FIXED-EFFECT REGRESSIONS ON NETWORK DATA

SUPPLEMENT

Koen Jochmans* Sciences Po, Paris

Martin Weidner‡ University College London

May 24, 2017

Contents

*Sciences Po, Département d'économie, 28 rue des Saints Pères, 75007 Paris, France. E-mail: koen.jochmans@sciencespo.fr

‡University College London, Department of Economics, Gower Street, London WC1E 6BT, United Kingdom, and CeMMAP. E-mail: m.weidner@ucl.ac.uk

S.1 Further Discussion and Illustrations

S.1.1 Relation to network-formation models

The existing literature on the estimation of fixed-effect models for network data typically assumes a complete graph, that is, that all possible pairs of outcomes are observed; see Simons and Yao (1999) and Yan and Xu (2013) for results on the Bradley and Terry (1952) model, Fernández-Val and Weidner (2016) for two-way models for panel data, and Dzemski (2014), Graham (2017), and Jochmans (2017) for work on network-formation models.

In the network-formation literature, sparsity means that the number of links between n vertices grows slower than n^2. This definition of sparsity is different than ours. Indeed, while many links may fail to form, the *decision* whether to link or not is observed for all $n(n-1)/2$ pairs of vertices. In our context, sparsity means to no outcome is observed for a pair of vertices.

S.1.2 Additional illustrations

Remember that our measure of global connectivity is λ_2, the second smallest eigenvalue of the normalized Laplacian matrix. In the following we provide some concrete examples of graphs for which λ_2 can be explicitly calculated, and we discuss the implications of our first-order variance bounds in Theorem 3 for those examples. The first example illustrates that even for λ_2 converging to zero asymptotically (the graph becoming less and less strongly connected in that sense) we may still have the variance of $\widehat{\alpha}_i$ converging to zero at the rate $1/d_i$.

Example S.1 (Hypercube graph)**.** *Consider the N-dimensional hypercube, where each of $n = 2^N$ vertices is involved in N edges; see the left hand side of Figure S.1. This is an N-regular graph — that is, $d_i = h_i = N$ for all i — with the total number of edges in the graph equaling 2^{N-1}. Here,*

$$\lambda_2 = \frac{2}{N} = O((\ln n)^{-1}).$$

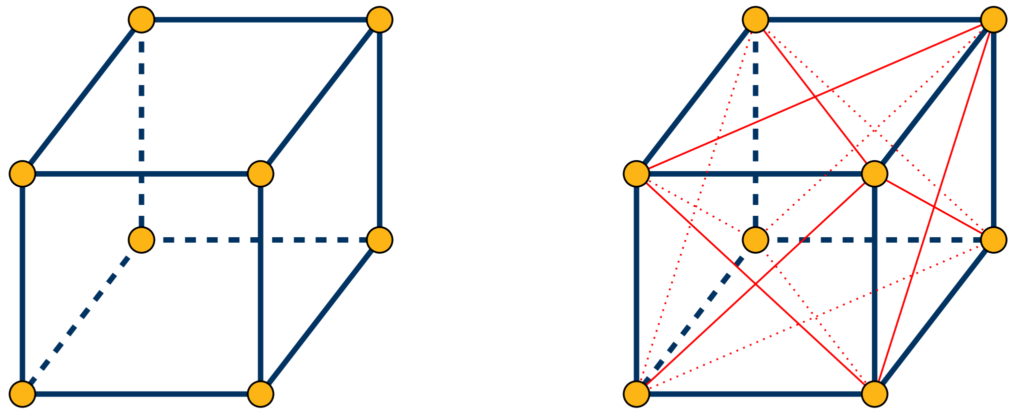

Figure S.1: three-dimensional hypercube (left) and extended hypercube (right).

Thus, $\lambda_2 \, h_i$ is constant in n. An application of Theorem 3 yields

$$1 + o(1) \leq \frac{N \operatorname{var}(\widehat{\alpha}_i)}{\sigma^2} \leq \frac{3}{2} + o(1).$$

From this, we obtain the convergence rate result $(\widehat{\alpha}_i - \alpha_i) = O_p\left((\ln n)^{-1/2}\right)$, but the bounds are not sufficient to determine the leading order asymptotic variance of $\widehat{\alpha}_i$. However, using the refined bound in Theorem S.3 given below in this supplement one obtains

$$\operatorname{var}(\widehat{\alpha}_i) = \sigma^2/N + O(N^{-2}),$$

that is, (3.5) holds.

Theorem 3 allows to establish the convergence rate for the hypercube, but the conditions are too stringent to obtain (3.5). This is so because h_i does not increase fast enough to ensure that $\lambda_2 \, h_i \to \infty$. The following example illustrates that despite $\lambda_2 \to 0$ we can still have $\lambda_2 \, h_i \to \infty$.

Example S.2 (Extended Hypercube graph)**.** *Start with the N-dimensional hypercube $\mathcal{G}$ from the previous example and add edges between all path-two neighbors in $\mathcal{G}$; see the right hand side of Figure S.1 for an example. The resulting graph still has $n = 2^N$ vertices, but now has $N(N+1)\,2^{N-1}$ edges. Here,*

$$d_i = h_i = \frac{N(N+1)}{2}, \qquad \lambda_2 = \frac{4}{N+1},$$

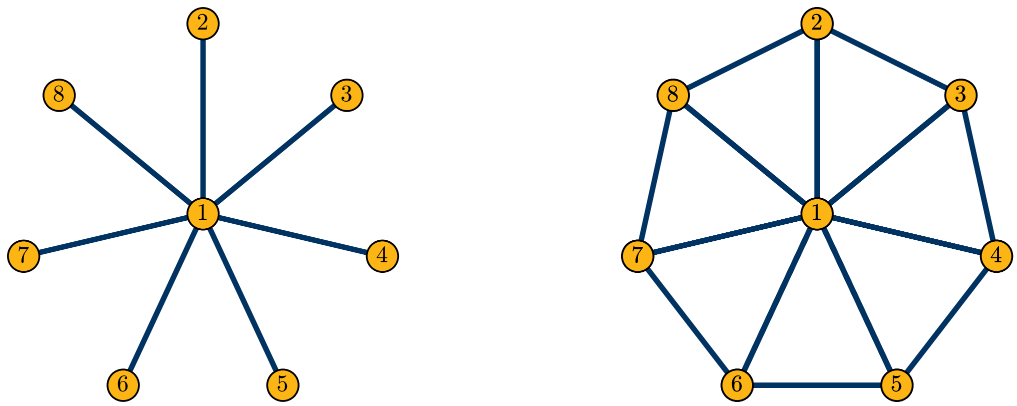

Figure S.2: Star graph (left) and Wheel graph (right) for $n = 8$.

so that $\lambda_2 h_i \to \infty$ *holds, despite* $\lambda_2 \to 0$ *as* $n \to \infty$*. Theorem 3 therefore implies* (3.5) *in this example.*

The next example illustrates that the first-order bounds can still be informative in situations where h_i does not converge to infinity.

Example S.3 (Star graph)**.** *Consider a Star graph around the central vertex* 1*, that is, the graph with* n *vertices and edges*

$$E = \{(1, j) : 2 \le j \le n\};$$

see the left hand side of Figure S.2. Here, $\lambda_2 = 1$ *for any* n *while* $d_1 = n - 1$*,* $h_1 = 1$ *and* $d_i = 1$*,* $h_i = n - 1$ *for* $i \neq 1$*. For* $i = 1$ *one finds that the bounds in Theorem 3 imply that* $\mathrm{var}(\widehat{\alpha}_1) = O(n^{-1})$*, and so*

$$(\widehat{\alpha}_1 - \alpha_1) = O_p\left(n^{-1/2}\right).$$

In contrast, for $i \neq 1$ *we find* $\lambda_2 h_i \to \infty$ *and thus, although* (3.5) *holds, these* α_i *cannot be estimated consistently as* $d_i = 1$*.*

The star graph also illustrates that λ_2 can be large despite having many vertices with small degrees. As mentioned in the main text, it is largely due to this property that we prefer to measure global connectivity by $\lambda_2 = \lambda_2(\boldsymbol{S})$ and not by the "algebraic connectivity" (the second smallest eigenvalue of $\boldsymbol{L}$), which is studied more extensively in the graph theory literature.

Our last example shows the effect on the upper bound in Theorem 3 when neighbors themselves are more strongly connected.

Example S.4 (Wheel graph)**.** *The Wheel graph is obtained on combining a Star graph centered at vertex 1 with a Cycle graph on the remaining $n-1$ vertices; see the right hand side of Figure S.2. Thus, a Wheel graph contains strictly more edges than the underlying Star graph, although none of these involve the central vertex directly. From Butler (2016), we have*

$$\lambda_2 = \min\left\{\frac{4}{3}, 1 - \frac{2}{3}\cos\left(\frac{2\pi}{n}\right)\right\},$$

which satisfies $\lambda_2 \geq 1$ only for $n \leq 4$, and converges to $1/3$ at an exponential rate. However, while, as in the Star graph, $d_1 = n-1$, we now have that $h_i = 3$ for all $i \neq 1$. Hence, $\lambda_2 h_1 > 1$ for any finite n and the upper bound in Theorem 3 is strictly smaller than in the Star graph.

The last two examples also illustrate that adding edges to the graph (to obtain the Wheel graph from the Star graph) can result in a decrease of our measure of global connectivity λ_2. This is not a problem, however, because what really matters to us is not the exact value of λ_2, but that λ_2 is sufficiently different from zero. The Wheel graph with $\lambda_2 \geq 1/3$, for example, clearly describes a very well globally connected graph by that measure.

S.2 Variance bounds for differences

Our focus thus far has been inference on the α_i, under the constraint in (2.2), $\sum_i \alpha_i = 0$. An alternative to normalizing the parameters that may be useful in certain applications is to focus directly on the differences $\alpha_i - \alpha_j$ for all $i \neq j$ (Bradley and Terry, 1952). We give corresponding versions of Theorem 2 and Theorem 3 here.

The resistance distance between vertices i and j in $\mathcal{G}$ is

$$r_{ij} := \left(\boldsymbol{L}^{\dagger}\right)_{ii} + \left(\boldsymbol{L}^{\dagger}\right)_{jj} - 2\left(\boldsymbol{L}^{\dagger}\right)_{ij} \tag{S.1}$$

(Klein and Randić, 1993), and is a metric on the set V (Klein, 2002). It is linked to the commute distance, say c_{ij}, which is the expected time it takes for a random walk to travel from i to j and back again, through the relation

$$c_{ij} = 2m\, r_{ij}\,,$$

see, e.g., von Luxburg, Radl and Hein (2010). For example, vertices in different clusters of a graph have a large commute distance, relative to vertices in the same cluster of the graph. The precise connection between the magnitude of these quantities and the precision of statistical inference is

$$\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) = \sigma^2\, r_{ij} = \frac{\sigma^2}{2}\frac{c_{ij}}{m}. \tag{S.2}$$

This is the equivalent of (3.1) for differences.

The counterpart to Theorem 2 is intuitive.

Theorem S.1 (Global bound for differences)**.** *Let $\mathcal{G}$ be connected. Then*

$$\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \leq \left(\frac{1}{d_i} + \frac{1}{d_j}\right)\frac{\sigma^2}{\lambda_2},$$

for all $i \neq j$.

Let $d_{ij} := |[i] \cap [j]|$ be the number of vertices that are neighbors of both i and j. Write

$$h_{ij} := \begin{cases} \left(\dfrac{1}{d_{ij}} \displaystyle\sum_{k \in [i]\cap[j]} \dfrac{1}{d_k}\right)^{-1} & \text{for } d_{ij} \neq 0, \\ \infty & \text{for } d_{ij} = 0, \end{cases}$$

for the corresponding harmonic mean of the degrees of the vertices $k \in [i] \cap [j]$. For weighted graphs we need to change the definition of h_{ij} for $d_{ij} \neq 0$ to

$$h_{ij} = \left(\frac{1}{d_{ij}} \sum_{k \in V} \frac{(\boldsymbol{A})_{ik}\,(\boldsymbol{A})_{jk}}{d_k}\right)^{-1},$$

with $d_{ij} = \sum_{k\in V} (\boldsymbol{A})_{ik}\,(\boldsymbol{A})_{jk}$. We have the following theorem.

