FIXED-EFFECT REGRESSIONS ON NETWORK DATA

Koen Jochmans*
Sciences Po, Paris

Martin Weidner‡
University College London

August 5, 2016

Abstract

This paper studies inference on fixed effects in a linear regression model estimated from network data. We derive bounds on the variance of the fixed-effect estimator that uncover the importance of the smallest non-zero eigenvalue of the (normalized) Laplacian of the network and of the degree structure of the network. The eigenvalue is a measure of connectivity, with smaller values indicating less-connected networks. These bounds yield conditions for consistent estimation and convergence rates, and allow to evaluate the accuracy of first-order approximations to the variance of the fixed-effect estimator.

Keywords: fixed effects, graph, Laplacian, network data, variance bound

JEL classification: C23, C55

*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.

We are particularly grateful to Jean-Marc Robin for stimulating discussions and for insightful comments on earlier drafts of this paper. We also thank Bryan Graham for providing us with the iPython notebook we used to make our plots, and Áureo de Paula for comments and suggestions. Jochmans gratefully acknowledges support from Sciences Po's Scientific Advisory Board through the grant 'Econometric analysis of linked data'. Weidner gratefully acknowledges support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 and from the European Research Council grant ERC-2014-CoG-646917-ROMIA.

1 Introduction

A substantial literature approaches dyadic interactions between agents by means of models featuring agent-specific parameters. The Bradley-Terry model for paired comparisons of Zermelo (1929) and Bradley and Terry (1952) and the β-model for network formation with degree heterogeneity (Yan and Xu, 2013; Graham, 2015) are two classic examples. The use of fixed-effect models for network data is now widespread in empirical work. Applications include studies of risk sharing (Fafchamps and Gubert, 2007), sorting between workers and firms in the labor market (Abowd, Kramarz and Margolis, 1999), and the interaction between students and teachers (Aaronson, Barrow and Sander, 2007; Rivkin, Hanushek and Kain, 2005), as well as the analysis of trade flows (Harrigan, 1996; Anderson and van Wincoop, 2003).

The structure of the network, that is, who interacts with whom and to which extent, differs strongly across applications. While rather dense networks might be observed in the analysis of financial markets (Acemoglu, Ozdaglar and Tahbaz-Salehi, 2015), sparse networks—i.e., networks with relatively few links—are typically the norm in networks of friendship or trust (Jackson, Rodriguez-Barraquer and Tan, 2012). The network structure is also an important determinant of the accuracy of statistical inference. One important illustration is given by fixed-effect regressions of log wages on matched employer-employee data (Abowd, Kramarz and Margolis, 1999). There, estimated worker and firm effects are typically found to be negatively correlated (Goux and Maurin, 1999; Barth and Dale-Olsen, 2003), which is in contrast with economic intuition. The origin of this negative assortative matching puzzle is limited-mobility bias (Abowd, Kramarz, Lengermann and Perez-Duarte, 2004; Andrews, Gill, Schank and Upward, 2008), that is, the fact that, throughout their working history, workers are employed in only few firms. Moreover, even though linked data sets are typically very large, the worker fixed effects are estimated from very small subsamples.

We are not aware of studies of the statistical accuracy of fixed-effect estimators in the network literature. A chief reason would appear to be that the structure of a network becomes complex very fast and so it is rather difficult to see how data carries information about certain parameters. In this paper we analyze this issue in the context of a linear version of the Bradley and Terry (1952) model. The linear regression model contains all the main features of the typical models for network data, yet is sufficiently simple to lend itself to careful analysis.

We use results from graph theory to show that the variance of the fixed-effect estimator is related to the Laplacian of the network. A bound on the variance of the fixed-effect estimator is obtained that depends inversely on the smallest non-zero eigenvalue of the (normalized) Laplacian. This eigenvalue is a measure of connectivity of the network. The larger it is, the more dense is the network. One interesting consequence of this bound is that consistent estimation is possible even if the network becomes less connected as the sample grows. Eigenvalues of network matrices have previously been found to be important in determining equilibrium conditions in games on networks (Bramboullé, Kranton and D'Amours, 2014) but our result seems the first to uncover their importance for statistical inference.

We next refine the variance bound to uncover how the local structure of the network around a given vertex influences the variance of the vertex-specific parameter estimator. Clearly, the variance of such an estimator is decreasing in the degree of the vertex—the number of edges that originate or arrive in it—that is, the number of neighbors of the vertex. The improved bounds, however, uncover the sensitivity of the variance with respect to the degree of the neighbors of the vertex.

A potential issue with a global connectivity measure such as the smallest non-zero eigenvalue is that it can lead to variance bounds that are overly conservative. A leading situation where this will be the case is when the network consists of clusters, so that units within a cluster are strongly connected, but the clusters are connected by relatively few links with each other. To deal with such cases we consider within-between decompositions of the network as a way to characterize the variance in terms of the eigenvalues within each cluster and the number of links across clusters.

In Section 2 we introduce the model and estimator under study. In Section 3 we derive bounds on the variance of the estimator. In Section 4 we provide corresponding bounds for parameter differences. In Section 5 we present our results on within-between decompositions of the network. In Section 6 we discuss weighted graphs. Concluding remarks end the paper. An Appendix contains additional results. All technical proofs are available as supplementary material.

2 Model and estimator

Consider a graph $\mathcal{G} := \mathcal{G}(V, E)$ where $m := |E|$ edges are placed between $n := |V|$ vertices. We will work with a simple undirected graph without loops. Without loss of generality we label the vertices by natural numbers, so $V = \{1, \ldots, n\}$. The set E contains the $m \leq n(n-1)/2$ unordered pairs (i, j) from the product set $V \times V$ that are connected by an edge, where we assume throughout that $m > 0$. Vertices i and j are said to be connected if $\mathcal{G}$ contains a path from i to j, and the graph $\mathcal{G}$ is said to be connected if every pair of vertices in the graph is connected.

2.1 A fixed-effect model

Our interest lies in estimating a linear regression model where outcomes are labelled by elements of E. For each $(i, j) \in E$, we observe the real-valued outcome

$$y_{ij} = -y_{ji} = \alpha_i - \alpha_j + u_{ij}, \qquad u_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \tag{2.1}$$

where $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ are vertex-specific parameters to be estimated and the $u_{ij} \in \mathbb{R}$ are unobserved disturbances with unknown variance σ^2. Equation (2.1) is overparametrized, so we impose that

$$\sum_{i=1}^{n} \alpha_i = 0. \tag{2.2}$$

The choice of normalization on the α_i is not unique but (2.1) is conventional (see, e.g., Simons and Yao, 1999) and will prove convenient for our purposes.

Equation (2.1) is similar to a regression version of the classic Bradley and Terry (1952) model for paired comparisons.

Example 1 (Inversion of market shares)**.** Consider an extended version of the classic Bradley and Terry (1952) model, where the probability that team i wins against team j equals

$$p_{ij} := \Pr(i \text{ beats } j) = \Lambda(\alpha_i - \alpha_j + u_{ij}),$$

for $\Lambda(a) := (1 + e^{-a})^{-1}$. The odds ratio is

$$\frac{p_{ij}}{1 - p_{ij}} = e^{\alpha_i - \alpha_j + u_{ij}}.$$

This equation fits (2.1) with $y_{ij} = \ln(p_{ij}/(1 - p_{ij}))$ and is estimable provided the p_{ij} are observed (or estimable). One situation where this model arises is in repeated interactions in the Bradley-Terry setting. Suppose that teams i and j meet multiple times and that, at encounter k,

$$i \text{ beats } j \quad \text{if} \quad (\alpha_i - \alpha_j) + u_{ij} > \varepsilon_{ijk},$$

where $\varepsilon_{ijk} \sim$ i.i.d. Λ. Then p_{ij} can be recovered nonparametrically (Berry, 1994). Note that, here, α_i and α_j represent team-specific heterogeneity while u_{ij} captures heterogeneity that is specific to the match-up. □

Example 2 (Matched employer-employee data)**.** Partition V as $V_1 \cup V_2$ and consider a bipartite graph. That is, suppose that E is a subset of the product set $V_1 \times V_2$. Then edges are formed between the vertex sets V_1 and V_2 but not within V_1 and V_2. So, for an edge (i, j) we necessarily have that $i \in V_1$ and $j \in V_2$. A leading example of a regression model here are wage regressions as in Abowd, Kramarz and Margolis (1999), where the log wage of worker i in firm j decomposes as

$$y_{ij} = \mu_i + \eta_j + u_{ij},$$

for worker effects μ_i and firm effects η_j. To obtain (2.1) we set

$$\alpha_i = \begin{cases} \mu_i & \text{if } i \in V_1, \\ -\eta_i & \text{if } i \in V_2. \end{cases}$$

Choosing the sign in front of η_i is without loss of generality here because the graph under consideration is bipartite; links are only formed between, but never within, V_1 and V_2. We extend this example to panel data, where workers and firms are observed over multiple time periods, later. □

The literature on estimation of fixed-effect models for network data typically assumes that $m = n(n-1)/2$, that is, that each vertex is connected to all other vertices by a path of length one; see Simons and Yao (1999) and Yan and Xu (2013) for results on the Bradley and Terry (1952) model, Dzemski (2014) and Graham (2015) for work on network-formation models, and Fernández-Val and Weidner (2016) for two-way models for panel data. In this case, distribution theory for the maximum-likelihood estimator of the α_i in (2.1) would be rather standard, with the estimator of each of the α_i being unbiased, normally distributed, and converging at the $\sqrt{n}$-rate. In this paper we specifically study the case of an incomplete graph. Our aim is to see how the structure of $\mathcal{G}$ affects the precision of statistical inference. As of yet, this is an unexplored issue in the literature. Allowing for incomplete graphs is important, as data, where all vertices interact, is rare. In country-level data on bilateral trade, for example, only around half of the potential trade flows are realized. Similarly, in the bipartite graphs of workers and firms in Example 2, each worker is related to at most a handful of firms. Finally, friendship networks are typically sparse; see, for example, the data of Jackson, Rodriguez-Barraquer and Tan (2012).

While the model in (2.1) may appear overly restrictive, we note that certain features are not essential to the following analysis. For example, the presumption of normality and the assumption of homoskedastic disturbances could easily be dispensed with. They are introduced here as they allow us to focus on exact finite-sample inference. Also, everything to follow can be modified to hold for weighted graphs. One example would be a situation where we observe multiple outcomes for each $(i, j) \in E$. We will come back on each of these issues in more detail at a later stage. Our choice of (2.1) is motivated by a desire to concentrate on a setting that contains all essential features of a fixed-effect model for random graphs while at the same time connecting as much as possible to the literature on graph theory.

2.2 Estimation and inference

In the following we will work under the convention that $i < j$ for $(i, j) \in E$. This choice imposes an orientation on the graph $\mathcal{G}$, and the corresponding oriented incidence matrix of $\mathcal{G}$ is the $m \times n$ matrix $\boldsymbol{B}$ with entries

$$(\boldsymbol{B})_{ei} := \begin{cases} 1 & \text{if the } e^{\text{th}} \text{ edge is given by } (i, j) \in E \text{ for some } j \in V, \\ -1 & \text{if the } e^{\text{th}} \text{ edge is given by } (j, i) \in E \text{ for some } j \in V, \\ 0 & \text{otherwise.} \end{cases}$$

The incidence matrix fully describes $\mathcal{G}$. Note that the oriented incidence matrix is unique up to negation of any of the columns, since negating the entries of a rows corresponds to reversing the orientation of an edge. Moreover, the analysis to follow is invariant to our choice of orientation. Indeed, changing the orientation of the edge (i, j) jointly with the sign of y_{ij} leaves model (2.1) invariant. Throughout, the network structure is treated as fixed, that is, $\boldsymbol{B}$ is conditioned on.

Let $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)'$. Collect all outcomes in the m-vector $\boldsymbol{y}$ and all regression errors in the m-vector $\boldsymbol{u}$. Write $\boldsymbol{\iota}_n$ for the n-vector of ones and $\boldsymbol{I}_m$ for the $m \times m$ identity matrix. Equations (2.1)–(2.2) can then be written as

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{\alpha} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_m),$$

subject to

$$\boldsymbol{\alpha}'\boldsymbol{\iota}_n = 0.$$

Because of normality of $\boldsymbol{u}$, the maximum-likelihood estimator of $\boldsymbol{\alpha}$ is equal to the (ordinary) least-squares estimator, that is,

$$\widehat{\boldsymbol{\alpha}} := (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_n)' = \underset{\boldsymbol{a} \in \{\boldsymbol{a} \in \mathbb{R}^n : \boldsymbol{a}'\boldsymbol{\iota}_n = 0\}}{\arg\min} \|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}\|^2, \tag{2.3}$$

where $\|\cdot\|$ denotes the Euclidean norm.

We first address existence and uniqueness of $\widehat{\boldsymbol{\alpha}}$. Here and later, we let $\boldsymbol{M}^\dagger$ denote the Moore-Penrose pseudoinverse of matrix $\boldsymbol{M}$.

Lemma 1 (Existence)**.** *Let $\mathcal{G}$ be connected. Then*

$$\widehat{\boldsymbol{\alpha}} = (\boldsymbol{B}'\boldsymbol{B})^{\dagger}\boldsymbol{B}'\boldsymbol{y}$$

and is unique.[1]

The need for a pseudoinverse arises because $\boldsymbol{B}'\boldsymbol{B}$ is singular, as $\boldsymbol{B}\boldsymbol{\iota}_n = 0$. The use of the Moore-Penrose pseudoinverse follows from our normalization choice on $\boldsymbol{\alpha}$, that is, $\boldsymbol{\alpha}'\boldsymbol{\iota}_n = 0$. The result of the lemma is intuitive. If $\mathcal{G}$ is connected, then $m \geq n-1$ must hold, and the zero eigenvalue of $\boldsymbol{B}'\boldsymbol{B}$ has multiplicity one and corresponding eigenvector $\boldsymbol{\iota}_n$; see our discussion of the Laplacian matrix below. If $\mathcal{G}$ is disconnected our analysis for $\widehat{\boldsymbol{\alpha}}$ could be applied separately to each connected component.

The following theorem is immediate.

Theorem 1 (Sampling distribution)**.** *Let $\mathcal{G}$ be connected. Then*

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

for any n.

The main implication of Theorem 1 is the sampling distribution of the conventional F-statistic for testing linear hypotheses on $\boldsymbol{\alpha}$.

