INFERENCE FOR SENDER-RECEIVER ALIGNMENT IN DIRECTED NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

In a directed network, does understanding a node's sending behaviour give us any information about its receiving behaviour? To address this question, we propose the effective shared dimension, a real-valued count of how many directions the two roles share, alongside a degree-free counterpart which removes the effect of a node's overall activity level. We define both for general latent position models, and demonstrate that under a directed random dot product graph model they are explicit functions of moments of the underlying node positions. For each quantity we give an estimator, for which we prove consistency and, provided the average node degree grows faster than $n^{1/2}$, asymptotic normality, before discussing a bias correction to potentially extend the latter to sparser regimes. Finally, we apply this methodology to estimate the effective shared dimension across 10 real directed networks spanning social, information, infrastructure and biological data, and ask whether the answer differs systematically between these domains.

1 INTRODUCTION

In a directed network each node exhibits two behaviours: one based on who it sends edges to, and one on who it receives edges from. The two can differ significantly depending on the network in question; in communication networks one would expect them to be largely similar, while in a food web a species' prey and predators typically have minimal overlap. While closely related to the notion of reciprocity, which counts how many edges are returned (Garlaschelli & Loffredo, 2004), the alignment between these two behaviours is subtly different; it determines how likely nodes that send to similar places are to also receive from similar places.

To quantify the alignment between the two behaviours, we assume that our network can be modelled as a directed random dot product graph (DRDPG), in which node i has sending and receiving positions $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^d$, and an edge $i \to j$ appears with probability $\rho_n \mathbf{x}_i^\top \mathbf{y}_j$ (Young & Scheinerman, 2007; Athreya et al., 2018; Marenco et al., 2024). We define subspaces describing the sending and receiving behaviour of nodes under this model, and then measure their alignment using a measure proposed by Krzanowski (1979): $\sum_{k=1}^{d} \cos^2(\theta_k)$, where the θ_k are the principal angles between the subspaces. This measure, which in our setting we call the *effective shared dimension* Ψ_F, ranges from 0 for unrelated roles to d for identical ones. We also propose a degree-free version, Ψ_F°, which normalises for node activity. Both quantities can be estimated from the singular value decomposition of the adjacency matrix.

The novelty of this paper lies in its theoretical contribution: we give closed-form expressions for Ψ_F and Ψ_F° in our chosen setting (Section 4) and prove consistency of spectral estimates, together with asymptotic normality whenever the average degree grows faster than $n^{1/2}$ (Section 5). We demonstrate that this threshold is sharp: below it, the estimate keeps its Gaussian shape and variance but is shifted by a bias term. Finally, by deriving the leading bias term in closed-form, we propose a corrected estimate that we theorise should extend our results to whenever the average degree grows faster than $n^{1/4}$ (Section 6).

Our work fits within a well-established body of literature on the statistical theory of networks. Latent position models (Hoff et al., 2002), such as random dot product graphs (Athreya et al., 2018; Rubin-Delanchy et al., 2022) play a prominent role in the study of network embeddings, which can be used to understand node behaviour. For directed graphs in particular, spectral embedding gives separate

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

sending and receiving node representations (Sussman et al., 2012); the stochastic co-blockmodel lets sending and receiving communities differ (Rohe et al., 2016); effects models include sender and receiver factors (Hoff, 2021); and machine-learning methods learn distinct source and target embeddings (Ou et al., 2016; Khosla et al., 2019; Zhang et al., 2021). While these methods estimate the two roles, they make no claims as to how aligned they are, nor do they provide inference for such a measure. Works such as Arroyo et al. (2021) address a question similar in spirit to ours, but look at roles shared across a number of networks, rather than the alignment of two different roles within the same network.

The quantity $\sum_{k=1}^{d} \cos^2(\theta_k)$ is also known in the literature as the Grassmann projection kernel (Hamm & Lee, 2008) and appears in the RV coefficient of Robert & Escoufier (1976). Related indices are widely used to compare neural representations (Raghu et al., 2017; Kornblith et al., 2019), and Giaffar et al. (2024) propose a variance-weighted effective number of shared dimensions for paired datasets which has a similar flavour to our method. Our index counts shared directions equally, and, unlike in these settings, the two representations are not observed, instead being estimated from a single noisy network.

Lastly, our technical results build on existing subspace perturbation theory (Davis & Kahan, 1970; Cape et al., 2019), entrywise expansions (Abbe et al., 2020; Xia, 2021) and limit theorems for spectral embeddings (Athreya et al., 2016; Tang & Priebe, 2018). Removal of leading bias terms is central to inference for singular subspaces (Xia, 2021), network memberships (Fan et al., 2022) and principal components (Yan et al., 2024); in our case this bias arises from comparing two estimated subspaces, and our correction depends on the sparsity of the network.

2 Effective shared dimension

We assume a general latent position model (Hoff et al., 2002) for the adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, in which there exist compact subsets $\mathcal{X}$ and $\mathcal{Y}$ of Euclidean space, a continuous function $f : \mathcal{X} \times \mathcal{Y} \to [0,1]$ and a Borel probability measure F on $\mathcal{X} \times \mathcal{Y}$ such that the entries $\mathbf{A}_{ij}$ for $i \neq j$ are conditionally independent given the latent positions and satisfy

$$\mathbf{A}_{ij} \sim \text{Bernoulli}\big(f(\mathbf{x}_i, \mathbf{y}_j)\big)$$

where $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)$ are i.i.d. pairs of random vectors distributed according to F.

For any $(\mathbf{x}, \mathbf{y}) \in \text{supp}(F)$ (representing a potential node in the network) we can define a *sending profile* $f_{\mathbf{x}} := f(\mathbf{x}, \cdot\,)$, and a *receiving profile* $f_{\mathbf{y}} := f(\,\cdot\,, \mathbf{y})$, which show how likely the node is to send an edge to or receive an edge from each possible partner. Both functions lie in $L^2(F)$, and we can therefore define *sending and receiving subspaces* $\mathcal{S}_{\text{out}}$ and $\mathcal{S}_{\text{in}}$ as the spans of these functions as $(\mathbf{x}, \mathbf{y})$ ranges over $\text{supp}(F)$.

To quantify the alignment between these two subspaces, we make the following assumption:

A1. $\mathcal{S}_{\text{out}}$ and $\mathcal{S}_{\text{in}}$ have finite dimension d.

Note that the dimensions of $\mathcal{S}_{\text{in}}$ and $\mathcal{S}_{\text{out}}$ are both equal to the rank of the integral operator with kernel f, so the finiteness assumption forces them to both be equal to d. While Assumption **A1** does rule out infinite-rank kernels, any such kernel has a best rank-d approximation (see, for example, Hsing & Eubank, 2015, Theorem 4.4.7), so one can view a finite-rank f as a truncation of the true kernel.

Under this assumption, we make the following definition:

Definition 1 (Effective shared dimension)**.** Let $\theta_1 \leq \cdots \leq \theta_d$ be the principal angles between the subspaces $\mathcal{S}_{\text{out}}$ and $\mathcal{S}_{\text{in}}$. The *effective shared dimension* is

$$\Psi_F := \sum_{k=1}^{d} \cos^2(\theta_k).$$

The effective shared dimension therefore counts how many independent directions the sending and receiving roles agree on, with partial agreement counted fractionally. The quantity $\Psi_F/d \in [0,1]$ can then be viewed informally as the fraction of a typical sending direction that is explained by the receiving subspace.

For real-world data, there is no guarantee that node degree is distributed homogeneously across a network: for example, brain activity networks often display a number of nodes with very high degree (Fornito et al., 2016, Chapter 4). As such, it can be helpful to have an alternative measure of alignment that separates node activity from the two roles. To achieve this, we rescale each profile by how busy its partner is. Let

$$d_{\text{out}}(\mathbf{x}) = \int f(\mathbf{x}, \mathbf{y}) \, \mathrm{d}F_y(\mathbf{y}), \qquad d_{\text{in}}(\mathbf{y}) = \int f(\mathbf{x}, \mathbf{y}) \, \mathrm{d}F_x(\mathbf{x})$$

be the expected out- and in-degrees of a node at position $\mathbf{x}$ or $\mathbf{y}$, where F_x and F_y are the marginals of F on $\mathcal{X}$ and $\mathcal{Y}$, and make the following additional assumption:

A2. The expected degrees $d_{\text{out}}(\cdot)$ and $d_{\text{in}}(\cdot)$ are bounded away from zero on $\text{supp}(F)$.

Under this assumption, we may define the *normalised* profiles

$$f^{\circ}_{\mathbf{x}} := \frac{f(\mathbf{x}, \cdot)}{d_{\text{in}}(\cdot)}, \qquad f^{\circ}_{\mathbf{y}} := \frac{f(\cdot, \mathbf{y})}{d_{\text{out}}(\cdot)}$$

so that for example $f^{\circ}_{\mathbf{x}}(\mathbf{y}')$ is the node's *share* of the edges arriving at $\mathbf{y}'$, a relative measure against all other senders rather than an absolute one. Writing $\mathcal{S}^{\circ}_{\text{out}}$ and $\mathcal{S}^{\circ}_{\text{in}}$ as the spans of these functions as before (these both have dimension d by Assumptions **A1** and **A2**), we make the following definition:

Definition 2 (Degree-free effective shared dimension)**.** Let $\theta^{\circ}_1 \leq \cdots \leq \theta^{\circ}_d$ be the principal angles between the subspaces $\mathcal{S}^{\circ}_{\text{out}}$ and $\mathcal{S}^{\circ}_{\text{in}}$. The *degree-free effective shared dimension* is

$$\Psi^{\circ}_F := \sum_{k=1}^{d} \cos^2(\theta^{\circ}_k).$$

The degree-free effective shared dimension counts how many independent directions the sending and receiving roles agree on *after* normalising to account for differing activity levels among nodes.

3 The directed random dot product graph

Under the finite-dimensional restriction of Assumption **A1**, our latent position model can be seen to be a directed instance of the *random dot product graph* (RDPG) of Young & Scheinerman (2007), in which the kernel function $f(\mathbf{x}_i, \mathbf{y}_j)$ is simply the dot product between the vectors $\mathbf{x}_i$ and $\mathbf{y}_j$, which are assumed to live in some compact set in $\mathbb{R}^d$. For our purposes it is useful to retain a distinction between the domains of the two separate latent positions, and so we use the following definition:

Definition 3 (Directed random dot product graph)**.** Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$ be compact sets with $\mathbf{x}^\top \mathbf{y} \in [0, 1]$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$. A random matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$ is said to follow a *directed random dot product graph model* if $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)$ are i.i.d. random vectors distributed according to a Borel probability measure F supported on $\mathcal{X} \times \mathcal{Y}$ and, conditionally on $\mathbf{X} = [\mathbf{x}_1 | \cdots | \mathbf{x}_n]^\top$ and $\mathbf{Y} = [\mathbf{y}_1 | \cdots | \mathbf{y}_n]^\top$, the entries $\mathbf{A}_{ij}$ satisfy

$$\mathbf{A}_{ij} \sim \text{Bernoulli}\left(\rho_n \mathbf{x}_i^\top \mathbf{y}_j\right) \text{ for all } i \neq j \quad \text{and} \quad \mathbf{A}_{ii} = 0 \text{ for all } i$$

where the entries $\mathbf{A}_{ij}$ are mutually independent and ρ_n is a sparsity factor satisfying either $\rho_n = 1$ or $\rho_n \to 0$ as $n \to \infty$. We write $\mathbf{A} \sim \text{DRDPG}(F, \rho_n)$ as shorthand.

Note that the latent positions $(\mathbf{x}_i, \mathbf{y}_i)$ are identifiable only up to a transformation $(\mathbf{x}_i, \mathbf{y}_i) \mapsto (\mathbf{G}\mathbf{x}_i, \mathbf{G}^{-\top}\mathbf{y}_i)$ for some invertible matrix $\mathbf{G} \in \mathbb{R}^{d \times d}$. While this may cause some slight unease, we will establish later that the effective shared dimension is unaffected by this ambiguity.

The following result (whose proof, along with those of the other results in this section, we present in Appendix A.2) shows that Assumption **A1** means we are implicitly working with a DRDPG:

Proposition 1. *Under Assumption **A1** there exist continuous maps $\xi : \mathcal{X} \to \mathbb{R}^d$ and $\eta : \mathcal{Y} \to \mathbb{R}^d$ with $f(\mathbf{x}, \mathbf{y}) = \xi(\mathbf{x})^\top \eta(\mathbf{y})$ for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$, so that $\mathbf{A}_{ij} \sim \text{Bernoulli}\big(\xi(\mathbf{x}_i)^\top \eta(\mathbf{y}_j)\big)$ follows a directed random dot product graph model with latent positions $\big(\xi(\mathbf{x}_1), \eta(\mathbf{y}_1)\big), \ldots, \big(\xi(\mathbf{x}_n), \eta(\mathbf{y}_n)\big)$.*

Many studies of statistical inference on RDPGs restrict their attention to the symmetric case, which covers classical models such as the stochastic blockmodel (Holland et al., 1983) and its mixed-membership (Airoldi et al., 2008) and degree-corrected (Karrer & Newman, 2011) variants. In this case a node has only a single latent position $\mathbf{x}_i$ and the effective shared dimension is trivially equal to d, and thus yields no extra insight into the graph's structure. Under the DRDPG, we have access to a more general version of the classical stochastic blockmodel:

The stochastic co-blockmodel. Fix positive integers K and L and let $\mathbf{\Pi} \in [0,1]^{K\times L}$ be a matrix whose entries sum to 1 and $\mathbf{B} \in [0,1]^{K\times L}$ an arbitrary matrix. We assign to each node a 'sending' community $z_i \in \{1,\dots,K\}$ and a 'receiving' community $w_i \in \{1,\dots,L\}$ where $(z_1,w_1),\dots,(z_n,w_n)$ are i.i.d. random pairs with $\mathbb{P}(z_i = k, w_i = l) = \mathbf{\Pi}_{kl}$ and then define an adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$ for a given sparsity factor ρ_n by

$$\mathbf{A}_{ij} \sim \text{Bernoulli}\left(\rho_n \mathbf{B}_{z_i w_j}\right) \text{ for all } i \neq j \quad \text{and} \quad \mathbf{A}_{ii} = 0 \text{ for all } i.$$

To see that this is indeed a DRDPG, let $d = \text{rank}(\mathbf{B})$, construct the compact singular value decomposition $\mathbf{B} = \mathbf{U_B}\mathbf{\Sigma_B}\mathbf{V_B}^\top$, and define $\mathbf{x}_i = \mathbf{\Sigma_B}^{1/2}\mathbf{U_B}^\top \mathbf{e}_{z_i}$ and $\mathbf{y}_j = \mathbf{\Sigma_B}^{1/2}\mathbf{V_B}^\top \mathbf{f}_{w_j}$ for each $i,j \in \{1,\dots,n\}$, where $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_j\}$ are the standard bases for $\mathbb{R}^K$ and $\mathbb{R}^L$ respectively. Since $\mathbf{x}_i^\top \mathbf{y}_j = \mathbf{B}_{z_i w_j}$, we see that $\mathbf{A}_{ij} \sim \text{Bernoulli}(\rho_n \mathbf{x}_i^\top \mathbf{y}_j)$ as required.

Note: Conditional on the community assignments this is precisely the stochastic co-blockmodel defined by Rohe et al. (2016), and so we retain their name for the sake of continuity.

We can define an analogue of the degree-corrected stochastic blockmodel similarly:

The degree-corrected stochastic co-blockmodel. Let H be a distribution on $[a,1]^2$ for some $a > 0$, and let $(\theta_1,\phi_1),\dots,(\theta_n,\phi_n)$ be i.i.d. random pairs distributed according to H. Given a set of i.i.d. pairs $(z_1,w_1),\dots,(z_n,w_n)$ distributed as in the standard stochastic co-blockmodel, we define an adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$ for a given sparsity factor ρ_n by

$$\mathbf{A}_{ij} \sim \text{Bernoulli}\left(\rho_n \theta_i \phi_j \mathbf{B}_{z_i w_j}\right) \text{ for all } i \neq j \quad \text{and} \quad \mathbf{A}_{ii} = 0 \text{ for all } i.$$

As before, this can be realised as a DRDPG, this time by setting $\mathbf{x}_i = \theta_i \mathbf{\Sigma_B}^{1/2}\mathbf{U_B}^\top \mathbf{e}_{z_i}$ and $\mathbf{y}_j = \phi_j \mathbf{\Sigma_B}^{1/2}\mathbf{V_B}^\top \mathbf{f}_{w_j}$ for each $i,j \in \{1,\dots,n\}$.

4 The effective shared dimension of directed random dot product graphs

Identifying our model as a DRDPG allows us to express the objects of Section 2 in a more concrete form. Let $\mathbf{A} \sim \text{DRDPG}(F,\rho_n)$, so that the kernel is $f(\mathbf{x},\mathbf{y}) = \rho_n \mathbf{x}^\top \mathbf{y}$, and write $\Delta_x = \mathbb{E}[\mathbf{x}\mathbf{x}^\top]$, $\Delta_{xy} = \mathbb{E}[\mathbf{x}\mathbf{y}^\top]$, $\Delta_{yx} = \Delta_{xy}^\top$, $\Delta_y = \mathbb{E}[\mathbf{y}\mathbf{y}^\top]$, $\boldsymbol{\mu}_x = \mathbb{E}[\mathbf{x}]$ and $\boldsymbol{\mu}_y = \mathbb{E}[\mathbf{y}]$ for the second moments and means of a pair of random vectors $(\mathbf{x},\mathbf{y}) \sim F$. Under our model, we then have the following alternative statements for Assumptions **A1** and **A2**:

Proposition 2. *Let* $\mathbf{A} \sim \text{DRDPG}(F,\rho_n)$. *Then Assumptions **A1** and **A2** are equivalent to:*

A1′. The second moment matrices Δ_x and Δ_y are of maximal rank.