Theorem S.2 (First-order bound for differences)**.** *Let $\mathcal{G}$ be connected. Then*

$$\sigma^2\left(\frac{1}{d_i} + \frac{1}{d_j} - \frac{2(\boldsymbol{A})_{ij}}{d_i d_j}\right) \leq \operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \leq \sigma^2\left(\frac{1}{d_i} + \frac{1}{d_j} - \frac{2(\boldsymbol{A})_{ij}}{d_i d_j}\right) + \frac{\sigma^2}{\lambda_2}\left(\frac{1}{d_i h_i} + \frac{1}{d_j h_j} - \frac{2\, d_{ij}}{d_i d_j h_{ij}}\right)$$

One implication of the theorem is that, when $[i] = [j]$ but $i \notin [j]$ and $i \notin [j]$, that is, when vertices i and j share exactly the same neighbors and are not connected themselves, we have

$$\mathrm{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) = \sigma^2 \left(\frac{1}{d_i} + \frac{1}{d_j} \right), \tag{S.3}$$

as, in that case, both $(\boldsymbol{A})_{ij}$ and the second term in the upper bound in Theorem S.2 are zero.

Theorem S.2 is related to work on the amplified commute distance by von Luxburg, Radl and Hein (2014), which they propose as an alternative to the commute distance in large graphs. However, their results are restricted to the class of random geometric graphs and are purely asymptotic in nature. Here, we provide finite-sample bounds for arbitrary connected graphs, using λ_2 as a measure of global connectivity.

S.3 Second-order bound

This section discusses an improvement on the bounds in Theorem 3. Recall that $d_{ij} = |[i] \cap [j]|$ denotes the number of vertices that are direct neighbors of both vertex i and vertex j. For $j \in [i]$, let $\underline{d}_{j,i} := d_j - d_{ij}$, the number of direct neighbors of j that are not also direct neighbors of i. The following example illustrates that $\underline{d}_{j,i}$ can be a more relevant measure than d_j for the dependence of $\mathrm{var}(\widehat{\alpha}_i)$ on the connectedness of a neighbor j of i.

Example S.5. *Both for the Star and for the Wheel graph example from above one finds*

$$\mathrm{var}(\widehat{\alpha}_1) = \frac{\sigma^2}{n} \frac{n-1}{n}$$

by direct calculation. Thus, the additional edges in the Wheel graph between the neighbors of vertex $i = 1$ relative to the Star graph do not lower the variance of $\widehat{\alpha}_1$. For $i \neq 1$ we have $d_i = 1$ for the Star graph but $d_i = 3$ for the Wheel graph, while for both graphs we have $\underline{d}_{i,1} = 1$.

Let

$$[i]_2 := \bigcup_{j \in [i]} [j] \setminus \{i\},$$

the set of all path-two neighbors of vertex i. Analogous to the definition of the harmonic mean h_i above we let

$$\underline{h}_i := \left(\frac{1}{d_i} \sum_{j \in [i]} \frac{1}{\underline{d}_{j,i}} \right)^{-1}, \qquad h_{i;2} := \left(\frac{1}{|[i]_2|} \sum_{j \in [i]_2} \frac{1}{d_j} \right)^{-1}.$$

In addition, for $i \in V$ we define the set

$$W_i = \left\{ (j,k,\ell) \in V^3 \ : \ k \neq i \ \& \ (i,j) \in E \ \& \ (j,k) \in E \ \& \ (k,\ell) \in E \right\},$$

which is the set of all triplets (j,k,ℓ) such that (i,j,k,ℓ,i) is a closed walk in $\mathcal{G}$ that reaches distance two from i (thus ruling out $k = i$). Notice that we may have $j = \ell$, that is, the closed walk need not be a simple cycle.

Theorem S.3 (Second-order bound)**.** *Let $\mathcal{G}$ be connected and let $\underline{h}_i > 1$. Then*

$$\frac{\sigma^2}{d_i(1 - h_i^{-1})} \left(1 - \frac{2}{n} - \frac{2}{n} \frac{d_i}{\underline{h}_i} \right) \ \leq \ \mathrm{var}(\widehat{\alpha}_i) \ \leq \ \frac{\sigma^2}{d_i(1 - \underline{h}_i^{-1})} \left(\left(1 - \frac{2}{n} - \frac{2}{n} \frac{d_i}{\underline{h}_i} \right) + \frac{C_i}{\lambda_2 \, h_{i;2} (\underline{h}_i - 1)} \right)$$

where $C_i := \underline{h}_i \ h_{i;2} \ d_i^{-1} \sum_{(j,k,\ell) \in W_i} \left(d_k \, \underline{d}_{j,i} \, \underline{d}_{\ell,i} \right)^{-1}$.

Including the factor $\underline{h}_i \ h_{i;2} \ d_i^{-1}$ in the definition of C_i guarantees that C_i is naturally scaled in many examples; see below.

An asymptotic implication of Theorem S.3 is that

$$\frac{\sigma^2}{d_i(1 - h_i^{-1})} + O\left(\frac{1}{\min(d_i, \underline{h}_i) \, n} \right) \ \leq \ \mathrm{var}(\widehat{\alpha}_i) \ \leq \ \frac{\sigma^2}{d_i(1 - \underline{h}_i^{-1})} + O\left(\frac{1}{\min(d_i, \underline{h}_i) \, n} \right) + o(d_i^{-1} \underline{h}_i^{-1}), \tag{S.4}$$

provided $\lambda_2 h_{i;2} / C_i \to \infty$ as $n \to \infty$ and $\underline{h}_i \geq 1 + \epsilon$ for some constant $\epsilon > 0$ independent of n. Notice that this does not require that $\underline{h}_i \to \infty$, and the refinement obtained here relative to the first order asymptotic result (3.5) is in fact particularly important for those cases where h_i and $\underline{h}_i$ are small.

The term C_i requires further discussion. Notice that $(j,k,\ell) \in W_i$ implies $j, \ell \in [i]$ and $k \in [i]_2$, and for any tensor $a_{ijk\ell}$ we have

$$\sum_{(j,k,\ell) \in W_i} a_{ijk\ell} = \sum_{k \in [i]_2} \ \sum_{j \in [i] \cap [k]} \ \sum_{\ell \in [i] \cap [k]} a_{ijk\ell}. \tag{S.5}$$

Applying this to $a_{ijk\ell} = 1$ and using that $\sum_{j \in [i] \cap [k]} = d_{ik}$ we obtain

$$|W_i| = \sum_{k \in [i]_2} d_{ik}^2.$$

Thus, the number of elements in W_i depends on the number of path-two neighbors of i and on the typical number of neighbors that i has in common with one of its path-two neighbors. The cases of interest in the following are those where the typical value of d_{ik} is small compared to the degree d_i for $k \in [i]_2$, so that the ratio between $|W_i|$ and $|[i]_2|$ is not large. This is true in many interesting examples. Applying (S.5) to C_i gives

$$\begin{aligned} C_i &= \underline{h}_i \, h_{i;2} \, \frac{1}{d_i} \sum_{k \in [i]_2} \frac{1}{d_k} \left(\sum_{j \in [i] \cap [k]} \frac{1}{\underline{d}_{j,i}} \right)^2 \\ &= \frac{|[i]_2|}{d_i \, \underline{h}_i} \left[\frac{1}{|[i]_2|} \sum_{k \in [i]_2} d_{ik}^2 \left(\frac{h_{i;2}}{d_k} \right) \left(\frac{1}{d_{ik}} \sum_{j \in [i] \cap [k]} \frac{\underline{h}_i}{\underline{d}_{j,i}} \right)^2 \right]. \end{aligned}$$

Using the last result we want to argue that C_i is of order one in cases where d_{ik} is not large for $k \in [i]_2$. To do so, first notice that the sums in the last expression for C_i are all self-normalized (i.e., divided by the number of terms that is summed over). We also typically have $\frac{|[i]_2|}{d_i \, \underline{h}_i} = O(1)$, because

$$\frac{|[i]_2|}{d_i} \leq \frac{1}{d_i} \sum_{j \in [i]} \underline{d}_{j,i},$$

and one expects the arithmetic mean $\left(\frac{1}{d_i} \sum_{j \in [i]} \underline{d}_{j,i} \right)$ to be of the same order as the harmonic mean $\underline{h}_i$.

In the following we present concrete examples where d_{ik} is relatively small for $k \in [i]_2$ and thus C_i is of order one asymptotically.

Example 2 (cont'd). *Consider the Erdős and Rényi (1959) random-graph model with $p_n = c\,(\ln n)/n$. Let $c > 1$ to guarantee that the graph is connected as $n \to \infty$. In this model for randomly picked $(i,j) \in E$ we have $\underline{d}_{j,i} = d_i[1 + O(p_n)]$, that is, the difference between $\underline{d}_{j,i}$ and d_i is typically very small. Also, for randomly picked $i \in V$ and $k \in [i]_2$*

we have $d_{ik} = 1 + O(np_n^2)$*, and therefore* $|W_i| = |[i]_2|\,[1 + O(np_n^2)] = n^2p_n^2 + O(n^3p_n^4)$*. We therefore have* $\lambda_2 \to 1$*,* $d_i/(\ln n) \to c$*,* $\underline{d}_{j,i}/(\ln n) \to c$*,* $h_i/(\ln n) \to c$*,* $\underline{h}_i/(\ln n) \to c$*,* $\underline{h}_{i;2}/(\ln n) \to c$ *and* $C_i \to 1$*, almost surely, as* $n \to \infty$*. Applying Theorem S.3 thus gives*

$$\operatorname{var}(\widehat{\alpha}_i) = \frac{\sigma^2}{d_i(1 - h_i^{-1})} + O\left(d_i^{-1} h_i^{-1} h_{i;2}^{-1}\right),$$

which is simpler than (S.4), because in this example 3-cycles are relatively rare, implying that h_i *and* $\underline{h}_i$ *are typically very close to each other.* □

Example 1 (cont'd) (Bipartite graph applied to matched amployer-employee data)**.** *In the worker-firm example the graph* $\mathcal{G}$ *is bipartite, so that two neighboring vertices have no direct neighbors in common, implying that* $\underline{d}_{i,j} = d_i$ *and* $\underline{h}_i = h_i$*. Let* $i \in V_2$ *be a firm. Then,* $j \in [i]$ *are workers, and the number of observations* d_j *for workers are typically small in this application, so that the harmonic mean* h_i *is typically small. Also,* $j \in [i]_2$ *are firms, and the number of observations* d_j *for firms are often large in this application, so the harmonic mean* $h_{i;2}$ *is often large. Therefore, the second-order bound in Theorem S.3 is particularly simple in this example (because the distinction between* $\underline{d}_{i,j}$ *and* d_i *is irrelevant), and is also particularly important (because* $\underline{h}_i = h_i$ *is small, so that the improvement relative to the first-order bound is very relevant). For simplicity, we consider the case of a simple graph where* $d_j = 2$ *for all workers* $j \in V_1$*.*[1] *Then, for* $i \in V_2$ *the bounds in Theorem S.3 become*

$$\frac{2\,\sigma^2}{d_i}\left(1 - \frac{2}{n} - \frac{d_i}{n}\right) \le \operatorname{var}(\widehat{\alpha}_i) \le \frac{2\sigma^2}{d_i}\left(1 - \frac{2}{n} - \frac{d_i}{n}\right) + \frac{2\,\sigma^2\,C_i}{\lambda_2\,d_i\,h_{i;2}},$$

where

$$C_i = \frac{h_{i;2}}{2}\frac{1}{d_i}\sum_{j\in[i]_2}\frac{d_{ij}^2}{d_j} \le \frac{1}{2}\max_{j\in[i]_2} d_{ij}^3\,,$$

where for the last inequality we used the definition of $h_{i;2}$ *and* $|[i]_2| \le d_i \max_{j\in[i]_2} d_{ij}$*. For example, if any two firms are connected by at most two workers, then we have* $d_{ij} \le 1$ *and*

[1] This occurs if we observe wages annually for two years, and we drop workers from the dataset that do not change firms in those two years, because their observations are not informative for the firm fixed effects. For all remaining workers we then have exactly $d_j = 2$ log wage observations and the graph is simple.

therefore $C_i = 1/2$. Thus, the leading order asymptotic variance is increased by a factor of two compared to the first order result in (3.5). □

It is also possible that Theorem S.3 cannot be used to obtain a refinement of the variance as in (S.4) but that it can justify the first-order rate in (3.5) for cases where this first-order asymptotic variance of $\widehat{\alpha}_i$ does not follow from Theorem 3. The following example illustrates this in the context of the hypercube example from above.