Corollary 1 (Inference)**.** *Let $\boldsymbol{R}$ be an $n \times r$ matrix of maximal column rank that is linearly independent of $\boldsymbol{\iota}_n$. If $\mathcal{G}$ is connected, then*

$$\frac{m-(n-1)}{r}\,\frac{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})'\boldsymbol{R}\,(\boldsymbol{R}'(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\boldsymbol{R})^{-1}\boldsymbol{R}'\,(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})}{(\boldsymbol{y}-\boldsymbol{B}\widehat{\boldsymbol{\alpha}})'(\boldsymbol{y}-\boldsymbol{B}\widehat{\boldsymbol{\alpha}})}$$

follows an F-distribution with parameters r and $m-(n-1)$.

The F-statistic can be used to test the null hypothesis that $\boldsymbol{R}\boldsymbol{\alpha} = \boldsymbol{0}$ against the alternative that $\boldsymbol{R}\boldsymbol{\alpha} \neq \boldsymbol{0}$. The requirement that $\boldsymbol{R}$ is linearly independent of $\boldsymbol{\iota}_n$ is needed because

[1] Note that $(\boldsymbol{B}'\boldsymbol{B})^{\dagger}\boldsymbol{B}'$ equals $\boldsymbol{B}^{\dagger}$, and so the expression for the estimator could be shortened. Our choice to highlight the longer formulation in the lemma is motivated by the developments to follow, where the matrix $\boldsymbol{B}'\boldsymbol{B}$ features prominently.

$\boldsymbol{\iota}_n'\boldsymbol{\alpha} = 0$ holds by construction. The degrees of freedom being $m-(n-1)$ rather than $m-n$ is for the same reason.

Corollary 1 shows that test statistics and confidence bounds constructed in the usual way will have correct coverage. This is a direct consequence of Theorem 1. These results, however, do not aid in understanding when test statistics will have low power or when confidence bounds will be wide. In the sequel we aim to understand how the structure of the network affects the standard error of the least-squares estimator. Such an analysis is also a useful aid when setting up sampling designs. Furthermore, it also yields conditions for consistent estimation and asymptotically-valid inference under non-normality for sequences of growing networks.

3 Network structure and variance bounds

Theorem 1 shows that, up to the scalar factor σ^2, the variance of $\widehat{\boldsymbol{\alpha}}$ is completely determined by the $n \times n$ Laplacian matrix of $\mathcal{G}$,

$$\boldsymbol{L} := \boldsymbol{B}'\boldsymbol{B} = \boldsymbol{D} - \boldsymbol{A},$$

where $\boldsymbol{D} := \operatorname{diag}(d_1, \ldots, d_n) = \operatorname{diag}(\boldsymbol{B}'\boldsymbol{B})$ is the degree matrix and $\boldsymbol{A}$ is the $n \times n$ adjacency matrix of $\mathcal{G}$, with entries

$$(\boldsymbol{A})_{ij} := \begin{cases} 1 & \text{if } (i,j) \in E \text{ or } (j,i) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Note that d_i, the degree of i, equals the number of vertices that vertex i is connected to.

It will be convenient to work with the normalized Laplacian

$$\boldsymbol{S} := \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{I}_n - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}.$$

We have $(\boldsymbol{L}^\dagger)_{ii} = d_i^{-1}(\boldsymbol{S}^\dagger)_{ii}$, and so

$$\operatorname{var}(\widehat{\alpha}_i) = \mathbb{E}((\widehat{\alpha}_i - \alpha_i)^2) = \frac{\sigma^2}{d_i}\,(\boldsymbol{S}^\dagger)_{ii}. \tag{3.1}$$

Equation (3.1) highlights the importance of the degree d_i, which is the effective number of observations that are used to infer α_i. However, (3.1) does not imply that $\text{var}(\widehat{\alpha}_i)$ shrinks as $d_i \to \infty$, nor would it give a convergence rate if it did, as the normalized Laplacian matrix of $\mathcal{G}$, too, changes when n grows.

3.1 Zero-order bound

To make progress on bounding the variance, let λ_i denote the ith eigenvalue of $\boldsymbol{S}$, arranged in increasing order; so, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. From Chung (1997, Lemma 1.7) we have $\min_i \lambda_i = 0$ and $\max_i \lambda_i \leq 2$. Zero is always an eigenvalue of $\boldsymbol{S}$ because $\boldsymbol{B}\boldsymbol{\iota}_n = 0$, but, if $\mathcal{G}$ is connected, it has multiplicity one. That is, $\lambda_2 > 0$ is the smallest non-zero eigenvalue of the normalized Laplacian when $\mathcal{G}$ is connected. As a simple example, $\lambda_2 = n/(n-1)$ when $\mathcal{G}$ is complete, that is, when $m = n(n-1)/2$.

The following result bounds the variance of $\widehat{\boldsymbol{\alpha}}$ as a function of λ_2.

Theorem 2 (Global bound)**.** *Let $\mathcal{G}$ be connected. Then*

$$\text{var}(\widehat{\alpha}_i) \leq \frac{1}{d_i}\frac{\sigma^2}{\lambda_2}.$$

The theorem follows from (3.1) and the fact that $(\boldsymbol{S}^{\dagger})_{ii} \leq \|\boldsymbol{S}^{\dagger}\|_2 = \lambda_2$, where $\|.\|_2$ refers to the spectral norm; see the proof in the supplementary material for further details. We note that, analogous to Lemma 2, we can also show that

$$\text{var}(\widehat{\alpha}_i) \leq \frac{\sigma^2}{\tilde{\lambda}_2},$$

where $\tilde{\lambda}_2$ is the smallest non-zero eigenvalue of the (unnormalized) Laplacian $\boldsymbol{L}$. In the graph literature, the spectrum of $\boldsymbol{L}$ has been the subject of more study than that of $\boldsymbol{S}$. However,

$$\tilde{\lambda}_2 \leq \frac{n}{n-1} \min_{i \in V} d_i.$$

Thus, $\tilde{\lambda}_2$ may be very small—and the corresponding bound on $\text{var}(\widehat{\alpha}_i)$ very large—as soon as a single vertex in V has a small degree, making it an unattractive quantity for our purposes.

To interpret the bound it is useful to connect it to the Cheeger constant,

$$C := \min_{U \in \left\{U \subset V : 0 < \sum_{i \in U} d_i \leq m\right\}} \frac{\sum_{i \in U} \sum_{j \notin U} (\boldsymbol{A})_{ij}}{\sum_{i \in U} d_i}.$$

The constant $C \in [0,1]$ reflects how difficult it is to disconnect $\mathcal{G}$ by removing edges. Moreover, a larger value of C implies a more strongly-connected graph. From Chung (1997, Theorems 2.1 and 2.3),

$$2C \geq \lambda_2 \geq 1 - \sqrt{1 - C^2} \geq \frac{1}{2} C^2. \tag{3.2}$$

Hence, Theorem 2 states that inference will be more precise when the graph is more strongly connected.

Theorem 2 also allows to derive some asymptotic properties under sequences of growing networks $\mathcal{G}$. First, we find the pointwise consistency result

$$(\widehat{\alpha}_i - \alpha_i) \xrightarrow{p} 0 \quad \text{if} \quad \lambda_2 \, d_i \to \infty.$$

This result allows $\lambda_2 \to 0$ as $n \to \infty$. Second, letting h be the harmonic mean of the sequence $d_1, \ldots, d_n$, we have

$$\frac{\mathbb{E}(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2)}{n} \leq \frac{1}{h} \frac{\sigma^2}{\lambda_2},$$

and so

$$\frac{\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|}{\sqrt{n}} \xrightarrow{p} 0 \quad \text{if} \quad \lambda_2 \, h \to \infty \quad \text{as} \quad n \to \infty,$$

by an application of Markov's inequality.

Example 3 (Erdős-Rényi graph)**.** Consider the Erdős and Rényi (1959) random-graph model, where edges between n vertices are formed independently with probability p_n. The threshold on p_n for $\mathcal{G}$ to be connected is $\ln(n)/n$. That is, if

$$p_n = c \, \frac{\ln n}{n}$$

for a constant c, then, as $n \to \infty$, with probability approaching one, $\mathcal{G}$ is disconnected if $c < 1$ and connected if $c > 1$ (Erdős and Rényi, 1960). In the former case, $\lambda_2 \to 0$ while, in the latter case, $\lambda_2 \to 1$, almost surely; see Hoffman, Kahle and Paquette (2013, Theorem 1.1). and Kolokolnikov, Osting and von Brecht (2014, Corollary 1.2). □

We next proceed by refining the variance bound in Theorem 2 to take into account the local structure of the graph around vertex i.

3.2 First-order bound

A refinement of Theorem 2 takes into account the connectivity of the direct neighbors of i. Here, we call a direct neighbor, or a path-one neighbor, a vertex to which i is connected via a path of length one. Similarly, we will call those vertices that have geodesic distance equal to two from i path-two neighbors of i. The collection of direct neighbors of vertex i is

$$[i] := \{j \in V : (i,j) \in E \text{ or } (j,i) \in E\};$$

note that $|[i]| = d_i$. Let

$$h_i := \left(\frac{1}{d_i} \sum_{j \in [i]} \frac{1}{d_j} \right)^{-1}, \tag{3.3}$$

the harmonic mean of the degrees of all $j \in [i]$. Note that, for a given vertex i, h_i is increasing in the degree of its direct neighbors.

Theorem 3 (First-order bound)**.** *Let $\mathcal{G}$ be connected. Then*

$$\frac{\sigma^2}{d_i}\left(1 - \frac{2}{n}\right) \;\leq\; \operatorname{var}(\widehat{\alpha}_i) \;\leq\; \frac{\sigma^2}{d_i}\left(1 - \frac{2}{n} + \frac{1}{\lambda_2 h_i}\right).$$

Theorem 3 states that, for a given degree d_i and global connectivity measure λ_2, the upper bound on the variance of $\widehat{\alpha}_i$ is smaller if the direct neighbors of vertex i are themselves more strongly connected to other vertices in the network. Another implication of the theorem is the rate refinement

$$\operatorname{var}(\widehat{\alpha}_i) = \frac{\sigma^2}{d_i} + o(d_i^{-1}), \tag{3.4}$$

provided that $\lambda_2 h_i \to \infty$ as $n \to \infty$. Furthermore, the parametric rate is achievable even if λ_2 is not treated as fixed.

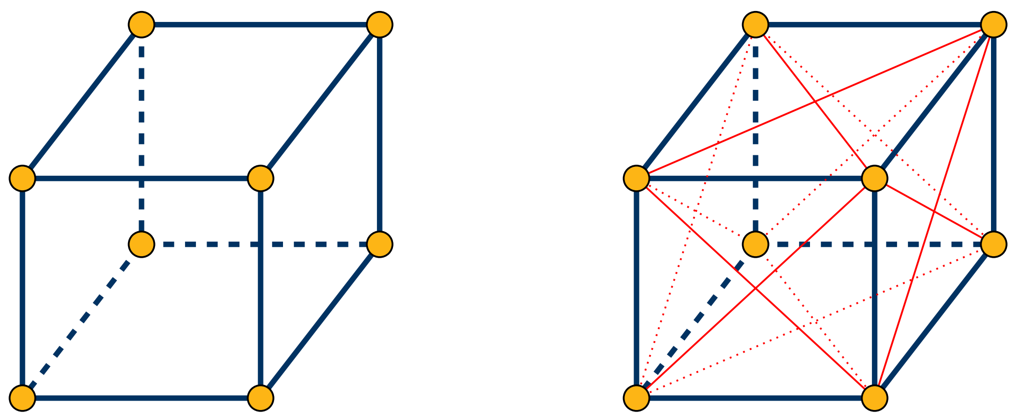

Figure 1: three-dimensional hypercube (left) and extended hypercube (right).

In the Appendix we present a refinement of Theorem 3 that accounts for the dependence on h_i in the lower bound as well, and also adjusts the upper bound for overlap between $[i]$ and the sets $[j_1], \ldots, [j_{d_i}]$ for $j_1, \ldots, j_{d_i} \in [i]$, that is, for common neighbors. These bounds can be particularly useful when h_i is small, but are vacuous when all path-two neighbors of vertex i are also path-one neighbors. This is the case, for example, in the complete graph, where all vertices are direct neighbors.

We illustrate the usefulness of improving on Theorem 2 in our running example of a random graph.

Example 3 (cont'd). Consider the Erdős and Rényi (1959) random-graph model with $p_n = c \ln(n)/n$ for $c > 1$. Let i be a randomly chosen vertex. Then, as $n \to \infty$, we have, almost surely,

$$\lambda_2 \to 1, \qquad \frac{d_i}{\ln n} \to c, \qquad \frac{h_i}{\ln n} \to c.$$

Consequently,

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

follows from Theorem 3. □

The next example deals with an analytically-tractable case where $\lambda_2 \to 0$ as $n \to \infty$.

Example 4 (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 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,

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

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

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

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 bound in Theorem A.1 of the Appendix one obtains $\mathrm{var}(\widehat{\alpha}_i) = \sigma^2/N + O(N^{-2})$, that is, (3.4) holds. See the Appendix for details. □

Theorem 3 allows to establish the convergence rate for the hypercube, but the conditions are too stringent to obtain (3.4). 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 5 (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 1 for an example. The resulting graph still has $n = 2^N$ vertices, but now has $N(N+1)\,2^{N-1}$ edges. Here,

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

so that $\lambda_2 \, h_i \to \infty$ holds, despite $\lambda_2 \to 0$ as $n \to \infty$. Theorem 3 therefore implies (3.4) 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.

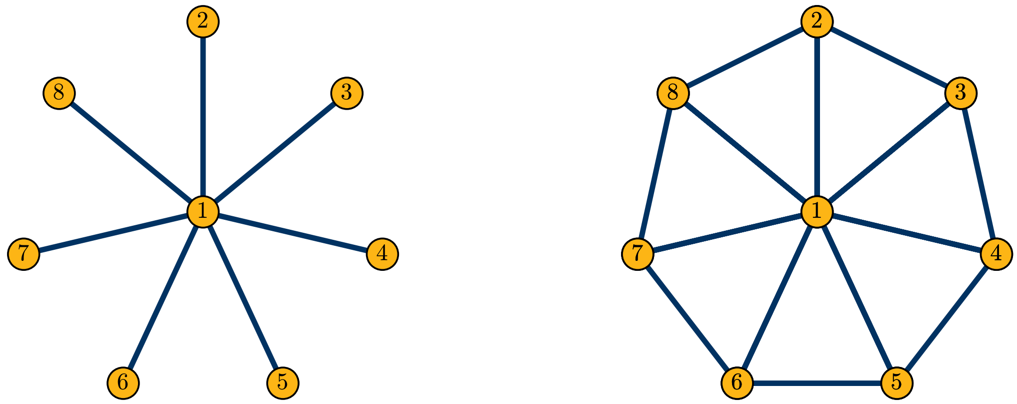

Figure 2: Star graph (left) and Wheel graph (right) for $n = 8$.

