8694	0.8784
	5000	0.8658	0.8808	0.8694	0.8750
Dyadic-robust	500	0.8912	0.8808	0.9142	0.9058
	1000	0.8936	0.8852	0.9124	0.9160
	5000	0.8940	0.9020	0.9152	0.9176
Network-robust	500	0.9084	0.8928	0.9394	0.9368
	1000	0.9246	0.9190	0.9424	0.9494
	5000	0.9404	0.9436	0.9450	0.9512
		Average Length of the C.I.			
Eicker-Huber-White	500	0.5282	0.5577	0.3088	0.3230
	1000	0.3964	0.4155	0.2183	0.2323
	5000	0.1944	0.2132	0.0992	0.1045
Dyadic-robust	500	0.5580	0.5841	0.3471	0.3601
	1000	0.4211	0.4380	0.2465	0.2595
	5000	0.2085	0.2254	0.1124	0.1172
Network-robust	500	0.6099	0.6428	0.3825	0.4022
	1000	0.4751	0.4966	0.2743	0.2927
	5000	0.2449	0.2660	0.1259	0.1331

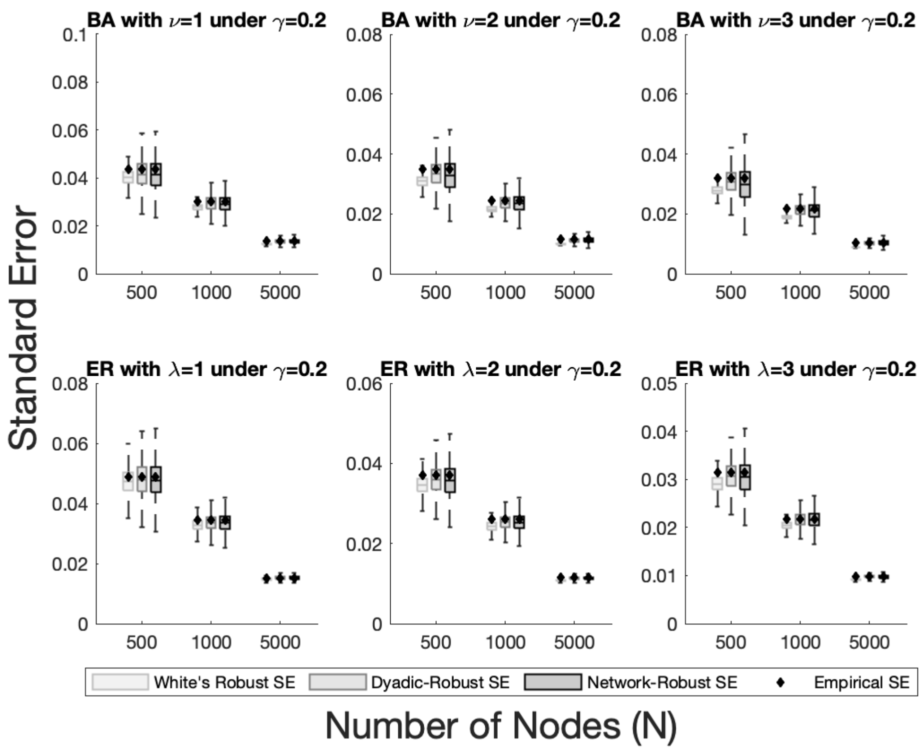
Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β, and the lower-half showcases the average length of the estimated confidence intervals. As the sample size (N) increases, the empirical coverage probability approaches 0.95, the nominal level. This convergence is accompanied by the shrinking average length of confidence intervals.

Table 7: The empirical coverage probability and average length of confidence intervals for β at 95% nominal level: $S = 2$, $\gamma = 0.2$.