Example S.6. *For $N \geq 2$ consider the N-dimensional hypercube graph, which has $n = 2^N$ edges, as introduced above. In that case, firstly, we have $d_i = N$ for all $i \in V$. Secondly, there are no edges among the vertices in $[i]$, implying that $\underline{d}_{i,j} = d_i = N$ and $h_i = \underline{h}_i = N$ for all possible $i, j \in V$. Thirdly, we have $|[i]_2| = N(N-1)/2$, and for all $i \in V$ and $k \in [i]_2$ we have $d_{ik} = 2$ implying that $|W_i| = 4\,|[i]_2| = 2N(N-1)$. We thus find $C_i = 2(N-1)/N$. The bounds in Theorem S.3 thus become*

$$\frac{\sigma^2}{N\,(1-N^{-1})}\left(1-\frac{4}{2^N}\right) \;\leq\; \mathrm{var}(\widehat{\alpha}_i) \;\leq\; \frac{\sigma^2}{N\,(1-N^{-1})}\left(1-\frac{4}{2^N}+\frac{2}{\lambda_2\,N^2}\right).$$

Because $\lambda_2 = 2/N$ we thus find,

$$\mathrm{var}(\widehat{\alpha}_i) = \frac{\sigma^2}{N} + O(N^{-2}),$$

as $N \to \infty$.

S.4 Estimation of moments

S.4.1 Consistency

Let $k \in \{2, 3, 4, \ldots\}$. The kth moment of the α_i equals

$$\theta := n^{-1} \sum_{i=1}^{n} \alpha_i^k.$$

The variance discussed in the main text is obtained on setting $k = 2$. The plug-in estimator is

$$\widehat{\theta} := n^{-1} \sum_{i=1}^{n} \widehat{\alpha}_i^k.$$

The following theorem gives sufficient conditions for consistency.

Theorem S.4 (Consistency of moment estimator)**.** *Let $\mathcal{G}$ be connected. Let $k \in \{2, 3, 4, \ldots\}$. Assume that $\limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} |\alpha_i|^k < \infty$ and that*

$$\frac{\lambda_2}{\left(n^{-1} \sum_{i=1}^{n} \frac{1}{d_i^{k/2}}\right)^{2/k}} \to \infty$$

as $n \to \infty$. Then $(\widehat{\theta} - \theta) \xrightarrow{p} 0$ as $n \to \infty$.

For $k = 2$ the condition in the theorem is simply that $\lambda_2 h \to \infty$ as $n \to \infty$. For $k > 2$ the asymptotic requirement on the degrees sequence becomes stronger, because a large variance of $\widehat{\alpha}_i$ for a single vertex i can have a stronger effect on the moment plug-in estimator when k is large. We have

$$\lim_{k\to\infty} \frac{\lambda_2}{\left(n^{-1} \sum_{i=1}^{n} \frac{1}{d_i^{k/2}}\right)^{2/k}} = \lambda_2 \min_i d_i$$

which is completely driven by the least-connected vertex.

S.4.2 Bias and variance of $\widehat{\vartheta}$

In the following we discuss the case $k = 2$, with $\vartheta = \theta$ and $\widehat{\vartheta} = \widehat{\theta}$ already introduced in the main text for this case. We are going to provide additional detail on the bias and variance of the variance estimator $\widehat{\vartheta}$.

Lemma S.5 (Bias and variance of $\widehat{\vartheta}$)**.** *Let $\mathcal{G}$ be connected. Then,*

$$\mathbb{E}(\widehat{\vartheta} - \vartheta) = \sigma^2 \frac{\operatorname{tr}(\boldsymbol{L}^\dagger)}{n}, \qquad \operatorname{var}(\widehat{\vartheta}) = 4\,\sigma^2 \frac{\boldsymbol{\alpha}' \boldsymbol{L}^\dagger \boldsymbol{\alpha}}{n^2} + 2\,\sigma^4 \frac{\|\boldsymbol{L}^\dagger\|^2}{n^2}.$$

With r_{ij} the resistance distance as defined in (S.1) the bias expression in the lemma is proportional to

$$n \operatorname{tr}\left(\boldsymbol{L}^\dagger\right) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij},$$

which is called the resistance or "Kirchhoff index" of the graph (see e.g. Klein and Randić 1993; Boumal and Cheng 2014).

Because

$$\sigma^2 \frac{\mathrm{tr}(\boldsymbol{L}^\dagger)}{n} = n^{-1} \sum_{i=1}^{n} \mathrm{var}(\widehat{\alpha}_i)$$

we can apply the bounds on $\mathrm{var}(\widehat{\alpha}_i)$ in Theorem 3 to obtain the bounds on the bias of $\widehat{\vartheta}$ given in the following corollary.

Corollary S.6 (Bounds on Bias of $\widehat{\vartheta}$)**.** *Let $\mathcal{G}$ be connected. Then,*

$$\frac{\sigma^2}{h}\left(1 - \frac{2}{n}\right) \leq \mathbb{E}(\widehat{\vartheta} - \vartheta) \leq \frac{\sigma^2}{h}\left(1 - \frac{2}{n} + \frac{1}{\lambda_2 H}\right).$$

Moving on, because $\widehat{\alpha}_i$ is normally distributed with mean α_i we have $\mathrm{var}\,(\widehat{\alpha}_i^2) = 2[\mathrm{var}\,(\widehat{\alpha}_i)]^2 + 4\alpha_i^2 \mathrm{var}\,(\widehat{\alpha}_i)$. Using this we find the following simple bound on the variance

$$\begin{aligned}
\mathrm{var}(\widehat{\vartheta}) &= \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\alpha}_i^2\right) \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n} \sqrt{\mathrm{var}\,(\widehat{\alpha}_i^2)}\right)^2 \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} \sqrt{2[\mathrm{var}\,(\widehat{\alpha}_i)]^2 + 4\alpha_i^2 \mathrm{var}\,(\widehat{\alpha}_i)}\right)^2 \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n} \sqrt{2\,[\mathrm{var}\,(\widehat{\alpha}_i)]^2} + \frac{1}{n}\sum_{i=1}^{n} 2\,\alpha_i \sqrt{\mathrm{var}\,(\widehat{\alpha}_i)}\right)^2 \\
&\leq \left(\sqrt{2}\,\frac{1}{n}\sum_{i=1}^{n} \mathrm{var}\,(\widehat{\alpha}_i) + 2\,\left(\frac{1}{n}\sum_{i=1}^{n} \alpha_i^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{var}\,(\widehat{\alpha}_i)\right)^{1/2}\right)^2 \\
&= \left(\sqrt{2}\,\mathbb{E}(\widehat{\vartheta} - \vartheta) + 2\,\vartheta^{1/2}\left(\mathbb{E}(\widehat{\vartheta} - \vartheta)\right)^{1/2}\right)^2.
\end{aligned}$$

From this bound on $\mathrm{var}(\widehat{\vartheta})$ together with Corollary S.6 above we find that for asymptotic sequences where ϑ converges to a constant and $\lambda_2 H \to \infty$, as $n \to \infty$, we have

$$\mathbb{E}(\widehat{\vartheta} - \vartheta) = \frac{\sigma^2}{h} + o(h^{-1}), \qquad \mathrm{var}(\widehat{\vartheta}) = O(h^{-1}).$$

However, this upper bound on $\mathrm{var}(\widehat{\vartheta})$ of order h^{-1} is potentially crude.

In the following we provide further bounds on $\mathrm{var}(\widehat{\vartheta})$. We first introduce some additional notation. Let $\boldsymbol{v} = (v_1, \ldots, v_n)'$ be an eigenvector of $\boldsymbol{S}$ corresponding to the eigenvalue λ_2

and let

$$\kappa := \frac{\sum_{i=1}^{n} v_i\, \alpha_i\, d_i^{-1/2}}{\sqrt{\left(\sum_{i=1}^{n} v_i^2\right)\left(\sum_{i=1}^{n} \alpha_i^2\, d_i^{-1}\right)}}.$$

Note that $\kappa \in [-1, 1]$ is the uncentered correlation coefficient of v_i and $\alpha_i/\sqrt{d_i}$. We also introduce

$$h_\alpha := \left(\sum_{i=1}^{n} \frac{(\alpha_i^2/\vartheta)/n}{d_i}\right)^{-1}, \qquad h_{\text{sqr}} := \left(\sum_{i=1}^{n} \frac{1/n}{d_i^2}\right)^{-1/2}.$$

Theorem S.7 (Bounds on variance of $\widehat{\vartheta}$)**.** *Let $\mathcal{G}$ be connected. Then,*

$$\frac{\kappa^2}{n}\,\frac{4\,\vartheta\,\sigma^2}{\lambda_2\, h_\alpha} \le \operatorname{var}(\widehat{\vartheta}) \le \frac{1}{n}\left(\frac{4\,\vartheta\,\sigma^2}{\lambda_2\, h_\alpha} + \frac{2\,\sigma^4}{(\lambda_2\, h_{\text{sqr}})^2}\right).$$

An implication of these bounds is that, if, as $n \to \infty$, (i) ϑ and κ converge to non-zero constants; (ii) h, h_α, and h_{sqr} grow at the same rate; and (iii) $\lambda_2 h \to \infty$, then

$$\operatorname{var}(\widehat{\vartheta}) \asymp (\lambda_2\, n h)^{-1}.$$

It is not difficult to construct data generating processes for $\mathcal{G}$ and $\boldsymbol{\alpha}$ satisfying the above conditions and also $\lambda_2 \to 0$, implying that the convergence rate of $\operatorname{var}(\widehat{\vartheta})$ is indeed given by $(\lambda_2\, n\, h)^{-1}$.

The above conclusion requires that κ converges to a non-zero constant. In situations where κ is small or converges to zero, the lower bound in Theorem S.7 is uninformative and the variance may converge faster than $(\lambda_2\, n\, h)^{-1}$. Such a small correlation may arise when either v_i or $\alpha_i/\sqrt{d_i}$ have approximately mean zero and the correlation between v_i and $\alpha_i/\sqrt{d_i}$ is small. This will be the case, for example, if $\mathcal{G}$ and $\boldsymbol{\alpha}$ are independent. We therefore now also present an alternative version of the variance bounds. Let

$$h_{\text{cub}} := \left(\sum_{i=1}^{n} \frac{1/n}{d_i^3}\right)^{-1/3}.$$

For $\ell = 0, 1, 2, \ldots$ let

$$\begin{aligned}
\boldsymbol{\alpha}^{(\ell)} &:= \left(\boldsymbol{A}\boldsymbol{D}^{-1}\right)^{\ell} \boldsymbol{\alpha}, \\
\vartheta^{(\ell)} &:= \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{\alpha}_i^{(\ell)}\right)^2, \\
h_\alpha^{(\ell)} &:= \left(\frac{1}{\sum_{i=1}^{n} (\vartheta_i^{(\ell)})^2} \sum_{i=1}^{n} (\vartheta_i^{(\ell)})^2 d_i^{-1} \right)^{-1}, \\
\gamma^{(\ell)} &:= \operatorname{corr}\left(\frac{\alpha_i^{(\ell)}}{d_i}, \frac{\alpha_j^{(\ell)}}{d_j} \,\middle|\, (i,j) \in E \right) = \frac{\sum_{(i,j)\in E} \frac{\alpha_i^{(\ell)}}{d_i} \frac{\alpha_j^{(\ell)}}{d_j}}{\sqrt{\sum_{(i,j)\in E} \left(\frac{\alpha_i^{(\ell)}}{d_i}\right)^2} \sqrt{\sum_{(i,j)\in E} \left(\frac{\alpha_j^{(\ell)}}{d_j}\right)^2}},
\end{aligned}$$

where in the last expression for the sample correlation we do not have to subtract any sample mean terms, because we have $\sum_{(i,j)\in E} d^{-1}\alpha_i^{(\ell)} = \frac{1}{2}\sum_{i,j=1}^{n} (\boldsymbol{A})_{ij} d^{-1}\alpha_i^{(\ell)} = \frac{1}{2}\sum_{i=1}^{n} \alpha_i^{(\ell)} = 0$. If $\boldsymbol{\alpha}^{(\ell)} = 0$, then the above expression for $h_\alpha^{(\ell)}$ and $\gamma^{(\ell)}$ are ill-defined, and we therefore set $h_\alpha^{(\ell)} = h$ and $\gamma^{(\ell)} = 0$ in that case. We have $\vartheta^{(\ell)} \geq 0$ and $\gamma^{(\ell)} \in [-1, 1]$.