Example 6 (Star graph)**.** Consider a Star graph around the central vertex 1, that is, the graph with n vertices and edges

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

see the left hand side of Figure 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 $\text{var}(\widehat{\alpha}_1) = O(n^{-1})$, 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 (3.4) holds, these α_i cannot be estimated consistently as $d_i = 1$. □

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

Example 7 (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 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. $\square$

3.3 Asymptotic linearity under weaker assumptions

The bounds in Theorem 3 continue to hold when the errors in (2.1) are non-normal, as the variance of $\widehat{\alpha}_i$ depends only on the first and second moments of the data. The asymptotic statements obtained in the previous subsection, too, carry over. We now want to briefly discuss how the results can be extended to also allow for heteroskedasticity and correlation in the error term.

If we only assume that

$$\mathbb{E}(\boldsymbol{u}) = \boldsymbol{0}, \qquad \|\mathbb{E}(\boldsymbol{u}\boldsymbol{u}')\|_2 \leq \overline{\sigma}^2, \tag{3.5}$$

we have the following result.

Theorem 4 (Generalized first-order approximation)**.** *Suppose that (2.1) is weakened by imposing only (3.5). Let $\mathcal{G}$ be connected. Then*

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

where $\mathbb{E}(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) \leq \overline{\sigma}^2/(\lambda_2 h_i)$.

It follows that

$$\widehat{\alpha}_i \overset{a}{\sim} \mathcal{N}\left(\alpha_i, \frac{\omega_i^2}{d_i}\right)$$

if $d_i^{-1/2} \sum_{j \in [i]} u_{ij} \overset{d}{\to} \mathcal{N}(0, \omega_i^2)$ for finite ω_i^2, provided $\mathbb{E}(\epsilon_i^2) = o(1)$, which follows from $\lambda_2 h_i \to \infty$ and $\overline{\sigma}^2 < \infty$. Thus, the key asymptotic condition that $\lambda_2 h_i \to \infty$ is unchanged compared to the previous subsection. The corresponding discussion and examples are thus also applicable to the more general situation of heteroscedastic and weakly correlated errors, but now with ω_i^2 featuring in the asymptotic variance.

4 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

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

(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

$$\mathrm{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) = \sigma^2\, r_{ij} = \frac{\sigma^2}{2}\frac{c_{ij}}{m}. \tag{4.1}$$

This is the equivalent of (3.1) for differences.

The counterpart to Theorem 2 is intuitive.

Theorem 5 (Global bound for differences)**.** *Let $\mathcal{G}$ be connected. Then*

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

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 $k \in [i] \cap [j]$. We have the following theorem.

Theorem 6 (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

$$\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) = \sigma^2\left(\frac{1}{d_i}+\frac{1}{d_j}\right), \tag{4.2}$$

as, in that case, both $(\boldsymbol{A})_{ij}$ and the second term in the upper bound in Theorem 6 are zero.

Theorem 6 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.

5 Variance bounds using graph partitioning

The variance bounds obtained so far depend on λ_2, which is a global measure of connectivity. Moreover, for a given vertex i, we require $\lambda_2\, h_i \to \infty$ for our bounds to yield the first-order asymptotic variance. The value of λ_2 may be rather low even if most vertices are rather densely connected. Consequently, Theorem 3 and Theorem 6 may be overly conservative. A leading situation where this will be the case is when the network consists of clusters, so that units within a cluster are strongly connected but the clusters themselves are connected by relatively few links. As a remedy, in this section we obtain bounds on the variance of $\widehat{\alpha}_i$ that are based on partitioning the graph into subgraphs.

Consider a graph $\mathcal{G} = \mathcal{G}(V, E)$. Partition V into q non-empty subsets $V_1, \ldots, V_q$. For each $r = 1, \ldots, q$, let $E_r := E \cap (V_r \times V_r)$, the set of edges connecting all vertices in V_r, so

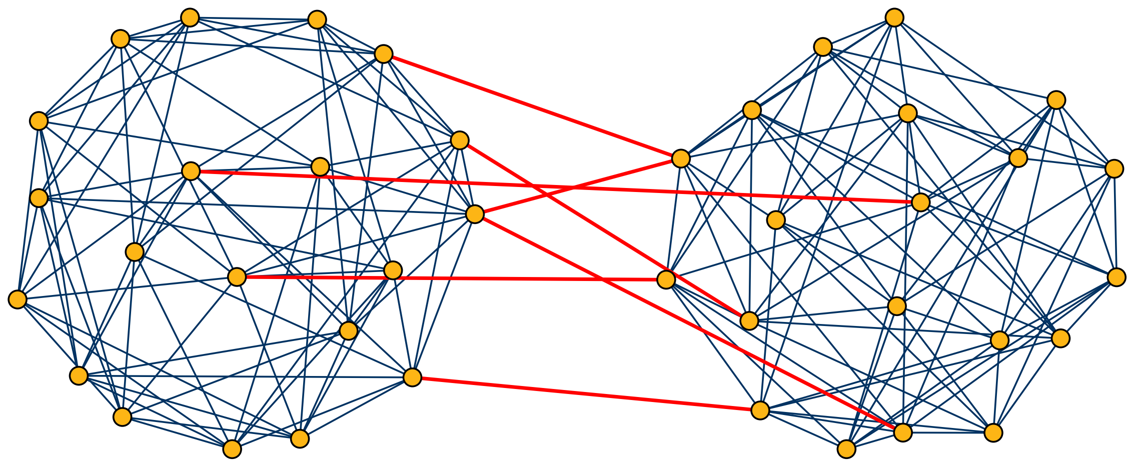

Figure 3: An approximate decomposition of a graph into two components, with many connections inside the two components (blue edges), but few across components (red edges).

that

$$\mathcal{G}_r := \mathcal{G}(V_r, E_r)$$

denotes the subgraph of $\mathcal{G}$ that is induced by V_r. Note that none of these subgraphs have a vertex in common. Throughout this section we assume that $\mathcal{G}$ and each of $\mathcal{G}_1, \ldots, \mathcal{G}_q$ are connected. Let $E_W := \cup_r E_r$ and $E_B := E - E_W$. The graph $\mathcal{G}$ can then be decomposed as

$$\mathcal{G} = \mathcal{G}_W \cup \mathcal{G}_B, \tag{5.1}$$

for $\mathcal{G}_W := \mathcal{G}(V, E_W)$ and $\mathcal{G}_B := \mathcal{G}(V, E_B)$. This is what we call a within-between decomposition of the graph $\mathcal{G}$; the graph $\mathcal{G}_W$ consists of q connected components, $\mathcal{G}_1, \ldots, \mathcal{G}_q$, and the graph $\mathcal{G}_B$ connects these q isolated components. We will let $n_r := |V_r|$ and denote by m_B the number of edges in E_B.

Our variance bounds in terms of λ_2 turn out to be conservative when m_B/n_r is small.

Example 8 (Partition into two sets)**.** Partition V into two sets V_1 and V_2. An example is given in Figure 3. Using, (3.2) we find

$$\lambda_2 \leq \frac{m_B}{2\min\left(\sum_{i\in V_1} d_i, \sum_{i\in V_2} d_i\right)} \leq \frac{m_B}{2\min(n_1, n_2)\ \min_{i\in V} d_i}.$$

Suppose that $\max_{i\in V} d_i / \min_{i\in V} d_i = O(1)$, so that all degrees grow at the same rate as $n \to \infty$. Then $\lambda_2 h_i \to \infty$ requires that

$$\frac{m_B}{\min(n_1, n_2)} \to \infty. \tag{5.2}$$

In this section we show that the actually-required condition for the rate result in (3.4) in this setting is

$$\frac{m_B}{d_i} \to \infty, \tag{5.3}$$

which is considerably weaker. □

Example 3 (cont'd). Specialize Example 8 by assuming that $\mathcal{G}_1$ and $\mathcal{G}_2$ are both Erdős-Rényi graphs with $p_n = c\ln(n)/n$ of equal size, that is, $n_1 = n_2$. Then

$$d_i = c\,\log(n_1)\,(1 + o(1)) \ll n_1,$$

which highlights the importance of the improvement of (5.3) over (5.2). □

To make use of the partitioning $\mathcal{G} = \mathcal{G}_W \cup \mathcal{G}_B$ we decompose $\alpha_1, \ldots, \alpha_n$ accordingly. We let

$$\beta_i := \alpha_i - \gamma_r, \qquad \gamma_r := \frac{1}{n_r} \sum_{i\in V_r} \alpha_i,$$

for all $i \in V_r$ and each $r = 1, \ldots, q$. We let $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_n)'$ and $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_q)'$. The relation between both these vectors and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$ can then be succinctly stated as

$$\boldsymbol{\alpha} = \boldsymbol{\beta} + \boldsymbol{P}'\boldsymbol{\gamma}, \qquad (\boldsymbol{P})_{ri} := \begin{cases} 1 & \text{if } i \in V_r, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

This decomposition gives an alternative (infeasible) estimator of $\boldsymbol{\alpha}$ based on estimators of the within parameter $\boldsymbol{\beta}$ and the between parameter $\boldsymbol{\gamma}$. The estimator of $\boldsymbol{\beta}$ is simply the least-squares estimator applied to the subgraph $\mathcal{G}_W$, subject to the proper normalization constraints, namely

$$\widehat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{b}\in\mathbb{R}^n} \sum_{(i,j)\in E_W} (y_{ij} - (b_i - b_j))^2 \quad \text{s.t.} \quad \sum_{i\in V_r} b_i = 0, \quad r = 1, \ldots, q.$$

Similarly, an (infeasible) least-squares estimator for $\boldsymbol{\gamma}$, which assumes $\boldsymbol{\beta}$ to be known, equals

$$\widehat{\boldsymbol{\gamma}} := \arg\min_{\boldsymbol{g}\in\mathbb{R}^q} \sum_{(i,j)\in E_B} \big(y_{ij} - (\beta_i + g_{r(i)}) + (\beta_j + g_{r(j)})\big)^2 \qquad \text{s.t.} \ \sum_{r=1}^{q} n_r\, g_r = 0,$$

where we write $r(\cdot)$ to denote the map $V \to \{1,\dots,q\}$ that satisfies $i \in V_{r(i)}$. Note that these estimators are statistically independent of each other. The sampling variability of these estimators can be studied, and can be used to sharpen our results on the statistical accuracy of $\widehat{\boldsymbol{\alpha}}$.

Let $\boldsymbol{L}_W$ and $\boldsymbol{L}_B$ be the $n\times n$ Laplacian matrices of the graphs $\mathcal{G}_W$ and $\mathcal{G}_B$, respectively. We also introduce $\boldsymbol{H} := \operatorname{diag}(n_1,\dots,n_q)$ and the $q\times q$ matrix

$$\boldsymbol{L}_* := \boldsymbol{P}\,\boldsymbol{L}_B\boldsymbol{P}',$$

which is the Laplacian matrix of the multigraph $\mathcal{G}_*$ with vertex set $\{1,\dots,q\}$, obtained from $\mathcal{G}$ by edge contraction of the subgraphs $\mathcal{G}_r$, $r\in\{1,\dots,q\}$. Analogous to Theorem 1 for $\widehat{\boldsymbol{\alpha}}$ one can show the following lemma.

Lemma 2 (Variances of component estimators)**.** *Let $\mathcal{G}$ and $\mathcal{G}_1,\dots,\mathcal{G}_q$ be connected. Then*

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta},\ \sigma^2\,\boldsymbol{L}_W^\dagger), \qquad \textit{and} \qquad \widehat{\boldsymbol{\gamma}} \sim \mathcal{N}(\boldsymbol{\gamma},\ \sigma^2\,\boldsymbol{L}_*^{\text{inv}}),$$

for $\boldsymbol{L}_*^{\text{inv}} := \boldsymbol{H}^{-1/2}(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2})^\dagger\boldsymbol{H}^{-1/2}$.[2]

If we label the elements of V such that $V_1 = \{1,\dots,n_1\}$, $V_2 = \{n_1+1,\dots,n_2\}$, etc, then $\boldsymbol{L}_W$ is a block-diagonal matrix with q non-zero blocks given by $\boldsymbol{L}_r$, the Laplacian of $\mathcal{G}_r$, $r = 1,\dots,q$. The Moore-Penrose inverse $\boldsymbol{L}_W^\dagger$ is also block-diagonal with q non-zero blocks given by $\boldsymbol{L}_r^\dagger$. The first part of Lemma 2 therefore is simply Theorem 1 applied separately to each of the connected components of $\mathcal{G}_W$, and all our results from the previous sections apply. For example, with d_i^W the degree of vertex i in $\mathcal{G}_W$, h_i^W the corresponding harmonic mean, and λ_2^r the second-smallest eigenvalue of the normalized Laplacian of $\mathcal{G}_r$, we have

$$\operatorname{var}(\widehat{\beta}_i) = \frac{\sigma^2}{d_i^W} + o\left(\frac{1}{d_i^W}\right), \tag{5.5}$$

[2] $\boldsymbol{L}_*^{\text{inv}}$ is a pseudoinverse of $\boldsymbol{L}_*$, but unless $n_1 = n_2 = \ldots = n_q$ it is not the Moore-Penrose pseudoinverse.

provided that $\lambda_2^{r(i)} h_i^W \to \infty$ as $n \to \infty$. The second part of the lemma deals with the between component of the graph. The following example illustrates the result for the case where $q = 2$.