		Specification 1			Specification 2		
	N	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
		Coverage Probability					
Eicker-Huber-White	500	0.9286	0.9186	0.9108	0.9434	0.9292	0.9308
	1000	0.9320	0.9158	0.9148	0.9338	0.9322	0.9322
	5000	0.9200	0.9110	0.9124	0.9434	0.9382	0.9308
Dyadic-robust	500	0.9342	0.9350	0.9336	0.9454	0.9368	0.9422
	1000	0.9454	0.9376	0.9458	0.9398	0.9422	0.9486
	5000	0.9446	0.9448	0.9432	0.9486	0.9490	0.9472
Network-robust	500	0.9284	0.9246	0.9162	0.9428	0.9360	0.9384
	1000	0.9414	0.9294	0.9370	0.9392	0.9410	0.9456
	5000	0.9454	0.9476	0.9418	0.9494	0.9492	0.9470
		Average Length of the Confidence Intervals					
Eicker-Huber-White	500	0.1578	0.1214	0.1092	0.1860	0.1360	0.1141
	1000	0.1088	0.0846	0.0743	0.1290	0.0955	0.0799
	5000	0.0486	0.0388	0.0346	0.0579	0.0423	0.0357
Dyadic-robust	500	0.1648	0.1316	0.1213	0.1890	0.1410	0.1205
	1000	0.1158	0.0931	0.0833	0.1319	0.0994	0.0848
	5000	0.0532	0.0439	0.0398	0.0594	0.0443	0.0381
Network-robust	500	0.1637	0.1291	0.1174	0.1885	0.1404	0.1196
	1000	0.1154	0.0922	0.0825	0.1318	0.0993	0.0848
	5000	0.0533	0.0440	0.0401	0.0594	0.0444	0.0382

Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β, and the lower-half showcases the average length of the estimated confidence intervals. One computational issue that plagues the Monte Carlo simulation is the potential lack of positive-semi-definiteness of the estimated variance-covariance matrix. In general, this problem prevails only when the sample size (N) is small. In our case, when $N = 500$, four variance estimates out of five thousands take negative values. We deal with this issue by first applying the eigenvalue decomposition to the estimated variance-covariance matrix and then augmenting the diagonal matrix of eigenvalues by a small constant, followed by pre- and post-multiplications by the matrix of eigenvectors to obtain the updated estimate for the variance-covariance matrix. As the sample size (N) increases, the empirical coverage probability approaches 0.95, the nominal level. This convergence is accompanied by the shrinking average length of confidence intervals.

Figure 3: Boxplots of Standard Errors for Specifications 1 and 2 ($S = 2$, $\gamma = 0.2$)



Note: This figure shows boxplots describing the estimated standard errors and the empirical standard errors for various combinations of parameters under Specification 1 (Barabási-Albert networks) and Specification 2 (Erdös-Renyi networks). The horizontal axis shows the number of nodes and the vertical axis represents the the standard error of the coefficient. The shaded boxes represent the 25th, 50th and 75th percentiles of estimated standard errors with the whiskers indicating the most extreme values that are not considered as outliers. The light-gray box illustrates the Eicker-Huber-White standard error, the medium-gray one the dyadic-robust standard error and the dark-gray one the network-robust standard error. The diamonds stand for the empirical standard error, defined as the standard deviation of the estimates of the regression coefficient. This figure showcases the boxplots for the case when $\gamma = 0.2$.

C.6 $S = 1$

For comparison purposes, this subsection explores the results for $S = 1$. If $S = 1$, there are no higher-order correlations beyond direct (adjacent) neighbors. Then, the network-robust variance estimator ought to coincide with the dyadic-robust variance estimator by definition, for any γ, as pointed out in Example 2.1. This is verified below for the case of $\gamma = 0.8$. Table 8 shows the simulation results.

Table 8: The empirical coverage probability and average length of confidence intervals for β at 95% nominal level: $S = 1$, $\gamma = 0.8$.