Theorem S.8 (Bounds on variance of $\widehat{\vartheta}$)**.** *Let $\mathcal{G}$ be connected. Let $k \in \{1, 2, 3, \ldots\}$. Then,*

$$0 \leq \operatorname{var}(\widehat{\vartheta}) - \frac{4\,\sigma^2}{n} \sum_{\ell=0}^{k-1} \vartheta^{(\ell)} \frac{1 + \gamma^{(\ell)}}{h_\alpha^{(\ell)}} \leq \frac{4\,\sigma^2\,\vartheta^{(k)}}{\lambda_2\, n\, h_\alpha^{(k)}} + \frac{2\,\sigma^4}{n} \left(\frac{1}{h_{\text{sqr}}} + \frac{1}{\lambda_2 h_{\text{cub}}^{3/2}} \right)^2.$$

Comment: For the first term of the variance expression in Lemma S.5 the theorem uses the bounds

$$0 \leq n^{-1}\boldsymbol{\alpha}'\boldsymbol{L}^{\dagger}\boldsymbol{\alpha} - \sum_{\ell=0}^{k-1} \vartheta^{(\ell)} \frac{1 + \gamma^{(\ell)}}{h_\alpha^{(\ell)}} \leq \frac{\vartheta^{(k)}}{\lambda_2\, h_\alpha^{(k)}}. \tag{S.6}$$

If $\mathcal{G}$ is <u>not</u> bipartite, then we have $\frac{\vartheta^{(k)}}{\lambda_2\, h_\alpha^{(k)}} \to 0$ as $k \to \infty$, that is, those bounds on $n^{-1}\boldsymbol{\alpha}'\boldsymbol{L}^{\dagger}\boldsymbol{\alpha}$ provide an exact expansion for $k \to \infty$.[2] However, if $\mathcal{G}$ is bipartite this is usually not the

[2]For non-bipartite graphs we can expand the inverse of $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ to find $\boldsymbol{\alpha}'\boldsymbol{L}^{\dagger}\boldsymbol{\alpha} = \boldsymbol{\alpha}'\boldsymbol{D}^{-1}\sum_{\ell=0}^{\infty}\left(\boldsymbol{A}\boldsymbol{D}^{-1}\right)^{\ell}\boldsymbol{\alpha}$, which gives rise to the above expansion of $\operatorname{var}(\widehat{\vartheta})$ as $k \to \infty$. This expansion is convergent, because $\|\boldsymbol{A}\boldsymbol{D}^{-1}\|_2 \leq 1$, -1 is not an eigenvalue of $\boldsymbol{A}\boldsymbol{D}^{-1}$, and the nondegenerate eigenvalue 1 of $\boldsymbol{A}\boldsymbol{D}^{-1}$ has eigenvector $\boldsymbol{\iota}_n$, orthogonal to $\boldsymbol{\alpha}$. We know this, because $\boldsymbol{A}\boldsymbol{D}^{-1}$ and $\boldsymbol{I}_n - \boldsymbol{S}$ have the same spectrum, and the spectrum of $\boldsymbol{S}$ is discussed in Lemma 1.7 of Chung (1997). Part (v) of that

case. For example, let $\mathcal{G}$ be bipartite with vertex components V_1 and V_2 and set $\alpha_i = d_i$ for $i \in V_1$ and $\alpha_i = -d_i$ for $i \in V_2$. We then have $\boldsymbol{\alpha}^{(\ell)} = (-1)^\ell \boldsymbol{\alpha}$, $\vartheta^{(\ell)} = \vartheta$, $h_\alpha^{(\ell)} = h_\alpha$ and $\gamma^{(\ell)} = -1$, so that the above bounds on $n^{-1}\boldsymbol{\alpha}'\boldsymbol{L}^\dagger\boldsymbol{\alpha}$ become completely independent from the choice of k. Nevertheless, the bounds are valid for bipartite $\mathcal{G}$, and for generic choices of $\boldsymbol{\alpha}$ are also more informative the larger is k. □

Consider Theorem S.8 with $k = 0$. Assume that ϑ and $1 + \gamma$ converge to non-zero constants, and h, h_α, $h_\alpha^{(1)}$ h_{sqr} and h_{cub} all grow to infinity at the same rate, and $\lambda_2 \, h \to \infty$, and $\lambda_2^{-1}\vartheta^{(1)} \to 0$, as $n \to \infty$. Then we have

$$\operatorname{var}(\widehat{\vartheta}) = \frac{4\,\vartheta\,\sigma^2\,(1+\gamma)}{n\,h_\alpha} + o((n\,h_\alpha)^{-1}).$$

This conclusion requires that $\vartheta^{(1)}$ is small, that is, $\boldsymbol{\alpha}^{(1)}$ is close to zero. We have $\alpha_i^{(1)} = \sum_{j\in[i]} d_j^{-1}\alpha_j$, which can be expected to be small if d_i and the typical d_j, $j \in [i]$, are similarly large, and if the α_j, $j \in [i]$, have a mean close to the overall sample mean of zero, and are not strongly correlated with d_j^{-1}. This would, for example, be the case if $\mathcal{G}$ and $\boldsymbol{\alpha}$ are independent, and the degrees of $\mathcal{G}$ are not too heterogeneous.

More generally, one can show that $\vartheta = \vartheta^{(0)} \geq \vartheta^{(1)} \geq \vartheta^{(2)} \geq \ldots$, but it is not necessarily the case that $\vartheta^{(\ell)}$, for $\ell \geq 1$, is much smaller than ϑ, so $\frac{4\,\sigma^2}{n}\vartheta^{(\ell)}\frac{1+\gamma^{(\ell)}}{h_\alpha^{(\ell)}}$ may also give contributions to $\operatorname{var}(\widehat{\vartheta})$ of order $(n\,h)^{-1}$. Theorem S.8 may then still be employed to show that $\operatorname{var}(\widehat{\vartheta})$ is of order $(nh)^{-1}$ by letting $k = k_n$ grow to infinity as $n \to \infty$, but we will not discuss this any further here.

To conclude, for generic realizations of $\mathcal{G}$ and $\boldsymbol{\alpha}$ we do not expect the correlation γ to be close to -1, so that the lower bound in Theorem S.8 (for $k = 1$) shows that $\operatorname{var}(\widehat{\vartheta})$ is at least of order $(n\,h)^{-1}$, while the upper bound in Theorem S.7 shows that it is at most of order $(\lambda_2\,n\,h)^{-1}$. However, if $\lambda_2 \to 0$, then it depends very much on the detailed data generating process for $\mathcal{G}$ and $\boldsymbol{\alpha}$ which of these two rates is the actual convergence rate of $\operatorname{var}(\widehat{\vartheta})$, and any rate in between is also possible in general.

---

lemma states that $\boldsymbol{S}$ has an eigenvalue 2 if and only if $\mathcal{G}$ is bipartitie. Thus, for bipartite graphs $\boldsymbol{A}\boldsymbol{D}^{-1}$ has an eigenvalue -1 and the above expansion for $\boldsymbol{\alpha}'\boldsymbol{L}^\dagger\boldsymbol{\alpha}$ is not convergent, unless $\boldsymbol{\alpha}$ is orthogonal to the corresponding eigenvector.

S.5 Proofs

PROOF OF LEMMA 1 (EXISTENCE)

The estimator is defined by the constraint minimization problem in (2.4). For convenience we express the constraint in quadratic form, $(\boldsymbol{a}'\boldsymbol{\iota}_n)^2 = 0$. By introducing the Lagrange multiplier $\lambda > 0$ we can write

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{a}\in\mathbb{R}^n} \; (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}) + \lambda\,(\boldsymbol{a}'\boldsymbol{\iota}_n)^2 .$$

Solving the corresponding first-order condition we obtain

$$\widehat{\boldsymbol{\alpha}} = (\boldsymbol{B}'\boldsymbol{B} + \lambda\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n')^{-1}\,\boldsymbol{B}'\boldsymbol{y}.$$

Here, the matrix $\boldsymbol{B}'\boldsymbol{B} + \lambda\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'$ is invertible, because $\boldsymbol{L} = \boldsymbol{B}'\boldsymbol{B}$ only has a single zero eigenvalue (because we assume the graph to be connected) with eigenvector $\boldsymbol{\iota}_n$, so that adding $\lambda\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'$ gives a non-degenerate matrix. The matrices $\boldsymbol{B}'\boldsymbol{B}$ and $\boldsymbol{\iota}_n\boldsymbol{\iota}_n'$ commute, and by properties of the Moore-Penrose inverse we thus have

$$(\boldsymbol{B}'\boldsymbol{B} + \lambda\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n')^{-1} = (\boldsymbol{B}'\boldsymbol{B})^{\dagger} + \lambda^{-1}\,(\boldsymbol{\iota}_n\boldsymbol{\iota}_n')^{\dagger} .$$

We furthermore have $(\boldsymbol{\iota}_n\boldsymbol{\iota}_n')^{\dagger} = n^{-2}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'$ and, because $\boldsymbol{B}\boldsymbol{\iota}_n = 0$, the contribution from $(\boldsymbol{\iota}_n\boldsymbol{\iota}_n')^{\dagger}$ drops out of the above formula for $\widehat{\boldsymbol{\alpha}}$, and we obtain $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{B}'\boldsymbol{y}$. This concludes the proof. ■

PROOF OF THEOREM 1 AND EQ. (2.5) (SAMPLING DISTRIBUTION)

As $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{\alpha} + \boldsymbol{u}$, Lemma 1 gives

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + (\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{B}'\boldsymbol{u}.$$

Conditional on $\boldsymbol{B}$, $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2\,\boldsymbol{I}_n)$, and so

$$\widehat{\boldsymbol{\alpha}} \sim N\left(\boldsymbol{\alpha},\; \sigma^2\,(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\right),$$

where the variance expression follows from properties of the Moore-Penrose pseudoinverse. This concludes the proof of the theorem.

The result in display (2.5) follows from Theorem 1 by standard arguments on the F-statistic in linear regression models. Here, the degrees-of-freedom correction from $m-n$ to $m-(n-1)$ arises, because the projection matrix

$$\boldsymbol{I}_m - \boldsymbol{B}\left(\boldsymbol{B}'\boldsymbol{B}\right)^{\dagger}\boldsymbol{B}'$$

has rank $m-(n-1)$. Notice that although $\boldsymbol{B}$ has n columns, we have that rank $\boldsymbol{B} = (n-1)$. This concludes the proof. ■

PROOF OF THEOREMS 2 AND S.1 (ZERO-ORDER BOUNDS)

There are no isolated vertices, because $\mathcal{G}$ is connected and $n > 2$. That is, $d_i > 0$ for all i, and so $\boldsymbol{D}$ is invertible. From Theorem 1 and the definition of the normalized Laplacian $\boldsymbol{S}$ we find

$$\operatorname{var}(\widehat{\boldsymbol{\alpha}}) = \sigma^2\, \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{S}^{\dagger}\boldsymbol{D}^{-\frac{1}{2}}.$$

In the following we write $\boldsymbol{M}_1 \leq \boldsymbol{M}_2$ for symmetric matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ to indicate that $\boldsymbol{M}_2 - \boldsymbol{M}_1$ is positive semi-definite. We have $\boldsymbol{S}^{\dagger} \leq \lambda_2^{-1}\boldsymbol{I}_n$, because λ_2 is the smallest non-zero eigenvalue of $\boldsymbol{S}$. Therefore,

$$\operatorname{var}(\widehat{\boldsymbol{\alpha}}) \leq \frac{\sigma^2}{\lambda_2}\boldsymbol{D}^{-1}.$$

This result implies that, for any vector $\boldsymbol{v} \in \mathbb{R}^n$,

$$\operatorname{var}(\boldsymbol{v}'\widehat{\boldsymbol{\alpha}}) = \boldsymbol{v}'\operatorname{var}(\widehat{\boldsymbol{\alpha}})\,\boldsymbol{v} \leq \frac{\sigma^2}{\lambda_2}\boldsymbol{v}'\boldsymbol{D}^{-1}\boldsymbol{v} = \frac{\sigma^2}{\lambda_2}\boldsymbol{v}'\operatorname{diag}(d_1^{-1}, d_2^{-1}, \ldots, d_n^{-1})\,\boldsymbol{v}.$$

The bound in Theorem 2 follows on setting $\boldsymbol{v} = \boldsymbol{e}_i$, the ith unit vector. The corresponding bound for the differences in Theorem S.1 follows on setting $\boldsymbol{v} = \boldsymbol{e}_i - \boldsymbol{e}_j$ for $i \neq j$. This concludes the proof. ■

PROOF OF THEOREMS 3 AND S.2 (FIRST-ORDER BOUNDS)

We first show that, if $\mathcal{G}$ is connected, then

$$0 \leq \left[\operatorname{var}(\widehat{\boldsymbol{\alpha}}) - \sigma^2\left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1} - \frac{\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}}{n} - \frac{\boldsymbol{D}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'}{n}\right)\right] \leq \frac{\sigma^2}{\lambda_2}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}. \tag{S.7}$$

Theorems 3 and S.2 will then follow readily. First note that, because $\mathcal{G}$ is connected, we know that the zero eigenvalue of the Laplacian matrix $\boldsymbol{L}$ has multiplicity one, and the corresponding eigenvector is given by $\boldsymbol{\iota}$. The Moore-Penrose pseudoinverse of $\boldsymbol{L}$ therefore satisfies $\boldsymbol{L}^\dagger \boldsymbol{L} = \boldsymbol{I}_n - n^{-1}\,\boldsymbol{\iota}_n \boldsymbol{\iota}_n'$, where the right hand side is the idempotent matrix that projects orthogonally to $\boldsymbol{\iota}_n$. Using that $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ and solving this equation for $\boldsymbol{L}^\dagger$ gives

$$\boldsymbol{L}^\dagger = \boldsymbol{D}^{-1} + \boldsymbol{L}^\dagger \boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\,\boldsymbol{\iota}_n \boldsymbol{\iota}_n' \boldsymbol{D}^{-1}. \tag{S.8}$$