A2′. The quantities $\mathbf{x}^\top \boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x^\top \mathbf{y}$ are bounded away from zero on $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Under Assumption **A2′**, define the *degree-normalised positions*

$$\mathbf{x}^\circ = \frac{\mathbf{x}}{\mathbf{x}^\top \boldsymbol{\mu}_y}, \qquad \mathbf{y}^\circ = \frac{\mathbf{y}}{\boldsymbol{\mu}_x^\top \mathbf{y}},$$

and write $\Delta_{x^\circ} = \mathbb{E}[\mathbf{x}^\circ \mathbf{x}^{\circ\top}]$, $\Delta_{x^\circ y^\circ} = \mathbb{E}[\mathbf{x}^\circ \mathbf{y}^{\circ\top}]$, $\Delta_{y^\circ} = \mathbb{E}[\mathbf{y}^\circ \mathbf{y}^{\circ\top}]$ for their second moments. We then have a simple closed-form expression for Ψ_F and Ψ_F°:

Proposition 3. *Let* $\mathbf{A} \sim \text{DRDPG}(F,\rho_n)$.

(i) *Under Assumption **A1′**,*

$$\Psi_F = \left\|\Delta_x^{-1/2}\Delta_{xy}\Delta_y^{-1/2}\right\|_{\text{Frob}}^2 = \text{tr}\left(\Delta_x^{-1}\Delta_{xy}\Delta_y^{-1}\Delta_{yx}\right).$$

(ii) *Under Assumptions **A1′** and **A2′**, the matrices Δ_{x° and Δ_{y° are of maximal rank and*

$$\Psi_F^\circ = \left\|\Delta_{x^\circ}^{-1/2}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1/2}\right\|_{\text{Frob}}^2 = \operatorname{tr}\left(\Delta_{x^\circ}^{-1}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1}\Delta_{y^\circ x^\circ}\right).$$

Several properties of the effective shared dimension become clear from Proposition 3. Firstly, Ψ_F and Ψ_F° are invariant under any transformation of the form $(\mathbf{x}_i, \mathbf{y}_i) \mapsto (\mathbf{G}\mathbf{x}_i, \mathbf{G}^{-\top}\mathbf{y}_i)$ for an invertible matrix $\mathbf{G} \in \mathbb{R}^{d\times d}$, and so the ambiguity in the latent positions does not affect our measure.

Secondly, $\Psi_F = 0$ if, and only if, $\Delta_{xy} = \mathbf{0}$, which occurs if, and only if, F describes a bipartite population in which every node *either* sends or receives, and not both. At the other extreme, $\Psi_F = d$ if, and only if, the subspaces $\mathcal{S}_{\text{in}}$ and $\mathcal{S}_{\text{out}}$ coincide. The coordinate functions on $\mathcal{X}$ and $\mathcal{Y}$ span these spaces, and so they coincide if, and only if, $\mathbf{y} = \mathbf{G}\mathbf{x}$ almost surely for some invertible $\mathbf{G} \in \mathbb{R}^{d\times d}$.

Finally, since $\mathbf{x}^{\circ\top}\boldsymbol{\mu}_y = \boldsymbol{\mu}_x^\top\mathbf{y}^\circ = 1$ for every $(\mathbf{x}, \mathbf{y}) \in \operatorname{supp}(F)$, the constant function is a linear combination of the scaled coordinate functions, which form bases of $\mathcal{S}_{\text{out}}^\circ$ and $\mathcal{S}_{\text{in}}^\circ$. These subspaces therefore always share at least one direction, so $\Psi_F^\circ \geq 1$. Thus while Ψ_F/d is a natural quantity to consider when assessing the effective shared dimension of a graph, for the degree-free version it makes more sense to consider $(\Psi_F^\circ - 1)/(d-1)$.

In a stochastic co-blockmodel (or its degree-corrected counterpart), the second moment matrices reduce to sums of a finite number of rank-one matrices. In Appendix B.1 we demonstrate how to calculate the standard and degree-free effective shared dimension for these models.

Example: A stochastic co-blockmodel. Consider a 2-community stochastic co-blockmodel with

$$\boldsymbol{\Pi} = \begin{pmatrix} 0.25(1+p) & 0.25(1-p) \\ 0.25(1-p) & 0.25(1+p) \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$$

for $p \in [0, 1]$; when $p = 1$ a node's sending and receiving communities are the same, and as p decreases the two are increasingly likely to differ. For $p \in \{1, 0.9, 0.75, 0.5\}$ we generate an adjacency matrix $\mathbf{A}$ on $n = 1000$ nodes from the resulting co-blockmodel, and assign to each node two embeddings given by the rows of the matrices $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ of left and right singular vectors of $\mathbf{A}$. Figure 1 displays these embeddings after a Procrustes alignment: points are coloured red or blue according to community membership, with lighter hues for the sending embedding and darker hues for the receiving embedding, and lines are drawn between the two embeddings for each node.

If one were to only observe the two clusters for each embedding it would be easy to assume a full shared subspace regardless of p, as they appear well-aligned. As p decreases, however, so too does the effective shared dimension, and this discrepancy is represented visually by an increasing number of lines crossing between clusters.

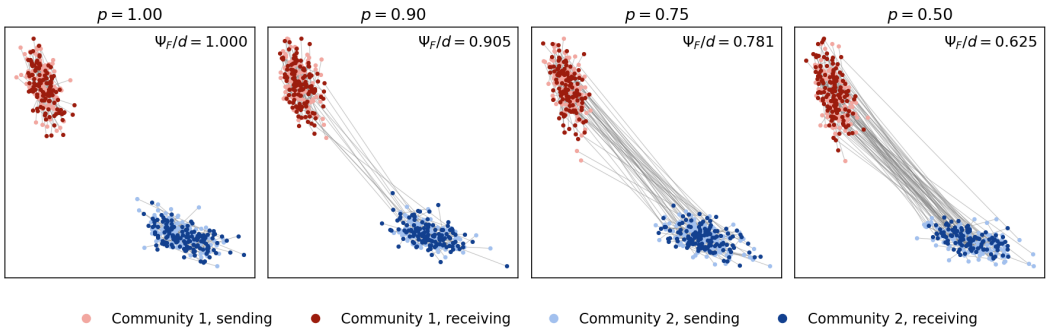


Figure 1: Procrustes-aligned embedding of a 2-community stochastic co-blockmodel, with decreasing effective shared dimension. Colour represents community membership, and hue differentiates between sending and receiving behaviour. Lines are drawn between a node's two embeddings.

5 Estimating the Effective Shared Dimension

As an estimate for the effective shared dimension, we consider the alignment between the subspaces of left and right singular vectors of the adjacency matrix $\mathbf{A}$. Let $\widehat{\mathbf{U}}, \widehat{\mathbf{V}} \in \mathbb{R}^{n\times d}$ be the matrices

whose columns are the top-d left and right singular vectors of $\mathbf{A}$ respectively, and define

$$\boxed{\widehat{\Psi} := \mathrm{tr}\big(\mathbf{\Pi}_{\widehat{\mathbf{U}}}\mathbf{\Pi}_{\widehat{\mathbf{V}}}\big) = \|\widehat{\mathbf{V}}^\top\widehat{\mathbf{U}}\|_{\mathrm{Frob}}^2}$$

where for a matrix $\mathbf{M} \in \mathbb{R}^{n\times d}$ of rank d we define $\mathbf{\Pi}_{\mathbf{M}}$ to be the orthogonal projection onto the column space of $\mathbf{M}$, given by $\mathbf{\Pi}_{\mathbf{M}} := \mathbf{M}(\mathbf{M}^\top\mathbf{M})^{-1}\mathbf{M}^\top$ (in particular, if $\mathbf{M}$ has orthonormal columns then $\mathbf{\Pi}_{\mathbf{M}} = \mathbf{M}\mathbf{M}^\top$).

For the degree-free version, let $\widehat{\mathbf{D}}_{\mathrm{out}}, \widehat{\mathbf{D}}_{\mathrm{in}} \in \mathbb{R}^{n\times n}$ be the diagonal matrices whose entries are the out- and in-degrees of each node respectively (set to 1 if the degree is zero), and define

$$\boxed{\widehat{\Psi}^\circ := \mathrm{tr}\big(\mathbf{\Pi}_{\widehat{\mathbf{D}}_{\mathrm{out}}^{-1}\widehat{\mathbf{U}}}\mathbf{\Pi}_{\widehat{\mathbf{D}}_{\mathrm{in}}^{-1}\widehat{\mathbf{V}}}\big) = \|\widehat{\mathbf{V}}_*^\top\widehat{\mathbf{U}}_*\|_{\mathrm{Frob}}^2}$$

where $\widehat{\mathbf{U}}_*, \widehat{\mathbf{V}}_* \in \mathbb{R}^{n\times d}$ are the matrices whose columns are the left singular vectors of $\widehat{\mathbf{D}}_{\mathrm{out}}^{-1}\widehat{\mathbf{U}}$ and $\widehat{\mathbf{D}}_{\mathrm{in}}^{-1}\widehat{\mathbf{V}}$ respectively. In other words, we divide the rows of $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ by their corresponding node's activity before comparing the two subspaces. The matrices $\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{U}}_*$ and $\widehat{\mathbf{V}}_*$ are only defined up to multiplication by a $d \times d$ orthogonal matrix, but $\widehat{\Psi}$ and $\widehat{\Psi}^\circ$ are invariant under such transformations, and thus are well-defined.

Theorems 4 and 5 provide asymptotic guarantees for the behaviour of these two estimators, namely that they are consistent and, in a suitably dense regime, asymptotically normal. We present proofs for these statements in Appendices A.3 and A.4. The following assumption ensures that the average degree of the network grows sufficiently quickly, which is necessary for the results to hold:

A3. The sparsity factor ρ_n satisfies $n\rho_n = \omega(\log n)$.

(Here we use the standard asymptotic notation that $f = \omega(g)$ if $f/g \to \infty$ as $n \to \infty$.)

Theorem 4. *Under Assumptions **A1′** and **A3**, let $\tau_F^2 = \mathrm{Var}\big(g(\mathbf{x}, \mathbf{y})\big)$ where*

$$g(x, y) = 2\mathbf{x}^\top\mathbf{T}\mathbf{y} - \mathbf{x}^\top\mathbf{T}\Delta_{yx}\Delta_x^{-1}\mathbf{x} - \mathbf{y}^\top\mathbf{T}^\top\Delta_{xy}\Delta_y^{-1}\mathbf{y}, \qquad \mathbf{T} = \Delta_x^{-1}\Delta_{xy}\Delta_y^{-1}.$$

Then:

(i) $|\widehat{\Psi} - \Psi_F| = \mathrm{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}} + \frac{1}{n\rho_n}\right)$.

(ii) *If $n\rho_n = \omega(n^{1/2})$ then $n^{1/2}(\widehat{\Psi} - \Psi_F) \to \mathcal{N}(0, \tau_F^2)$.*

Theorem 5. *Under Assumptions **A1′**, **A2′** and **A3**, let $\tau_F^{\circ 2} = \mathrm{Var}\big(g^\circ(\mathbf{x}, \mathbf{y})\big)$, where*

$$g^\circ(\mathbf{x}, \mathbf{y}) = \tfrac{1}{2}\big(q_x + q_y\big)(\mathbf{x}^\circ, \mathbf{y}^\circ) - \mathbb{E}\big[q_y(\mathbf{x}^\circ, \mathbf{y}^\circ)\,\mathbf{y}^\circ\big]^\top(\mathbf{x} - \boldsymbol{\mu}_x) - \mathbb{E}\big[q_x(\mathbf{x}^\circ, \mathbf{y}^\circ)\,\mathbf{x}^\circ\big]^\top(\mathbf{y} - \boldsymbol{\mu}_y)$$

and

$$q_x(\mathbf{u}, \mathbf{v}) = 2\mathbf{u}^\top\mathbf{T}^\circ\mathbf{v} - 2\mathbf{u}^\top\mathbf{T}^\circ\Delta_{y^\circ x^\circ}\Delta_{x^\circ}^{-1}\mathbf{u}, \qquad q_y(\mathbf{u}, \mathbf{v}) = 2\mathbf{u}^\top\mathbf{T}^\circ\mathbf{v} - 2\mathbf{v}^\top\mathbf{T}^{\circ\top}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1}\mathbf{v}.$$

Then:

(i) $|\widehat{\Psi}^\circ - \Psi_F^\circ| = \mathrm{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}} + \frac{1}{n\rho_n}\right)$.

(ii) *If $n\rho_n = \omega(n^{1/2})$ then $n^{1/2}(\widehat{\Psi}^\circ - \Psi_F^\circ) \to \mathcal{N}(0, \tau_F^{\circ 2})$.*

(The notation $f = \mathrm{O}_{\mathbb{P}}(g)$ here means that $f = O(g)$ with *overwhelming probability*: for any $\gamma > 0$ there exists a constant C_γ such that $\mathbb{P}(|f| \le C_\gamma g) \ge 1 - n^{-\gamma}$ for all sufficiently large n).

Example: The CLT for the effective shared dimension of a stochastic co-blockmodel. To demonstrate this theory in practice, we consider a stochastic co-blockmodel with two sending communities and three receiving communities, where:

$$\mathbf{B} = \begin{pmatrix} 0.9 & 0.1 & 0.2 \\ 0.1 & 0.9 & 0.1 \end{pmatrix}, \quad \mathbf{\Pi} = \begin{pmatrix} 0.3 & 0.1 & 0.15 \\ 0.05 & 0.25 & 0.15 \end{pmatrix}.$$

We consider only the standard estimate $\widehat{\Psi}$, and investigate two sparsity regimes, in which the average degree $n\rho_n$ is proportional to $n^{3/4}$ and $n^{1/2}$ respectively (the former is dense enough for Theorem 4(ii) to hold, while the latter is not covered by the theory). For each $n \in \{500, 2500, 10000\}$ we run 2000 simulations, in which the latent positions and edges are generated anew each time, and then plot the resulting histograms for the normalised quantity $n^{1/2}(\widehat{\Psi} - \Psi_F)/\tau_F$, which we present in Figure 2. Here the blue histogram represents the denser regime, which we observe converging towards the limiting distribution (in black) while the histogram for the sparser regime appears to remain offset to the left as n grows. The summary statistics presented in Table 1 confirm this, with the dense regime visibly converging to a standard normal distribution, while the sparse regime remains offset by a mean shift of around 7.4.

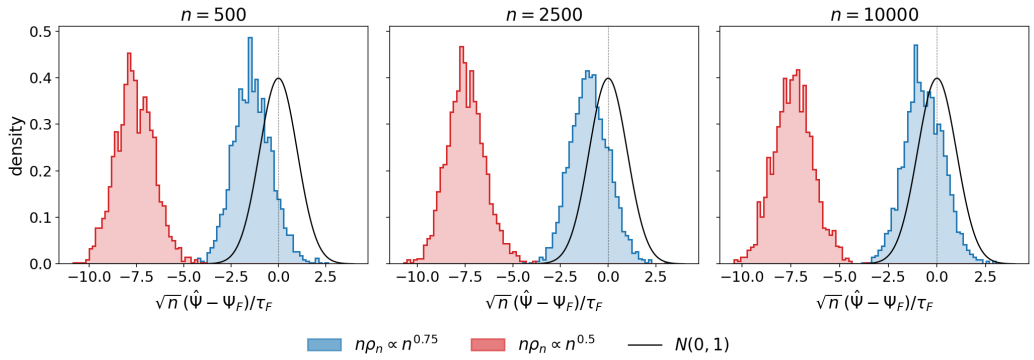


Figure 2: Sample histograms for the quantity $n^{1/2}(\widehat{\Psi} - \Psi_F)/\tau_F$ calculated for a stochastic co-blockmodel over 2000 trials under sparsity regimes $n\rho_n \propto n^{3/4}$ (blue) and $n\rho_n \propto n^{1/2}$ (red). The first regime meets the criteria of Theorem 4 and can be seen to converge to a standard normal distribution, while the second does not.

Table 1: Summary statistics for the quantity $n^{1/2}(\widehat{\Psi} - \Psi_F)/\tau_F$ for a stochastic co-blockmodel under different sparsity regimes. Columns display the mean, standard deviation, and proportion of simulated datasets in which the nominal 95% confidence interval contains Ψ_F, calculated from a set of 2000 trials.

	$n\rho_n \propto n^{3/4}$			$n\rho_n \propto n^{1/2}$		
n	mean	s.d.	coverage	mean	s.d.	coverage
500	−1.38	1.00	71.2%	−7.60	1.01	0.0%
2500	−0.97	0.99	83.5%	−7.48	0.98	0.0%
10000	−0.67	1.00	89.8%	−7.39	1.00	0.0%

6 A BIAS CORRECTION

The central limit theorem in Theorem 4(ii) holds if the average degree $n\rho_n$ grows faster than $n^{1/2}$, and the example above demonstrates the tightness of this bound. The same example seems to indicate that at this threshold a central limit theorem does still hold: the variance of the empirical distribution of $n^{1/2}(\widehat{\Psi} - \Psi_F)$ approaches τ_F^2, while the mean approaches a constant, non-zero value. In this section we attempt to explain this phenomenon, and justify a correction to the initial estimate to allow us to explore sparser regimes. Any results referenced can be found in Appendix A.3.

Define $f(\mathbf{M}) = \mathrm{tr}(\mathbf{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{\Pi}_{\widetilde{\mathbf{V}}})$ for a matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ with top-d singular vectors $\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}} \in \mathbb{R}^{n\times d}$, so that $\widehat{\Psi} = f(\mathbf{A})$. Define $\Psi := f(\mathbf{P})$, where $\mathbf{P} = \rho_n \mathbf{X}\mathbf{Y}^\top$ is the edge probability matrix from the DRDPG. One can show (Proposition 8) that $n^{1/2}(\Psi - \Psi_F) \to \mathcal{N}(0, \tau_F^2)$, and this holds under *any* sparsity regime. Any obstruction must therefore be caused by the discrepancy between the estimate $\widehat{\Psi}$ and the noiseless value Ψ, more specifically from its average over the edge noise.