Example 8 (cont'd). When V is partitioned into two sets V_1 and V_2 we have that

$$\boldsymbol{L}_* = \begin{pmatrix} m_B & -m_B \\ -m_B & m_B \end{pmatrix}, \qquad \boldsymbol{L}_*^{\text{inv}} = \frac{1}{n^2\, m_B} \begin{pmatrix} n_2^2 & -n_1 n_2 \\ -n_1 n_2 & n_1^2 \end{pmatrix}.$$

Lemma 2 then yields

$$\operatorname{var}(\widehat{\gamma}_1) = \left(\frac{n_2}{n}\right)^2 \frac{\sigma^2}{m_B}, \qquad \operatorname{var}(\widehat{\gamma}_2) = \left(\frac{n_1}{n}\right)^2 \frac{\sigma^2}{m_B}, \qquad \operatorname{var}(\widehat{\gamma}_1 - \widehat{\gamma}_2) = \frac{\sigma^2}{m_B}.$$

As one would expect, the variance of $\widehat{\boldsymbol{\gamma}}$ crucially depends on the magnitude of m_B, but also on the relative size n_1 and n_2 of the two graph components. □

The following theorem allows us to use Lemma 2 to bound the variance of our estimator of interest, $\widehat{\boldsymbol{\alpha}}$. To state the result we introduce

$$\kappa := \max_{r: n_r > 1} \; \min_{i \in V_r} \frac{2}{\lambda_2^r} \frac{d_i^B}{d_i^W},$$

where, analogous to d_i^W, we denote by d_i^B the degree of vertex i in the graph $\mathcal{G}_B$. If $n_r = 1$ for all $r \in \{1, \dots, q\}$, then we define $\kappa = 0$.

Theorem 7 (Variance bounds from graph partitioning)**.** *Let $\mathcal{G}$ and $\mathcal{G}_1, \dots, \mathcal{G}_q$ be connected. Then, for any $\boldsymbol{v} \in \mathbb{R}^n$,*

$$\begin{aligned} -\kappa \operatorname{var}\left(\boldsymbol{v}'\widehat{\boldsymbol{\beta}}\right) - 2\kappa^{1/2} \left[\operatorname{var}\left(\boldsymbol{v}'\widehat{\boldsymbol{\beta}}\right) \operatorname{var}(\boldsymbol{v}'\boldsymbol{P}'\widehat{\boldsymbol{\gamma}})\right]^{1/2} \\ \le \operatorname{var}(\boldsymbol{v}'\widehat{\boldsymbol{\alpha}}) - \left[\operatorname{var}\left(\boldsymbol{v}'\widehat{\boldsymbol{\beta}}\right) + \operatorname{var}(\boldsymbol{v}'\boldsymbol{P}'\widehat{\boldsymbol{\gamma}})\right] \\ \le \kappa \operatorname{var}(\boldsymbol{v}'\boldsymbol{P}'\widehat{\boldsymbol{\gamma}}) + 2\kappa^{1/2} \left[\operatorname{var}\left(\boldsymbol{v}'\widehat{\boldsymbol{\beta}}\right) \operatorname{var}(\boldsymbol{v}'\boldsymbol{P}'\widehat{\boldsymbol{\gamma}})\right]^{1/2}. \end{aligned}$$

The theorem shows that, if κ is small, the variance of $\widehat{\boldsymbol{\alpha}}$ is close the the variance of an infeasible estimator of $\boldsymbol{\alpha}$ constructed from (5.4), which equals

$$\sigma^2(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}).$$

One can also show that not only their variances but actually the estimators $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}+\boldsymbol{P}'\widehat{\boldsymbol{\gamma}}$ themselves are close to each other when κ is small; see Section S.2 of the supplementary material.

Only components with $n_r > 1$ enter into the definition of κ. For those components we need $d_i^B \gg \lambda_2^r d_i^W$, uniformly over $i \in V_r$, for κ to be small. This requires that for every vertex $i \in V_r$ the number of between edges is much smaller than the number of within edges, that is, the vertices need to be much more connected within components than between components.

Example 8 (cont'd). Consider the example with two components, $V = V_1 \cup V_2$. Assume that the m_B edges between V_1 and V_2 are chosen such that $\max_{i \in V} d_i^B = O(1)$. For example, if the vertices of those edges are drawn without repetition from V_1 and V_2, then we have $\max_{i \in V} d_i^B = 1$, but this requires $m_B \leq \max(n_1, n_2)$. If we furthermore assume that $\min_{i \in V} \lambda_2^{r(i)} d_i^W \to \infty$, then we have $\kappa \to 0$. This also implies that $\lambda_2^{r(i)} h_i^W \to \infty$, so that (5.5) holds. Applying Theorem 7 we then find, as $n \to \infty$,

$$\operatorname{var}(\widehat{\alpha}_i) \asymp \sigma^2 \left(\frac{1}{d_i} + \left(\frac{n_2}{n}\right)^2 \frac{1}{m_B} \right), \qquad \operatorname{var}(\widehat{\alpha}_j) \asymp \sigma^2 \left(\frac{1}{d_j} + \left(\frac{n_1}{n}\right)^2 \frac{1}{m_B} \right),$$

and

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

where we write d_i instead of d_i^W, because $d_i / d_i^W = 1 + d_i^B / d_i^W \to 1$ under our assumptions. Thus, if $m_B / d_i \to \infty$, then our original first-order results (3.4) still holds. For m_B values that are smaller the asymptotic variance needs to be adjusted. Finally, if $n_1/n_2 \to c \in (0, \infty)$ and $m_B / d_i \to 0$, then the asymptotic variance is completely dominated by the weak global connectivity of $\mathcal{G}$, and the local structure of the graph is no longer of first-order relevance. □

Example 9 (Partition into three sets)**.** Consider an analogous situation as in Example 8, only now with $q = 3$ partitions, each of which containing many vertices; as in Figure 4.

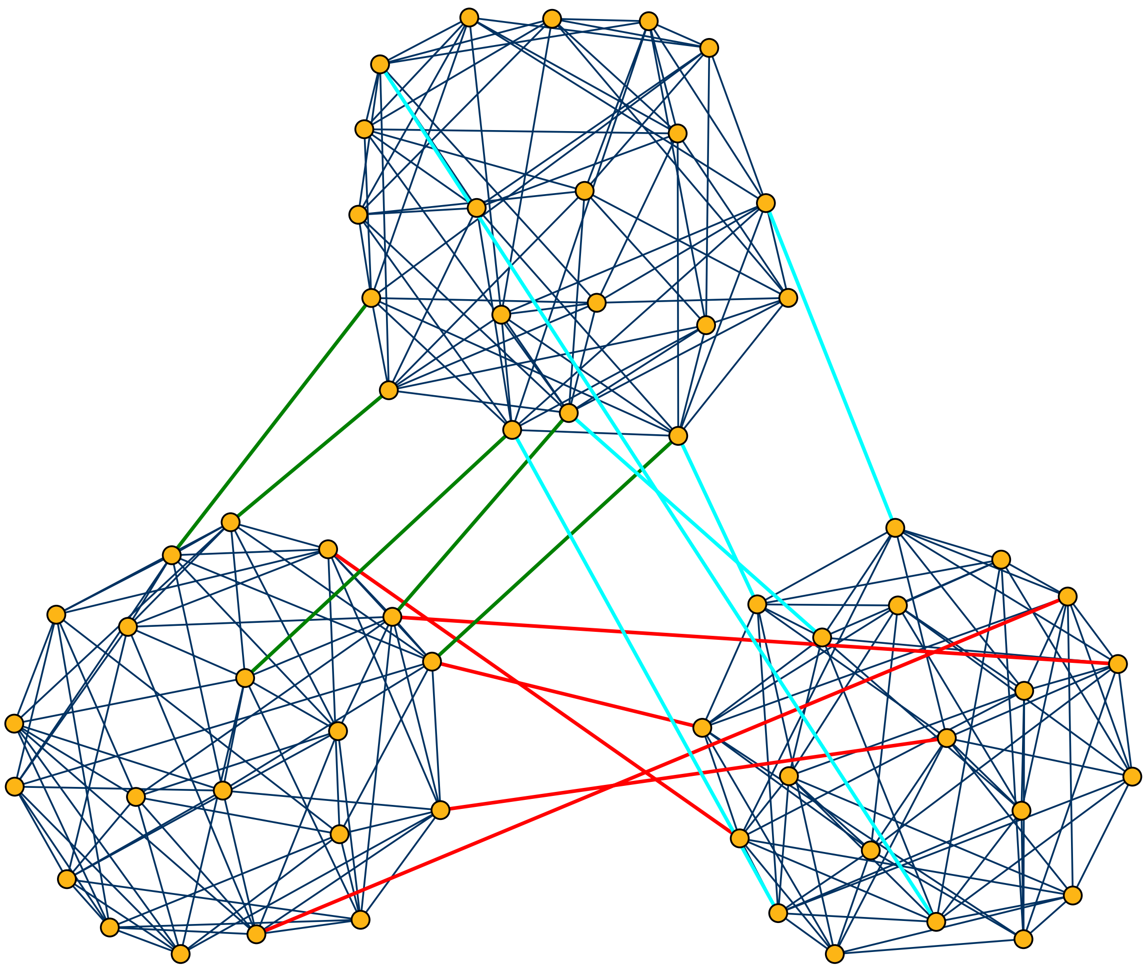

Figure 4: An approximate decomposition of a graph into three components, with many connections inside each of the three components (blue edges), but few across components (red, green and cyan edges).

For $r, s \in \{1, 2, 3\}$, with $r \neq s$, let m_{rs} be the number of edges between V_r and V_s. Then

$$\boldsymbol{L}_* = \begin{pmatrix} m_{12} + m_{13} & -m_{12} & -m_{13} \\ -m_{12} & m_{12} + m_{23} & -m_{23} \\ -m_{13} & -m_{23} & m_{13} + m_{23} \end{pmatrix}.$$

For the pseudo-inverse $\boldsymbol{L}_*^{\text{inv}}$ we calculate

$$\left(\boldsymbol{L}_*^{\text{inv}}\right)_{11} = \frac{n_3^2\, m_{12} + n_2^2\, m_{13} + (n_2 + n_3)^2 m_{23}}{n^2\left(m_{12}m_{13} + m_{12}m_{23} + m_{13}m_{23}\right)},$$

and

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}' \boldsymbol{L}_*^{\text{inv}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \frac{m_{12} + m_{23}}{m_{12}m_{13} + m_{12}m_{23} + m_{13}m_{23}}.$$

Thus, if we again assume that $\max_{i \in V} d_i^B = O(1)$ and $\min_{i \in V} \lambda_2^{r(i)} d_i^W \to \infty$, applying Theorem 7 for $i \in V_1$ and $j \in V_3$ we have

$$\operatorname{var}(\widehat{\alpha}_i) \asymp \sigma^2 \left(\frac{1}{d_i} + \frac{n_3^2\, m_{12} + n_2^2\, m_{13} + (n_2 + n_3)^2 m_{23}}{n^2\left(m_{12}m_{13} + m_{12}m_{23} + m_{13}m_{23}\right)} \right),$$

and

$$\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \asymp \sigma^2 \left(\frac{1}{d_i} + \frac{1}{d_j} + \frac{m_{12} + m_{23}}{m_{12}m_{13} + m_{12}m_{23} + m_{13}m_{23}} \right).$$

The asymptotic result for $\operatorname{var}(\widehat{\alpha}_i)$ again depends not only on m_{12}, m_{13} and m_{23}, but also on the relative component sizes n_1, n_2 and n_3, while those component sizes do not matter at all for the asymptotic result on $\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j)$. Our original first-order results (4.2) for $\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j)$ still holds if, for example, either $m_{13}/\min(d_i, d_j) \to \infty$ or $\min(m_{12}, m_{23})/\min(d_i, d_j) \to \infty$. An interesting special case is $m_{12} = 0$, where the result simplifies to

$$\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \asymp \sigma^2 \left(\frac{1}{d_i} + \frac{1}{d_j} + \frac{1}{m_{12}} + \frac{1}{m_{23}} \right).$$

This simple formula generalizes to four and more components. For example, for $q = 4$ with $m_{13} = m_{14} = m_{24} = 0$ we find for $i \in V_1$ and $j \in V_4$ under the asymptotic assumptions above that $\operatorname{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \asymp \sigma^2(1/d_i + 1/d_j + 1/m_{12} + 1/m_{23} + 1/m_{34})$. □

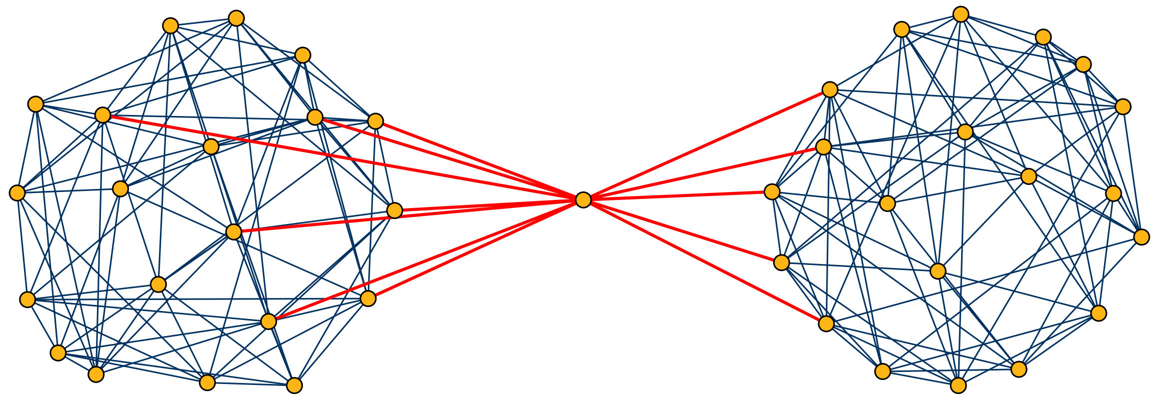

Figure 5: An approximate decomposition of a graph into three components with $n_2 = 1$ and $m_{13} = 0$.

Theorem 7 is also applicable if $n_r = 1$ for some $r \in \{1, \dots, q\}$, as the following example illustrates.

Example 10 (Subgraphs connected by individual vertices)**.** Take the setup of Example 9 but now set

$$n_2 = 1, \qquad n_1, n_3 = \text{ large }, \qquad m_{13} = 0.$$

This situation is depicted in Figure 5. This is a setting where one vertex $i \in V_2$ provides the only connection between V_1 and V_3; the degree of this connecting vertex is $m_{12} + m_{23}$. For $r \in \{1, 3\}$ we again assume $\max_{i \in V_r} d_i^B = O(1)$ and $\min_{i \in V_r} \lambda_2^r d_i^W \to \infty$. Application of Theorem 7 then gives the same asymptotic conclusions as in Examples 9, in particular for $i \in V_1$ and $j \in V_3$ we again find

$$\text{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \asymp \sigma^2 \left(\frac{1}{d_i} + \frac{1}{d_j} + \frac{1}{m_{12}} + \frac{1}{m_{23}} \right).$$

This example can again be extended. If we introduce an additional vertex $i \in V_4$ that also connects the subgraphs V_1 and V_3, and we have $m_{24} = 0$ and $n_4 = 1$, then applying Theorem 7 with $q = 4$, $i \in V_1$ and $j \in V_3$ yields

$$\text{var}(\widehat{\alpha}_i - \widehat{\alpha}_j) \asymp \sigma^2 \left(\frac{1}{d_i} + \frac{1}{d_j} + \left(\left(\frac{1}{m_{12}} + \frac{1}{m_{23}} \right)^{-1} + \left(\frac{1}{m_{14}} + \frac{1}{m_{34}} \right)^{-1} \right)^{-1} \right).$$

The result for three and more connecting vertices is analogous. □

6 Weighted graphs

So far we have considered simple graphs. Our variance bounds generalize to weighted graphs. Let $\mathcal{G}$ be an undirected weighted graph with associated (weighted) adjacency matrix $\boldsymbol{A}$. A simple example is a multigraph, which differs from a simple graph in that multiple edges may exist between vertices. In this case, $(\boldsymbol{A})_{ij}$ equals the number of edges between i and j. More generally, $\boldsymbol{A}$ is symmetric, has diagonal entries equal to zero, and has off-diagonal entries that are non-negative.