The Laplacian is symmetric, and so transposition gives

$$\boldsymbol{L}^\dagger = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^\dagger - n^{-1}\,\boldsymbol{D}^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n'. \tag{S.9}$$

Replacing $\boldsymbol{L}^\dagger$ on the right-hand side of (S.8) by the expression for $\boldsymbol{L}^\dagger$ given by (S.9) yields

$$\boldsymbol{L}^\dagger = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^\dagger\boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n' \boldsymbol{D}^{-1} - n^{-1}\boldsymbol{D}^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n', \tag{S.10}$$

where we have also used the fact that $\boldsymbol{D}^{-1}\boldsymbol{A}\,\boldsymbol{\iota}_n = \boldsymbol{\iota}_n$. Re-arranging this equation allows us to write

$$\boldsymbol{L}^\dagger - \left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n' \boldsymbol{D}^{-1} - n^{-1}\boldsymbol{D}^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n'\right) = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^\dagger\boldsymbol{A}\boldsymbol{D}^{-1}.$$

Because $\boldsymbol{L} \geq 0$ and by the arguments in the preceding proof we also have the bounds

$$\boldsymbol{0} \leq \boldsymbol{L}^\dagger \leq \lambda_2^{-1}\boldsymbol{D}^{-1}.$$

Put together this yields

$$0 \leq \boldsymbol{L}^\dagger - \left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n' \boldsymbol{D}^{-1} - n^{-1}\boldsymbol{D}^{-1}\boldsymbol{\iota}_n \boldsymbol{\iota}_n'\right) \leq \lambda_2^{-1}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1},$$

and multiplication with σ^2 gives the bounds stated in (S.7).

To show Theorems 3 and S.2 we calculate, for $i \neq j$,

$$\begin{aligned}
\boldsymbol{e}_i'\,\boldsymbol{D}^{-1}\,\boldsymbol{e}_i &= d_i^{-1}, & \boldsymbol{e}_i'\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\,\boldsymbol{e}_i &= d_i^{-1}h_i^{-1}, \\
\boldsymbol{e}_i'\,\boldsymbol{D}^{-1}\,\boldsymbol{e}_j &= 0, & \boldsymbol{e}_i'\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\,\boldsymbol{e}_j &= d_i^{-1}d_j^{-1}d_{ij}h_{ij}^{-1}, \\
\boldsymbol{e}_i'\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\,\boldsymbol{e}_i &= 0, & \boldsymbol{e}_i'\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}\,\boldsymbol{e}_i = \boldsymbol{\iota}_n'\boldsymbol{D}^{-1}\,\boldsymbol{e}_i &= d_i^{-1}, \\
\boldsymbol{e}_i'\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\,\boldsymbol{e}_j &= d_i^{-1}d_j^{-1}(\boldsymbol{A})_{ij}, & \boldsymbol{e}_i'\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}\,\boldsymbol{e}_j = \boldsymbol{\iota}_n'\boldsymbol{D}^{-1}\,\boldsymbol{e}_j &= d_j^{-1}.
\end{aligned}$$

Combining these results with (S.7) gives the bounds on $\text{var}(\widehat{\alpha}_i) = \boldsymbol{e}_i'\text{var}(\widehat{\boldsymbol{\alpha}})\boldsymbol{e}_i$ and $\text{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) = (\boldsymbol{e}_i - \boldsymbol{e}_j)'\text{var}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{e}_i - \boldsymbol{e}_j)$ stated in the theorems and concludes the proof. ∎

PROOF OF THEOREM 4 (COVARIATES)

Define the $n \times n$ matrix

$$\boldsymbol{C} := (\boldsymbol{B}'\boldsymbol{B})^{\dagger} \boldsymbol{B}'\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{B}.$$

Let $\lambda_i(\boldsymbol{C})$ denote the ith eigenvalue of $\boldsymbol{C}$, arranged in ascending order. $\boldsymbol{C}$ is similar to the positive semi-definite matrix

$$(\boldsymbol{X}'\boldsymbol{X})^{-1/2} \boldsymbol{X}'\boldsymbol{B} (\boldsymbol{B}'\boldsymbol{B})^{\dagger} \boldsymbol{B}'\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1/2},$$

and so $\lambda_1(\boldsymbol{C}) \geq 0$. $\boldsymbol{C}$ is also similar to the matrix

$$\boldsymbol{B} (\boldsymbol{B}'\boldsymbol{B})^{\dagger} \boldsymbol{B}'\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}',$$

which is the product of two projection matrices, whose spectral norm is thus bounded by one. Hence, $\lambda_n(\boldsymbol{C}) \leq 1$. In addition, we must have $\lambda_i(\boldsymbol{C}) \neq 1$ for any $1 < i < n$ because, otherwise, $\operatorname{rank}(\boldsymbol{I}_n - \boldsymbol{C}) < n$, which implies that $\operatorname{rank}(\boldsymbol{B}'\boldsymbol{M_X}\boldsymbol{B}) < n-1$, contradicting our non-collinearity assumption. We therefore have $\|\boldsymbol{C}\|_2 < 1$. Noting that we equally have $\boldsymbol{C}\boldsymbol{\iota}_n = 0$ and $\boldsymbol{\iota}_n' C = 0$, and therefore also $(\boldsymbol{I}_m - \boldsymbol{C})^{-1} \boldsymbol{\iota}_n = \boldsymbol{\iota}_n$,

$$\begin{aligned}
(\boldsymbol{B}'\boldsymbol{M_X}\boldsymbol{B})^{\dagger} &= \left(\boldsymbol{B}'\boldsymbol{M_X}\boldsymbol{B} + n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right)^{-1} - n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' \\
&= \left[(\boldsymbol{B}'\boldsymbol{B})(\boldsymbol{I}_m - \boldsymbol{C}) + n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right]^{-1} - n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' \\
&= \left[\left(\boldsymbol{B}'\boldsymbol{B} + n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right)(\boldsymbol{I}_m - \boldsymbol{C})\right]^{-1} - n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' \\
&= (\boldsymbol{I}_m - \boldsymbol{C})^{-1} \left(\boldsymbol{B}'\boldsymbol{B} + n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right)^{-1} - n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' \\
&= (\boldsymbol{I}_m - \boldsymbol{C})^{-1} \left[\left(\boldsymbol{B}'\boldsymbol{B} + n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right)^{-1} - n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right] \\
&= (\boldsymbol{I}_m - \boldsymbol{C})^{-1} (\boldsymbol{B}'\boldsymbol{B})^{\dagger}.
\end{aligned}$$

Using again $\|\boldsymbol{C}\|_2 < 1$ we can expand $(\boldsymbol{I}_m - \boldsymbol{C})^{-1}$ in powers of $\boldsymbol{C}$, as

$$(\boldsymbol{B}'\boldsymbol{M_X}\boldsymbol{B})^{\dagger} = \sum_{r=0}^{\infty} \boldsymbol{C}^r (\boldsymbol{B}'\boldsymbol{B})^{\dagger}.$$

Define the $p \times p$ matrix

$$\widetilde{\boldsymbol{C}} := (\boldsymbol{X}'\boldsymbol{X})^{-1/2} \boldsymbol{X}'\boldsymbol{B} (\boldsymbol{B}'\boldsymbol{B})^{\dagger} \boldsymbol{B}'\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1/2}.$$

Then, the above expansion can be written as

$$(\boldsymbol{B}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{B})^{\dagger} = (\boldsymbol{B}'\boldsymbol{B})^{\dagger} + (\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\left(\sum_{r=0}^{\infty}\widetilde{\boldsymbol{C}}^{\,r}\right)(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\,\boldsymbol{X}'\boldsymbol{B}\,(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,.$$

The matrix $\widetilde{\boldsymbol{C}}$ is positive definite and satisfies $\|\widetilde{\boldsymbol{C}}\|_2 = 1-\rho$. Therefore,

$$\sum_{r=0}^{\infty}\widetilde{\boldsymbol{C}}^{\,r} \leq \sum_{r=0}^{\infty}(1-\rho)^r\ \boldsymbol{I}_p = \rho^{-1}\,\boldsymbol{I}_p.$$

We thus have

$$\begin{aligned}
\operatorname{var}(\widetilde{\alpha}_i) - \operatorname{var}(\widehat{\alpha}_i) &= \sigma^2\,\boldsymbol{e}_i'\left[(\boldsymbol{B}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{B})^{\dagger} - (\boldsymbol{B}'\boldsymbol{B})^{\dagger}\right]\boldsymbol{e}_i \\
&= \sigma^2\,\boldsymbol{e}_i'\left[(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\left(\sum_{r=0}^{\infty}\widetilde{\boldsymbol{C}}^{\,r}\right)(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\,\boldsymbol{X}'\boldsymbol{B}\,(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\right]\boldsymbol{e}_i \\
&\leq \frac{\sigma^2}{\rho}\,\boldsymbol{e}_i'\left[(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{B}\,(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\right]\boldsymbol{e}_i.
\end{aligned}$$

We have already shown that $(\boldsymbol{B}'\boldsymbol{B})^{\dagger} = \boldsymbol{L}^{\dagger}$ satisfies

$$\boldsymbol{L}^{\dagger} = \boldsymbol{D}^{-1} + \boldsymbol{L}^{\dagger}\boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}.$$

and

$$\boldsymbol{L}^{\dagger} = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^{\dagger} - n^{-1}\,\boldsymbol{D}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'.$$

Using this we obtain

$$\begin{aligned}
&\boldsymbol{e}_i'\,(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{B}\,(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\,\boldsymbol{e}_i \\
&= \boldsymbol{e}_i'\boldsymbol{L}^{\dagger}\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{B}\boldsymbol{L}^{\dagger}\boldsymbol{e}_i \\
&= \boldsymbol{e}_i'\left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^{\dagger}\right)\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{B}\left(\boldsymbol{D}^{-1} + \boldsymbol{L}^{\dagger}\boldsymbol{A}\boldsymbol{D}^{-1}\right)\boldsymbol{e}_i \\
&\leq T_i^{(1)} + T_i^{(2)} + 2\sqrt{T_i^{(1)}\,T_i^{(2)}},
\end{aligned}$$

where

$$\begin{aligned}
T_i^{(1)} &= \boldsymbol{e}_i'\boldsymbol{D}^{-1}\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{e}_i \\
T_i^{(2)} &= \boldsymbol{e}_i'\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^{\dagger}\boldsymbol{B}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{B}\boldsymbol{L}^{\dagger}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{e}_i,
\end{aligned}$$

and we used the Cauchy-Schwarz inequality to bound the mixed term. Again, because similar matrices have the same eigenvalues we have

$$\|\left(\boldsymbol{L}^{\dagger}\right)^{1/2} \boldsymbol{B}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{B}\left(\boldsymbol{L}^{\dagger}\right)^{1/2}\|_2 = |\widetilde{\boldsymbol{C}}|_2 = 1-\rho,$$

and therefore,

$$\begin{aligned}
T_i^{(2)} &= \boldsymbol{e}_i'\boldsymbol{D}^{-1}\boldsymbol{A}\left(\boldsymbol{L}^{\dagger}\right)^{1/2}\left[\left(\boldsymbol{L}^{\dagger}\right)^{1/2}\boldsymbol{B}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{B}\left(\boldsymbol{L}^{\dagger}\right)^{1/2}\right]\left(\boldsymbol{L}^{\dagger}\right)^{1/2}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{e}_i \\
&\leq (1-\rho)\,\boldsymbol{e}_i'\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^{\dagger}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{e}_i \\
&\leq \frac{1-\rho}{\lambda_2}\boldsymbol{e}_i'\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{e}_i \\
&= \frac{1-\rho}{\lambda_2\,d_i\,h_i},
\end{aligned}$$

where the last two steps follow from previous proofs. Using the definition of $\boldsymbol{B}$ we find that the $p \times n$ matrix $\boldsymbol{B}'\boldsymbol{X}$ has columns

$$\begin{aligned}
(\boldsymbol{X}'\boldsymbol{B})_{\cdot\, i} &= \sum_{j=1}^{n} 1\{(i,j)\in E\}\left(\boldsymbol{x}_{ij}-\boldsymbol{x}_{ij}\right) \\
&= \frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{A}\right)_{ij}\left(\boldsymbol{x}_{ij}-\boldsymbol{x}_{ij}\right) \\
&= \sum_{j=1}^{n}\left(\boldsymbol{A}\right)_{ij}\boldsymbol{x}_{ij},
\end{aligned}$$

and therefore

$$\boldsymbol{X}'\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{e}_i = \frac{1}{d_i}\sum_{j=1}^{n}\left(\boldsymbol{A}\right)_{ij}\boldsymbol{x}_{ij} = \overline{\boldsymbol{x}}_i.$$