For latent positions $\mathbf{X}$ and $\mathbf{Y}$, define the *conditional bias* $b_n := \mathbb{E}\big[\widehat{\Psi} \mid \mathbf{X}, \mathbf{Y}\big] - \Psi$. The scaled difference $n^{1/2}(\widehat{\Psi} - \Psi_F)$ then decomposes as:

$$n^{1/2}\big(\widehat{\Psi} - \Psi_F\big) = n^{1/2}\big(\Psi - \Psi_F\big) + n^{1/2}\, b_n + n^{1/2}\big(\widehat{\Psi} - \Psi - b_n\big),$$

where, as already stated, the first term converges in distribution to $\mathcal{N}(0, \tau_F^2)$, while the last term is by definition zero-mean and has a leading part which can be seen (by Proposition 13) to have conditional variance of order $1/(n^2\rho_n)$. After scaling by $n^{1/2}$ this leading part is of order $(n\rho_n)^{-1/2}$ and thus vanishes whenever $n\rho_n \to \infty$, and so the behaviour of $\widehat{\Psi}$ below the $n^{1/2}$ threshold is governed by the bias.

To see where the bias comes from, we expand f around $\mathbf{P}$. By Theorem 1 of Xia (2021) (which we apply to the Hermitian dilations of the matrices concerned as in the proof of Proposition 11):

$$f(\mathbf{A}) - f(\mathbf{P}) = \sum_{k \geq 1} f_k(\mathbf{A} - \mathbf{P})$$

for n sufficiently large, where each f_k is homogeneous of degree k in the entries of $\mathbf{A} - \mathbf{P}$.

The first-order term has conditional mean zero (Proposition 12) so does not contribute to the bias. The second-order term has conditional mean:

$$Q(\mathbf{P}) = \frac{1}{2} \sum_{i \neq j} \mathbf{P}_{ij}(1 - \mathbf{P}_{ij})\, \frac{\partial^2 f}{\partial \mathbf{P}_{ij}^2}(\mathbf{P})$$

and we define an estimate $\hat{b}_n$ of the conditional bias and a corrected estimate $\widetilde{\Psi}$ of Ψ_F by

$$\boxed{\hat{b}_n = Q(\hat{\mathbf{P}})} \qquad \text{and} \qquad \boxed{\widetilde{\Psi} = \widehat{\Psi} - \hat{b}_n}$$

with $\widehat{\mathbf{P}} := \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^\top$ the rank-d approximation to $\mathbf{A}$. (For a full derivation of Q, see Appendix C.)

To investigate this bias correction, we sampled 50 random stochastic co-blockmodels with 2 sending and 3 receiving communities. The matrix $\mathbf{B}$ was drawn randomly with $\mathbf{B}_{11}, \mathbf{B}_{22} \sim U(0.8, 0.95)$, $\mathbf{B}_{12}, \mathbf{B}_{21} \sim U(0.02, 0.15)$ and $\mathbf{B}_{13}, \mathbf{B}_{23} \sim U(0.05, 0.2)$, and each node's sending community was chosen uniformly, with its receiving community equal to its sending community with probability $\lambda \sim U(0.2, 0.9)$, and otherwise chosen uniformly. For each model we generated 400 networks for $n \in \{500, 2500, 10000\}$ and sparsity $n\rho_n \propto n^{3/4}, n^{2/5}$ and $n^{1/4}$, and computed $\widehat{\Psi}$ and $\widetilde{\Psi}$.

Studying the normalised quantities $\sqrt{n}(\widehat{\Psi} - \Psi_F)/\tau_F$ and $\sqrt{n}(\widetilde{\Psi} - \Psi_F)/\tau_F$ allows us to make a direct comparison between the different models, and Table 2 presents summary statistics for these quantities. From this we observe that, as expected from Theorem 4(ii), we have convergence to the standard normal distribution for the uncorrected estimate only when $n\rho_n \propto n^{3/4}$, with the other regimes being too sparse. We note, however, that the *corrected* estimate also appears to converge when $n\rho_n \propto n^{2/5}$, which is sparser than the boundary rate of $n^{1/2}$ given in the theorem.

Table 2: Summary statistics for the quantities $n^{1/2}(\widehat{\Psi} - \Psi_F)/\tau_F$ and $n^{1/2}(\widetilde{\Psi} - \Psi_F)/\tau_F$ across 50 random stochastic co-blockmodels under different sparsity regimes. Columns display the mean, standard deviation, and proportion of simulated datasets in which the nominal 95% confidence interval contains Ψ_F, averaged over all models with 400 trials per model.

		$n\rho_n \propto n^{3/4}$			$n\rho_n \propto n^{2/5}$			$n\rho_n \propto n^{1/4}$		
	n	mean	s.d.	coverage	mean	s.d.	coverage	mean	s.d.	coverage
$\widehat{\Psi}$	500	-1.50	0.99	66.7%	-7.80	1.01	0.0%	-20.39	1.37	0.0%
	2500	-1.05	0.99	81.7%	-9.09	0.99	0.0%	-31.28	1.17	0.0%
	10000	-0.75	1.00	88.0%	-10.31	0.99	0.0%	-43.76	1.05	0.0%
$\widetilde{\Psi}$	500	-0.16	1.04	93.9%	-1.85	1.21	53.4%	-11.25	2.00	0.0%
	2500	-0.05	1.01	94.8%	-1.07	1.10	78.1%	-11.23	1.48	0.0%
	10000	-0.02	1.01	94.7%	-0.68	1.06	87.6%	-11.09	1.29	0.0%

In the sparsest regime the corrected estimate no longer converges to the standard normal, but its mean remains steady at around -11.2 as n increases, and its standard deviation appears to converge to 1. This is predicted by the bias expansion: the removal of the leading term leaves a remainder of order $\sqrt{n}/(n\rho_n)^2$, which is constant when $n\rho_n \propto n^{1/4}$, suggesting another, smaller, bias obstruction. We hypothesise that removing further terms of the expansion could push valid inference into sparser regimes still, with the kth correction lowering the threshold to $n\rho_n \propto n^{1/(2k+2)}$.

7 How aligned are real-world networks?

We now calculate the effective shared dimension $\widehat{\Psi}$, its correction $\widetilde{\Psi}$ and its degree-free variant $\widehat{\Psi}^{\circ}$ for 10 real-world directed networks, obtained from SNAP (Leskovec & Krevl, 2014) (email-Eu-core, CollegeMsg, wiki-Vote, Epinions and HEP-Ph), igraphdata (Csárdi, 2026) (Florida Bay food web) and ConnectomeToolbox (OpenWorm, 2026) (*C. elegans*); the remaining networks come from the sources cited in Table 3.

Each network was binarised and restricted to its largest weakly connected component and then its $(1, 1)$-core, the largest subgraph in which every node has an in- and an out-edge (Giatsidis et al., 2011); Epinions was restricted to its $(5, 5)$-core, and HEP-Ph to papers citing, and cited by, at least five papers in the full dataset. We set d to one plus the profile-likelihood elbow (Zhu & Ghodsi, 2006) of the singular values after the first, which typically just reflects overall activity in the network.

Table 3: Effective shared dimension of 10 directed networks. Here r is the proportion of reciprocated links in the network, d the chosen dimension, and $\widehat{\Psi}$, $\widetilde{\Psi}$ and $\widehat{\Psi}^{\circ}$ the uncorrected, corrected and degree-free estimates.

network	source	category	n	r	d	$\widehat{\Psi}/d$	$\widetilde{\Psi}/d$	$\frac{\widehat{\Psi}^{\circ}-1}{d-1}$
email-Eu-core	Leskovec et al. (2007) Yin et al. (2017)	social	803	0.73	7	0.86	0.93	0.72
CollegeMsg	Panzarasa et al. (2009)	social	1302	0.68	2	0.74	0.81	0.51
Epinions	Richardson et al. (2003)	social	8182	0.46	6	0.58	0.64	0.59
wiki-Vote	Leskovec et al. (2010)	social	1300	0.15	5	0.28	0.31	0.39
political blogs	Adamic & Glance (2005)	information	813	0.29	2	0.39	0.40	0.56
HEP-Ph citations	Leskovec et al. (2005) Gehrke et al. (2003)	information	11962	0.01	4	0.20	0.23	0.43
OpenFlights routes	Patokallio (2014)	infrastructure	3362	0.98	4	1.00	1.04	0.97
US faculty hiring	Wapman et al. (2022)	infrastructure	363	0.53	4	0.55	0.60	0.29
C. elegans connectome	Cook et al. (2019)	biology	293	0.37	8	0.51	0.59	0.58
Florida Bay food web	Ulanowicz et al. (1998)	biology	103	0.04	4	0.27	0.29	0.25

The results are presented in Table 3. We observe that exchange networks (email, messaging and flight data) share the most, whereas attention and flow networks (citations, food webs, wiki votes and political blogs) share the least, with the other networks falling in between. This does not always hold for the degree-free estimate, however, which can either raise or lower the estimate. For example, it is much lower for the US faculty hiring dataset (0.29 against 0.60 for the standard estimate) suggesting that once we account for size, where an institution hires from is different from where its graduates go.

Applying the correction consistently increases the estimate. In particular, the corrected estimate for OpenFlights actually moves above 1, which highlights a potential limitation of our method: the assumption of independence between the two directions of edges between a given pair of nodes. We expect that this assumption could be weakened, at the cost of having to estimate the covariances between edge pairs.

Finally, and perhaps most importantly, *reciprocity alone does not explain the shared dimension*. Each network shares more than its reciprocity alone would suggest, which highlights the subtle difference between the two that was mentioned in the introduction: we ask whether a node sends to and receives from the same *kinds* of node, not whether it hears back from every node it talks to, which are two fundamentally different questions.

AI USE STATEMENT

In this work, we used generative AI tools for formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, implementing methods, cleaning and reformatting datasets and supporting qualitative and thematic data analysis.

We have not used generative AI tools for developing theoretical models or conceptual frameworks, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, or interpreting results. Generating synthetic data sets and assisting with translation are not applicable to this work.

Additionally, we used generative AI tools for creating scientific figures or images, analysing existing literature, brainstorming, sourcing/searching for information, and identifying relevant literature.

We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness, and any recommended proof strategies were worked through, verified and refined by the authors. Any papers suggested during a literature review were read by the authors to verify claims made. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The authors do not feel that this submission raises any questions regarding the Code of Ethics.

REPRODUCIBILITY STATEMENT

Proofs for all results claimed in the main body of text are given in Appendix A, with Appendices B and C giving explicit derivations of various quantities mentioned in the text (Ψ_F and the variance terms in the central limit theorems for stochastic co-blockmodels in the case of the former, and the quadratic bias term for the latter). These are present in the code used to produce the figures and tables in the paper, all of which can be found in the supplementary file in the form of Jupyter notebooks.

REFERENCES

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, 48(3):1452–1474, 2020.

Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, pp. 36–43, 2005.

Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(65):1981–2014, 2008.

Jesús Arroyo, Avanti Athreya, Joshua Cape, Guodong Chen, Carey E. Priebe, and Joshua T. Vogelstein. Inference for multiple heterogeneous networks with a common invariant subspace. *Journal of Machine Learning Research*, 22(142):1–49, 2021.

Avanti Athreya, Carey E. Priebe, Minh Tang, Vince Lyzinski, David J. Marchette, and Daniel L. Sussman. A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, 78 (1):1–18, 2016.

Avanti Athreya, Donniell E. Fishkind, Minh Tang, Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, Keith Levin, Vince Lyzinski, Yichen Qin, and Daniel L. Sussman. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.

Joshua Cape, Minh Tang, and Carey E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, 2019.

Steven J. Cook, Travis A. Jarrell, Christopher A. Brittin, Yi Wang, Adam E. Bloniarz, Maksim A. Yakovlev, Ken C. Q. Nguyen, Leo T.-H. Tang, Emily A. Bayer, Janet S. Duerr, Hannes E. Bülow, Oliver Hobert, David H. Hall, and Scott W. Emmons. Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature*, 571(7763):63–71, 2019.

Gábor Csárdi. igraphdata: A collection of network data sets for the 'igraph' package. `https://CRAN.R-project.org/package=igraphdata`, 2026. R package, accessed September 2026.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. SIMPLE: Statistical inference on membership profiles in large networks. *Journal of the Royal Statistical Society: Series B*, 84(2):630–653, 2022.

Alex Fornito, Andrew Zalesky, and Edward T. Bullmore. *Fundamentals of Brain Network Analysis*. Academic Press, San Diego, 2016.

Diego Garlaschelli and Maria I. Loffredo. Patterns of link reciprocity in directed networks. *Physical Review Letters*, 93(26):268701, 2004.

Johannes Gehrke, Paul Ginsparg, and Jon Kleinberg. Overview of the 2003 KDD cup. *ACM SIGKDD Explorations Newsletter*, 5(2):149–151, 2003.

Hamza Giaffar, Camille Rullán Buxó, and Mikio Aoi. The effective number of shared dimensions between paired datasets. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *PMLR*, pp. 4249–4257, 2024.

Christos Giatsidis, Dimitrios M. Thilikos, and Michalis Vazirgiannis. D-cores: Measuring collaboration of directed graphs based on degeneracy. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, ICDM '11, pp. 201–210, USA, 2011. IEEE Computer Society.

Jihun Hamm and Daniel D. Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 376–383, 2008.

Peter D. Hoff. Additive and multiplicative effects network models. *Statistical Science*, 36(1):34–50, 2021.

Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

Roger Horn and Charles Johnson. *Matrix Analysis (Second Edition)*. Cambridge University Press, New York, NY, 2012.

Tailen Hsing and Randall Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, Ltd, 2015.

Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

Megha Khosla, Jurek Leonhardt, Wolfgang Nejdl, and Avishek Anand. Node representation learning for directed graphs. In *Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2019)*, 2019.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *PMLR*, pp. 3519–3529, 2019.

W. J. Krzanowski. Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74(367):703–707, 1979. doi: 10.2307/2286995.

Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43:215–237, 2015.

Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, 2014.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pp. 177–187, 2005.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1):1–41, 2007.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pp. 1361–1370, 2010.

Bernardo Marenco, Paola Bermolen, Marcelo Fiori, Federico Larroca, and Gonzalo Mateos. A random dot product graph model for weighted and directed networks. In *58th Asilomar Conference on Signals, Systems, and Computers*, pp. 1155–1159, 2024.

OpenWorm. ConnectomeToolbox: Python tools for connectomics datasets of *C. elegans* and other organisms. `https://github.com/openworm/ConnectomeToolbox`, 2026. Python package `cect`, accessed September 2026.

Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. Asymmetric transitivity preserving graph embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1105–1114, 2016.

Pietro Panzarasa, Tore Opsahl, and Kathleen M. Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932, 2009.

Jani Patokallio. OpenFlights: Airport, airline and route data. `https://openflights.org/data`, 2014.

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. SVCCA: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *The Semantic Web – ISWC 2003*, pp. 351–368, 2003.

Paul Robert and Yves Escoufier. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 25(3):257–265, 1976.

Karl Rohe, Tai Qin, and Bin Yu. Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences*, 113(45):12679–12684, 2016.

Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society: Series B*, 84(4):1446–1473, 2022.

Daniel L. Sussman, Minh Tang, Donniell E. Fishkind, and Carey E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

Minh Tang and Carey E. Priebe. Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *The Annals of Statistics*, 46(5):2360–2415, 2018.

Robert E. Ulanowicz, Cristina Bondavalli, and Michael S. Egnotovich. Network analysis of trophic dynamics in South Florida ecosystem, FY 97: The Florida Bay ecosystem. Annual Report to the United States Geological Service Biological Resources Division [UMCES]CBL 98-123, Chesapeake Biological Laboratory, University of Maryland, 1998.

K. Hunter Wapman, Sam Zhang, Aaron Clauset, and Daniel B. Larremore. Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature*, 610(7930):120–127, 2022.

Dong Xia. Normal approximation and confidence region of singular subspaces. *Electronic Journal of Statistics*, 15(2):3798–3851, 2021.

Yuling Yan, Yuxin Chen, and Jianqing Fan. Inference for heteroskedastic PCA with missing data. *The Annals of Statistics*, 52(2):729–756, 2024.

Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 555–564, 2017.

Stephen J. Young and Edward R. Scheinerman. Random dot product graph models for social networks. In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-Graph*, pp. 138–149. Springer-Verlag, 2007.

Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, and Matthew Hirn. MagNet: A neural network for directed graphs. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Mu Zhu and Ali Ghodsi. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis*, 51(2):918–930, 2006.

A Proofs

A.1 Notation

Throughout these proofs, we will assume that $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n) \overset{\text{i.i.d.}}{\sim} F$ as in Definition 3, and we retain the notation

$$\Delta_x = \mathbb{E}[\mathbf{x}\mathbf{x}^\top], \quad \Delta_{xy} = \mathbb{E}[\mathbf{x}\mathbf{y}^\top], \quad \Delta_y = \mathbb{E}[\mathbf{y}\mathbf{y}^\top], \quad \mu_x = \mathbb{E}[\mathbf{x}], \quad \boldsymbol{\mu}_y = \mathbb{E}[\mathbf{y}]$$

where $(\mathbf{x}, \mathbf{y}) \sim F$.

Setting $\mathbf{X} = [\mathbf{x}_1|\cdots|\mathbf{x}_n]^\top$ and $\mathbf{Y} = [\mathbf{y}_1|\cdots|\mathbf{y}_n]^\top$, we define the underlying probability matrix $\mathbf{P} = \rho_n \mathbf{X}\mathbf{Y}^\top$, so that $\mathbf{P}_{ij} = \mathbb{E}[\mathbf{A}_{ij}]$ for any $i \neq j$.