		Specification 1			Specification 2		
	N	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
		Coverage Probability					
Eicker-Huber-White	500	0.8804	0.8676	0.8734	0.8906	0.8768	0.8692
	1000	0.8678	0.8810	0.8710	0.8984	0.8864	0.8856
	5000	0.8752	0.8652	0.8742	0.8996	0.8910	0.8778
Dyadic-robust	500	0.9292	0.9304	0.9384	0.9366	0.9416	0.9368
	1000	0.9364	0.9426	0.9432	0.9428	0.9454	0.9484
	5000	0.9474	0.9414	0.9498	0.9452	0.9518	0.9506
Network-robust	500	0.9292	0.9304	0.9384	0.9366	0.9416	0.9368
	1000	0.9364	0.9426	0.9432	0.9428	0.9454	0.9484
	5000	0.9474	0.9414	0.9498	0.9452	0.9518	0.9506
		Average Length of the Confidence Intervals					
Eicker-Huber-White	500	0.3282	0.2901	0.2881	0.2664	0.2377	0.2235
	1000	0.2321	0.2088	0.1964	0.1887	0.1665	0.1564
	5000	0.1131	0.1042	0.0980	0.0844	0.0742	0.0704
Dyadic-robust	500	0.3934	0.3591	0.3603	0.3104	0.2888	0.2776
	1000	0.2853	0.2625	0.2500	0.2227	0.2037	0.1950
	5000	0.1428	0.1330	0.1259	0.0998	0.0913	0.0882
Network-robust	500	0.3934	0.3591	0.3603	0.3104	0.2888	0.2776
	1000	0.2853	0.2625	0.2500	0.2227	0.2037	0.1950
	5000	0.1428	0.1330	0.1259	0.0998	0.0913	0.0882

Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β, and the lower-half showcases the average length of the estimated confidence intervals. As the sample size (N) increases, the empirical coverage probability approaches 0.95, the nominal level. This convergence is accompanied by the shrinking average length of confidence intervals.

C.7 $S = \infty$ with the Parzen kernel

In this subsection, we investigate the consequences of adaptively choosing the value of the lag-truncation parameter following the rule outlined in the main text. To this end, we set $S = \infty$ (i.e., spillovers may propagate to all neighbors), with the magnitude of the spillovers controlled by $\gamma = 0.8$ (the same as in the main text). In this environment, the spillovers are never truncated while decaying as they propagate farther. With regards to estimation, we consider the Parzen kernel, letting the lag-truncation parameter be chosen on the basis of Kojevnikov et al. (2021). The simulation results are given in Table 9, while the selected lag-truncation parameters are shown in Table 10.

The empirical coverage probability based on the network-robust variance estimator approaches to 95%, as expected. On the other hand, both the Eicker-Huber-White and dyadic-robust variance estimator understate the targeted nominal level, as claimed in the main text. It should be noted that these biases can become larger when the decay rate is slower. We focus on Specification 2, as it likely satisfies the assumptions above under $S = \infty$. After all, with $S = \infty$ and a very dense network, Assumption 3.4 is violated.

Table 9: The empirical coverage probability and average length of confidence intervals for β at 95% nominal level, Specification 2: $S = \infty$, $\gamma = 0.8$, the Parzen kernel.

	N	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
		Coverage Probability		
Eicker-Huber-White	500	0.8884	0.8768	0.8718
	1000	0.8892	0.8832	0.8806
	5000	0.8966	0.8820	0.8806
Dyadic-robust	500	0.9282	0.9150	0.8964
	1000	0.9300	0.9214	0.9126
	5000	0.9384	0.9272	0.9118
Network-robust	500	0.9366	0.9302	0.9180
	1000	0.9382	0.9386	0.9362
	5000	0.9480	0.9510	0.9462
		Average Length of C.I.		
Eicker-Huber-White	500	0.2890	0.3085	0.3658
	1000	0.2103	0.2241	0.2570
	5000	0.0933	0.0994	0.1188
Dyadic-robust	500	0.3293	0.3483	0.3966
	1000	0.2405	0.2526	0.2809
	5000	0.1075	0.1127	0.1300
Network-robust	500	0.3387	0.3705	0.4233
	1000	0.2502	0.2729	0.3070
	5000	0.1120	0.1233	0.1457

Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β, and the lower-half showcases the average length of the estimated confidence intervals. As the sample size (N) increases, the empirical coverage probability approaches 0.95, the nominal level. This convergence is accompanied by the shrinking average length of confidence intervals.

Table 10: The lag-truncation parameters for Table 9 based on the Kojevnikov et al.'s (2021) rule.

N	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
500	224.3186	17.5262	12.0174
1000	254.5841	20.0388	13.4822
5000	320.3268	24.1851	16.0915

Note: This table displays the lag-truncation parameters b_M for the simulations in Table 9, selected using the rule: $b_M = 2\log(M)/\log(\max(\textit{average degree}, 1.05))$, with M denoting the number of active dyads.