Our variance bounds generalize to situations where an estimator $\check{\boldsymbol{\alpha}}$, constructed from $\mathcal{G}$, has variance $\boldsymbol{L}^\dagger$ for

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A},$$

where, as before, $\boldsymbol{D}$ is a diagonal (weighted) degree matrix with entries $d_i = \sum_{j=1}^n (\boldsymbol{A})_{ij}$. A symmetric matrix $\boldsymbol{L}$ is such a Laplacian matrix if and only if

(i) All off-diagonal elements of $\boldsymbol{L}$ are negative;

(ii) All column sums of $\boldsymbol{L}$ are equal to zero;

(iii) $\operatorname{rank}(\boldsymbol{L}) = n - 1$.

The variance bounds in Theorems 2–7 continue to apply, on setting $\sigma = 1$ and redefining the harmonic means

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

with $d_{ij} = \sum_{k \in V} (\boldsymbol{A})_{ik}\,(\boldsymbol{A})_{jk}$. Our proofs of the theorems fully cover the weighted-graph case.

We give some examples of weighted graphs.

Example 11 (Weighted least squares)**.** We generalize the least-squares estimator in (2.3) to situations where $(i, j) \in E$ interact on $m_{ij} \geq 1$ occasions and errors are heteroskedastic across meetings. Using obvious notation, the weighted least-squares estimator of $\boldsymbol{\alpha}$ equals

$$\check{\boldsymbol{\alpha}} := \arg \min_{\boldsymbol{a} \in \{\boldsymbol{a} \in \mathbb{R}^n : \boldsymbol{a}' \boldsymbol{\iota}_n = 0\}} \sum_{(i,j)\in E} \sum_{k=1}^{m_{ij}} \left(\frac{y_{ijk} - (a_i - a_j)}{\sigma_k} \right)^2 .$$

Let $\boldsymbol{A}$ be the weighted adjacency matrix with entries

$$(\boldsymbol{A})_{ij} := \begin{cases} \sum_{k=1}^{m_{ij}} \sigma_k^{-2} & \text{if } (i,j) \in E \text{ or } (j,i) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

and let $\boldsymbol{L}$ be the associated Laplacian matrix. Then Theorem 1 can be suitably extended to yield $\check{\boldsymbol{\alpha}} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{L}^\dagger)$. □

Example 12 (Profiled estimator for bipartite graph)**.** Consider a bipartite graph $\mathcal{G}$, where V is partitioned as $V_1 \cup V_2$ and edges are formed between V_1 and V_2 but not within these sets. Let $n_1 := |V_1|$ and $n_2 := |V_2|$. The Laplacian is

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A} = \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_2 \end{pmatrix} - \begin{pmatrix} \boldsymbol{0} & \boldsymbol{C} \\ \boldsymbol{C}' & \boldsymbol{0} \end{pmatrix},$$

where $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are $n_1 \times n_1$ and $n_2 \times n_2$ diagonal degree matrices and $\boldsymbol{C}$ is the $n_1 \times n_2$ upper-right block of the adjacency matrix of the graph. Decompose $\boldsymbol{\alpha}$ accordingly as $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2')'$. The corresponding estimator $\widehat{\boldsymbol{\alpha}}$ is defined in (2.3) for the case of a simple graph, but the following construction works for any estimator that satisfies $\mathrm{var}(\widehat{\boldsymbol{\alpha}}) = \sigma^2 \boldsymbol{L}^\dagger$, with $\boldsymbol{L}$ being the Laplacian matrix of a simple, weighted or multigraph, as described above (we may simply have $\sigma = 1$). We also define

$$\check{\boldsymbol{\alpha}}_2 := \widehat{\boldsymbol{\alpha}}_2 - \overline{\widehat{\boldsymbol{\alpha}}}_2, \qquad \overline{\widehat{\boldsymbol{\alpha}}}_2 := \frac{1}{n_2} \sum_{i \in V_2} \widehat{\alpha}_i.$$

corresponding to the natural normalization $\boldsymbol{\iota}_{n_2}' \check{\boldsymbol{\alpha}}_2 = 0$. By the block-inversion formula we find

$$\mathrm{var}(\check{\boldsymbol{\alpha}}_2) = \check{\boldsymbol{L}}^\dagger, \qquad \check{\boldsymbol{L}} := \sigma^{-2} \left(\boldsymbol{D}_2 - \boldsymbol{C}' \boldsymbol{D}_1^{-1} \boldsymbol{C} \right).$$

This is the variance formula after profiling-out all the parameters corresponding to vertices in V_1. It can be verified that $\check{\boldsymbol{L}}$ satisfies the Conditions (i)–(iii). The adjacency matrix

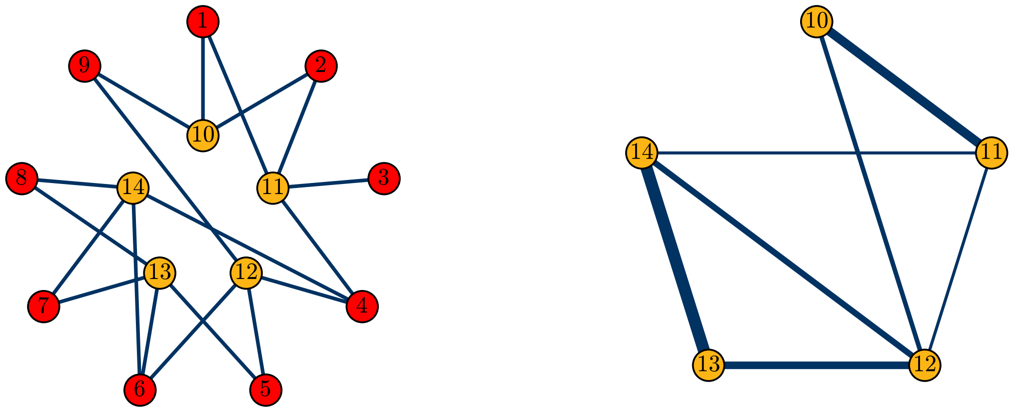

Figure 6: A simple bipartite graph $\mathcal{G}$ (left) with links between V_1 (red vertices) and V_2 (yellow vertices), and the induced weighted graph $\check{\mathcal{G}}$ (right) on V_2 alone resulting from profiling out the parameters associated with V_1.

of the corresponding graph, say $\check{\mathcal{G}}$, that involves only the vertices in V_2 is given by the off-diagonal part of $\sigma^{-2}\,\boldsymbol{C}'\boldsymbol{D}_1^{-1}\boldsymbol{C}$. Thus, even when starting with a simple bipartite graph $\mathcal{G}$ we naturally obtain a weighted graph $\check{\mathcal{G}}$ when profiling out some of the parameters. Moreover, two vertices in $\check{\mathcal{G}}$ are connected if and only if they are path-two neighbors in the original graph $\mathcal{G}$. □

An interesting application of Example 12 is Example 2.

Example 2 (cont'd) (Matched employer-employee data)**.** Consider the wage regression with panel data, where the log wage of worker i in firm j at time t equals

$$y_{ijt} = \mu_i + \eta_j + u_{ijt}, \qquad t = 1, \ldots, m_{ij}.$$

To maintain focus, assume that the u_{ijt} are i.i.d. Then, with $\boldsymbol{\alpha} = (\boldsymbol{\mu}', -\boldsymbol{\eta}')'$ as discussed before, the pooled (ordinary) least squares estimator satisfies

$$\mathrm{var}(\widehat{\boldsymbol{\alpha}}) = \sigma^2\,\boldsymbol{L}^{\dagger},$$

where $\boldsymbol{L}$ is the Laplacian associated with the adjacency matrix

$$(\boldsymbol{A})_{ij} = (\boldsymbol{A})_{ji} = \begin{cases} m_{ij} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

This illustration is interesting because, here, the μ_i cannot be estimated precisely due to limited cross-firm mobility (Andrews, Gill, Schank and Upward, 2008). It therefore makes sense to focus on the η_j, that is, on the firm effects. Profiling-out $\boldsymbol{\mu}$ and letting

$$\check{\boldsymbol{\eta}} := \widehat{\boldsymbol{\eta}} - \overline{\widehat{\boldsymbol{\eta}}}, \qquad \overline{\widehat{\boldsymbol{\eta}}} := \frac{1}{n_2} \sum_{i \in V_2} \widehat{\eta}_i,$$

where $n_2 := |V_2|$ is the number of firms, application of Example 12 gives

$$\operatorname{var}(\check{\boldsymbol{\eta}}) = \check{\boldsymbol{L}}^{\dagger},$$

where $\check{\boldsymbol{L}}$ is the $n_2 \times n_2$ Laplacian matrix associated with the weighted $n_2 \times n_2$ adjacency matrix

$$(\check{\boldsymbol{A}})_{jk} := \begin{cases} \sigma^{-2} \displaystyle\sum_{i \in [j] \cap [k]} \frac{m_{ij}\, m_{ik}}{d_i} & \text{for } j \neq k, \\ 0 & \text{for } j = k, \end{cases}$$

where $d_i = \sum_{j \in V} m_{ij}$ is the degree of $i \in V_1$, that is, the total number of observations for that worker, and $[j] \cap [k]$ is the set of all workers for which wages are observed both in firm j and in firm k. In this example the vertex set of of the weighted graph $\check{\mathcal{G}}$ are the firms. Two firms are connected by an edge if there is at least one worker who has worked in both firms. The weight $(\check{\boldsymbol{A}})_{jk}$ of the edge is larger the more workers there are connecting firms j and k, and the longer these workers have worked in both firms. Figure 6 provides an illustration of a simple bipartite graph (with all $m_{ij} = 1$) for workers (red vertices) and firms (yellow vertices), given in the left plot, and the induced weighted graph featuring only firms, given in the right plot. The thickness of the edge between (j, k) in the plot of $\check{\mathcal{G}}$ reflects the magnitude of the weight $(\check{\boldsymbol{A}})_{jk}$. □

7 Conclusions and outlook

The model we have discussed has the feature that each observed outcome depends on exactly two fixed-effect parameters, and we accordingly consider the graph $\mathcal{G}$ where each parameter is a vertex and observations are edges connecting those vertices. Examples are the Bradley and Terry (1952) model for paired comparisons, wage regressions with worker and firm fixed effects (e.g. Abowd, Kramarz and Margolis, 1999), gravity equations with exporter and importer fixed effects (e.g. Harrigan, 1996), and panel data models with individual specific fixed effects and time dummies. In applications of such models it is often the case that not all possible pairings of parameters are actually observed in the data, implying that the underlying $\mathcal{G}$ is not a complete graph or a complete bipartite graph, but has a more complicated connectivity structure.

We have derived bounds on the variance of the fixed-effect estimators for such network data applications. The bounds highlight the role of both global and local measures of network connectivity on the precision of statistical inference.

The local-connectivity measures that are relevant for our first-order variance bounds on the estimator $\widehat{\alpha}_i$ of the fixed effect α_i are the degree d_i, that is, the number of observations that depend on the parameter α_i, and the harmonic mean h_i of the degrees of the direct neighbors of vertex i. For the second-order bound (given in the appendix) the degrees of the path-two neighbors are also important as local measures of connectivity. These are very natural descriptors of the local connectivity of the vertex i.

For most of our variance bounds the global connectivity of the graph is described by the second-smallest eigenvalue of the normalized graph Laplacian matrix, which is well studied in the graph theory literature (e.g. Chung, 1997), and is closely related to other conventional connectivity measures like the Cheeger constant. We also discuss cases where our bounds based on those global connectivity measures are crude, and we derive more precise variance bounds for situations where the graph consists of well-connected clusters that are only connected by relatively few observations with each other.

Our variance bounds provide new insight into the potential for accurate statistical inference from network data that highlight the structure of the network. This can aid when deciding on sampling design or when performing sample selection. The bounds also readily yield conditions for consistent estimation and asymptotically valid inference under non-normality.

We have focused on linear models in this paper. In ongoing work we are extending our analysis to nonlinear models, such as the original Bradley and Terry (1952) model. In that case, again, the variance of the estimator takes the form of (the inverse of) a weighted Laplacian. A complication, however, is the presence of bias in the estimator coming from the nonlinearity. Like the variance, the magnitude of the bias is driven by the structure of the network, and so requires careful analysis. For example, it is not guaranteed that, even in regular problems, the bias is small relative to the standard deviation.

A restriction of our model is that each observation involves only two model parameters, which enter complementarily (that is, the off-diagonal Hessian elements have the opposite sign from the diagonal Hessian elements, implying that the Hessian of the log-likelihood can be interpreted as a graph Laplacian). Focusing on such a model allows us to connect very closely with the graph-theory literature, in particular with the results on global-connectivity measures for graphs. Models where more than two fixed-effect parameters determine one observation would lead to hypergraphs. Extrapolating our results, one would again expect that the precision of statistical inference in such models is governed by the local and global connectivity of the underlying hypergraph, but formalizing this relation is left for future research.

Appendix 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 $\operatorname{var}(\widehat{\alpha}_i)$ on the connectedness of a neighbor j of i.