We define the quantities

$$\Delta = (\Delta_x, \Delta_{xy}, \Delta_y), \quad \bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i\mathbf{x}_i^\top, \mathbf{x}_i\mathbf{y}_i^\top, \mathbf{y}_i\mathbf{y}_i^\top), \quad \bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \quad \bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i,$$

and note that by an application of Hoeffding's inequality each of the quantities $\|\bar{\Delta} - \Delta\|$, $\|\bar{\mathbf{x}} - \boldsymbol{\mu}_x\|$, and $\|\bar{\mathbf{y}} - \boldsymbol{\mu}_y\|$ are $\mathrm{O}_{\mathbb{P}}\big(\sqrt{\frac{\log n}{n}}\big)$, where for a triple $(\mathbf{B}, \mathbf{C}, \mathbf{D})$ of matrices and an arbitrary matrix norm $\|\cdot\|$ we define $\|(\mathbf{B}, \mathbf{C}, \mathbf{D})\| = \max\{\|\mathbf{B}\|, \|\mathbf{C}\|, \|\mathbf{D}\|\}$.

We also define

$$M := \sup_{(\mathbf{x},\mathbf{y})\in\mathrm{supp}(F)} \max\{\|\mathbf{x}\|, \|\mathbf{y}\|\} \quad \text{and} \quad \lambda = \min\{\lambda_{\min}(\Delta_x), \lambda_{\min}(\Delta_y)\},$$

where $\lambda_{\min}(\mathbf{B})$ denotes the smallest eigenvalue of the matrix $\mathbf{B}$, and we note that M is finite by compactness and λ is positive by Assumption $\mathbf{A1}'$.

Throughout the proof we shall use the following matrix norms:

- The operator norm $\|\mathbf{M}\|$: the largest singular value of $\mathbf{M}$.
- The Frobenius norm $\|\mathbf{M}\|_{\mathrm{Frob}} := \mathrm{tr}(\mathbf{M}\mathbf{M}^\top)^{1/2}$.
- The 2-to-∞ norm $\|\mathbf{M}\|_{2\to\infty}$: the maximal Euclidean norm of the rows of $\mathbf{M}$.
- The nuclear norm $\|\mathbf{M}\|_*$: the sum of the singular values of $\mathbf{M}$.

At various points in the proof we will define equivalent quantities for the adjacency matrix $\mathbf{A}$, the underlying probability matrix $\mathbf{P}$, and an unspecified matrix $\mathbf{M}$ (which will usually lie close to $\mathbf{P}$). We will apply the convention that the quantities use the same notation for the three matrices, marked with a $\wedge$, left plain, or marked with a $\sim$ for the three matrices respectively. Thus for example $\widehat{\sigma}_i$, σ_i and $\widetilde{\sigma}_i$ denote the ith singular value of $\mathbf{A}$, $\mathbf{P}$ and $\mathbf{M}$ respectively.

Finally, we shall always assume that a singular value decomposition is truncated at rank d, and denote with a $\perp$ the remaining terms from the full decomposition. Thus for example

$$\mathbf{A} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^\top + \widehat{\mathbf{U}}^\perp\widehat{\mathbf{\Sigma}}^\perp(\widehat{\mathbf{V}}^\perp)^\top$$

where $\widehat{\mathbf{U}}, \widehat{\mathbf{V}} \in \mathbb{R}^{n\times d}$ have orthonormal columns, as do $\widehat{\mathbf{U}}^\perp, \widehat{\mathbf{V}}^\perp \in \mathbb{R}^{n\times(n-d)}$, and $\widehat{\mathbf{U}}^\top\widehat{\mathbf{U}}^\perp = \mathbf{0}$. For an orthogonal projection $\mathbf{\Pi}_{\mathbf{U}}$, we let $\mathbf{\Pi}_{\mathbf{U}}^\perp = (\mathbf{I} - \mathbf{\Pi}_{\mathbf{U}}) = \mathbf{\Pi}_{\mathbf{U}^\perp}$.

A.2 Proofs of results in Sections 3 and 4

Proof of Proposition 1. We retain the notation of Section 2, so that $\mathcal{S}_{\mathrm{out}}$ is the linear span of the continuous functions $\{f_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$, and $\mathcal{S}_{\mathrm{in}}$ is the span of the functions $\{f_{\mathbf{y}} : \mathbf{y} \in \mathcal{Y}\}$, where

$$f_{\mathbf{x}} := f(\mathbf{x}, \cdot), \qquad f_{\mathbf{y}} := f(\cdot, \mathbf{y}).$$

Let $\eta_1, \ldots, \eta_d$ be a basis for $\mathcal{S}_{\mathrm{out}}$. Each η_k is a finite linear combination of the functions $f_{\mathbf{x}}$ and thus is clearly continuous, and we let $\eta = (\eta_1, \ldots, \eta_d)$. Now, for any $\mathbf{x} \in \mathcal{X}$, $f_{\mathbf{x}} \in \mathcal{S}_{\mathrm{out}}$ by definition, so $f_{\mathbf{x}} = \sum_{k=1}^{d}\xi_k(\mathbf{x})\eta_k$ for some function $\xi : \mathcal{X} \to \mathbb{R}^d$, and so

$$f(\mathbf{x}, \mathbf{y}) = \xi(\mathbf{x})^\top\eta(y)$$

for all $\mathbf{x} \in \mathcal{X}$, $y \in \mathcal{Y}$.

It remains to prove continuity of ξ. Since the η_k are linearly independent as functions, there must exist $\mathbf{y}_1, \ldots, \mathbf{y}_d \in \mathcal{Y}$ such that the matrix $\mathbf{M}$ with $\mathbf{M}_{ij} = \eta_i(\mathbf{y}_j)$ is invertible. Now, by definition $\big(f(\mathbf{x}, \mathbf{y}_1), \ldots, f(\mathbf{x}, \mathbf{y}_d)\big) = \mathbf{M}^\top\xi(\mathbf{x})$, and so

$$\xi(\mathbf{x}) = \mathbf{M}^{-\top}\big(f(\mathbf{x}, \mathbf{y}_1), \ldots, f(\mathbf{x}, \mathbf{y}_d)\big)^\top.$$

Since the right-hand side is a fixed linear map applied to a vector of continuous functions of $\mathbf{x}$, ξ is continuous as required. □

Proof of Proposition 2. In a mild abuse of notation, let $x_1, \dots, x_d, y_1, \dots, y_d \in L^2(F)$ denote the coordinate functions on $\mathcal{X} \times \mathcal{Y}$, so that $x_k\big((\mathbf{x}, \mathbf{y})\big)$ is the kth coordinate of $\mathbf{x}$, and let $\mathcal{S}_x$ and $\mathcal{S}_y$ be their spans. The sending profile of a node at $(\mathbf{x}, \mathbf{y})$ is

$$f_{\mathbf{x}}\big((\mathbf{x}', \mathbf{y}')\big) = \rho_n \mathbf{x}^\top \mathbf{y}' = \rho_n \sum_{k=1}^{d} x_k\big((\mathbf{x}, \mathbf{y})\big) y_k\big((\mathbf{x}', \mathbf{y}')\big),$$

a linear combination of the coordinate functions $y_1, \dots, y_d$, so $\mathcal{S}_{\text{out}} \subseteq \mathcal{S}_y$, while by definition of the inner product on $L^2(F)$ Δ_y is the Gram matrix of the y_i, and so the dimension of their span is equal to the rank of Δ_y, that is, $\dim(\mathcal{S}_y) = \operatorname{rank}(\Delta_y)$. Similarly we have $\mathcal{S}_{\text{in}} \subseteq \mathcal{S}_x$ and $\dim(\mathcal{S}_x) = \operatorname{rank}(\Delta_x)$. If $\mathcal{S}_{\text{out}}$ and $\mathcal{S}_{\text{in}}$ both have dimension d, then clearly Δ_x and Δ_y must have rank d. For the converse, if $\operatorname{rank}(\Delta_x) = \operatorname{rank}(\Delta_y) = d$ then by an identical line of reasoning to the previous proof we may express the coordinates $x_1, \dots, x_d$ and $y_1, \dots, y_d$ as linear combinations of elements of $\mathcal{S}_{\text{in}}$ and $\mathcal{S}_{\text{out}}$ respectively, so $\mathcal{S}_y \subseteq \mathcal{S}_{\text{out}}$ and $\mathcal{S}_x \subseteq \mathcal{S}_{\text{in}}$, thus proving the first equivalence.

The second equivalence follows directly from the observation that

$$d_{\text{out}}(\mathbf{x}) = \int \rho_n \mathbf{x}^\top \mathbf{y} \, \mathrm{d}F_y(\mathbf{y}) = \rho_n \mathbf{x}^\top \boldsymbol{\mu}_y,$$

and likewise for d_{in}. □

Proof of Proposition 3.

(i) The principal angles between two subspaces $\mathcal{U}$ and $\mathcal{V}$ are defined by $\cos(\theta_k) = \sigma_k$, where σ_k is the kth singular value of the matrix $\mathbf{M}$ of inner products $\mathbf{M}_{ij} := \langle u_i, v_j \rangle$ for orthonormal bases $\{u_i\}$ of $\mathcal{U}$ and $\{v_j\}$ of $\mathcal{V}$. The scaled coordinate functions $\{\Delta_x^{-1/2} x_i\}$ and $\{\Delta_y^{-1/2} y_j\}$ form such bases for $\mathcal{S}_x$ and $\mathcal{S}_y$ respectively, and the corresponding matrix of inner products is precisely $\Delta_x^{-1/2} \Delta_{xy} \Delta_y^{-1/2}$, so the first identity follows from the fact that the squared Frobenius norm of a matrix is equal to the sum of its squared singular values. The second follows from the fact that $\|\mathbf{M}\|_{\text{Frob}}^2 = \operatorname{tr}(\mathbf{M}\mathbf{M}^\top)$ for any matrix $\mathbf{M}$ and the cyclic property of the trace.

(ii) $\Delta_{x^\circ} = \mathbb{E}[\mathbf{x}\mathbf{x}^\top / (\mathbf{x}^\top \boldsymbol{\mu}_y)^2] \succeq M^{-4} \Delta_x$, where M bounds $\|\mathbf{x}\|, \|\mathbf{y}\|$ on the compact support of F, so that $(\mathbf{x}^\top \boldsymbol{\mu}_y)^2 \leq M^4$, and thus Δ_{x° (and likewise Δ_{y°) is of maximal rank. The result follows by applying the argument of part (i) to $\mathcal{S}_{x^\circ}$ and $\mathcal{S}_{y^\circ}$.

□

A.3 Proof of Theorem 4

For simplicity, we will assume throughout that the matrix $\mathbf{A}$ is not hollow, and so has diagonal entries $\mathbf{A}_{ii} \sim \text{Bernoulli}(\mathbf{P}_{ii})$. Proposition 15 will show that this choice does not affect the validity of our claims.

We begin with a technical result:

Proposition 6. *Let $\mathbf{z}_1, \mathbf{z}_2, \dots$ be random elements of $\mathbb{R}^m$ with $\|\mathbf{z}_n\| = \mathrm{O}_{\mathbb{P}}(a_n)$ for some sequence $a_n \to 0$, and let $\mathcal{B}$ be an open ball around a point $\mathbf{z} \in \mathbb{R}^m$. Then:*

(i) *$\mathbf{z} + \mathbf{z}_n \in \mathcal{B}$ with overwhelming probability.*

(ii) *If $h : \mathcal{B} \to \mathbb{R}$ is twice continuously differentiable on $\mathcal{B}$ then*

$$|h(\mathbf{z} + \mathbf{z}_n) - h(\mathbf{z})| = \mathrm{O}_{\mathbb{P}}(a_n), \qquad h(\mathbf{z} + \mathbf{z}_n) - h(\mathbf{z}) - \nabla h(\mathbf{z}) \mathbf{z}_n = \mathrm{O}_{\mathbb{P}}(a_n^2).$$

Proof.

(i) This follows automatically from the definition of overwhelming probability.

(ii) If $\mathcal{B}$ has radius r, then derivatives ∇h and $\nabla^2 h$ (both being continuous) have bounded norm on the closed ball of radius $r/2$ around $\mathbf{z}$, which is contained within $\mathcal{B}$, and so the result follows by applying Taylor's theorem.

□

Proposition 7. *Let* $\Psi := \mathrm{tr}(\mathbf{\Pi_X \Pi_Y}) = \mathrm{tr}(\mathbf{\Pi_U \Pi_V})$ *where* $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times d}$ *are the matrices of left and right singular vectors of* $\mathbf{P}$*, and for a triple* $(\mathbf{B}, \mathbf{C}, \mathbf{D})$ *of* $d \times d$ *matrices with* $\mathbf{B}$ *and* $\mathbf{D}$ *invertible, write*

$$h(\mathbf{B}, \mathbf{C}, \mathbf{D}) = \mathrm{tr}\big(\mathbf{B}^{-1}\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^\top\big).$$

Then under Assumption ***A1′****,* $\Psi_F = h(\Delta)$*, and with overwhelming probability* $\Psi = h(\bar{\Delta})$*.*

Proof. The first identity is Proposition 3(i). For the second, since $\|\bar{\Delta} - \Delta\| = \mathrm{O}_{\mathbb{P}}(\sqrt{\log n/n})$ and Δ_x, Δ_y are invertible, $\bar{\Delta}_x = \frac{1}{n}\mathbf{X}^\top\mathbf{X}$ and $\bar{\Delta}_y = \frac{1}{n}\mathbf{Y}^\top\mathbf{Y}$ are invertible with overwhelming probability; on that event $\mathbf{X}$ and $\mathbf{Y}$ have full column rank and so

$$\mathbf{\Pi_X} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top, \qquad \mathbf{\Pi_Y} = \mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top.$$

By the cyclic property of the trace

$$\Psi = \mathrm{tr}\big((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{X}\big) = h\Big(\tfrac{1}{n}\mathbf{X}^\top\mathbf{X},\ \tfrac{1}{n}\mathbf{X}^\top\mathbf{Y},\ \tfrac{1}{n}\mathbf{Y}^\top\mathbf{Y}\Big) = h(\bar{\Delta}),$$

as required. □

Proposition 8. *Let* $\tau_F^2 = \mathrm{Var}\big(g(\mathbf{x}, \mathbf{y})\big)$ *where*

$$g(\mathbf{x}, \mathbf{y}) = 2\mathbf{x}^\top\mathbf{T}\mathbf{y} - \mathbf{x}^\top\mathbf{T}\Delta_{yx}\Delta_x^{-1}\mathbf{x} - \mathbf{y}^\top\mathbf{T}^\top\Delta_{xy}\Delta_y^{-1}\mathbf{y}, \qquad \mathbf{T} = \Delta_x^{-1}\Delta_{xy}\Delta_y^{-1}.$$

Then under Assumption ***A1′****:*

(i) $|\Psi - \Psi_F| = \mathrm{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right)$.

(ii) $\Psi - \Psi_F = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}_i, \mathbf{y}_i) + R_n$ *with* $R_n = \mathrm{O}_{\mathbb{P}}\left(\frac{\log n}{n}\right)$.

(iii) $n^{1/2}\,(\Psi - \Psi_F) \to \mathcal{N}(0, \tau_F^2)$.

Proof.

(i) Since the matrices Δ_x and Δ_y are invertible, there is a suitably small open ball $\mathcal{B}$ around Δ on which any triple $(\mathbf{B}, \mathbf{C}, \mathbf{D})$ has $\mathbf{B}$ and $\mathbf{D}$ invertible, and thus h, being a polynomial in the entries of $\mathbf{B}^{-1}$, $\mathbf{C}$ and $\mathbf{D}^{-1}$ is smooth. Since $\|\bar{\Delta} - \Delta\| = \mathrm{O}_{\mathbb{P}}(\sqrt{\log n/n})$, $\bar{\Delta} \in \mathcal{B}$ with overwhelming probability, and so applying Proposition 6 with $\mathbf{z}_n = \bar{\Delta} - \Delta$ and $a_n = \sqrt{\log n/n}$ gives the result.

(ii) The partial derivatives of h are

$$\begin{aligned}
\partial_{\mathbf{B}} h &= -\mathbf{B}^{-\top}\mathbf{C}\mathbf{D}^{-\top}\mathbf{C}^\top\mathbf{B}^{-\top} \\
\partial_{\mathbf{C}} h &= \mathbf{B}^{-1}\mathbf{C}\mathbf{D}^{-1} + \mathbf{B}^{-\top}\mathbf{C}\mathbf{D}^{-\top} \\
\partial_{\mathbf{D}} h &= -\mathbf{D}^{-\top}\mathbf{C}^\top\mathbf{B}^{-\top}\mathbf{C}\mathbf{D}^{-\top}
\end{aligned}$$

and evaluating these at Δ gives

$$\partial_{\mathbf{B}} h(\Delta) = -\mathbf{T}\Delta_{yx}\Delta_x^{-1}, \quad \partial_{\mathbf{C}} h(\Delta) = 2\mathbf{T}, \quad \partial_{\mathbf{D}} h(\Delta) = -\mathbf{T}^\top\Delta_{xy}\Delta_y^{-1}$$

using symmetry of Δ_x and Δ_y. The function h is homogeneous of degree 0, and so by Euler's theorem $\langle \nabla h(\mathbf{M}), \mathbf{M}\rangle = 0$ for any triple $\mathbf{M}$, where we use the matrix inner product $\langle \mathbf{M}, \mathbf{N}\rangle = \mathrm{tr}(\mathbf{M}^\top\mathbf{N})$ extended to triples. Thus for any $i = 1, \ldots, n$ we have

$$\nabla h(\Delta) \cdot [(\mathbf{x}_i\mathbf{x}_i^\top, \mathbf{x}_i\mathbf{y}_i^\top, \mathbf{y}_i\mathbf{y}_i^\top) - \Delta] = g(\mathbf{x}_i, \mathbf{y}_i)$$

and so

$$\nabla h(\Delta) \cdot (\bar{\Delta} - \Delta) = \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{x}_i, \mathbf{y}_i).$$

Thus

$$\Psi - \Psi_F = \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{x}_i, \mathbf{y}_i) + \left[h(\bar{\Delta}) - h(\Delta) - \nabla h(\Delta) \cdot (\bar{\Delta} - \Delta) \right]$$

where the latter term is $\mathrm{O}_{\mathbb{P}}(\log n/n)$ by Proposition 6 with $\mathbf{z}_n = \bar{\Delta} - \Delta$ and $a_n = \sqrt{\log n/n}$ as before.