D Additional Information for the Empirical Illustration

D.1 Seating Arrangement at the European Parliament

Figure 4 exhibits an example of the seating arrangement at the European Parliament, and describes how we construct an adjacency relationship among MEPs within their EPG groups.

D.2 Summary Statistics of the Seating Arrangement

Table 11 lists the summary statistics of the seating arrangement (for Strasbourg at term 7) when viewed as a network over pairs of MEPs. Its summary statistics are consistent with those from the Erdös-Renyi random network with $\lambda = 1$ to $\lambda = 3$ (see Table 4). This suggests that our empirical illustration should perform well with the Parzen kernel and bandwidth choice proposed in Kojevnikov et al. (2021).

Table 11: Summary Statistics of The Seating Arrangement: Strasbourg, Term 7

d_{act}	d_{max}	d_{ave}	e_{direct}	$e_{indirect}$
602	2	1.7076	514	3136

See Table 4 for the definition of the first three indicators. The last two represent the number of adjacent and connected dyads, respectively.

D.3 Data Construction

Data construction for our empirical exercise in Section 5 proceeds in multiple steps:

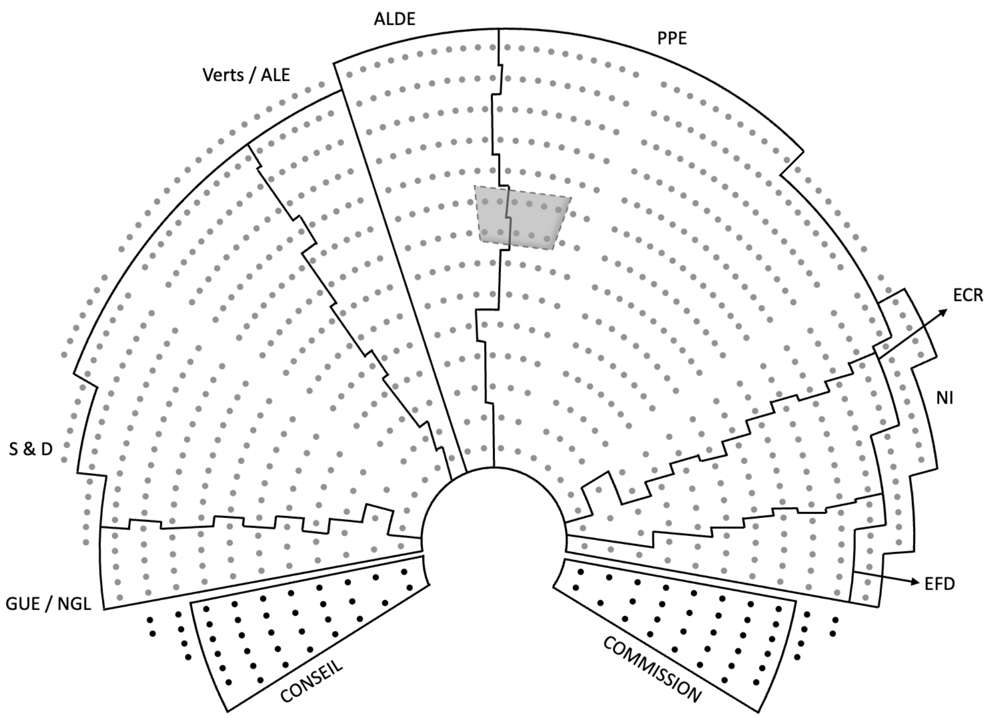
Step 1: Our subsample consists of the location of interest (i.e., Strasbourg) for the period of interest (i.e., Term 7). We select a further subset of the extracted data by seating arrangement (i.e., we focus on Pattern 1 for the present analysis - see Table 12).

Step 2: Since our analysis is concerned with voting concordance, we follow the original authors in dropping entries with missing data or "abstain" in the variable "vote."[30]

Step 3: The resulting data still contains individuals belonging to "Identity, Tradition and Sovereignty (ITS)," one of the European Political Groups that dissolved in November 7, during the sixth term. We drop such MEPs from our analysis.

[30] This amounts to assuming that those observations are missing completely at random (MCAR).