Example 6 and 7 (cont'd). Both for the Star and for the Wheel graph example above one finds

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

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 $\underline{d}_{i,1} = 1$. $\square$

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

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

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 A.1 is that

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

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 (3.4) 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

$$\sum_{(j,k,\ell)\in W_i} a_{ijk\ell} = \sum_{k\in[i]_2} \sum_{j\in[i]\cap[k]} \sum_{\ell\in[i]\cap[k]} a_{ijk\ell}. \tag{A.2}$$

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 (A.2) 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

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

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 3 (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 A.1 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 (A.1), 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 2 (cont'd) (Matched employer-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 A.1 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 $j \in V_1$.[3] Then, for $i \in V_2$ the bounds in Theorem A.1 become

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

[3]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.

where

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

where for the last inequality we used the definition of $h_{i;2}$ and $|[i]_2| \leq d_i \max_{j\in[i]_2} d_{ij}$. For example, if any two firms are connected by at most two workers, then we have $d_{ij} \leq 1$ and 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 (3.4). □

It is also possible that Theorem A.1 cannot be used to obtain a refinement of the variance as in (A.1) but that it can justify the first-order rate in (3.4) for cases where this first-order asymptotic variance of $\widehat{\alpha}_i$ does not follow from Theorem 3. The following example illustrates this.

Example 4 (cont'd). 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 A.1 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 $N \to \infty$. □

References

Aaronson, D. L., L. Barrow, and W. Sander (2007). Teacher and student achievement in the Chicago public high schools. *Journal of Labor Economics 25*, 95–135.

Abowd, J., F. Kramarz, P. Lengermann, and S. Perez-Duarte (2004). Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States. Mimeo.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*, 251–333.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic risk and stability in financial networks. *American Economic Review 105*, 564–608.

Anderson, J. E. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review 93*, 170–192.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias. *Journal of the Royal Statistical Society, Series A 171*, 673–697.

Barth, E. and H. Dale-Olsen (2003). Assortative matching in the labour market? Stylised facts about workers and plants. Mimeo.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*, 242–262.

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

Bramboullé, Y., R. Kranton, and M. D'Amours (2014). Strategic interaction and networks. *American Economic Review 104*, 898–930.

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.

Erdős, P. and A. Rényi (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5*, 17–61.

Fafchamps, M. and F. Gubert (2007). Risk sharing and network formation. *American Economic Review 97*, 75–79.

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.

Goux, D. and E. Maurin (1999). Persistence of interindustry wage differentials: A re-examination using matched worker-firm data. *Journal of Labor Economics 17*, 492–533.

Graham, B. S. (2015). An econometric model of link formation with degree heterogeneity. Mimeo.

Harrigan, J. (1996). Openness to trade in manufactures in the OECD. *Journal of International Economics 40*, 23–39.

Hoffman, C., M. Kahle, and E. Paquette (2013). Spectral gaps of random graphs and applications to random topology. Mimeo.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review 102*, 1857–1897.

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.

Kolokolnikov, T., B. Osting, and J. von Brecht (2014). Algebraic connectivity of Erdős-Rényi graphs near the connectivity threshold. Mimeo.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica 73*, 417–458.

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.

Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift 29*, 436–460.

SUPPLEMENTARY MATERIAL FOR
'FIXED-EFFECT REGRESSIONS ON NETWORK DATA'

S.1 Proofs

PROOF OF LEMMA 1 (EXISTENCE)

The estimator is defined by the constraint minimization problem in (2.3). 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 (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 $\boldsymbol{B}$, $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\,\boldsymbol{I}_n)$, and so

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

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

PROOF OF COROLLARY 1 (INFERENCE)

The result 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

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

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 5 (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

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

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,

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

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

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

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 5 follows on setting $\boldsymbol{v} = \boldsymbol{e}_i - \boldsymbol{e}_j$ for $i \neq j$. This concludes the proof. □

PROOF OF THEOREMS 3 AND 6 (FIRST-ORDER BOUNDS)

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

$$0 \leq \left[\text{var}(\widehat{\boldsymbol{\alpha}}) - \sigma^2 \left(\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1} - \frac{\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}}{n} - \frac{\boldsymbol{D}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'}{n}\right)\right] \leq \frac{\sigma^2}{\lambda_2}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}. \tag{S.3}$$

Theorems 3 and 6 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

$$\boldsymbol{L}^\dagger = \boldsymbol{D}^{-1} + \boldsymbol{L}^\dagger\boldsymbol{A}\boldsymbol{D}^{-1} - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{D}^{-1}. \tag{S.4}$$

The Laplacian is symmetric, and so transposition gives

$$\boldsymbol{L}^\dagger = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{L}^\dagger - n^{-1}\,\boldsymbol{D}^{-1}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'. \tag{S.5}$$

Replacing $\boldsymbol{L}^\dagger$ on the right-hand side of (S.4) by the expression for $\boldsymbol{L}^\dagger$ given by (S.5) yields

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

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.3).

To show Theorems 3 and 6 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.3) 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 (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 LEMMA 2 (VARIANCE OF COMPONENT ESTIMATORS)

Additional notation. Without loss of generality we relabel the elements of V such that

$$V_1 = \{1, \dots, n_1\}, \quad V_2 = \{n_1 + 1, \dots, n_2\}, \quad \dots \quad V_q = \{n - n_q + 1, \dots, n\}.$$

We decompose $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \dots, \boldsymbol{\beta}_q')'$ and $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1', \widehat{\boldsymbol{\beta}}_2', \dots, \widehat{\boldsymbol{\beta}}_q')'$, where each $\boldsymbol{\beta}_r$ and $\widehat{\boldsymbol{\beta}}_r$ are n_r column vectors. Note that the Laplacian matrix $\boldsymbol{L}_W$ is block-diagonal; moreover,

$$\boldsymbol{L}_W = \operatorname{diag}\left(\boldsymbol{L}_1, \boldsymbol{L}_2, \dots, \boldsymbol{L}_q\right),$$

where $\boldsymbol{L}_r$ is the Laplacian of the graph $\mathcal{G}_r$. We also decompose $\boldsymbol{A} = \boldsymbol{A}_W + \boldsymbol{A}_B$, where $\boldsymbol{A}_W$ and $\boldsymbol{A}_B$ are the adjacency matrix of $\mathcal{G}_W$ and $\mathcal{G}_B$, respectively, and write $\boldsymbol{D}_W = \operatorname{diag}\left(\boldsymbol{A}_W \boldsymbol{\iota}_n\right)$ and $\boldsymbol{D}_B = \operatorname{diag}\left(\boldsymbol{A}_B \boldsymbol{\iota}_n\right)$ for the corresponding degree matrices. We have $\boldsymbol{L}_W = \boldsymbol{D}_W - \boldsymbol{A}_W$ and $\boldsymbol{L}_B = \boldsymbol{D}_B - \boldsymbol{A}_B$. We also relabel the elements of E such that

$$E_B = \{1, \dots, m_B\}, \qquad E_W = \{m_B + 1, \dots, m\},$$

and correspondingly we decompose $\boldsymbol{B} = (\boldsymbol{B}_B', \boldsymbol{B}_W')'$, where $\boldsymbol{B}_B$ and $\boldsymbol{B}_W$ are $m_B \times n$ and $m_W \times n$ matrices, respectively, whose rows correspond to edges in $\mathcal{G}_B$ and $\mathcal{G}_W$, respectively. We then have $\boldsymbol{L} = \boldsymbol{B}'\boldsymbol{B} = \boldsymbol{B}_W'\boldsymbol{B}_W + \boldsymbol{B}_B'\boldsymbol{B}_B = \boldsymbol{L}_W + \boldsymbol{L}_B$.

Inverse expressions. Notice that, under the conventions from above, $\boldsymbol{P}$ is simply given by

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{\iota}_{n_1}' & 0 & \dots & 0 \\ 0 & \boldsymbol{\iota}_{n_2}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{\iota}_{n_q}' \end{pmatrix}.$$

We define the block-diagonal $n \times n$ matrix

$$\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{I}_{n_1} - n_1^{-1}\boldsymbol{\iota}_{n_1}\boldsymbol{\iota}_{n_1}' & 0 & \dots & 0 \\ 0 & \boldsymbol{I}_{n_2} - n_2^{-1}\boldsymbol{\iota}_{n_2}\boldsymbol{\iota}_{n_2}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{I}_{n_q} - n_q^{-1}\boldsymbol{\iota}_{n_q}\boldsymbol{\iota}_{n_q}' \end{pmatrix}.$$

Some useful relations are $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{P}'$, $\boldsymbol{M}^2 = \boldsymbol{M}$, $\boldsymbol{P}\boldsymbol{M} = 0$, $\boldsymbol{P}'\boldsymbol{\iota}_q = \boldsymbol{\iota}_n$, $\boldsymbol{P}\boldsymbol{\iota}_n = \boldsymbol{H}\boldsymbol{\iota}_q$, and thus also $\boldsymbol{H}^{-1}\boldsymbol{P}\boldsymbol{\iota}_n = \boldsymbol{\iota}_q$. The various pseudo-inverses that appear in the following satisfy

$$\begin{aligned} \boldsymbol{L}_W^{\dagger}\boldsymbol{L}_W &= \boldsymbol{M}, \\ \left(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2}\right)^{\dagger}\left(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2}\right) &= \boldsymbol{I}_q - n^{-1}\,\boldsymbol{H}^{1/2}\boldsymbol{\iota}_q\boldsymbol{\iota}_q'\boldsymbol{H}^{1/2}, \\ \left(\boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right)^{\dagger}\left(\boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right) &= \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{P} - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n', \\ \left(\boldsymbol{L}_W + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right)^{\dagger}\left(\boldsymbol{L}_W + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right) &= \boldsymbol{I}_n - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n', \end{aligned} \tag{S.6}$$

where on the right-hand side always appears the projector orthogonal to the null-space of the respective matrix, e.g. we have $\boldsymbol{L}_W\boldsymbol{M} = \boldsymbol{L}_W$. Using that $\left(\boldsymbol{H}^{-1/2}\boldsymbol{P}\right)^{\dagger} = \boldsymbol{P}'\boldsymbol{H}^{-1/2}$ and the definition of $\boldsymbol{L}_*^{\text{inv}}$ we find that

$$\left(\boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right)^{\dagger} = \boldsymbol{P}'\boldsymbol{H}^{-1/2}\left(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2}\right)^{\dagger}\boldsymbol{H}^{-1/2}\boldsymbol{P} = \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}.$$

Proof of Lemma 2. We derive the result for $\widehat{\boldsymbol{\beta}}$ first. By applying Theorem 1 to each $\mathcal{G}_r$ separately we obtain

$$\widehat{\boldsymbol{\beta}}_r \sim \mathcal{N}(\boldsymbol{\beta}_r,\ \sigma^2\,\boldsymbol{L}_r^{\dagger}),$$

for $r = 1, \dots, q$. Note that we do not rule out $n_r = 1$ (i.e., $\mathcal{G}_r$ may be a graph with one vertex and no edges), but in this case we simply have $\widehat{\boldsymbol{\beta}}_r = \boldsymbol{\beta}_r = \boldsymbol{L}_r = \boldsymbol{L}_r^{\dagger} = \boldsymbol{0}$, so the result for $\widehat{\boldsymbol{\beta}}_r$ holds trivially. Independence of the errors u_{ij} across observations implies independence of $\widehat{\boldsymbol{\beta}}_r$ and $\widehat{\boldsymbol{\beta}}_s$ for all $r \neq s$. We thus find $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta},\ \sigma^2\,\boldsymbol{L}_W^{\dagger})$, which is the result in the theorem.

Now turn to $\widehat{\boldsymbol{\gamma}}$. Analogous to the proof of Lemma 1 we can write the minimization problem for $\widehat{\boldsymbol{\gamma}}$ as

$$\widehat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{g}\in\mathbb{R}^q}\ (\boldsymbol{y} - \boldsymbol{B}_W\boldsymbol{\beta} - \boldsymbol{B}_B\boldsymbol{P}'\boldsymbol{g})'(\boldsymbol{y} - \boldsymbol{B}_W\boldsymbol{\beta} - \boldsymbol{B}_B\boldsymbol{P}'\boldsymbol{g}) + \lambda\left(\boldsymbol{g}'\boldsymbol{P}\boldsymbol{\iota}_n\right)^2,$$

where $\lambda > 0$ is a Lagrange multiplier. Solving the corresponding first-order condition gives

$$\begin{aligned}\widehat{\boldsymbol{\gamma}} &= \left(\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{B}_B\boldsymbol{P}' + \lambda\,\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\right)^{-1}\boldsymbol{P}\boldsymbol{B}_B'(\boldsymbol{y} - \boldsymbol{B}_W\boldsymbol{\beta}) \\ &= \left(\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{B}_B\boldsymbol{P}' + \lambda\,\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\right)^{-1}\left[\left(\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{B}_B\boldsymbol{P}' + \lambda\,\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\right)\boldsymbol{\gamma} + \boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{u}\right], \\ &= \boldsymbol{\gamma} + \left(\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{B}_B\boldsymbol{P}' + \lambda\,\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\right)^{-1}\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{u},\end{aligned}$$

where in the second step we used the model $\boldsymbol{y} = \boldsymbol{B}_W\boldsymbol{\beta} + \boldsymbol{B}_B\boldsymbol{P}'\boldsymbol{\gamma} + \boldsymbol{u}$, and we added a term proportional to λ in the square brackets, which is zero due to the normalization of $\boldsymbol{\gamma}$, which can be written as $\boldsymbol{\iota}_n'\boldsymbol{P}'\boldsymbol{\gamma} = 0$. Notice that $\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{B}_B\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{L}_B\boldsymbol{P}' = \boldsymbol{L}_*$. However, compared to the proof of Lemma 1 the difficulty is that, here, the matrices $\boldsymbol{L}_*$ and $\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'$ do not commute. To resolve this problem we rewrite the last result as

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = \boldsymbol{H}^{-1/2}\left(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2} + \lambda\,\boldsymbol{H}^{-1/2}\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\boldsymbol{H}^{-1/2}\right)^{-1}\boldsymbol{H}^{-1/2}\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{u}.$$