(iii) Note that by linearity we have $\mathbb{E}[g(\mathbf{x}, \mathbf{y})] = \nabla h(\Delta) \cdot [(\mathbb{E}[\mathbf{x}\mathbf{x}^\top], \mathbb{E}[\mathbf{x}\mathbf{y}^\top], \mathbb{E}[\mathbf{y}\mathbf{y}^\top]) - \Delta] = 0$, and that the function g is uniformly bounded by compactness. Thus, applying the central limit theorem to the bounded, mean-zero i.i.d. random variables $g(\mathbf{x}_i, \mathbf{y}_i)$ and using Slutsky's theorem in combination with the fact that $n^{1/2} R_n \to 0$ gives the desired result.

□

Proposition 9. *Under Assumptions **A1′** and **A3** the following hold:*

(i) $\|\mathbf{A} - \mathbf{P}\| = \mathrm{O}_{\mathbb{P}}(\sqrt{n\rho_n})$.

(ii) $\sigma_d^{-1} = \mathrm{O}_{\mathbb{P}}\big((n\rho_n)^{-1}\big)$ *and* $\sigma_1 = \mathrm{O}_{\mathbb{P}}(n\rho_n)$, *hence* $\kappa := \sigma_1/\sigma_d = OP(1)$.

(iii) $\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\} = \mathrm{O}_{\mathbb{P}}(\sqrt{d/n})$.

Proof.

(i) We apply Theorem 5.2 of Lei & Rinaldo (2015) to the Hermitian dilation $\mathcal{H}(\mathbf{A}) = \left(\begin{smallmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{smallmatrix}\right)$, viewed as the adjacency matrix of a graph on $2n$ nodes with edge probability matrix $\mathcal{H}(\mathbf{P}) = \left(\begin{smallmatrix} \mathbf{0} & \mathbf{P} \\ \mathbf{P}^\top & \mathbf{0} \end{smallmatrix}\right)$. Since $2n \max\{\mathbf{P}_{ij}\} \le 2n\rho_n$ and $2n\rho_n \ge \log n$ for n sufficiently large, this result tells us that $\|\mathbf{A} - \mathbf{P}\| = \|\mathcal{H}(\mathbf{A} - \mathbf{P})\| = \mathrm{O}_{\mathbb{P}}(\sqrt{n\rho_n})$, as required.

(ii) Note that the non-zero eigenvalues of $\mathbf{P}^\top\mathbf{P} = \rho_n^2 \mathbf{Y}\mathbf{X}^\top\mathbf{X}\mathbf{Y}^\top$ coincide with those of $n^2\rho_n^2 \bar{\Delta}_x^{1/2} \bar{\Delta}_y \bar{\Delta}_x^{1/2}$, and that $\bar{\Delta}$ lies in some fixed open ball around Δ with overwhelming probability, and are consequently bounded above and below, which yields the first three claims.

(iii) $\mathbf{U} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1/2}\mathbf{W}$ for some orthogonal matrix $\mathbf{W}$, so (by Cape et al. (2019), Proposition 6.5) $\|\mathbf{U}\|_{2\to\infty} \le \|\mathbf{X}\|_{2\to\infty} \|(\mathbf{X}^\top\mathbf{X})^{-1/2}\| = \mathrm{O}_{\mathbb{P}}(\sqrt{d/n})$, and likewise for $\mathbf{V}$.

□

Note: For a constant $C > 0$, denote by $\mathcal{E}_C$ the intersection of the events

$$\mathcal{E}_C = \{\sigma_d \ge n\rho_n/C\} \cap \{\sigma_1 \le Cn\rho_n\} \cap \left\{ \max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\} \le C\sqrt{d/n} \right\}.$$

Proposition 9(ii) and (iii) say that for any $\gamma > 0$ there exists C_γ such that $\mathcal{E}_{C_\gamma}$ holds with probability $1 - O(n^{-\gamma})$. Note that for a sequence of random variables X_n and a deterministic sequence a_n, if for any $\gamma > 0$ we have $\mathbb{P}(|X_n| > K a_n | \mathbf{X}, \mathbf{Y}) = O(n^{-\gamma})$ for all $(\mathbf{X}, \mathbf{Y})$ satisfying $\mathcal{E}_{C_\gamma}$ for an appropriately chosen C_γ, then $X_n = \mathrm{O}_{\mathbb{P}}(a_n)$.

For a matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, set

$$f(\mathbf{M}) := \mathrm{tr}\left(\mathbf{\Pi}_{\widetilde{\mathbf{U}}} \mathbf{\Pi}_{\widetilde{\mathbf{V}}}\right)$$

so that $\widehat{\Psi} = f(\mathbf{A})$ and $\Psi = f(\mathbf{P})$, and let $\mathcal{B} = \{\mathbf{M} : \|\mathbf{M} - \mathbf{P}\| \le \sigma_d/4\}$. The following result says that if $\mathbf{M}$ is close enough to $\mathbf{P}$, then there is a gap between the top-d singular values of $\mathbf{M}$ and the rest.

Proposition 10. *For* $\mathbf{M} \in \mathcal{B}$,

$$\widetilde{\sigma}_d \geq \tfrac{3}{4}\sigma_d, \qquad \widetilde{\sigma}_{d+1} \leq \tfrac{1}{4}\sigma_d.$$

Proof. A corollary of Weyl's inequalities (see, for example, Horn & Johnson (2012), Corollary 7.3.5) states that $|\widetilde{\sigma}_i - \sigma_i| \leq \|\mathbf{M} - \mathbf{P}\|$, and so the result follows from the fact that $\mathbf{P}$ is of rank d and the definition of $\mathcal{B}$. □

The next result gives a first-order expansion for the projection $\mathbf{\Pi}_{\widetilde{\mathbf{U}}}$ for matrices $\mathbf{M} \in \mathcal{B}$:

Proposition 11. *For* $\mathbf{M} \in \mathcal{B}$*:*

$$\mathbf{\Pi}_{\widetilde{\mathbf{U}}} - \mathbf{\Pi}_{\mathbf{U}} = \mathbf{G}_{\widetilde{\mathbf{U}}} + \mathbf{R}_{\widetilde{\mathbf{U}}}$$

where

$$\mathbf{G}_{\widetilde{\mathbf{U}}} = \mathbf{\Pi}_{\mathbf{U}}^{\perp}(\mathbf{M} - \mathbf{P})\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\top} + \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}(\mathbf{M} - \mathbf{P})^{\top}\mathbf{\Pi}_{\mathbf{U}}^{\perp},$$

and

$$\|\mathbf{R}_{\widetilde{\mathbf{U}}}\| \leq \frac{C\|\mathbf{M} - \mathbf{P}\|^2}{\sigma_d^2}, \qquad \|\mathbf{\Pi}_{\widetilde{\mathbf{U}}} - \mathbf{\Pi}_{\mathbf{U}}\| \leq \frac{C\|\mathbf{M} - \mathbf{P}\|}{\sigma_d},$$

with absolute constants C*. The same holds for* $\mathbf{V}$ *with the roles of* $\mathbf{U}$ *and* $\mathbf{V}$ *interchanged and* $\mathbf{M} - \mathbf{P}$ *replaced with* $(\mathbf{M} - \mathbf{P})^{\top}$*.*

Proof. This follows directly from Theorem 1 of Xia (2021), with the Hermitian dilations $\mathcal{H}(\mathbf{P})$ and $\mathcal{H}(\mathbf{M} - \mathbf{P})$ of $\mathbf{P}$ and $\mathbf{M} - \mathbf{P}$ fulfilling the roles of A and X in their notation. □

The next result uses Proposition 11 to give an expansion for f around $\mathbf{P}$:

Proposition 12. *For* $\mathbf{M} \in \mathcal{B}$*:*

$$f(\mathbf{M}) - f(\mathbf{P}) = 2\langle \mathbf{M} - \mathbf{P}, \mathbf{G} \rangle + r(\mathbf{M})$$

where

$$\mathbf{G} = \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{\Pi}_{\mathbf{V}}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top} + \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}\mathbf{\Pi}_{\mathbf{U}}\mathbf{\Pi}_{\mathbf{V}}^{\perp} \quad \text{and} \quad |r(\mathbf{M})| \leq \frac{Cd\|\mathbf{M} - \mathbf{P}\|^2}{\sigma_d^2}$$

and $|f(\mathbf{M}) - f(\mathbf{P})| \leq Cd\,\|\mathbf{M} - \mathbf{P}\|/\sigma_d$ *for some absolute constant* $C > 0$*.*

Proof. Write $\boldsymbol{\delta}_{\widetilde{\mathbf{U}}} = \mathbf{\Pi}_{\widetilde{\mathbf{U}}} - \mathbf{\Pi}_{\mathbf{U}}$ and $\boldsymbol{\delta}_{\widetilde{\mathbf{V}}} = \mathbf{\Pi}_{\widetilde{\mathbf{V}}} - \mathbf{\Pi}_{\mathbf{V}}$. Since each of the projectors are of rank d, $\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}$ and $\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}$ are of rank at most $2d$, and we can write

$$f(\mathbf{M}) - f(\mathbf{P}) = \operatorname{tr}\big(\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{\Pi}_{\mathbf{V}}\big) + \operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{U}}\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}\big) + \operatorname{tr}\big(\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}\big).$$

By Proposition 11, $\boldsymbol{\delta}_{\widetilde{\mathbf{U}}} = \mathbf{G}_{\widetilde{\mathbf{U}}} + \mathbf{R}_{\widetilde{\mathbf{U}}}$ with $\|\mathbf{R}_{\widetilde{\mathbf{U}}}\| \leq C\|(\mathbf{M} - \mathbf{P})\|^2/\sigma_d^2$ and $\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\| \leq C\|(\mathbf{M} - \mathbf{P})\|/\sigma_d$, and similarly for $\mathbf{V}$.

For any matrices $\mathbf{B}, \mathbf{C}$ we have $|\operatorname{tr}(\mathbf{BC})| \leq \operatorname{rank}(\mathbf{B})\|\mathbf{B}\|\|\mathbf{C}\|$, and so it follows that

$$|\operatorname{tr}\big(\mathbf{R}_{\widetilde{\mathbf{U}}}\mathbf{\Pi}_{\mathbf{V}}\big)| \leq d\|\mathbf{R}_{\widetilde{\mathbf{U}}}\| \leq \frac{Cd\|\mathbf{M} - \mathbf{P}\|^2}{\sigma_d^2},$$

and similarly for $|\operatorname{tr}\big(\mathbf{\Pi}_{\mathbf{U}}\mathbf{R}_{\widetilde{\mathbf{V}}}|$, while

$$|\operatorname{tr}\big(\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}\big)| \leq 2d\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\|\|\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}\| \leq \frac{2C^2d\|\mathbf{M} - \mathbf{P}\|^2}{\sigma_d^2}.$$

and so we may write

$$f(\mathbf{M}) - f(\mathbf{P}) = \operatorname{tr}\big(\mathbf{G}_{\widetilde{\mathbf{U}}}\mathbf{\Pi}_{\mathbf{V}} + \mathbf{\Pi}_{\mathbf{U}}\mathbf{G}_{\widetilde{\mathbf{V}}}\big) + r(\mathbf{M}) \text{ with } |r(\mathbf{M})| \leq \frac{2C(C+1)d\|\mathbf{M} - \mathbf{P}\|^2}{\sigma_d^2}.$$

Expanding out $\mathbf{G}_{\widetilde{\mathbf{U}}}$ and $\mathbf{G}_{\widetilde{\mathbf{V}}}$ using their definitions from Proposition 11 and using the definition $\langle \mathbf{B}, \mathbf{C} \rangle = \operatorname{tr}(\mathbf{B}^{\top}\mathbf{C})$ for the matrix inner product and then shows us that

$$\operatorname{tr}\big(\mathbf{G}_{\widetilde{\mathbf{U}}}\mathbf{\Pi}_{\mathbf{V}} + \mathbf{\Pi}_{\mathbf{U}}\mathbf{G}_{\widetilde{\mathbf{V}}}\big) = 2\langle \mathbf{M} - \mathbf{P}, \mathbf{G} \rangle$$

as required. Finally, using similar reasoning to before,

$$|f(\mathbf{M}) - f(\mathbf{P})| \leq d\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\| + d\|\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}\| + 2d\max\{\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\|, \|\boldsymbol{\delta}_{\widetilde{\mathbf{V}}}\|\} \leq \frac{4Cd\|\mathbf{M} - \mathbf{P}\|}{\sigma_d},$$

completing the proof. □

Proposition 13. *On the event $\mathcal{E}_C$, the matrix $\mathbf{G}$ from Proposition 12 satisfies*

$$\|\mathbf{G}\|_{\mathrm{Frob}}^2 \le \frac{4d}{\sigma_d^2} \quad \text{and} \quad \max_{i,j} |\mathbf{G}_{ij}| \le \frac{Cd}{n\sigma_d}.$$

Proof. Expanding out $\mathbf{G}\mathbf{G}^\top$ gives us four matrix products, each of the form $\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^\top\mathbf{B}$ or $\mathbf{B}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^\top$ for a rank d matrix $\mathbf{B}$. Since $\|\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^\top\| = \sigma_d^{-1}$, the Frobenius norm bound follows from the definition $\|\mathbf{G}\|_{\mathrm{Frob}}^2 = \mathrm{tr}(\mathbf{G}\mathbf{G}^\top)$. For the entries, write

$$\mathbf{G} = \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{V}(\mathbf{V}^\top\mathbf{V}\mathbf{\Sigma}^{-1})\mathbf{V}^\top + \mathbf{U}(\mathbf{\Sigma}^{-1}\mathbf{V}^\top\mathbf{U})\mathbf{U}^\top\mathbf{\Pi}_{\mathbf{V}}^{\perp}$$

so we see that each entry is a product of a row of $\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{V}$ or of $\mathbf{U}$ which is of norm at most $2C\sqrt{d/n}$ on the event $\mathcal{E}_C$, a $d \times d$ matrix of norm at most σ_d^{-1}, and a row of $\mathbf{V}$ or of $\mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{U}$, again of norm at most $2C\sqrt{d/n}$ on the event $\mathcal{E}_C$, and the result follows. □

Corollary 14. *Under Assumptions **A1′** and **A3**:*

$$\widehat{\Psi} - \Psi = 2\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle + r(\mathbf{A})$$

where

$$\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle = \mathrm{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n^2\rho_n}}\right), \quad |r(\mathbf{A})| = \mathrm{O}_{\mathbb{P}}\big((n\rho_n)^{-1}\big).$$

Proof. By Proposition 9 $\|\mathbf{A} - \mathbf{P}\|/\sigma_d = \mathrm{O}_{\mathbb{P}}((n\rho_n)^{-1/2})$ and so $\|\mathbf{A} - \mathbf{P}\| \le \sigma_d/8$ with overwhelming probability, and thus $\mathbf{A} \in \mathcal{B}$. Proposition 12 with $\mathbf{M} = \mathbf{A}$ then gives $\widehat{\Psi} - \Psi = 2\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle + r(\mathbf{A})$ with $|r(\mathbf{A})| \le Cd\|\mathbf{A} - \mathbf{P}\|^2/\sigma_d^2$, and so $|r(\mathbf{A})| = \mathrm{O}_{\mathbb{P}}\big((n\rho_n)^{-1}\big)$.

Conditional on $\mathbf{X}$ and $\mathbf{Y}$ the matrix $\mathbf{G}$ is deterministic and $\mathbf{A} - \mathbf{P}$ has mean zero, so $\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle$ has conditional mean zero and conditional variance $\sum_{ij} \mathbf{P}_{ij}(1 - \mathbf{P}_{ij})\mathbf{G}_{ij}^2 \le \rho_n\|\mathbf{G}\|_{\mathrm{Frob}}^2$.