We thus obtain

$$T_i^{(1)} = \boldsymbol{e}_i'\boldsymbol{D}^{-1}\boldsymbol{B}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{e}_i = \frac{1}{m}\,\overline{\boldsymbol{x}}_i'\,\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i.$$

Combining the above results we find

$$\begin{aligned}
\operatorname{var}\left(\widetilde{\alpha}_i\right)-\operatorname{var}\left(\widehat{\alpha}_i\right) &\leq \frac{\sigma^2}{\rho}\left(T_i^{(1)}+T_i^{(2)}+2\sqrt{T_i^{(1)}\,T_i^{(2)}}\right) \\
&\leq \frac{\sigma^2}{\rho}\left(\frac{1}{m}\,\overline{\boldsymbol{x}}_i'\,\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i+\frac{1-\rho}{\lambda_2\,d_i\,h_i}+2\sqrt{\frac{1}{m}\,\overline{\boldsymbol{x}}_i'\,\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i\,\frac{1-\rho}{\lambda_2\,d_i\,h_i}}\right).
\end{aligned}$$

For any $a, b \geq 0$ we have $a + b + 2\sqrt{ab} \leq 2(a + b)$. Thus, a slightly cruder but simpler bound is given by

$$\|\text{var}\,(\widehat{\alpha}_i^*) - \text{var}\,(\widehat{\alpha}_i)\| \;\leq\; \frac{2\,\sigma^2}{\rho}\left(\frac{\overline{\boldsymbol{x}}_i'\,\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i}{m} + \frac{1-\rho}{\lambda_2\,d_i\,h_i}\right),$$

which is the bound given in the theorem. ■

PROOF OF THEOREM 5 (GENERALIZED APPROXIMATION)

From the proof of Lemma 1, the least-squares estimator satisfies the first-order condition

$$\boldsymbol{L}\,\widehat{\boldsymbol{\alpha}} = \boldsymbol{B}'\boldsymbol{y}.$$

Using that $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{\alpha} + \boldsymbol{u}$ and that $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ this yields $\boldsymbol{D}^{1/2}\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \boldsymbol{D}^{-1/2}\boldsymbol{B}'\boldsymbol{u} + \boldsymbol{\epsilon}$, where

$$\boldsymbol{\epsilon} := \boldsymbol{D}^{-1/2}\boldsymbol{A}\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\,.$$

Note that this is the vector version of the expression for $\sqrt{d_i}(\widehat{\alpha}_i - \alpha_i)$ as given in the theorem. From $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = (\boldsymbol{B}'\boldsymbol{B})^{\dagger}\boldsymbol{B}'\boldsymbol{u}$ it follows that $\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ while from the assumption that $\mathbb{E}(\boldsymbol{u}\boldsymbol{u}') \leq \overline{\sigma}^2\boldsymbol{I}_n$ we have that

$$\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{D}^{-1/2}\boldsymbol{A}(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\boldsymbol{B}'\,\mathbb{E}\,(\boldsymbol{u}\boldsymbol{u}')\,\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\boldsymbol{A}\boldsymbol{D}^{-1/2} = \overline{\sigma}^2\,\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{L}^{\dagger}\boldsymbol{A}\boldsymbol{D}^{-1/2}.$$

As in the preceding proofs, we still have that $\boldsymbol{L}^{\dagger} \leq \lambda_2^{-1}\boldsymbol{D}^{-1}$, and so

$$\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \leq \overline{\sigma}^2\lambda_2^{-1}\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1/2}.$$

From this we find

$$\mathbb{E}(\epsilon_i^2) \leq \frac{\overline{\sigma}^2}{\lambda_2\,h_i}.$$

Thus, if $\overline{\sigma}^2\lambda_2^{-1}h_i^{-1} \to 0$ as $n \to \infty$, then by Markov's inequality we have $\epsilon_i \to_p 0$. By the continuous mapping theorem we therefore have

$$\sqrt{d_i}\,(\widehat{\alpha}_i - \alpha_i) \to_p \frac{1}{\sqrt{d_i}}\sum_{j\in[i]} u_{ij}.$$

Moreover, if $\frac{1}{\sqrt{d_i}}\sum_{j\in[i]} u_{ij}$ is asymptotically normal, then so is $\sqrt{d_i}\,(\widehat{\alpha}_i - \alpha_i)$. This concludes the proof. ■

PROOF OF THEOREM S.3 (SECOND-ORDER BOUND)

Proof of Theorem S.3. We start with the lower bound given in the theorem. Let $V_o := \{i\} \cup [i]$; then $n_o := |V_o| = 1 + d_i$. Without loss of generality we fix $i = 1$ and relabel the elements of V so that $V_o = \{1, 2, \ldots, 1 + d_i\}$. Let

$$\boldsymbol{L}_o := \begin{pmatrix} d_i & -\boldsymbol{\iota}'_{d_i} \\ -\boldsymbol{\iota}'_{d_i} & \boldsymbol{L}_{[i]} \end{pmatrix}, \qquad \boldsymbol{L}_{[i]} := \boldsymbol{D}_{[i]} - \boldsymbol{A}_{[i]},$$

using obvious notation for the $d_i \times d_i$ degree and adjacency matrices in the latter definition. Now, by the inversion formula for partitioned matrices,

$$\boldsymbol{L}_\circ^{-1} = \frac{1}{d_i - \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i}} \begin{pmatrix} 1 & \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1} \\ \boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i} & \left[\frac{\boldsymbol{L}_{[i]} - d_i^{-1}\boldsymbol{\iota}_{d_i}\boldsymbol{\iota}'_{d_i}}{d_i - \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i}}\right]^{-1} \end{pmatrix}.$$

Below we show that

$$0 \leq \left\{ \operatorname{var}(\widehat{\alpha}_i) - \frac{\sigma^2\left[1 - \frac{2}{n}\left(1 + \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i}\right)\right]}{d_i - \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i}} \right\} \leq \frac{\sigma^2\, \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}(\boldsymbol{A}_{\circ\#})\boldsymbol{D}_{\#}^{-1}(\boldsymbol{A}_{\circ\#})'\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i}}{\lambda_2\left(d_i - \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i}\right)^2}, \tag{S.11}$$

where $\boldsymbol{L}_\circ$ is the upper left $n_\circ \times n_\circ$ block of $\boldsymbol{L}$, $\boldsymbol{A}_{\circ\#}$ is the upper right $n_\circ \times n_\#$ block of $\boldsymbol{A}$, and $\boldsymbol{D}_\#$ is the lower right $n_\# \times n_\#$ block of $\boldsymbol{D}$. To make further progress, note that the expansion

$$\boldsymbol{L}_{[i]}^{-1} = \sum_{q=0}^{\infty} \left(\boldsymbol{D}_{[i]}^{-1}\boldsymbol{A}_{[i]}\right)^q \boldsymbol{D}_{[i]}^{-1}$$

is convergent, because we have $\|\boldsymbol{D}_{[i]}^{-1}\boldsymbol{A}_{[i]}\|_\infty < 1$, where $\|.\|_\infty$ denotes the maximum absolute row sum matrix norm. We therefore have

$$\begin{aligned} \boldsymbol{\iota}'_{d_i}\boldsymbol{L}_{[i]}^{-1}\boldsymbol{\iota}_{d_i} &= \boldsymbol{\iota}'_{d_i}\boldsymbol{D}_{[i]}^{-1}\boldsymbol{\iota}_{d_i} + \boldsymbol{\iota}'_{d_i}\sum_{q=1}^{\infty}\left(\boldsymbol{D}_{[i]}^{-1}\boldsymbol{A}_{[i]}\right)^q \boldsymbol{D}_{[i]}^{-1}\boldsymbol{\iota}_{d_i} \\ &\geq \boldsymbol{\iota}'_{d_i}\boldsymbol{D}_{[i]}^{-1}\boldsymbol{\iota}_{d_i} = \sum_{j\in[i]} d_j^{-1}, \end{aligned} \tag{S.12}$$

where we used that $\boldsymbol{\iota}_{d_i}' \sum_{q=1}^{\infty} \left(\boldsymbol{D}_{[i]}^{-1}\boldsymbol{A}_{[i]}\right)^q \boldsymbol{D}_{[i]}^{-1}\boldsymbol{\iota}_{d_i} \geq 0$, because this is a product and sum of vector and matrices that all have non-negative entries. Define the $n_\circ \times n_\circ$ diagonal matrix $\underline{\boldsymbol{D}}_{[i]} = \operatorname{diag}(\underline{d}_{j,i} : j \in [i])$. We have

$$\boldsymbol{L}_{[i]} - \underline{\boldsymbol{D}}_{[i]} = \operatorname{diag}(\boldsymbol{A}_{[i]}\boldsymbol{\iota}_{d_i}) - \boldsymbol{A}_{[i]} \geq 0, \tag{S.13}$$

because $\operatorname{diag}(\boldsymbol{A}_{[i]}\boldsymbol{\iota}_{d_i}) - \boldsymbol{A}_{[i]}$ can be expressed as a sum of matrices of the form

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \geq 0,$$

embedded into an $n_\circ \times n_\circ$ matrix. We therefore have $\boldsymbol{L}_{[i]}^{-1} \leq \underline{\boldsymbol{D}}_{[i]}^{-1}$, implying

$$\boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} \leq \boldsymbol{\iota}_{d_i}' \underline{\boldsymbol{D}}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} = \sum_{j \in [i]} \underline{d}_{j,i}^{-1}. \tag{S.14}$$

Combining (S.11), (S.12) and (S.14) gives

$$\operatorname{var}(\widehat{\alpha}_i) \geq \frac{\sigma^2 \left[1 - \frac{2}{n}\left(1 + \sum_{j\in[i]} \underline{d}_{j,i}^{-1}\right)\right]}{\sum_{j\in[i]} \left(1 - d_j^{-1}\right)} = \frac{\sigma^2}{d_i(1 - h_i^{-1})}\left(1 - \frac{2}{n} - \frac{2}{n}\frac{d_i}{\underline{h}_i}\right),$$

which is the lower bound stated in the theorem.

To show the upper bound, consider the the graph $\widetilde{\mathcal{G}} := (V, \widetilde{E})$, with $\widetilde{E} := E \setminus [i] \times [i]$. That is, we construct $\widetilde{G}$ by deleting all edges between neighbors of i from $\mathcal{G}$. Note that $\widetilde{\mathcal{G}}$ is still connected, because all vertices in $[i]$ are connected through i. Let $\widetilde{\boldsymbol{\alpha}}$ be the estimator for $\boldsymbol{\alpha}$ obtained for $\widetilde{\mathcal{G}}$, in the same way that $\widehat{\boldsymbol{\alpha}}$ was obtained for $\mathcal{G}$. Let $\widetilde{\boldsymbol{L}}$ be the Laplacian matrix of $\widetilde{\mathcal{G}}$. Analogous to (S.13) we have $\widetilde{\boldsymbol{L}} \leq \boldsymbol{L}$, and therefore $\widetilde{\boldsymbol{L}}^\dagger \geq \boldsymbol{L}^\dagger$. The result (S.11) holds for any connected graph, and so can equally be applied to $\widetilde{\mathcal{G}}$, we only need to replace $\operatorname{var}(\widehat{\alpha}_i)$ by $\operatorname{var}(\widetilde{\alpha}_i)$ and $\boldsymbol{L}$ by $\widetilde{\boldsymbol{L}}$. The matrices $\boldsymbol{A}_{\circ\#}$ and $\boldsymbol{D}_{\#}^{-1}$ are identical for $\widetilde{\mathcal{G}}$ and $\mathcal{G}$. However, for $\widetilde{\mathcal{G}}$ we find $\widetilde{\boldsymbol{D}}_{[i]} = \underline{\boldsymbol{D}}_{[i]}$, because the degree of vertex j is given by $\underline{d}_{j,i}$, and we have $\widetilde{\boldsymbol{A}}_{[i]} = 0$, because there are no edges that connect elements in $[i]$. We thus have $\widetilde{\boldsymbol{L}}_{[i]} = \widetilde{\boldsymbol{D}}_{[i]} - \widetilde{\boldsymbol{A}}_{[i]} = \underline{\boldsymbol{D}}_{[i]}$. Therefore,

$$\operatorname{var}(\widehat{\alpha}_i) \leq \operatorname{var}(\widetilde{\alpha}_i) \leq \frac{\sigma^2 \left[1 - \frac{2}{n}\left(1 + \boldsymbol{\iota}_{d_i}' \underline{\boldsymbol{D}}_{[i]}^{-1} \boldsymbol{\iota}_{d_i}\right)\right]}{d_i - \boldsymbol{\iota}_{d_i}' \underline{\boldsymbol{D}}_{[i]}^{-1} \boldsymbol{\iota}_{d_i}} + \frac{\sigma^2\, \boldsymbol{\iota}_{d_i}' \underline{\boldsymbol{D}}_{[i]}^{-1} (\boldsymbol{A}_{\circ\#}) \boldsymbol{D}_{\#}^{-1} (\boldsymbol{A}_{\circ\#})' \underline{\boldsymbol{D}}_{[i]}^{-1} \boldsymbol{\iota}_{d_i}}{\lambda_2 \left(d_i - \boldsymbol{\iota}_{d_i}' \underline{\boldsymbol{D}}_{[i]}^{-1} \boldsymbol{\iota}_{d_i}\right)^2},$$

and evaluating the right-hand side of the last inequality gives the upper bound on $\operatorname{var}(\widehat{\alpha}_i)$ in the theorem. This concludes the proof. ■