Figure 4: Seating Plan at the European Parliament: Strasbourg, September 14, 2009



		ALDE				PPE			
Row 10	⋯	⋮	⋮	⋮	⋮	⋮	⋮	⋯	
Row 9	⋯	Seat 36: ALFANO	Seat 37: ALVARO	Seat 38: CAVADA	Seat 39: COELHO	Seat 40: COLLINO	Seat 41: COMI	⋯	
Row 8	⋯	Seat 32: OJULAND	Seat 33: OVIIR	Seat 34: BACH	Seat 35: BALDASSARRE	Seat 36: BALZ	⋯		
Row 7	⋯	⋮	⋮	⋮	⋮	⋮	⋮	⋯	

Note: The upper panel illustrates a zoomed-out view of a seating plan for the European parliament in Strasbourg on September 14, 2009. gray circles are individual MEPs, while black circles embody members of conseil and commission. The associated party (EPG) is denoted at the top. The lower panel provides a zoomed-in view elaborating on the part of the upper panel marked by the dotted trapezoid shaded in gray. Alafano and Alvaro are treated as adjacent because they are sitting next to each other and belong to the same political party, i.e., ALE. Similarly, Ojuland and Oviir are considered to be adjacent. On the other hand, following the original authors, Alvaro and Cavada are not regarded as adjacent though they are seated together because they belong to different political parties, i.e., ALE and PPE, respectively. In terms of dyad-level adjacency, Cavada-Coelho and Coelho-Collino are adjacent dyads as they share Coelho, whereas Cavada-Coelho and Collino-Comi are not adjacent, but they are still connected as they have indirect paths to one another along the dyadic network.

Step 4: The selected data is used to form the dyadic data registering the pair-of-MEPs-specific information. When pairing two MEPs, we follow Harmon et al. (2019) in focusing on those pairs of MEPs, both of whom are

(i) in the same EPG;

(ii) from an alphabetically-seated EPG; and

(iii) non-leaders at the time of voting.

Our dyadic data consists of two types of variables: binary variables and numerical variables. The dyad-level binary (i.e., indicator) variables are defined to be one if the individual-level binary variables are the same, and zero otherwise. The dyadic-specific numerical variables in our analysis are the differences between the individual-level numerical variables, such as age and tenure. When calculating the differences in ages and tenures, we take the absolute values as we do not consider directional dyads, and we then rescale them into ten-year units. See the note below Table 2 for details.

Table 12: Patterns of Seating Arrangements: Strasbourg, Term 7

Pattern	Date			Number of Proposals
1	7/14/2009	~	7/16/2009	116
2	8/18/2009	~	8/21/2009	72
3	9/23/2009	~	9/25/2009	114
4	10/13/2009	~	10/16/2009	40
5	11/19/2009	~	12/11/2009	94
6	1/5/2010	~	1/8/2010	79
7	3/17/2010	~	3/19/2010	45
8	4/14/2010	~	4/16/2010	120
9	5/5/2010	~	5/7/2010	79
10	7/7/2010	~	7/9/2010	34
11	7/21/2010	~	7/22/2010	50
12	8/18/2010	~	8/20/2010	118

Note: This table presents patterns of seating arrangements with the corresponding dates and the number of total observations for each pattern. Since voting may be taken place for multiple proposals within the same day, the total number proposals tends to be higher than that of days in a single pattern. For example, the first line indicates that 116 proposals were discussed and votes were cast over the three days (from the 14th of July, 2009 to the 16th of July, 2009).

D.4 Full Results

Table 13 reports the detailed result of the empirical illustration. As explained in Section 5.3, Panel A reports the estimates of the parameter of interest, while Panel B lists the standard errors based on the different variance estimators. In particular, we carry out the estimation using both the network-robust variance estimator with the mean-shifted rectangular kernel, and the one with the Parzen kernel, generating the same estimates. Panel C collects the parameter estimates for other covariates accompanied by the standard errors obtained from our proposed variance estimator from equation (10), and indicates the presence or absence of day-level fixed effects.