Working on the event $\mathcal{E}_C$ and conditioning on the positions $\mathbf{X}$ and $\mathbf{Y}$, the summands $(\mathbf{A}_{ij} - \mathbf{P}_{ij})\mathbf{G}_{ij}$ are independent and mean-zero, with total variance at most $v = \rho_n\|\mathbf{G}\|_{\mathrm{Frob}}^2 \le 4d\rho_n/\sigma_d^2 \le 4C^2d/(n^2\rho_n)$ and each bounded by $b = \max_{ij}|\mathbf{G}_{ij}| \le Cd/(n\sigma_d) \le C^2d/(n^2\rho_n)$, by Proposition 13 and the fact that $\sigma_d \ge n\rho_n/C$. Applying Bernstein's inequality then shows that

$$\mathbb{P}\big(|\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle| > t \mid \mathbf{X}, \mathbf{Y}\big) \le 2\exp\left\{\frac{-t^2}{(2v + 2bt/3)}\right\},$$

and setting $t = K(\sqrt{v\log n} + b\log n)$ the right-hand side is at most $2n^{-c\min(K,K^2)}$ for an absolute $c > 0$. Since $\log n \le n^2\rho_n$, $\sqrt{v\log n} + b\log n \le C\sqrt{\log n/(n^2\rho_n)}$, and the bound for $\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle$ follows. □

Proof of Theorem 4. Write $\widehat{\Psi} - \Psi_F = (\Psi - \Psi_F) + (\widehat{\Psi} - \Psi)$. By Proposition 8, the first term is $\mathrm{O}_{\mathbb{P}}(\sqrt{\log n/n})$, with the Gaussian limit of part (iii). By Corollary 14 the second is $2\langle \mathbf{A} - \mathbf{P}, \mathbf{G}\rangle + r(\mathbf{A}) = \mathrm{O}_{\mathbb{P}}\big(\sqrt{\log n/(n^2\rho_n)} + (n\rho_n)^{-1}\big)$, and

$$\sqrt{\frac{\log n}{n^2\rho_n}} = \sqrt{\frac{\log n}{n}} \cdot \frac{1}{\sqrt{n\rho_n}} \le \min\left\{\sqrt{\frac{\log n}{n}},\ \frac{\log n}{n} + \frac{1}{n\rho_n}\right\},$$

the second bound by the inequality of arithmetic and geometric means. This proves (i). For (ii), $n^{1/2}(\widehat{\Psi} - \Psi) = \mathrm{O}_{\mathbb{P}}\big(\sqrt{\log n/(n\rho_n)} + n^{1/2}/(n\rho_n)\big)$, which tends to zero in probability when $n\rho_n = \omega(n^{1/2})$, and Slutsky's theorem completes the proof. □

Finally, we justify our choice to consider a non-hollow matrix with the following result:

Proposition 15. *Let $\mathbf{A}'$ denote the matrix obtained by removing the diagonal of $\mathbf{A}$, and let $\widehat{\Psi}'$ be the resulting estimate. Then, under Assumptions **A1′** and **A3***

$$|\widehat{\Psi}' - \widehat{\Psi}| = \mathrm{O}_{\mathbb{P}}\big((n\rho_n)^{-1}\big),$$

and consequently the results of Theorem 4 hold for $\mathbf{A}'$.

Proof. Note that $\|\mathbf{A}' - \mathbf{P}\| \le \|\mathbf{A} - \mathbf{P}\| + 1$, and consequently both $\mathbf{A}'$ and $\mathbf{A}$ lie in $\mathcal{B}$ with overwhelming probability. We can therefore apply Proposition 12 to both, from which we see that

$$\widehat{\Psi}' - \widehat{\Psi} = -2\langle \mathrm{diag}(\mathbf{A}), \mathbf{G}\rangle + (r(\mathbf{A}') - r(\mathbf{A})),$$

where

$$|r(\mathbf{A})| \le \frac{Cd\|\mathbf{A} - \mathbf{P}\|^2}{\sigma_d^2}, \quad |r(\mathbf{A}')| \le \frac{Cd(\|\mathbf{A} - \mathbf{P}\| + 1)^2}{\sigma_d^2}.$$

For the first term, Hölder's inequality for the operator and nuclear norms gives $|\langle \mathrm{diag}(\mathbf{A}), \mathbf{G}\rangle| \le \|\mathrm{diag}(\mathbf{A})\|\,\|\mathbf{G}\|_* \le 1 \cdot 2d/\sigma_d$, since each of the two terms of $\mathbf{G}$ has rank at most d and operator norm at most σ_d^{-1}. Working on the event $\mathcal{E}_C$ and conditioning on $\mathbf{X}$ and $\mathbf{Y}$ gives the bound on $|\widehat{\Psi}' - \widehat{\Psi}|$. Since $\widehat{\Psi}' - \Psi_F = (\widehat{\Psi}' - \widehat{\Psi}) + (\widehat{\Psi} - \Psi_F)$, part (i) of Theorem 4 holds automatically and part (ii) holds by applying Slutsky's theorem and noting that $n^{1/2}|\widehat{\Psi}' - \widehat{\Psi}| \to 0$ when $n\rho_n = \omega(n^{1/2})$. □

A.4 Proof of Theorem 5

We retain the notation of the proof of Theorem 4, and again assume that the matrix $\mathbf{A}$ is not hollow. In addition, we let $c_0 > 0$ be the lower bound from Assumption **A2′**, so that

$$c_0 < \mathbf{x}^\top \boldsymbol{\mu}_y,\ \boldsymbol{\mu}_x^\top \mathbf{y} \le M^2,$$

where we recall that

$$M := \sup_{(\mathbf{x},\mathbf{y}) \in \mathrm{supp}(F)} \max\{\|\mathbf{x}\|, \|\mathbf{y}\|\}.$$

Let $\mathbf{D}_{\mathrm{out}}$ and $\mathbf{D}_{\mathrm{in}}$ be the diagonal matrices whose entries are the row and column sums of $\mathbf{P}$ respectively, so that $(\mathbf{D}_{\mathrm{out}})_{ii} = n\rho_n \mathbf{x}_i^\top \bar{\mathbf{y}}$ and $(\mathbf{D}_{\mathrm{in}})_{jj} = n\rho_n \bar{\mathbf{x}}^\top \mathbf{y}_j$, and define $\widehat{\mathbf{D}}_{\mathrm{out}}, \widehat{\mathbf{D}}_{\mathrm{in}}$ and $\widetilde{\mathbf{D}}_{\mathrm{out}}, \widetilde{\mathbf{D}}_{\mathrm{in}}$ similarly for $\mathbf{A}$ and an unspecified matrix $\mathbf{M}$.

For a matrix $\mathbf{M}$ with positive row and column sums, let

$$f^\circ(\mathbf{M}) = \mathrm{tr}\big(\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}} \boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{in}}^{-1}\widetilde{\mathbf{V}}}\big).$$

and define $\Psi^\circ := f^\circ(\mathbf{P})$.

Finally, write

$$\mathbf{x}_i^\bullet = \frac{\mathbf{x}_i}{\mathbf{x}_i^\top \bar{\mathbf{y}}}, \qquad \mathbf{y}_i^\bullet = \frac{\mathbf{y}_i}{\bar{\mathbf{x}}^\top \mathbf{y}_i}$$

for the positions normalised by the *sample* means, and $\bar{\Delta}^\bullet = \frac{1}{n}\sum_i (\mathbf{x}_i^\bullet \mathbf{x}_i^{\bullet\top}, \mathbf{x}_i^\bullet \mathbf{y}_i^{\bullet\top}, \mathbf{y}_i^\bullet \mathbf{y}_i^{\bullet\top})$.

We are now in a position to begin the proof. Let $h(\mathbf{B}, \mathbf{C}, \mathbf{D}) = \mathrm{tr}(\mathbf{B}^{-1}\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^\top)$ as in Proposition 7.

Proposition 16. *$\Psi_F^\circ = h(\Delta^\circ)$, and with overwhelming probability $\Psi^\circ = h(\bar{\Delta}^\bullet)$.*

Proof. The first identity is Proposition 3(ii). For the second, since $\|\bar{\mathbf{y}} - \boldsymbol{\mu}_y\| = \mathrm{O}_{\mathbb{P}}(\sqrt{\log n/n})$ it follows that $\mathbf{x}_i^\top \bar{\mathbf{y}} \ge c_0/2$ for all i with overwhelming probability, and likewise for $\bar{\mathbf{x}}^\top \mathbf{y}_j$, so the row and column sums of $\mathbf{P}$ are positive and $\mathbf{X}^\bullet, \mathbf{Y}^\bullet$, the matrices with rows $\mathbf{x}_i^{\bullet\top}$, $\mathbf{y}_i^{\bullet\top}$, are defined. By definition we have $\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{X}$ has the same column space as $\mathbf{X}^\bullet$, and so in particular $\Psi^\circ = f^\circ(\mathbf{P}) = \mathrm{tr}(\boldsymbol{\Pi}_{\mathbf{X}^\bullet}\boldsymbol{\Pi}_{\mathbf{Y}^\bullet})$, and

$$\Psi^\circ = \mathrm{tr}\big(\mathbf{X}^\bullet(\mathbf{X}^{\bullet\top}\mathbf{X}^\bullet)^{-1}\mathbf{X}^{\bullet\top}\mathbf{Y}^\bullet(\mathbf{Y}^{\bullet\top}\mathbf{Y}^\bullet)^{-1}\mathbf{Y}^{\bullet\top}\big) = h\Big(\tfrac{1}{n}\mathbf{X}^{\bullet\top}\mathbf{X}^\bullet, \tfrac{1}{n}\mathbf{X}^{\bullet\top}\mathbf{Y}^\bullet, \tfrac{1}{n}\mathbf{Y}^{\bullet\top}\mathbf{Y}^\bullet\Big) = h(\bar{\Delta}^\bullet),$$

as required. □

Proposition 17. *Assume Assumptions **A1′**, **A2′** and **A3**, write $\mathbf{x}^\circ = \mathbf{x}/(\mathbf{x}^\top\boldsymbol{\mu}_y)$, $\mathbf{y}^\circ = \mathbf{y}/(\boldsymbol{\mu}_x^\top\mathbf{y})$ for the degree-normalised positions, $\Delta^\circ = (\Delta_{x^\circ}, \Delta_{x^\circ y^\circ}, \Delta_{y^\circ})$ for their second moment matrices and $\mathbf{T}^\circ = \Delta_{x^\circ}^{-1}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1}$, and let $\tau_F^{\circ 2} = \mathrm{Var}\big(g^\circ(\mathbf{x}, \mathbf{y})\big)$, where*

$$g^\circ(\mathbf{x}, \mathbf{y}) = \tfrac{1}{2}\big(q_x + q_y\big)(\mathbf{x}^\circ, \mathbf{y}^\circ) - \mathbb{E}\big[q_y(\mathbf{x}^\circ, \mathbf{y}^\circ)\,\mathbf{y}^\circ\big]^\top(\mathbf{x} - \boldsymbol{\mu}_x) - \mathbb{E}\big[q_x(\mathbf{x}^\circ, \mathbf{y}^\circ)\,\mathbf{x}^\circ\big]^\top(\mathbf{y} - \boldsymbol{\mu}_y)$$

and

$$q_x(\mathbf{u}, \mathbf{v}) = 2\mathbf{u}^\top\mathbf{T}^\circ\mathbf{v} - 2\mathbf{u}^\top\mathbf{T}^\circ\Delta_{y^\circ x^\circ}\Delta_{x^\circ}^{-1}\mathbf{u}, \qquad q_y(\mathbf{u}, \mathbf{v}) = 2\mathbf{u}^\top\mathbf{T}^\circ\mathbf{v} - 2\mathbf{v}^\top\mathbf{T}^{\circ\top}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1}\mathbf{v}.$$

Then:

(i) $|\Psi^\circ - \Psi_F^\circ| = \mathrm{O}_\mathbb{P}\left(\sqrt{\frac{\log n}{n}}\right)$.

(ii) $\Psi^\circ - \Psi_F^\circ = \frac{1}{n}\sum_{i=1}^n g^\circ(\mathbf{x}_i, \mathbf{y}_i) + R_n$ *with* $R_n = \mathrm{O}_\mathbb{P}\left(\frac{\log n}{n}\right)$.

(iii) $n^{1/2}\left(\Psi^\circ - \Psi_F^\circ\right) \to \mathcal{N}(0, \tau_F^{\circ 2})$.

Proof.

(i) Let $\mathcal{B} = \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^d \times \mathbb{R}^d : \|\mathbf{a} - \boldsymbol{\mu}_x\|, \|\mathbf{b} - \boldsymbol{\mu}_y\| \le c_0/(2M)\}$, and note that for any $(\mathbf{a}, \mathbf{b}) \in \mathcal{B}$ we have $\mathbf{x}^\top \mathbf{b} \ge c_0/2$ and $\mathbf{a}^\top \mathbf{y} \ge c_0/2$ for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{supp}(F)$. Define

$$\Theta(\mathbf{a}, \mathbf{b}) = \frac{1}{n}\sum_{i=1}^n \left(\frac{\mathbf{x}_i \mathbf{x}_i^\top}{(\mathbf{x}_i^\top \mathbf{b})^2},\ \frac{\mathbf{x}_i \mathbf{y}_i^\top}{(\mathbf{x}_i^\top \mathbf{b})(\mathbf{a}^\top \mathbf{y}_i)},\ \frac{\mathbf{y}_i \mathbf{y}_i^\top}{(\mathbf{a}^\top \mathbf{y}_i)^2}\right), \qquad \theta(\mathbf{a}, \mathbf{b}) = \mathbb{E}[\Theta(\mathbf{a}, \mathbf{b})],$$

on $\mathcal{B}$ so that $\bar{\Delta}^\bullet = \Theta(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ and $\Delta^\circ = \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$ (noting that $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y) \in \mathcal{B}$ trivially and $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \mathcal{B}$ with overwhelming probability since $\|\bar{\mathbf{x}} - \boldsymbol{\mu}_x\|, \|\bar{\mathbf{y}} - \boldsymbol{\mu}_y\| = \mathrm{O}_\mathbb{P}(\sqrt{\log n/n})$. Each summand of Θ is a smooth function of $(\mathbf{a}, \mathbf{b})$ on this ball, with all derivatives up to order two bounded by a constant that only depends on M and c_0, and so by expanding out the Taylor series of Θ around $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$ we see that:

$$\Theta(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - \Theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y) = \partial_a \Theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)[\bar{\mathbf{x}} - \boldsymbol{\mu}_x] + \partial_b \Theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)[\bar{\mathbf{y}} - \boldsymbol{\mu}_y] + \mathrm{O}_\mathbb{P}\left(\frac{\log n}{n}\right).$$

Noting that $\Theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y) - \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$, $\partial_a \Theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y) - \partial_a \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$ and $\partial_b \Theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y) - \partial_b \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$ are averages of zero-mean, bounded, i.i.d. terms, and thus $\mathrm{O}_\mathbb{P}(\sqrt{\log n/n})$ by Hoeffding's inequality, we find that

$$\bar{\Delta}^\bullet - \Delta^\circ = \frac{1}{n}\sum_{i=1}^n \zeta(\mathbf{x}_i, \mathbf{y}_i) + \mathrm{O}_\mathbb{P}\left(\frac{\log n}{n}\right),$$

where

$$\zeta(\mathbf{x}, \mathbf{y}) = \left(\mathbf{x}^\circ \mathbf{x}^{\circ\top}, \mathbf{x}^\circ \mathbf{y}^{\circ\top}, \mathbf{y}^\circ \mathbf{y}^{\circ\top}\right) - \Delta^\circ + \partial_a \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)[\mathbf{x} - \boldsymbol{\mu}_x] + \partial_b \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)[\mathbf{y} - \boldsymbol{\mu}_y]$$

is a triple of zero-mean, bounded, i.i.d. random matrices. In particular, $\|\bar{\Delta}^\bullet - \Delta^\circ\| = \mathrm{O}_\mathbb{P}(\sqrt{\log n/n})$, and the result follows from Proposition 16 and Proposition 6 applied to h on a ball around Δ° on which its first two arguments are invertible, exactly as in Proposition 8(i).

(ii) Applying Proposition 6 once more,

$$\begin{aligned}\Psi^\circ - \Psi_F^\circ = h(\bar{\Delta}^\bullet) - h(\Delta^\circ) &= \nabla h(\Delta^\circ) \cdot (\bar{\Delta}^\bullet - \Delta^\circ) + \mathrm{O}_\mathbb{P}\left(\frac{\log n}{n}\right) \\ &= \frac{1}{n}\sum_{i=1}^n \nabla h(\Delta^\circ) \cdot \zeta(\mathbf{x}_i, \mathbf{y}_i) + \mathrm{O}_\mathbb{P}\left(\frac{\log n}{n}\right),\end{aligned}$$

and so it remains to show that $g^\circ(\mathbf{x}, \mathbf{y}) = \nabla h(\Delta^\circ) \cdot \zeta(\mathbf{x}, \mathbf{y})$. By the computation in the proof of Proposition 8(ii) with Δ° in place of Δ,

$$\nabla h(\Delta^\circ) = (-\mathbf{T}^\circ \Delta_{y^\circ x^\circ} \Delta_{x^\circ}^{-1},\ 2\mathbf{T}^\circ,\ -\mathbf{T}^{\circ\top} \Delta_{x^\circ y^\circ} \Delta_{y^\circ}^{-1}),$$

and Euler's theorem gives

$$\nabla h(\Delta^\circ) \cdot [(\mathbf{x}^\circ \mathbf{x}^{\circ\top}, \mathbf{x}^\circ \mathbf{y}^{\circ\top}, \mathbf{y}^\circ \mathbf{y}^{\circ\top}) - \Delta^\circ] = \frac{1}{2}(q_x + q_y)(\mathbf{x}^\circ, \mathbf{y}^\circ).$$

For the remaining terms, differentiating $\mathbf{x}/(\mathbf{x}^\top \mathbf{b})$ at $\mathbf{b} = \boldsymbol{\mu}_y$ in the direction $\mathbf{b}'$ gives $-\mathbf{x}^\circ(\mathbf{x}^{\circ\top}\mathbf{b}')$, so

$$\partial_a \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)[\mathbf{a}'] = -\mathbb{E}\left[(\mathbf{a}'^\top \mathbf{y}^\circ)\left(0,\ \mathbf{x}^\circ \mathbf{y}^{\circ\top},\ 2\mathbf{y}^\circ \mathbf{y}^{\circ\top}\right)\right]$$

and

$$\partial_b \theta(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)[\mathbf{b}'] = -\mathbb{E}\big[(\mathbf{x}^{\circ\top}\mathbf{b}')\big(2\mathbf{x}^\circ\mathbf{x}^{\circ\top},\ \mathbf{x}^\circ\mathbf{y}^{\circ\top},\ 0\big)\big]$$

and pairing with $\nabla h(\Delta^\circ)$, using the identities

$$\langle \partial_{\mathbf{B}} h, \mathbf{x}^\circ\mathbf{x}^{\circ\top}\rangle = -\mathbf{x}^{\circ\top}\mathbf{T}^\circ\Delta_{y^\circ x^\circ}\Delta_{x^\circ}^{-1}\mathbf{x}^\circ$$
$$\langle \partial_{\mathbf{C}} h, \mathbf{x}^\circ\mathbf{y}^{\circ\top}\rangle = 2\mathbf{x}^{\circ\top}\mathbf{T}^\circ\mathbf{y}^\circ$$
$$\langle \partial_{\mathbf{D}} h, \mathbf{y}^\circ\mathbf{y}^{\circ\top}\rangle = -\mathbf{y}^{\circ\top}\mathbf{T}^{\circ\top}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1}\mathbf{y}^\circ$$

shows us that

$$\nabla h(\Delta^\circ)\cdot\partial_a\theta(\boldsymbol{\mu}_x,\boldsymbol{\mu}_y)[\mathbf{a}'] = -\mathbb{E}\big[q_y(\mathbf{x}^\circ,\mathbf{y}^\circ)\,\mathbf{y}^\circ\big]^\top\mathbf{a}'$$

and

$$\nabla h(\Delta^\circ)\cdot\partial_b\theta(\boldsymbol{\mu}_x,\boldsymbol{\mu}_y)[\mathbf{b}'] = -\mathbb{E}\big[q_x(\mathbf{x}^\circ,\mathbf{y}^\circ)\,\mathbf{x}^\circ\big]^\top\mathbf{b}'$$

which gives us g°.