Proof of (S.11)**.** We prove the following more general result. Let $\mathcal{G}$ be connected. Choose $V_\circ \subset V$ with $0 < |V_\circ| < n$, and let $V_\# = V \setminus V_\circ$. Let $n_\circ = |V_\circ|$ and $n_\# = n - n_\circ$. Relabel the elements in V such that $V_\circ = \{1, 2, \dots, n_\circ\}$. Let $\widehat{\boldsymbol{\alpha}}_\circ = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_{n_\circ})'$, $\boldsymbol{L}_\circ$ be the upper left $n_\circ \times n_\circ$ block of $\boldsymbol{L}$, $\boldsymbol{A}_{\circ\#}$ be the upper right $n_\circ \times n_\#$ block of $\boldsymbol{A}$, and $\boldsymbol{D}_\#$ be the lower right $n_\# \times n_\#$ block of $\boldsymbol{D}$. Then,

$$0 \leq \left[\operatorname{var}(\widehat{\boldsymbol{\alpha}}_\circ) - \sigma^2 \left(\boldsymbol{L}_\circ^{-1} - \frac{\boldsymbol{\iota}_{n_\circ}\boldsymbol{\iota}_{n_\circ}' \boldsymbol{L}_\circ^{-1} + \boldsymbol{L}_\circ^{-1}\boldsymbol{\iota}_{n_\circ}\boldsymbol{\iota}_{n_\circ}'}{n} \right) \right] \leq \frac{\sigma^2}{\lambda_2} \boldsymbol{L}_\circ^{-1} (\boldsymbol{A}_{\circ\#}) \boldsymbol{D}_\#^{-1} (\boldsymbol{A}_{\circ\#})' \boldsymbol{L}_\circ^{-1}$$

holds.

To show the result, define the $n \times n$ matrices

$$\boldsymbol{L}_{\mathrm{b}} := \begin{pmatrix} \boldsymbol{L}_\circ & 0 \\ 0 & \boldsymbol{L}_\# \end{pmatrix}, \qquad \boldsymbol{A}_{\mathrm{b}} := \begin{pmatrix} 0 & \boldsymbol{A}_{\circ\#} \\ (\boldsymbol{A}_{\circ\#})' & 0 \end{pmatrix},$$

with obvious definition of $\boldsymbol{L}_\#$ such that $\boldsymbol{L} = \boldsymbol{L}_{\mathrm{b}} - \boldsymbol{A}_{\mathrm{b}}$. Because the graph is connected the pseudo-inverse $\boldsymbol{L}^\dagger$ satisfies $\boldsymbol{L}^\dagger \boldsymbol{L} = \boldsymbol{I}_n - n^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'$. Plugging $\boldsymbol{L} = \boldsymbol{L}_{\mathrm{b}} - \boldsymbol{A}_{\mathrm{b}}$ into this expression we obtain

$$\boldsymbol{L}^\dagger = \boldsymbol{L}_{\mathrm{b}}^{-1} \left(\boldsymbol{I}_n + \boldsymbol{A}_{\mathrm{b}} \boldsymbol{L}^\dagger - n^{-1}\, \boldsymbol{\iota}_n \boldsymbol{\iota}_n' \right).$$

Using the transposed of this last equation to replace $\boldsymbol{L}^\dagger = (\boldsymbol{L}^\dagger)'$ on the right-hand side of that same equation we obtain

$$\begin{aligned} \boldsymbol{L}^\dagger &= \boldsymbol{L}_{\mathrm{b}}^{-1} + \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{A}_{\mathrm{b}}\boldsymbol{L}_{\mathrm{b}}^{-1} + \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{A}_{\mathrm{b}}\boldsymbol{L}^\dagger\boldsymbol{A}_{\mathrm{b}}\boldsymbol{L}_{\mathrm{b}}^{-1} - n^{-1}\, \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' - n^{-1}\, \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{A}_{\mathrm{b}}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{L}_{\mathrm{b}}^{-1} \\ &= \boldsymbol{L}_{\mathrm{b}}^{-1} + \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{A}_{\mathrm{b}}\boldsymbol{L}_{\mathrm{b}}^{-1} + \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{A}_{\mathrm{b}}\boldsymbol{L}^\dagger\boldsymbol{A}_{\mathrm{b}}\boldsymbol{L}_{\mathrm{b}}^{-1} - n^{-1}\, \boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' - n^{-1}\, \boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{L}_{\mathrm{b}}^{-1}, \end{aligned}$$

where in the last step we have used that $\boldsymbol{L}_{\mathrm{b}}^{-1}\boldsymbol{A}_{\mathrm{b}}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n$, which follows from $0 = \boldsymbol{L}\,\boldsymbol{\iota}_n = (\boldsymbol{L}_{\mathrm{b}} - \boldsymbol{A}_{\mathrm{b}})\boldsymbol{\iota}_n$. Evaluating the last result for the upper left $n_\circ \times n_\circ$ block gives

$$(\boldsymbol{L}^\dagger)_\circ = \boldsymbol{L}_\circ^{-1} + \boldsymbol{L}_\circ^{-1}(\boldsymbol{A}_{\circ\#})(\boldsymbol{L}^\dagger)_\#(\boldsymbol{A}_{\circ\#})'\boldsymbol{L}_\circ^{-1} - n^{-1}\, \boldsymbol{L}_\circ^{-1}\boldsymbol{\iota}_{n_\circ}\boldsymbol{\iota}_{n_\circ}' - n^{-1}\, \boldsymbol{\iota}_{n_\circ}\boldsymbol{\iota}_{n_\circ}'\boldsymbol{L}_\circ^{-1},$$

with obvious definition of $(\boldsymbol{L}^\dagger)_\#$. We obtain the result searched for for $\operatorname{var}(\widehat{\boldsymbol{\alpha}}_\circ) = \sigma^2(\boldsymbol{L}^\dagger)_\circ$ by also using $0 \leq (\boldsymbol{L}^\dagger)_\# \leq \lambda_2^{-1}\boldsymbol{D}_\#^{-1}$. This concludes the proof. ■

PROOF OF THEOREM S.4 (Consistency of Moment Estimates)

Because $\widehat{\alpha}_i - \alpha_i \sim \mathcal{N}\left(0, \sigma^2 \left(\boldsymbol{L}^\dagger\right)_{ii}\right)$ there exists a constant $c_k > 0$ that only depends on k such that

$$\left[\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n |\widehat{\alpha}_i - \alpha_i|^k\right)\right]^{2/k} = c_k\,\sigma^2 \left(\frac{1}{n}\sum_{i=1}^n \left(\boldsymbol{L}^\dagger\right)_{ii}^{k/2}\right)^{2/k}$$
$$\leq c_k\,\sigma^2 \left(\frac{1}{n}\sum_{i=1}^n \left(\frac{1}{\lambda_2\, d_i}\right)^{k/2}\right)^{2/k}$$
$$= \frac{c_k\,\sigma^2}{\lambda_2 \left(\frac{1}{n}\sum_{i=1}^n d_i^{-k/2}\right)^{-2/k}} \to 0,$$

where we have applied Theorem 2 to bound $\left(\boldsymbol{L}^\dagger\right)_{ii} = \sigma^{-2}\mathrm{var}(\widehat{\alpha}_i)$, and in the last step we used our assumption $\lambda_2\left(n^{-1}\sum_{i=1}^n d_i^{-k/2}\right)^{-2/k} \to \infty$. Applying Markov's inequality we thus have

$$\frac{1}{n}\sum_{i=1}^n |\widehat{\alpha}_i - \alpha_i|^k \to_p 0. \tag{S.15}$$

By Taylor expanding $\widehat{\alpha}_i^k$ around α_i, and then applying Hölder's inequality we find

$$\left|\frac{1}{n}\sum_{i=1}^n \left(\widehat{\alpha}_i^k - \alpha_i^k\right)\right| = \left|\sum_{\ell=1}^k \binom{k}{\ell}\frac{1}{n}\sum_{i=1}^n \alpha_i^{k-\ell}\left(\widehat{\alpha}_i - \alpha_i\right)^\ell\right|$$
$$\leq \sum_{\ell=1}^k \binom{k}{\ell}\left(\frac{1}{n}\sum_{i=1}^n |\alpha_i|^k\right)^{(k-\ell)/k}\left(\frac{1}{n}\sum_{i=1}^n |\widehat{\alpha}_i - \alpha_i|^k\right)^{\ell/k} \to_p 0,$$

where in the last step we used (S.15) and the assumption that $\frac{1}{n}\sum_{i=1}^n |\alpha_i|^k$ is bounded asymptotically. We have thus shown that $\frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_i^k \;\to_p\; \frac{1}{n}\sum_{i=1}^n \alpha_i^k$. ■

PROOF OF LEMMA S.5 (Bias and Variance of $\widehat{\vartheta}$)

Using $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \boldsymbol{L}^\dagger\boldsymbol{B}'\boldsymbol{u}$, $\mathbb{E}(\boldsymbol{u}) = 0$, $\mathbb{E}(\boldsymbol{u}\boldsymbol{u}') = \sigma^2\boldsymbol{I}_m$, and $\boldsymbol{B}'\boldsymbol{B} = \boldsymbol{L}$, we find for $\vartheta = n^{-1}\boldsymbol{\alpha}'\boldsymbol{\alpha}$ and $\widehat{\vartheta} = n^{-1}\widehat{\boldsymbol{\alpha}}'\widehat{\boldsymbol{\alpha}}$ that

$$\mathbb{E}(\widehat{\vartheta} - \vartheta) = n^{-1}\mathbb{E}(\boldsymbol{u}'\boldsymbol{B}\boldsymbol{L}^\dagger\boldsymbol{L}^\dagger\boldsymbol{B}'\boldsymbol{u}) = n^{-1}\mathrm{tr}[\boldsymbol{L}^\dagger\boldsymbol{B}'\mathbb{E}(\boldsymbol{u}\boldsymbol{u}')\boldsymbol{B}\boldsymbol{L}^\dagger] = \sigma^2\, n^{-1}\,\mathrm{tr}(\boldsymbol{L}^\dagger),$$

which is the bias result in the lemma. Next, define the mean zero error terms

$$\varepsilon_1 = 2\,n^{-1}\,\boldsymbol{\alpha}'\boldsymbol{L}^\dagger\boldsymbol{B}'\boldsymbol{u}, \qquad \varepsilon_2 = n^{-1}\left[\boldsymbol{u}'\boldsymbol{B}\boldsymbol{L}^\dagger\boldsymbol{L}^\dagger\boldsymbol{B}'\boldsymbol{u} - \mathbb{E}(\boldsymbol{u}'\boldsymbol{B}\boldsymbol{L}^\dagger\boldsymbol{L}^\dagger\boldsymbol{B}'\boldsymbol{u})\right].$$

With those definitions we have $\widehat{\vartheta} = \mathbb{E}(\widehat{\vartheta}) + \varepsilon_1 + \varepsilon_2$. Because we assume $\boldsymbol{u}$ to be mean zero and normality distributed we have $\mathbb{E}(\varepsilon_1 \varepsilon_2) = 0$, and therefore

$$\operatorname{var}(\widehat{\vartheta}) = \mathbb{E}\varepsilon_1^2 + \mathbb{E}\varepsilon_2^2.$$