Now, the matrices $\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2}$ and $\boldsymbol{H}^{-1/2}\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\boldsymbol{H}^{-1/2}$ commute, because the zero eigenvalue of $\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2}$ has multiplicity one (as we assume $\mathcal{G}_B$ to be connected) with eigenvector given by $\boldsymbol{H}^{-1/2}\boldsymbol{P}\boldsymbol{\iota}_n$, namely we have $\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\boldsymbol{\iota}_n = \boldsymbol{L}_*\boldsymbol{\iota}_q = 0$. Here, we used $\boldsymbol{H}^{-1}\boldsymbol{P}\boldsymbol{\iota}_n = \boldsymbol{\iota}_q$, which follows from the definition of $\boldsymbol{H}$ and $\boldsymbol{P}$. We therefore have

$$\begin{aligned}\left(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2} + \lambda\,\boldsymbol{H}^{-1/2}\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\boldsymbol{H}^{-1/2}\right)^{-1} &= \left(\boldsymbol{H}^{-1/2}\boldsymbol{L}_*\boldsymbol{H}^{-1/2}\right)^{\dagger} + \frac{1}{\lambda}\left(\boldsymbol{H}^{-1/2}\boldsymbol{P}\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\boldsymbol{P}'\boldsymbol{H}^{-1/2}\right)^{\dagger} \\ &= \boldsymbol{H}^{1/2}\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{H}^{1/2} + \frac{1}{\lambda\,n^2}\boldsymbol{H}^{1/2}\boldsymbol{\iota}_q\boldsymbol{\iota}_q'\boldsymbol{H}^{1/2},\end{aligned}$$

and the last term does not contribute to $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ because we have $\boldsymbol{\iota}_q'\boldsymbol{P}\boldsymbol{B}_B' = \boldsymbol{\iota}_n'\boldsymbol{B}_B' = 0$. We therefore have, independent from the choice of λ, that

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = \boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{u}.$$

Using $\mathbb{E}(\boldsymbol{u}\boldsymbol{u}') = \sigma^2\boldsymbol{I}_m$ we thus find

$$\mathrm{var}(\widehat{\boldsymbol{\gamma}}) = \sigma^2\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\boldsymbol{B}_B'\boldsymbol{B}_B\boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}} = \sigma^2\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{L}_*\boldsymbol{L}_*^{\mathrm{inv}} = \sigma^2\boldsymbol{L}_*^{\mathrm{inv}}.$$

Because $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ is a linear combination of the jointly normal errors it is also normally distributed, so we have $\widehat{\boldsymbol{\gamma}} \sim \mathcal{N}(\boldsymbol{\gamma},\ \sigma^2\,\boldsymbol{L}_*^{\mathrm{inv}})$. This concludes the proof. □

PROOF OF THEOREM 7 (GRAPH PARTITIONING)

Throughout the proof we maintain the same notational conventions as for the proof of Lemma 2. Recall that the variance of $\widehat{\boldsymbol{\alpha}}$ is $\sigma^2 \boldsymbol{L}^\dagger$. The variance of the infeasible estimator based on (5.4) equals $\sigma^2(\boldsymbol{L}_W^\dagger + \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P})$. We show below that

$$- \boldsymbol{Z}_{\text{low}} - (\boldsymbol{Q} + \boldsymbol{Q}') \leq \boldsymbol{L}^\dagger - \left(\boldsymbol{L}_W^\dagger + \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\right) \leq \boldsymbol{Z}_{\text{up}} - (\boldsymbol{Q} + \boldsymbol{Q}')\,, \tag{S.7}$$

for matrices

$$\boldsymbol{Z}_{\text{low}} := \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{L}_W^\dagger, \qquad \boldsymbol{Z}_{\text{up}} := \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P},$$

and $\boldsymbol{Q} := \boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}$. We also establish that

$$\boldsymbol{Z}_{\text{low}} \leq \kappa \boldsymbol{L}_W^\dagger, \qquad \boldsymbol{Z}_{\text{up}} \leq \kappa \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P} = \kappa\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}, \tag{S.8}$$

and that

$$|\boldsymbol{v}'\boldsymbol{Q}\boldsymbol{v}| \leq \kappa^{1/2}\left(\boldsymbol{v}'\boldsymbol{L}_W^\dagger\boldsymbol{v}\right)^{1/2}\left(\boldsymbol{v}'\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{v}\right)^{1/2}, \tag{S.9}$$

for any $\boldsymbol{v} \in \mathbb{R}^n$. Combining these results yields that, for any $\boldsymbol{v} \in \mathbb{R}^n$,

$$\begin{aligned}
&- \kappa\, \boldsymbol{v}'\boldsymbol{L}_W^\dagger\boldsymbol{v} - 2\kappa^{1/2}\left[\left(\boldsymbol{v}'\boldsymbol{L}_W^\dagger\boldsymbol{v}\right)\left(\boldsymbol{v}'\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\,\boldsymbol{v}\right)\right]^{1/2} \\
&\qquad \leq \boldsymbol{v}'\left(\boldsymbol{L}^\dagger - \boldsymbol{L}_W^\dagger - \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\right)\boldsymbol{v} \leq \\
&\qquad\qquad \kappa\, \boldsymbol{v}'\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\,\boldsymbol{v} + 2\kappa^{1/2}\left[\left(\boldsymbol{v}'\boldsymbol{L}_W^\dagger\boldsymbol{v}\right)\left(\boldsymbol{v}'\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\,\boldsymbol{v}\right)\right]^{1/2}.
\end{aligned}$$

By Lemma 2 this is the result of Theorem 7. It remains only to show (S.7), (S.8), and (S.9), which we do, in turn, next. □

Proof of (S.7). Start with the upper bound. Because $\boldsymbol{L}_B \geq 0$ and $\boldsymbol{L}_W \geq 0$, it holds that

$$\begin{aligned}
0 \leq\; &\left(\boldsymbol{L}^\dagger - \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\right)\boldsymbol{L}_B\left(\boldsymbol{L}^\dagger - \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\right) \\
&+ \left(\boldsymbol{L}^\dagger - \boldsymbol{L}_W^\dagger + \boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger\right)\boldsymbol{L}_W\left(\boldsymbol{L}^\dagger - \boldsymbol{L}_W^\dagger + \boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\right).
\end{aligned}$$

Expanding those terms, and using that $\boldsymbol{L}_W + \boldsymbol{L}_B = \boldsymbol{L}$, and $\boldsymbol{L}_W^\dagger \boldsymbol{L}_W = \boldsymbol{M}$, and $\boldsymbol{L}_W^\dagger \boldsymbol{L}_W \boldsymbol{L}_W^\dagger = \boldsymbol{L}_W^\dagger$, we obtain

$$\begin{aligned} 0 \le \boldsymbol{L}^\dagger \boldsymbol{L} \boldsymbol{L}^\dagger &+ \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} - \boldsymbol{L}^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} - \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}^\dagger \\ &- \boldsymbol{M} \boldsymbol{L}^\dagger - \boldsymbol{L}^\dagger \boldsymbol{M} + \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{M} \boldsymbol{L}^\dagger + \boldsymbol{L}^\dagger \boldsymbol{M} \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \\ &+ \boldsymbol{L}_W^\dagger - \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger - \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} + \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P}. \end{aligned}$$

Using that $\boldsymbol{L}^\dagger \boldsymbol{L} \boldsymbol{L}^\dagger = \boldsymbol{L}^\dagger$, and $\boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} = \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P}$, and

$$\begin{aligned} -\boldsymbol{L}^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} + \boldsymbol{L}^\dagger \boldsymbol{M} \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} &= -\boldsymbol{L}^\dagger \left(\boldsymbol{I}_n - \boldsymbol{M} \right) \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \\ &= -\boldsymbol{L}^\dagger \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \\ &= -\boldsymbol{L}^\dagger \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P}, \end{aligned}$$

and also the transpose of the last result, we obtain

$$\begin{aligned} 0 \le \boldsymbol{L}^\dagger &+ \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} - \left(\boldsymbol{M} + \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P} \right) \boldsymbol{L}^\dagger - \boldsymbol{L}^\dagger \left(\boldsymbol{M} + \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P} \right) \\ &+ \boldsymbol{L}_W^\dagger - \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger - \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} + \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P}. \end{aligned}$$

Because $\boldsymbol{M} + \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P} = \boldsymbol{I}_n$ we thus find

$$\boldsymbol{L}^\dagger \le \boldsymbol{L}_W^\dagger + \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} - \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger - \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} + \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P},$$

which is the upper bound given in the lemma.

Now turn to the lower bound. Introduce

$$\boldsymbol{\Delta} := \boldsymbol{M} \boldsymbol{L}_B \boldsymbol{M} + \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P} \boldsymbol{L}_B \boldsymbol{M} + \boldsymbol{M} \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{P}.$$

We then have

$$\boldsymbol{L} = \boldsymbol{L}_W + \boldsymbol{L}_B = \boldsymbol{L}_W + \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{L}_* \boldsymbol{H}^{-1} \boldsymbol{P} + \boldsymbol{\Delta}.$$

Plugging this in the equality $\boldsymbol{L} \boldsymbol{L}^\dagger = \boldsymbol{I}_n - n^{-1} \boldsymbol{\iota}_n \boldsymbol{\iota}_n'$ we obtain

$$\left(\boldsymbol{L}_W + \boldsymbol{P}' \boldsymbol{H}^{-1} \boldsymbol{L}_* \boldsymbol{H}^{-1} \boldsymbol{P} + \boldsymbol{\Delta} \right) \boldsymbol{L}^\dagger = \boldsymbol{I}_n - n^{-1} \, \boldsymbol{\iota}_n \boldsymbol{\iota}_n'.$$

Bringing $\boldsymbol{\Delta}\boldsymbol{L}^{\dagger}$ to the right-hand side, multiplying with $(\boldsymbol{L}_W + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P})^{\dagger}$ from the left, and using the last equality in (S.6), we obtain

$$\left(\boldsymbol{I}_n - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right)\boldsymbol{L}^{\dagger} = \left(\boldsymbol{L}_W + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right)^{\dagger}\left(\boldsymbol{I}_n - \boldsymbol{\Delta}\boldsymbol{L}^{\dagger} - n^{-1}\,\boldsymbol{\iota}_n\boldsymbol{\iota}_n'\right). \tag{S.10}$$

The matrices $\boldsymbol{L}_W$ and $\boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}$ commute, and we therefore have

$$\left(\boldsymbol{L}_W + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right)^{\dagger} = \boldsymbol{L}_W^{\dagger} + \left(\boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P}\right)^{\dagger} = \boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}.$$

Using this as well as $\boldsymbol{L}^{\dagger}\boldsymbol{\iota}_n = 0$ and $(\boldsymbol{L}_W + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{L}_*\boldsymbol{H}^{-1}\boldsymbol{P})^{\dagger}\,\boldsymbol{\iota}_n = 0$ we find that the equation in (S.10) becomes

$$\boldsymbol{L}^{\dagger} = \boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P} - \left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right)\boldsymbol{\Delta}\boldsymbol{L}^{\dagger}. \tag{S.11}$$

Taking the transpose of this last equation gives

$$\boldsymbol{L}^{\dagger} = \boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P} - \boldsymbol{L}^{\dagger}\boldsymbol{\Delta}\left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right).$$

Replacing $\boldsymbol{L}^{\dagger}$ on the right-hand side of the last equation by the expression for $\boldsymbol{L}^{\dagger}$ in (S.11) we get

$$\boldsymbol{L}^{\dagger} = \boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P} - \left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right)\left(\boldsymbol{\Delta} - \boldsymbol{\Delta}\boldsymbol{L}^{\dagger}\boldsymbol{\Delta}\right)\left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right).$$

Using the definition of $\boldsymbol{\Delta}$ we obtain

$$\boldsymbol{\Delta}\left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right) = \boldsymbol{M}\boldsymbol{L}_B\left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right) + \boldsymbol{P}'\boldsymbol{H}^{-1}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^{\dagger},$$

and the last result on $\boldsymbol{L}^{\dagger}$ can therefore be rewritten as

$$\begin{aligned}\boldsymbol{L}^{\dagger} - \Big(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P} - \boldsymbol{L}_W^{\dagger}\boldsymbol{L}_B\boldsymbol{L}_W^{\dagger} - \boldsymbol{L}_W^{\dagger}\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P} - \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^{\dagger}\Big) \\ = \left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right)\boldsymbol{\Delta}\boldsymbol{L}^{\dagger}\boldsymbol{\Delta}\left(\boldsymbol{L}_W^{\dagger} + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\right).\end{aligned}$$

Because $\boldsymbol{L}^{\dagger} \geq 0$ the last expression is positive semi-definite, which gives the lower bound on $\boldsymbol{L}^{\dagger}$ in the lemma. This concludes the proof. □

Proof of (S.8). Let

$$\boldsymbol{\Lambda} := \mathrm{diag}(\lambda_2^r : i \in \boldsymbol{V}, \text{ with } r \text{ such that } i \in V_r),$$

where we set $\lambda_2^r = 0$ if $n_r = 1$. Then $\boldsymbol{L}_W^\dagger \leq (\boldsymbol{\Lambda}\boldsymbol{D}_W)^\dagger$.[1] Also define the symmetrically normalized Laplacian of $\mathcal{G}_B$[2]

$$\boldsymbol{S}_B := \left(\boldsymbol{D}_B^\dagger\right)^{1/2} \boldsymbol{L}_B \left(\boldsymbol{D}_B^\dagger\right)^{1/2}.$$

From Chung (1997, Lemma 1.7) we know $\lambda_n(\boldsymbol{S}_B) \leq 2$, which can also be written as $\boldsymbol{S}_B \leq 2\boldsymbol{I}_n$. We have $\boldsymbol{L}_B = \boldsymbol{D}_B^{1/2}\boldsymbol{S}_B\boldsymbol{D}_B^{1/2}$, and thus find $\boldsymbol{L}_B \leq 2\boldsymbol{D}_B$.

The diagonal matrix $\boldsymbol{D}_B^{1/2}(\boldsymbol{\Lambda}\boldsymbol{D}_W)^\dagger \boldsymbol{D}_B^{1/2}$ has ith diagonal element equal to $d_i^B/(\lambda_2^r\, d_i^W)$ for $n_r > 1$, $i \in V_r$, and equal to zero otherwise. From the definition of κ in the main text we thus find

$$\boldsymbol{D}_B^{1/2}(\boldsymbol{\Lambda}\boldsymbol{D}_W)^\dagger \boldsymbol{D}_B^{1/2} \leq \frac{\kappa}{2}\boldsymbol{I}_n,$$

and therefore

$$\boldsymbol{D}_B^{1/2}\boldsymbol{L}_W^\dagger\boldsymbol{D}_B^{1/2} \leq \boldsymbol{D}_B^{1/2}(\boldsymbol{\Lambda}\boldsymbol{D}_W)^\dagger \boldsymbol{D}_B^{1/2} \leq \frac{\kappa}{2}\boldsymbol{I}_n. \tag{S.12}$$

The matrix $\boldsymbol{D}_B^{1/2}\boldsymbol{L}_W^\dagger\boldsymbol{D}_B^{1/2}$ is similar to $\left(\boldsymbol{L}_W^\dagger\right)^{1/2}\boldsymbol{D}_B\left(\boldsymbol{L}_W^\dagger\right)^{1/2}$, and so they share the same eigenvalues.[3] We therefore have that

$$\left(\boldsymbol{L}_W^\dagger\right)^{1/2}\boldsymbol{D}_B\left(\boldsymbol{L}_W^\dagger\right)^{1/2} \leq \frac{\kappa}{2}\boldsymbol{I}_n \tag{S.13}$$

holds.