(iii) This follows from (ii) in the same manner as in Proposition 8(iii).

□

For a matrix $\mathbf{M}$ with positive row and column sums, let $\varepsilon_i^{\text{out}}$ be the ith row sum of $\mathbf{M}-\mathbf{P}$ divided by the ith row sum of $\mathbf{P}$ (that is, the relative error of the ith row sum of $\mathbf{M}$ as an estimate of that of $\mathbf{P}$), and let $\varepsilon_j^{\text{in}}$ be the same for the jth column. Define matrices

$$\boldsymbol{\Lambda}_{\text{out}} = \operatorname{diag}(\varepsilon^{\text{out}}), \qquad \boldsymbol{\Lambda}_{\text{in}} = \operatorname{diag}(\varepsilon^{\text{in}}),$$

so that $\widetilde{\mathbf{D}}_{\text{out}} = \mathbf{D}_{\text{out}}(\mathbf{I}+\boldsymbol{\Lambda}_{\text{out}})$ and $\widetilde{\mathbf{D}}_{\text{in}} = \mathbf{D}_{\text{in}}(\mathbf{I}+\boldsymbol{\Lambda}_{\text{in}})$. Finally let

$$\kappa_{\mathbf{D}} = \max\left\{\frac{\max_i(\mathbf{D}_{\text{out}})_{ii}}{\min_i(\mathbf{D}_{\text{out}})_{ii}},\ \frac{\max_j(\mathbf{D}_{\text{in}})_{jj}}{\min_j(\mathbf{D}_{\text{in}})_{jj}}\right\}$$

be the larger of the two ratios of largest to smallest expected degree.

Proposition 18. *Under Assumptions **A1′**, **A2′** and **A3**:*

(i) *With overwhelming probability all entries of $\mathbf{D}_{\text{out}}$ and $\mathbf{D}_{\text{in}}$ lie in $[c_0 n\rho_n/2,\ M^2 n\rho_n]$, so in particular $\kappa_{\mathbf{D}} \le 2M^2/c_0$.*

(ii) *For $\mathbf{M}=\mathbf{A}$, $\max_i|\varepsilon_i^{\text{out}}| + \max_j|\varepsilon_j^{\text{in}}| = \mathrm{O}_{\mathbb{P}}\big(\sqrt{\log n/(n\rho_n)}\big)$ and $\|\varepsilon^{\text{out}}\|^2 + \|\varepsilon^{\text{in}}\|^2 = \mathrm{O}_{\mathbb{P}}(\rho_n^{-1})$.*

(iii) *$\|\boldsymbol{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\|_{2\to\infty} \le \kappa_{\mathbf{D}}\|\mathbf{U}\|_{2\to\infty}$ and $\|\boldsymbol{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\|_{2\to\infty} \le \kappa_{\mathbf{D}}\|\mathbf{V}\|_{2\to\infty}$, so both are $\mathrm{O}_{\mathbb{P}}(\sqrt{d/n})$.*

Proof.

(i) $(\mathbf{D}_{\text{out}})_{ii} = n\rho_n\mathbf{x}_i^\top\bar{\mathbf{y}}$ lies between $n\rho_n(c_0 - M\|\bar{\mathbf{y}}-\boldsymbol{\mu}_y\|)$ and $n\rho_n M^2$, and $\|\bar{\mathbf{y}}-\boldsymbol{\mu}_y\| \le c_0/(2M)$ with overwhelming probability.

(ii) Conditional on the positions, $\sum_j(\mathbf{A}_{ij}-\mathbf{P}_{ij})$ is a sum of n independent zero-mean random variables bounded above by 1 with total variance $v_i = \sum_j \mathbf{P}_{ij}(1-\mathbf{P}_{ij}) \le n\rho_n$. Applying Bernstein's inequality gives $\mathbb{P}(|\sum_j(\mathbf{A}_{ij}-\mathbf{P}_{ij})| > t \mid \mathbf{X},\mathbf{Y}) \le 2\exp\{-t^2/(2v_i+2t/3)\}$, and with $t = K\sqrt{n\rho_n\log n}$, using $n\rho_n \ge \log n$, we find that the right side is at most $2n^{-cK}$. Taking a union bound over i and dividing through by $\sum_j \mathbf{P}_{ij} \ge c_0 n\rho_n/2$ gives us the first claim. For the second, let $Z_i = (\sum_j(\mathbf{A}_{ij}-\mathbf{P}_{ij}))^2$, so that, conditional on the latent positions, the Z_i are independent, with $\mathbb{E}[Z_i] = v_i \le n\rho_n$ and $\mathbb{E}[Z_i^2] \le v_i + 3v_i^2 \le 4(n\rho_n)^2$. On the event of the first claim, $Z_i \le K^2 n\rho_n\log n$ for all i, and so Bernstein's inequality for the truncated variables gives, with overwhelming probability,

$$\sum_i Z_i \le n\cdot n\rho_n + K'\big(\sqrt{n\cdot 4(n\rho_n)^2\log n} + K^2 n\rho_n\log^2 n\big) \le Cn\cdot n\rho_n,$$

since $\sqrt{n\log n} \le n$ and $\log^2 n \le n$. Dividing through by $(c_0 n\rho_n/2)^2$ gives $\|\varepsilon^{\text{out}}\|^2 \le C/\rho_n$, and an identical argument gives us the bound for $\|\varepsilon^{\text{in}}\|$.

(iii) Since $\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}$ is symmetric, we have

$$\|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\|_{2\to\infty} = \max_{i\in\{1,\dots,n\}} \|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\|$$

where $\mathbf{e}_i$ is the ith standard basis vector of $\mathbb{R}^n$. Writing $\mathbf{B} = \mathbf{D}_{\text{out}}^{-1}\mathbf{U}$, we have

$$\|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\|^2 = \mathbf{e}_i^\top \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{e}_i \le \sigma_{\min}(\mathbf{B})^{-2}\|\mathbf{B}^\top\mathbf{e}_i\|^2,$$

where $\sigma_{\min}(\mathbf{B}) \ge \sigma_{\min}(\mathbf{D}_{\text{out}}^{-1}) = 1/\max_i(\mathbf{D}_{\text{out}})_{ii}$ and $\|\mathbf{B}^\top\mathbf{e}_i\| = \|\mathbf{U}^\top\mathbf{e}_i\|/(\mathbf{D}_{\text{out}})_{ii}$. An identical argument holds for $\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}$ and so the result follows from Proposition 9(iii).

□

Before expanding the function f°, we require another technical result, concerning the orthogonal projection onto the column space of a perturbed matrix:

Proposition 19. *Let $\mathbf{B}_0 \in \mathbb{R}^{n\times d}$ have full column rank, let $\mathbf{E} \in \mathbb{R}^{n\times d}$, and suppose $\mathbf{K} = \mathbf{E}\mathbf{B}_0^+$ satisfies $\|\mathbf{K}\| \le \frac{1}{2}$, where $\mathbf{B}_0^+ = (\mathbf{B}_0^\top\mathbf{B}_0)^{-1}\mathbf{B}_0^\top$. Then $\mathbf{B} = \mathbf{B}_0 + \mathbf{E}$ has full column rank, and with $\mathbf{\Pi} = \mathbf{\Pi}_{\mathbf{B}_0}$,*

$$\|\mathbf{\Pi}_{\mathbf{B}} - \mathbf{\Pi}\| \le 2\|\mathbf{K}\|, \qquad \big\|\mathbf{\Pi}_{\mathbf{B}} - \mathbf{\Pi} - \mathbf{\Pi}^\perp\mathbf{E}\mathbf{B}_0^+ - (\mathbf{\Pi}^\perp\mathbf{E}\mathbf{B}_0^+)^\top\big\| \le 20\|\mathbf{K}\|^2.$$

Proof. Since $\mathbf{B}_0^+\mathbf{B}_0 = \mathbf{I}_d$, $\mathbf{B} = (\mathbf{I}+\mathbf{K})\mathbf{B}_0$ where $\mathbf{W} = \mathbf{I}+\mathbf{K}$ is invertible with $\|\mathbf{W}^{-1}\| \le 2$, so in particular the column space of $\mathbf{B}$ is the column space of $\mathbf{B}_0$ transformed by $\mathbf{W}$, and $\mathbf{B}$ has full column rank. Now,

$$\mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}} = \mathbf{\Pi}^\perp\mathbf{B}\mathbf{B}^+ = \mathbf{\Pi}^\perp\mathbf{E}\mathbf{B}^+ = \mathbf{\Pi}^\perp\mathbf{K}\mathbf{B}_0\mathbf{B}^+$$

and

$$\mathbf{B}_0\mathbf{B}^+ = \mathbf{W}^{-1}\mathbf{B}\mathbf{B}^+ = \mathbf{W}^{-1}\mathbf{\Pi}_{\mathbf{B}},$$

so $\|\mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}}\| \le 2\|\mathbf{K}\|$, and for two projectors of the same rank

$$\|\mathbf{\Pi}_{\mathbf{B}} - \mathbf{\Pi}\| = \|\mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}}\| = \|\mathbf{\Pi}\mathbf{\Pi}_{\mathbf{B}}^\perp\|.$$

Next, decompose

$$\mathbf{\Pi}_{\mathbf{B}} - \mathbf{\Pi} = \mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}}\mathbf{\Pi} + \mathbf{\Pi}\mathbf{\Pi}_{\mathbf{B}}\mathbf{\Pi}^\perp + \mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}}\mathbf{\Pi}^\perp - \mathbf{\Pi}\mathbf{\Pi}_{\mathbf{B}}^\perp\mathbf{\Pi}.$$

Since projections are symmetric and idempotent, the last two terms are $(\mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}})(\mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}})^\top$ and $-(\mathbf{\Pi}\mathbf{\Pi}_{\mathbf{B}}^\perp)(\mathbf{\Pi}\mathbf{\Pi}_{\mathbf{B}}^\perp)^\top$, of norm at most $4\|\mathbf{K}\|^2$ each. For the first, $\mathbf{\Pi}^\perp\mathbf{\Pi}_{\mathbf{B}}\mathbf{\Pi} = \mathbf{\Pi}^\perp\mathbf{K}\mathbf{W}^{-1}\mathbf{\Pi}_{\mathbf{B}}\mathbf{\Pi}$ by the identity above, while $\mathbf{\Pi}^\perp\mathbf{E}\mathbf{B}_0^+ = \mathbf{\Pi}^\perp\mathbf{K}\mathbf{B}_0\mathbf{B}_0^+ = \mathbf{\Pi}^\perp\mathbf{K}\mathbf{\Pi}$, so their difference is $\mathbf{\Pi}^\perp\mathbf{K}\big(\mathbf{W}^{-1}(\mathbf{\Pi}_{\mathbf{B}} - \mathbf{\Pi})\mathbf{\Pi} + (\mathbf{W}^{-1} - \mathbf{I})\mathbf{\Pi}\big)$, which is of norm at most $\|\mathbf{K}\|(2\cdot 2\|\mathbf{K}\| + 2\|\mathbf{K}\|) = 6\|\mathbf{K}\|^2$. The second term is the transpose of the first, and so the result follows by summing over all four terms. □

We are now in a position to expand f°:

Proposition 20. *Let $\mathbf{M} \in \mathcal{B}$ have positive row and column sums. If $\max_i|\varepsilon_i^{\text{out}}|, \max_j|\varepsilon_j^{\text{in}}| \le \frac{1}{2}$ and*

$$\kappa_{\mathbf{D}}\left(\frac{C\|\mathbf{M}-\mathbf{P}\|}{\sigma_d} + \kappa_{\mathbf{D}}\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}\big(\|\varepsilon^{\text{out}}\|^2 + \|\varepsilon^{\text{in}}\|^2\big)^{1/2}\right) \le \frac{1}{4},$$

with C the constant of Proposition 11, then

$$f^\circ(\mathbf{M}) - f^\circ(\mathbf{P}) = 2\langle \mathbf{M}-\mathbf{P}, \mathbf{G}^\circ\rangle + r^\circ(\mathbf{M})$$

for some absolute constant C', where

$$|r^\circ(\mathbf{M})| \le C' d\,\kappa_{\mathbf{D}}^2\Big(\frac{C\|\mathbf{M}-\mathbf{P}\|}{\sigma_d} + \kappa_{\mathbf{D}}\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}\big(\|\varepsilon^{\text{out}}\|^2 + \|\varepsilon^{\text{in}}\|^2\big)^{1/2}\Big)^2$$

where

$$\mathbf{G}^\circ = \mathbf{G}_{\mathbf{U}}^\circ + \mathbf{G}_{\mathbf{V}}^\circ - \mathbf{c}\mathbf{1}^\top - \mathbf{1}\mathbf{c}'^\top,$$

$$\mathbf{G}_{\mathbf{U}}^\circ = \mathbf{\Pi}_{\mathbf{U}}^\perp\mathbf{D}_{\text{out}}^{-1}\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}^\perp\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\big((\mathbf{D}_{\text{out}}^{-1}\mathbf{U})^+\big)^\top\mathbf{\Sigma}^{-1}\mathbf{V}^\top, \qquad \mathbf{c}_i = \frac{(\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}^\perp)_{ii}}{(\mathbf{P}\mathbf{1})_i},$$

$$\mathbf{G}_{\mathbf{V}}^\circ = \mathbf{U}\mathbf{\Sigma}^{-1}(\mathbf{D}_{\text{in}}^{-1}\mathbf{V})^+\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}^\perp\mathbf{D}_{\text{in}}^{-1}\mathbf{\Pi}_{\mathbf{V}}^\perp, \qquad \mathbf{c}'_j = \frac{(\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}^\perp)_{jj}}{(\mathbf{P}^\top\mathbf{1})_j}.$$

Proof. By Proposition 11 and the hypothesis, $\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\| \le C\|\mathbf{M}-\mathbf{P}\|/\sigma_d \le \frac{1}{4}$, where we recall that $\boldsymbol{\delta}_{\widetilde{\mathbf{U}}} = \boldsymbol{\Pi}_{\widetilde{\mathbf{U}}} - \boldsymbol{\Pi}_{\mathbf{U}}$. The matrix $\boldsymbol{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{U} = \mathbf{U} + \boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{U}$ therefore has full column rank, with the same column space as $\widetilde{\mathbf{U}}$, so $\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}}$ is the projector onto the column space of

$$\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\boldsymbol{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{U} = \mathbf{D}_{\mathrm{out}}^{-1}(\mathbf{I}+\boldsymbol{\Lambda}_{\mathrm{out}})^{-1}(\mathbf{U}+\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{U}) = \mathbf{B}_0 + \mathbf{E}$$

where

$$\mathbf{B}_0 = \mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}, \quad \mathbf{E} = \mathbf{D}_{\mathrm{out}}^{-1}(\mathbf{I}+\boldsymbol{\Lambda}_{\mathrm{out}})^{-1}(\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{U} - \boldsymbol{\Lambda}_{\mathrm{out}}\mathbf{U}).$$

Since $(\mathbf{I}+\boldsymbol{\Lambda}_{\mathrm{out}})^{-1} = \mathbf{I} - \boldsymbol{\Lambda}_{\mathrm{out}}(\mathbf{I}+\boldsymbol{\Lambda}_{\mathrm{out}})^{-1}$, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ with

$$\mathbf{E}_1 = \mathbf{D}_{\mathrm{out}}^{-1}\big(\boldsymbol{\Pi}_{\mathbf{U}}^{\perp}(\mathbf{M}-\mathbf{P})\mathbf{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Lambda}_{\mathrm{out}}\mathbf{U}\big)$$
$$\mathbf{E}_2 = \mathbf{D}_{\mathrm{out}}^{-1}\mathbf{R}_{\widetilde{\mathbf{U}}}\mathbf{U} - \mathbf{D}_{\mathrm{out}}^{-1}\boldsymbol{\Lambda}_{\mathrm{out}}(\mathbf{I}+\boldsymbol{\Lambda}_{\mathrm{out}})^{-1}\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{U} + \mathbf{D}_{\mathrm{out}}^{-1}\boldsymbol{\Lambda}_{\mathrm{out}}^2(\mathbf{I}+\boldsymbol{\Lambda}_{\mathrm{out}})^{-1}\mathbf{U},$$

where we have used the fact that $\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{U} = \mathbf{G}_{\widetilde{\mathbf{U}}}\mathbf{U} + \mathbf{R}_{\widetilde{\mathbf{U}}}\mathbf{U}$ and $\mathbf{G}_{\widetilde{\mathbf{U}}}\mathbf{U} = \boldsymbol{\Pi}_{\widetilde{\mathbf{U}}}^{\perp}(\mathbf{M}-\mathbf{P})\mathbf{V}\boldsymbol{\Sigma}^{-1}$. Write $D_{\max}, D_{\min}$ for the largest and smallest entries of $\mathbf{D}_{\mathrm{out}}$, so $D_{\max}/D_{\min} \le \kappa_{\mathbf{D}}$, $\|\mathbf{B}_0^{+}\| = \sigma_{\min}(\mathbf{B}_0)^{-1} \le D_{\max}$ and $\|\mathbf{D}_{\mathrm{out}}^{-1}\| = D_{\min}^{-1}$. Also

$$\|\boldsymbol{\Lambda}_{\mathrm{out}}\mathbf{U}\| \le \|\boldsymbol{\Lambda}_{\mathrm{out}}\mathbf{U}\|_{\mathrm{Frob}} = \Big(\sum_i (\varepsilon_i^{\mathrm{out}})^2\|\mathbf{U}^{\top}\mathbf{e}_i\|^2\Big)^{1/2} \le \|\mathbf{U}\|_{2\to\infty}\|\varepsilon^{\mathrm{out}}\|,$$

so that, with $\|\boldsymbol{\Lambda}_{\mathrm{out}}\| \le \frac{1}{2}$ and the hypothesis,

$$\|\mathbf{K}\| = \|\mathbf{E}\mathbf{B}_0^{+}\| \le D_{\min}^{-1}\cdot 2\big(\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\| + \|\boldsymbol{\Lambda}_{\mathrm{out}}\mathbf{U}\|\big)\cdot D_{\max} \le 2\kappa_{\mathbf{D}}\Big(\frac{C\|\mathbf{M}-\mathbf{P}\|}{\sigma_d} + \|\mathbf{U}\|_{2\to\infty}\|\varepsilon^{\mathrm{out}}\|\Big) \le \tfrac{1}{2},$$

and Proposition 19 applies, thus

$$\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}} = \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}} + \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{E}\mathbf{B}_0^{+} + (\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{E}\mathbf{B}_0^{+})^{\top} + \mathbf{R}_{\mathbf{U}}^{\circ}$$

with $\|\mathbf{R}_{\mathbf{U}}^{\circ}\| \le 20\|\mathbf{K}\|^2$ and $\|\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}} - \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}\| \le 2\|\mathbf{K}\|$. The same argument holds for $\mathbf{V}$ with $\mathbf{B}_0' = \mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}$, $\mathbf{E}'$ and $\mathbf{K}'$.