We calculate

$$\begin{aligned}
\mathbb{E}\varepsilon_1^2 &= 4\, n^{-2}\, \boldsymbol{\alpha}' \boldsymbol{L}^\dagger \boldsymbol{B}' \mathbb{E}(\boldsymbol{u}\boldsymbol{u}') \boldsymbol{B} \boldsymbol{L}^\dagger \boldsymbol{\alpha} = 4\, \sigma^2\, n^{-2}\, \boldsymbol{\alpha}' \boldsymbol{L}^\dagger \boldsymbol{\alpha}, \\
\mathbb{E}\varepsilon_2^2 &= n^{-2} \operatorname{var}(\boldsymbol{u}' \boldsymbol{B} \boldsymbol{L}^\dagger \boldsymbol{L}^\dagger \boldsymbol{B}' \boldsymbol{u}) = 2\, \sigma^4\, n^{-2} \operatorname{tr}(\boldsymbol{B} \boldsymbol{L}^\dagger \boldsymbol{L}^\dagger \boldsymbol{B}' \boldsymbol{B} \boldsymbol{L}^\dagger \boldsymbol{L}^\dagger \boldsymbol{B}') = 2\, \sigma^4\, n^{-2} \underbrace{\operatorname{tr}(\boldsymbol{L}^\dagger \boldsymbol{L}^\dagger)}_{= \|\boldsymbol{L}^\dagger\|^2},
\end{aligned}$$

where in the last line we used that for $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_m)$ we have for any symmetric $m \times m$ matrix $\boldsymbol{M}$ that $\operatorname{var}(\boldsymbol{u}' \boldsymbol{M} \boldsymbol{u}) = 2\sigma^4 \operatorname{tr}(\boldsymbol{M}^2)$. We have thus also shown the variance result in the lemma. ■

PROOF OF THEOREM S.7 (Bounds on Variance of $\widehat{\vartheta}$, Version 1)

Consider the spectral decomposition $\boldsymbol{S} = \sum_{i=2}^n \lambda_i \widetilde{\boldsymbol{v}}_i \widetilde{\boldsymbol{v}}_i'$, where $\widetilde{\boldsymbol{v}}_i$ are normalized eigenvectors, in particular $\widetilde{\boldsymbol{v}}_2 = \boldsymbol{v}/\|\boldsymbol{v}\|$. We then have $\boldsymbol{S}^\dagger = \sum_{i=2}^n \lambda_i^{-1} \widetilde{\boldsymbol{v}}_i \widetilde{\boldsymbol{v}}_i' \geq \lambda_2^{-1} \widetilde{\boldsymbol{v}}_2 \widetilde{\boldsymbol{v}}_2' = \lambda_2^{-1} \frac{\boldsymbol{v}\boldsymbol{v}'}{\boldsymbol{v}'\boldsymbol{v}}$. Using this we find

$$\begin{aligned}
n^{-1} \boldsymbol{\alpha}' \boldsymbol{L}^\dagger \boldsymbol{\alpha} &= n^{-1} \boldsymbol{\alpha}' \boldsymbol{D}^{-1/2} \boldsymbol{S}^\dagger \boldsymbol{D}^{-1/2} \boldsymbol{\alpha} \\
&\geq n^{-1} \boldsymbol{\alpha}' \boldsymbol{D}^{-1/2} \left(\frac{1}{\lambda_2} \frac{\boldsymbol{v}\boldsymbol{v}'}{\boldsymbol{v}'\boldsymbol{v}} \right) \boldsymbol{D}^{-1/2} \boldsymbol{\alpha} \\
&= \frac{1}{\lambda_2} \underbrace{\frac{(\boldsymbol{v}' \boldsymbol{D}^{-1/2} \boldsymbol{\alpha})^2}{(\boldsymbol{v}'\boldsymbol{v})(\boldsymbol{\alpha}' \boldsymbol{D}^{-1} \boldsymbol{\alpha})}}_{=\kappa^2} \underbrace{\frac{\boldsymbol{\alpha}' \boldsymbol{D}^{-1} \boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{\alpha}}}_{=h_\alpha^{-1}} \underbrace{\frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{n}}_{=\vartheta} = \frac{\kappa^2\, \vartheta}{\lambda_2\, h_\alpha}.
\end{aligned}$$

This result together with $\|\boldsymbol{L}^\dagger\|^2 \geq 0$ and Lemma S.5 gives the lower bound in the theorem.

Regarding the upper bound we have

$$n^{-1} \boldsymbol{\alpha}' \boldsymbol{L}^\dagger \boldsymbol{\alpha} \leq n^{-1} \lambda_2^{-1}\ \boldsymbol{\alpha}' \boldsymbol{D}^{-1} \boldsymbol{\alpha} = \frac{1}{\lambda_2} \underbrace{\frac{\boldsymbol{\alpha}' \boldsymbol{D}^{-1} \boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{\alpha}}}_{=h_\alpha^{-1}} \underbrace{\frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{n}}_{=\vartheta} = \frac{\vartheta}{\lambda_2\, h_\alpha},$$

where we used $\boldsymbol{L}^\dagger \leq \lambda_2^{-1} \boldsymbol{D}^{-1}$. This also implies $\left\|\boldsymbol{L}^\dagger\right\| \leq \lambda_2^{-1} \|\boldsymbol{D}^{-1}\|$, and therefore

$$n^{-1} \|\boldsymbol{L}^\dagger\|^2 \leq \lambda_2^{-2}\, n^{-1} \left\|\boldsymbol{D}^{-1}\right\|^2 = \lambda_2^{-2}\, n^{-1} \sum_{i=1}^n d_i^{-2} = \lambda_2^{-2}\, h_{\text{sqr}}^{-2},$$

Together with Lemma S.5 this gives the upper bound in the theorem. ■

PROOF OF THEOREM S.8 (Bounds on Variance of $\widehat{\vartheta}$, Version 2)

Using (S.8) and (S.9) as well as $\boldsymbol{L}^\dagger \leq \lambda_2^{-1}\boldsymbol{D}^{-1}$ and $\left\|\boldsymbol{L}^\dagger\right\| \leq \lambda_2^{-1}\left\|\boldsymbol{D}^{-1}\right\|$ we find

$$
\begin{aligned}
\|\boldsymbol{L}^\dagger\|^2 &= \operatorname{tr}\left(\boldsymbol{L}^\dagger\boldsymbol{L}^\dagger\right) \\
&= \operatorname{tr}\left[\left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^\dagger - n^{-1}\,\boldsymbol{D}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right)\left(\boldsymbol{D}^{-1} + \boldsymbol{L}^\dagger\boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}\right)\right] \\
&= \operatorname{tr}\left(\boldsymbol{D}^{-2}\right) + \operatorname{tr}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\left(\boldsymbol{L}^\dagger\right)^2\boldsymbol{A}\boldsymbol{D}^{-1}\right) + 2\operatorname{tr}\left(\boldsymbol{D}^{-2}\boldsymbol{A}\boldsymbol{L}^\dagger\right)\underbrace{-\boldsymbol{\iota}_n'\boldsymbol{D}^{-2}\boldsymbol{\iota}_n}_{\leq 0} \\
&\leq \left\|\boldsymbol{D}^{-1}\right\|^2 + (\lambda_2)^{-2}\underbrace{\operatorname{tr}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-2}\boldsymbol{A}\boldsymbol{D}^{-1}\right)}_{\leq\|\boldsymbol{D}^{-2}\boldsymbol{A}\|^2} + 2\underbrace{\left\|\boldsymbol{L}^\dagger\right\|}_{\leq\frac{\|\boldsymbol{D}^{-1}\|}{\lambda_2}}\left\|\boldsymbol{D}^{-2}\boldsymbol{A}\right\| \\
&\leq \left(\left\|\boldsymbol{D}^{-1}\right\| + \lambda_2^{-1}\left\|\boldsymbol{D}^{-2}\boldsymbol{A}\right\|\right)^2.
\end{aligned}
$$

We have $n^{-1/2}\left\|\boldsymbol{D}^{-2}\boldsymbol{A}\right\| = \left(n^{-1}\sum_{i,j=1}^n(\boldsymbol{A})_{ij}d_i^{-4}\right)^{1/2} = \left(n^{-1}\sum_{i,j=1}^n d_i^{-3}\right)^{1/2} = h_{\text{cub}}^{-3/2}$ and $n^{-1/2}\left\|\boldsymbol{D}^{-1}\right\| = (h_{\text{sqr}})^{-1}$. Combining those results we find

$$
0 \leq n^{-1/2}\|\boldsymbol{L}^\dagger\| \leq (h_{\text{sqr}})^{-1} + \lambda_2^{-1}\,h_{\text{cub}}^{-3/2}. \tag{S.16}
$$

Next, by iterating equation (S.10) $k-1$ times, that is, plugging the expression for $\boldsymbol{L}^\dagger$ in (S.10) back into the rhs of (S.10) itself, and using that $\boldsymbol{\alpha}'\boldsymbol{\iota}_n = 0$, we obtain

$$
\boldsymbol{\alpha}'\boldsymbol{L}^\dagger\boldsymbol{\alpha} = \boldsymbol{\alpha}'\left[\boldsymbol{D}^{-1}\sum_{\ell=0}^{2k-1}\left(\boldsymbol{A}\boldsymbol{D}^{-1}\right)^\ell + \left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^k\boldsymbol{L}^\dagger\left(\boldsymbol{A}\boldsymbol{D}^{-1}\right)^k\right]\boldsymbol{\alpha}
$$

Using $0 \leq \boldsymbol{L}^\dagger \leq \lambda_2^{-1}\boldsymbol{D}^{-1}$ and the definition of $\boldsymbol{\alpha}^{(\ell)}$ we thus find that

$$
0 \leq \boldsymbol{\alpha}'\boldsymbol{L}^\dagger\boldsymbol{\alpha} - \sum_{\ell=0}^{k-1}\boldsymbol{\alpha}^{(\ell)\prime}\left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\right)\boldsymbol{\alpha}^{(\ell)} \leq \lambda_2^{-1}\boldsymbol{\alpha}^{(k)\prime}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(k)}
$$

From this and

$$
\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(\ell)}}{n} = \underbrace{\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(\ell)}}{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{\alpha}^{(\ell)}}}_{=1/h_\alpha^{(\ell)}}\underbrace{\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{\alpha}^{(\ell)}}{n}}_{=\vartheta^{(\ell)}},
$$

$$
\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(\ell)}}{n} = \underbrace{\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(\ell)}}{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(\ell)}}}_{=\gamma^{(\ell)}}\underbrace{\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{D}^{-1}\boldsymbol{\alpha}^{(\ell)}}{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{\alpha}^{(\ell)}}}_{=1/h_\alpha^{(\ell)}}\underbrace{\frac{\boldsymbol{\alpha}^{(\ell)\prime}\boldsymbol{\alpha}^{(\ell)}}{n}}_{=\vartheta^{(\ell)}},
$$

we find that the statement in display (S.6) holds. Combining Lemma S.5 with the bounds in (S.6) and (S.16) gives the statement of the theorem. ■

References

Boumal, N. and X. Cheng (2014). Concentration of the Kirchhoff index for Erdős–Rényi graphs. *Systems & Control Letters 74*, 74–80.

Bradley, R. A. and M. E. Terry (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika 39*, 324–325.

Butler, S. (2016). Algebraic aspects of the normalized Laplacian. In A. Beveridge, J. R. Griggs, L. Hogben, G. Musiker, and P. Tetali (Eds.), *Recent Trends in Combinatorics*, pp. 295–315. Springer.

Chung, F. R. K. (1997). *Spectral Graph Theory.* Volume 92 of CBMS Regional Conference Series in Mathematics, American Mathematical Society.

Dzemski, A. (2014). An empirical model of dyadic link formation in a network with unobserved heterogeneity. Mimeo.

Erdős, P. and A. Rényi (1959). On random graphs. *Publicationes Mathematicae 6*, 290–297.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel data models with large N, T. *Journal of Econometrics 192*, 291–312.

Graham, B. S. (2017). An econometric model of link formation with degree heterogeneity. Forthcoming in Econometrica.

Jochmans, K. (2017). Semiparametric analysis of network formation. Forthcoming in *Journal of Business & Economic Statistics.*

Klein, D. J. (2002). Resistance-distance sum rules. *Croatica Chemica Acta 75*, 633–649.

Klein, D. J. and M. J. Randić (1993). Resistance distance. *Journal of Mathematical Chemistry 12*, 81–95.

Simons, G. and Y.-C. Yao (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Annals of Statistics 27*, 1041–1060.

von Luxburg, U., A. Radl, and M. Hein (2010). Getting lost in space: Large sample analysis of the resistance distance. In *Advances in Neural Information Processing Systems*, pp.

2622–2630.

von Luxburg, U., A. Radl, and M. Hein (2014). Hitting at commute times in large random neighborhood graphs. *Journal of Machine Learning Research 15*, 1751–1798.

Yan, T. and J. Xu (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika 100*, 519–524.