[1] The diagonal matrix $\boldsymbol{\Lambda}\boldsymbol{D}_W$ has non-negative elements but may be non-invertible as, for $n_r = 1$, we have $\lambda_2^r\, d_i^W = 0$, with $i \in V_r$. We therefore write $(\boldsymbol{\Lambda}\boldsymbol{D}_W)^\dagger$ instead of just $(\boldsymbol{\Lambda}\boldsymbol{D}_W)^{-1}$.

[2] Again we write $\boldsymbol{D}_B^\dagger$ because we may have $d_i^B = 0$ for some $i \in V$.

[3] Two square matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are similar if there exists an invertible matrix $\boldsymbol{M}_3$ such that $\boldsymbol{M}_1 = \boldsymbol{M}_3^{-1}\boldsymbol{M}_2\boldsymbol{M}_3$. Two similar matrices have the same eigenvalues.

Using the inequalities $\boldsymbol{S}_B \leq 2\boldsymbol{I}_n$ and $\boldsymbol{L}_B \leq 2\boldsymbol{D}_B$ along with (S.12) and (S.13) we obtain

$$\begin{aligned} \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{L}_W^\dagger &\leq 2\boldsymbol{L}_W^\dagger \boldsymbol{D}_B \boldsymbol{L}_W^\dagger \\ &= 2\left(\boldsymbol{L}_W^\dagger\right)^{1/2}\left(\boldsymbol{L}_W^\dagger\right)^{1/2}\boldsymbol{D}_B\left(\boldsymbol{L}_W^\dagger\right)^{1/2}\left(\boldsymbol{L}_W^\dagger\right)^{1/2} \\ &\leq \kappa\left(\boldsymbol{L}_W^\dagger\right)^{1/2}\left(\boldsymbol{L}_W^\dagger\right)^{1/2} \\ &\leq \kappa \boldsymbol{L}_W^\dagger, \end{aligned} \tag{S.14}$$

and

$$\begin{aligned} \boldsymbol{L}_B \boldsymbol{L}_W^\dagger \boldsymbol{L}_B &= \boldsymbol{D}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{D}_B^{1/2}\boldsymbol{L}_W^\dagger \boldsymbol{D}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{D}_B^{1/2} \\ &\leq \frac{\kappa}{2}\boldsymbol{D}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{S}_B\boldsymbol{S}_B^{1/2}\boldsymbol{D}_B^{1/2} \\ &\leq \kappa \boldsymbol{D}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{S}_B^{1/2}\boldsymbol{D}_B^{1/2} \\ &= \kappa \boldsymbol{L}_B. \end{aligned} \tag{S.15}$$

These yield the inequalities stated in (S.8). This concludes the proof. □

Proof of (S.9). Recall that

$$\boldsymbol{Q} = \boldsymbol{L}_W^\dagger \boldsymbol{L}_B \boldsymbol{P}' \boldsymbol{L}_*^{\text{inv}} \boldsymbol{P}.$$

Applying the Cauchy-Schwarz inequality $(\boldsymbol{x}'\boldsymbol{y})^2 \leq (\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{y}'\boldsymbol{y})$ with $\boldsymbol{x} = \boldsymbol{L}_B^{1/2}\boldsymbol{L}_W^\dagger \boldsymbol{v}$ and $\boldsymbol{y} = \boldsymbol{L}_B^{1/2}\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{v}$, and using (S.8) we find

$$|\boldsymbol{v}'\boldsymbol{Q}\boldsymbol{v}|^2 \leq \kappa\left(\boldsymbol{v}'\boldsymbol{L}_W^\dagger \boldsymbol{v}\right)\left(\boldsymbol{v}'(\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P})\boldsymbol{v}\right),$$

which gives (S.9). This concludes the proof. □

PROOF OF THEOREM A.1 (SECOND-ORDER BOUND)

Proof of Theorem A.1. 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

$$0 \leq \left\{ \mathrm{var}(\widehat{\alpha}_i) - \frac{\sigma^2 \left[1 - \frac{2}{n} \left(1 + \boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} \right) \right]}{d_i - \boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i}} \right\} \leq \frac{\sigma^2 \, \boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} (\boldsymbol{A}_{\circ\#}) \boldsymbol{D}_\#^{-1} (\boldsymbol{A}_{\circ\#})' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i}}{\lambda_2 \left(d_i - \boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} \right)^2}, \tag{S.16}$$

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

$$\begin{aligned} \boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} &= \boldsymbol{\iota}_{d_i}' \boldsymbol{D}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} + \boldsymbol{\iota}_{d_i}' \sum_{q=1}^{\infty} \left(\boldsymbol{D}_{[i]}^{-1} \boldsymbol{A}_{[i]} \right)^q \boldsymbol{D}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} \\ &\geq \boldsymbol{\iota}_{d_i}' \boldsymbol{D}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} = \sum_{j \in [i]} d_j^{-1}, \end{aligned} \tag{S.17}$$

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 $\underline{\boldsymbol{D}}_{[i]} = \mathrm{diag}(\underline{d}_{j,i} : j \in [i])$. We have

$$\boldsymbol{L}_{[i]} - \underline{\boldsymbol{D}}_{[i]} = \mathrm{diag}(\boldsymbol{A}_{[i]} \boldsymbol{\iota}_{d_i}) - \boldsymbol{A}_{[i]} \geq 0, \tag{S.18}$$

because $\mathrm{diag}(\boldsymbol{A}_{[i]} \boldsymbol{\iota}_{d_i}) - \boldsymbol{A}_{[i]}$ 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

$$\boldsymbol{\iota}_{d_i}' \boldsymbol{L}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} \leq \boldsymbol{\iota}_{d_i}' \underline{\boldsymbol{D}}_{[i]}^{-1} \boldsymbol{\iota}_{d_i} = \sum_{j \in [i]} \underline{d}_{j,i}^{-1}. \tag{S.19}$$

Combining (S.16), (S.17) and (S.19) 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.18) we have $\widetilde{\boldsymbol{L}} \leq \boldsymbol{L}$, and therefore $\widetilde{\boldsymbol{L}}^\dagger \geq \boldsymbol{L}^\dagger$. The result (S.16) 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.16). 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 $V_\circ = \{1, 2, \ldots, n_\circ\}$. Let $\widehat{\boldsymbol{\alpha}}_\circ = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{n_\circ})'$, $\boldsymbol{L}_\circ$ 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

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

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 $\mathrm{var}(\widehat{\boldsymbol{\alpha}}_\circ) = \sigma^2(\boldsymbol{L}^{\dagger})_{\circ}$ by also using $0 \leq (\boldsymbol{L}^{\dagger})_{\#} \leq \lambda_2^{-1}\boldsymbol{D}_{\#}^{-1}$. This concludes the proof. □

S.2 Component estimators from graph partitioning

Here we strenghten the result of Theorem 7 by showing that the estimator $\widehat{\boldsymbol{\alpha}}$ is close to the (infeasible) estimator $\widehat{\boldsymbol{\beta}} + \boldsymbol{P}'\widehat{\boldsymbol{\gamma}}$ when κ is small. We also provide a corresponding result for the feasible version $\widehat{\boldsymbol{\beta}} + \boldsymbol{P}'\widetilde{\boldsymbol{\gamma}}$, where

$$\widetilde{\boldsymbol{\gamma}} := \arg\min_{\boldsymbol{g}\in\mathbb{R}^q} \sum_{(i,j)\in E_B} \left(y_{ij} - (\widehat{\beta}_i + g_{r(i)}) + (\widehat{\beta}_j + g_{r(j)})\right)^2 \qquad \text{s.t.} \;\; \sum_{r=1}^{q} n_r\, g_r = 0.$$

Our focus in the main text is on the infeasible estimator. This is so because we use it as a devise to analyze the variance of $\widehat{\boldsymbol{\alpha}}$, and $\widehat{\boldsymbol{\gamma}}$ is independent of $\widehat{\boldsymbol{\beta}}$ while its feasible version is clearly not. If an alternative estimator to $\widehat{\boldsymbol{\alpha}}$ is desired, $\widehat{\boldsymbol{\beta}} + \boldsymbol{P}'\widetilde{\boldsymbol{\gamma}}$ will obviously be of interest. Note, however, that $\mathrm{var}(\widehat{\boldsymbol{\alpha}}_i) \leq \mathrm{var}(\widehat{\boldsymbol{\beta}} + \boldsymbol{P}'\widetilde{\boldsymbol{\gamma}})$ by the Gauss-Markov theorem (this, in fact, yields the upper bound given in (S.7)).

The following theorem is the main result of this section.

Theorem 1. Let $\mathcal{G}$ and $\mathcal{G}_1, \ldots, \mathcal{G}_q$ be connected. For $i \in V$ define $r_i, R_i \in \mathbb{R}$ by

$$\widehat{\alpha}_i = \widehat{\beta}_i + \widetilde{\gamma}_{r(i)} + r_i, \qquad\qquad \widehat{\alpha}_i = \widehat{\beta}_i + \widehat{\gamma}_{r(i)} + r_i + R_i.$$

We then have

$$\mathbb{E}(r_i^2) \;\leq\; \kappa \left[\mathrm{var}(\widehat{\beta}_i) + \mathrm{var}(\widehat{\gamma}_{r(i)})\right], \qquad\qquad \mathbb{E}(R_i^2) \;\leq\; \kappa\, \mathrm{var}(\widehat{\gamma}_{r(i)}).$$

The theorem shows that, if κ is small, then the differences between $\widehat{\alpha}_i$ and $\widehat{\beta}_i + \widetilde{\gamma}_{r(i)}$, and between $\widehat{\alpha}_i$ and $\widehat{\beta}_i + \widehat{\gamma}_{r(i)}$, are both small compared to the stochastic variability of $\widehat{\beta}_i$ and $\widehat{\gamma}_{r(i)}$ themselves. Thus, the result of Theorem 7 generalizes from the variances to the estimators themselves.

The result (and its proof) also immediately extends to a setting as in Theorem 4, where the errors u_{ij} can be non-normal, heteroscedastic, or correlated. One only needs to replace $\mathrm{var}(\widehat{\beta}_i)$ by $\overline{\sigma}^2(\boldsymbol{L}_W^\dagger)_{ii}$ and $\mathrm{var}(\widehat{\gamma}_{r(i)})$ by $\overline{\sigma}^2(\boldsymbol{L}^{\mathrm{inv}})_{rr}$, where $\overline{\sigma}^2$ is a bound on the largest eigenvalue of $\mathbb{E}\left(\boldsymbol{u}\boldsymbol{u}'\right)$.

Proof of Theorem 1. In vector notation the estimator decompositions reads

$$\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\beta}} + \boldsymbol{P}'\widetilde{\boldsymbol{\gamma}} + \boldsymbol{r}, \qquad\qquad \widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\beta}} + \boldsymbol{P}'\widehat{\boldsymbol{\gamma}} + \boldsymbol{r} + \boldsymbol{R}.$$

Analogous to the proof of Lemma 1 and Lemma 2 above we can use the first-order conditions of their respective minimization problem to obtain explicit formulas for $\widehat{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\gamma}}$. We thus find

$$\boldsymbol{r} = \boldsymbol{C}_1\boldsymbol{Y}, \qquad \boldsymbol{C}_1 = \left(\boldsymbol{L}^\dagger\boldsymbol{B}' - \boldsymbol{L}_W^\dagger\boldsymbol{B}_W' - \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\boldsymbol{B}_B' + \boldsymbol{P}'\boldsymbol{L}_*^{\mathrm{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger\boldsymbol{B}_W'\right),$$

and

$$\boldsymbol{r}+\boldsymbol{R}=\boldsymbol{C}_2\boldsymbol{Y}+\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{\beta}, \qquad \boldsymbol{C}_2=\left(\boldsymbol{L}^\dagger\boldsymbol{B}'-\boldsymbol{L}_W^\dagger\boldsymbol{B}_W'-\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{B}_B'\right).$$

It is easy to verify that $\boldsymbol{C}_1\boldsymbol{B}=0$ and $\boldsymbol{C}_2\boldsymbol{B}\boldsymbol{\alpha}+\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{\beta}=0$, and therefore

$$\boldsymbol{r}=\boldsymbol{C}_1\boldsymbol{U}, \qquad \boldsymbol{r}+\boldsymbol{R}=\boldsymbol{C}_2\boldsymbol{U}.$$

Using this we find

$$\begin{aligned}\sigma^{-2}\mathbb{E}\left(\boldsymbol{r}\boldsymbol{r}'\right)&=\boldsymbol{C}_1\boldsymbol{C}_1'\\&=-\boldsymbol{L}^\dagger+\boldsymbol{L}_W^\dagger+\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}+\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\\&\qquad-\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger-\boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\\&\leq\boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{L}_W^\dagger+\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\\&\leq\kappa\left(\boldsymbol{L}_W^\dagger+\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\right)=\kappa\left[\text{var}\big(\widehat{\boldsymbol{\beta}}\big)+\text{var}\left(\boldsymbol{P}'\widehat{\boldsymbol{\gamma}}\right)\right],\end{aligned}$$

where in the second to last inequality we used the lower bound for $\boldsymbol{L}^\dagger$ in (S.7) above, and in the last inequality we used results from the proof of Theorem 7. We have thus shown the result for $\mathbb{E}(r_i^2)$ in the theorem. Similarly we find

$$\begin{aligned}\sigma^{-2}\mathbb{E}\left(\boldsymbol{R}\boldsymbol{R}'\right)&=(\boldsymbol{C}_1-\boldsymbol{C}_2)(\boldsymbol{C}_1-\boldsymbol{C}_2)'\\&=\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\boldsymbol{L}_B\boldsymbol{L}_W^\dagger\boldsymbol{L}_B\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\\&\leq\kappa\boldsymbol{P}'\boldsymbol{L}_*^{\text{inv}}\boldsymbol{P}\leq\kappa\ \text{var}\left(\boldsymbol{P}'\widehat{\boldsymbol{\gamma}}^{\text{inf}}\right),\end{aligned}$$

which implies the result for $\mathbb{E}(R_i^2)$ in the theorem. This concludes the proof. □