Table 13: Spillovers in Legislative Voting – Main Analysis

	Specification (I)	Specification (II)	Specification (III)
Panel A: Parameter estimates for Seat neighbors			
Seat neighbors	0.0069	0.0060	0.0060
Panel B: Standard errors for Seat neighbors			
Eicker-Huber-White	0.0031	0.0030	0.0030
Dyadic-robust	0.0075	0.0082	0.0087
Network-robust (with the rectangular kernel)	0.0095	0.0104	0.0112
Network-robust (with the Parzen kernel)	0.0095	0.0104	0.0112
Panel C: Parameter estimates for other covariates			
Same country		0.0561	0.0562
		(0.0008)	(0.0008)
Same quality education		0.0030	0.0028
		(0.0007)	(0.0007)
Same freshman status		-0.0070	-0.0070
		(0.0008)	(0.0008)
Same gender		0.0004	0.0004
		(0.0007)	(0.0006)
Age difference		0.0007	0.0004
		(0.0004)	(0.0004)
Tenure difference		-0.0149	-0.0149
		(0.0006)	(0.0006)
Day-level FE	No	No	Yes

Note: Panel A displays the parameter estimates for the three different specifications; Panel B shows the standard errors for the regression coefficient of *SeatNeighbors* using different variance estimators; and Panel C collects the parameter estimates for other covariates accompanied by the standard errors obtained from our proposed variance estimator from equation (10), and indicates the presence or absence of day-level fixed effects. Adjacency of MEPs is defined at the level of a row-by-EP-by-EPG. (See the note below Figure 4.) Independent variables are as follows: *Seat neighbors* is an indicator variable denoting whether both MEPs sit together; *Same country* represents an indicator for whether both MEPs are from the same country; *Same quality education* is an indicator showing whether both MEPs have the same quality of education background, measured by if both have the degree from top 500 universities; *Same freshman status* encodes whether both MEPs are freshman or not; *Age difference* is the difference in the MEPs' ages; and *Tenure difference* measures the difference in the MEPs' tenures.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2022). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics Forthcoming*.

Aronow, P. M., C. Samii, and V. A. Assenova (2015). Cluster–robust variance estimation for dyadic data. *Political Analysis 23*(4), 564–577.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Chiang, H. D., B. E. Hansen, and Y. Sasaki (2022). Standard errors for two-way clustering with serially correlated time effects. Working Paper.

Chiang, H. D., K. Kato, Y. Ma, and Y. Sasaki (2021). Multiway cluster robust double/debiased machine learning. *Journal of Business & Economic Statistics 40*(3), 1–11.

Chiang, H. D., Y. Matsushita, and T. Otsu (2022). Multiway empirical likelihood. Working Paper.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2021). Empirical process results for exchangeable arrays. *The Annals of Statistics 49*(2), 845–862.

de Paula, A., S. Richards-Shubik, and E. Tamer (2018). Identifying preferences in networks with bounded degree. *Econometrica 86*(1), 263–288.

Harmon, N., R. Fisman, and E. Kamenica (2019). Peer effects in legislative voting. *American Economic Journal: Applied Economics 11*(4), 156–80.

Ibragimov, R. and U. K. Müller (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics 28*(4), 453–468.

Jenish, N. and I. R. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics 150*(1), 86–98.

Kojevnikov, D., V. Marmer, and K. Song (2021). Limit theorems for network dependent random variables. *Journal of Econometrics 222*(2), 882–908.

Leung, M. P. (2021). Network cluster-robust inference. Working Paper.

Leung, M. P. (2022). Causal inference under approximate neighborhood interference. *Econometrica 90*(1), 267–293.

Leung, M. P. and H. R. Moon (2021). Normal approximation in large network models. Working Paper.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2022a). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics* (Forthcoming).

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2022b). Fast and reliable jackknife and bootstrap methods for cluster-robust inference. Working Paper.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2022c). Leverage, influence, and the jackknife in clustered regression models: Reliable inference using summclust. Working Paper.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica 89*(5), 2143–2188.

Penrose, M. D. and J. E. Yukich (2003). Weak laws of large numbers in geometric probability. *Annals of Applied Probability 13*(1), 277–303.

Tabord-Meehan, M. (2019). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics 37*(4), 671–680.

Vainora, J. (2020). Network dependence. Working Paper.