As in Proposition 12,

$$f^{\circ}(\mathbf{M}) - f^{\circ}(\mathbf{P}) = \mathrm{tr}\big((\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}} - \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}})\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\big) + \mathrm{tr}\big(\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}(\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{in}}^{-1}\widetilde{\mathbf{V}}} - \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}})\big)$$
$$+ \mathrm{tr}\big((\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}} - \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}})(\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{in}}^{-1}\widetilde{\mathbf{V}}} - \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}})\big).$$

Since $|\mathrm{tr}(\mathbf{B}\mathbf{C})| \le \mathrm{rank}(\mathbf{B})\|\mathbf{B}\|\|\mathbf{C}\|$, the last term is at most $2d\cdot 2\|\mathbf{K}\|\cdot 2\|\mathbf{K}'\|$. For the first term, we have

$$\mathrm{tr}\big((\boldsymbol{\Pi}_{\widetilde{\mathbf{D}}_{\mathrm{out}}^{-1}\widetilde{\mathbf{U}}} - \boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}})\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\big) = 2\mathrm{tr}\big(\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{E}\mathbf{B}_0^{+}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\big) + \mathrm{tr}\big(\mathbf{R}_{\mathbf{U}}^{\circ}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\big)$$

with the contribution of $\mathbf{R}_{\mathbf{U}}^{\circ}$ being at most

$$20d\|\mathbf{K}\|^2 \le 20d^2\kappa_{\mathbf{D}}^2\Big(\frac{C\|\mathbf{M}-\mathbf{P}\|}{\sigma_d} + \|\mathbf{U}\|_{2\to\infty}\|\varepsilon^{\mathrm{out}}\|\Big)^2,$$

with a similar argument applying for $\mathbf{V}$.

It therefore remains to bound the term $2\,\mathrm{tr}(\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{E}\mathbf{B}_0^{+}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}})$ and its $\mathbf{V}$-counterpart. Split $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ as before, and note that since $\mathbf{U}\mathbf{B}_0^{+} = \mathbf{D}_{\mathrm{out}}\mathbf{B}_0\mathbf{B}_0^{+} = \mathbf{D}_{\mathrm{out}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}$ we have

$$2\,\mathrm{tr}(\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{E}_1\mathbf{B}_0^{+}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}) = 2\,\mathrm{tr}\big((\mathbf{M}-\mathbf{P})\,\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{B}_0^{+}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{D}_{\mathrm{out}}^{-1}\boldsymbol{\Pi}_{\mathbf{U}}^{\perp}\big)$$
$$- 2\,\mathrm{tr}\big(\boldsymbol{\Lambda}_{\mathrm{out}}\,\mathbf{D}_{\mathrm{out}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp}\mathbf{D}_{\mathrm{out}}^{-1}\big)$$
$$= 2\langle\mathbf{M}-\mathbf{P}, \mathbf{G}_{\mathbf{U}}^{\circ}\rangle - 2\sum_{i=1}^{n}\varepsilon_i^{\mathrm{out}}(\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp})_{ii},$$

where we obtain the final sum by noting that conjugating by $\mathbf{D}_{\mathrm{out}}$ does not affect the trace. Since $\varepsilon_i^{\mathrm{out}}(\mathbf{P}\mathbf{1})_i = ((\mathbf{M}-\mathbf{P})\mathbf{1})_i$, we find that

$$2\sum_{i=1}^{n}\varepsilon_i^{\mathrm{out}}(\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{in}}^{-1}\mathbf{V}}\boldsymbol{\Pi}_{\mathbf{D}_{\mathrm{out}}^{-1}\mathbf{U}}^{\perp})_{ii} = 2\sum_i((\mathbf{M}-\mathbf{P})\mathbf{1})_i c_i = 2\langle\mathbf{M}-\mathbf{P}, \mathbf{c}\mathbf{1}^{\top}\rangle,$$

and similarly obtain $2\langle \mathbf{M}-\mathbf{P}, \mathbf{G}^\circ_{\mathbf{V}}\rangle - 2\langle \mathbf{M}-\mathbf{P}, \mathbf{1}\mathbf{c}'^\top\rangle$ by considering the term associated to $\mathbf{V}$. Finally, let $\mathbf{E} = \mathbf{U}\mathbf{B}_0^+\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{D}_{\text{out}}^{-1} = \mathbf{D}_{\text{out}}\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{D}_{\text{out}}^{-1}$, so that

$$\begin{aligned}\|\mathbf{E}\mathbf{e}_i\| &\le \kappa_{\mathbf{D}}\|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\| \\ &\le \kappa_{\mathbf{D}}\big(\|\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{e}_i\| + \|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\|\big) \\ &\le 2\kappa_{\mathbf{D}}^2\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}\end{aligned}$$

and $|\mathbf{E}_{ii}| \le \|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\|\,\|\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\| \le 2\kappa_{\mathbf{D}}^2\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}^2$, by Proposition 18(iii). The three pieces of $2\operatorname{tr}(\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{E}_2\mathbf{B}_0^+\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}})$ are bounded as follows. The first is $2\operatorname{tr}(\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{D}_{\text{out}}^{-1}\mathbf{R}_{\widetilde{\mathbf{U}}}\mathbf{U}\mathbf{B}_0^+\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}})$, which is at most $2d\,D_{\min}^{-1}\|\mathbf{R}_{\widetilde{\mathbf{U}}}\|D_{\max} \le 2d\kappa_{\mathbf{D}}\,C^2\|\mathbf{M}-\mathbf{P}\|^2/\sigma_d^2$. The second is $-2\sum_i \frac{\varepsilon_i^{\text{out}}}{1+\varepsilon_i^{\text{out}}}\,\mathbf{e}_i^\top\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{E}\mathbf{e}_i$, which is at most

$$\begin{aligned}4\sum_i |\varepsilon_i^{\text{out}}|\,\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\mathbf{e}_i\|\,\|\mathbf{E}\mathbf{e}_i\| &\le 8\kappa_{\mathbf{D}}^2\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}\,\|\varepsilon^{\text{out}}\|\,\|\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}\|_{\text{Frob}} \\ &\le 8\sqrt{2d}\,\kappa_{\mathbf{D}}^2\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}\,\|\varepsilon^{\text{out}}\|\,\frac{C\|\mathbf{M}-\mathbf{P}\|}{\sigma_d},\end{aligned}$$

by the Cauchy–Schwarz inequality over i and because $\boldsymbol{\delta}_{\widetilde{\mathbf{U}}}$ has rank at most $2d$. The third is $2\sum_i \frac{(\varepsilon_i^{\text{out}})^2}{1+\varepsilon_i^{\text{out}}}\mathbf{E}_{ii}$, which is at most $4\|\varepsilon^{\text{out}}\|^2\cdot 2\kappa_{\mathbf{D}}^2\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}^2$. Each of these three is at most $Cd\kappa_{\mathbf{D}}^2$ times the square of the bracket in the proposition statement, and an identical result holds for the term associated to $\mathbf{V}$. The final bound on $r^\circ(\mathbf{M})$ is then obtained by summing all remainders. □

We now obtain bounds on the matrix $\mathbf{G}^\circ$, in a result analogous to Proposition 13:

Proposition 21. *On the event $\mathcal{E}_C$ intersected with the event of Proposition 18(i), the matrix $\mathbf{G}^\circ$ of Proposition 20 satisfies*

$$\|\mathbf{G}^\circ\|_{\text{Frob}} \le \frac{C'd}{\sigma_d} \qquad \textit{and} \qquad \max_{i,j}|\mathbf{G}^\circ_{ij}| \le \frac{C'd}{n\sigma_d},$$

with C' depending on C, M and c_0 only.

Proof. $\mathbf{G}^\circ_{\mathbf{U}}$ has rank at most d, and $\|\mathbf{G}^\circ_{\mathbf{U}}\| \le D_{\max}/(\sigma_d D_{\min}) \le \kappa_{\mathbf{D}}/\sigma_d$, so using the standard bound $\|\mathbf{B}\mathbf{C}\|_{\text{Frob}} \le \operatorname{rank}(\mathbf{B})\|\mathbf{C}\|$ we have $\|\mathbf{G}^\circ_{\mathbf{U}}\|_{\text{Frob}} \le \sqrt{d}\kappa_{\mathbf{D}}/\sigma_d$, and likewise for $\mathbf{G}^\circ_{\mathbf{V}}$.

The (i,j)th entry of $\mathbf{G}^\circ_{\mathbf{U}}$ is bounded by $\|\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{D}_{\text{out}}^{-1}\mathbf{\Pi}^\perp_{\widetilde{\mathbf{U}}}\mathbf{e}_i\|\cdot D_{\max}\sigma_d^{-1}\|\mathbf{V}^\top\mathbf{e}_j\|$, and

$$\begin{aligned}\|\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{D}_{\text{out}}^{-1}\mathbf{\Pi}^\perp_{\widetilde{\mathbf{U}}}\mathbf{e}_i\| &\le D_{\min}^{-1}\big(2\|\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\|_{2\to\infty} + 2\|\mathbf{U}^\top\mathbf{e}_i\|\big) \\ &\le \frac{4\kappa_{\mathbf{D}}}{D_{\min}}\cdot\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\},\end{aligned}$$

so the entry is at most $4\kappa_{\mathbf{D}}^2\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}^2/\sigma_d \le C'd/(n\sigma_d)$ on $\mathcal{E}_C$.

For the degree part,

$$\begin{aligned}|c_i| &\le D_{\min}^{-1}\|\mathbf{\Pi}_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\|\,\|\mathbf{\Pi}_{\mathbf{D}_{\text{in}}^{-1}\mathbf{V}}\mathbf{\Pi}^\perp_{\mathbf{D}_{\text{out}}^{-1}\mathbf{U}}\mathbf{e}_i\| \\ &\le \frac{2\kappa_{\mathbf{D}}^2}{D_{\min}}\max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}^2 \\ &\le C'd/(n\cdot n\rho_n) \\ &\le C'd/(n\sigma_d),\end{aligned}$$

using $D_{\min} \ge c_0 n\rho_n/2$ and $\sigma_d \le \sigma_1 \le Cn\rho_n$, and thus $\|c\mathbf{1}^\top\|_{\text{Frob}} = n^{1/2}\|c\| \le n\max_i|c_i| \le C'd/\sigma_d$. An identical argument applies to c'. □

Corollary 22. *Under Assumptions **A1′**, **A2′** and **A3**:*

$$\widehat{\Psi}^{\circ} - \Psi^{\circ} = 2\langle \mathbf{A} - \mathbf{P}, \mathbf{G}^{\circ} \rangle + r^{\circ}$$

where

$$\langle \mathbf{A} - \mathbf{P}, \mathbf{G}^{\circ} \rangle = \mathrm{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n^2 \rho_n}} \right), \qquad r^{\circ}(\mathbf{A}) = \mathrm{O}_{\mathbb{P}}\big((n\rho_n)^{-1}\big).$$

Proof. With overwhelming probability the following events all hold:

- $\|\mathbf{A} - \mathbf{P}\| \le \sigma_d/8$, so $\mathbf{A} \in \mathcal{B}$, by Proposition 9.
- The row and column sums of $\mathbf{A}$ are positive and $\max_i |\varepsilon_i^{\mathrm{out}}|, \max_j |\varepsilon_j^{\mathrm{in}}| \le \frac{1}{2}$, by Proposition 18(ii).
- $C\|\mathbf{A} - \mathbf{P}\|/\sigma_d \le C'(n\rho_n)^{-1/2}$, by Proposition 9.
- $\kappa_{\mathbf{D}} \max\{\|\mathbf{U}\|_{2\to\infty}, \|\mathbf{V}\|_{2\to\infty}\}\big(\|\varepsilon^{\mathrm{out}}\|^2 + \|\varepsilon^{\mathrm{in}}\|^2\big)^{1/2} \le C'\sqrt{d/(n\rho_n)}$, by Proposition 18(i), (ii) and the event $\mathcal{E}_C$.

Thus for n sufficiently large the hypotheses of Proposition 20 hold with $\mathbf{M} = \mathbf{A}$, and the remainder bound gives

$$|r^{\circ}(\mathbf{A})| \le C' d \kappa_{\mathbf{D}}^2 \big(C'(n\rho_n)^{-1/2} + C'\sqrt{d/(n\rho_n)}\big)^2 \le C' d^2/(n\rho_n).$$

Conditionally on the latent positions $\mathbf{G}^{\circ}$ is deterministic, so $\langle \mathbf{A} - \mathbf{P}, \mathbf{G}^{\circ} \rangle$ is a sum of independent zero-mean terms $(\mathbf{A}_{ij} - \mathbf{P}_{ij})\mathbf{G}^{\circ}_{ij}$ with total variance at most $\rho_n \|\mathbf{G}^{\circ}\|_{\mathrm{Frob}}^2 \le C'/(n^2\rho_n)$ and each bounded by $\max_{ij} |\mathbf{G}^{\circ}_{ij}| \le C'/(n^2\rho_n)$, by Proposition 21 and the fact that $\sigma_d \ge n\rho_n/C$. Applying Bernstein's inequality in the same manner as for Corollary 14 gives the bound on $\langle \mathbf{A} - \mathbf{P}, \mathbf{G}^{\circ} \rangle$. □

Proof of Theorem 5. Write $\widehat{\Psi}^{\circ} - \Psi^{\circ}_F = (\Psi^{\circ} - \Psi^{\circ}_F) + (\widehat{\Psi}^{\circ} - \Psi^{\circ})$. By Proposition 17 the first term is $\mathrm{O}_{\mathbb{P}}(\sqrt{\log n/n})$ with the Gaussian limit of part (iii); by Corollary 22 the second is $\mathrm{O}_{\mathbb{P}}\big(\sqrt{\log n/(n^2\rho_n)} + (n\rho_n)^{-1}\big)$. The rest of the argument follows in exactly the same way as the proof of Theorem 4. □

Finally, we show that assuming the matrix $\mathbf{A}$ was not hollow does not affect our result:

Proposition 23. *Let $\mathbf{A}'$ be the matrix obtained by removing the diagonal of $\mathbf{A}$, with its degree matrices computed from $\mathbf{A}'$, and let $\widehat{\Psi}^{\circ\prime} = f^{\circ}(\mathbf{A}')$ be the resulting estimate. Then under the assumptions of Theorem 5,*

$$|\widehat{\Psi}^{\circ\prime} - \widehat{\Psi}^{\circ}| = \mathrm{O}_{\mathbb{P}}\big((n\rho_n)^{-1}\big),$$

and consequently Theorem 5 holds for $\mathbf{A}'$.

Proof. $\|\mathbf{A}' - \mathbf{P}\| \le \|\mathbf{A} - \mathbf{P}\| + 1$, and each row and column sum of $\mathbf{A}'$ differs from that of $\mathbf{A}$ by at most one, so the relative degree errors of $\mathbf{A}'$ differ from those of $\mathbf{A}$ by at most $2/(c_0 n\rho_n)$ entrywise, which changes $\|\varepsilon^{\mathrm{out}}\|^2 + \|\varepsilon^{\mathrm{in}}\|^2$ by at most $Cn/(n\rho_n)^2 \le C/\rho_n$. Hence the hypotheses of Proposition 20 hold for $\mathbf{M} = \mathbf{A}'$ with overwhelming probability, as they do for $\mathbf{A}$, and

$$\widehat{\Psi}^{\circ\prime} - \widehat{\Psi}^{\circ} = -2\langle \mathrm{diag}(\mathbf{A}), \mathbf{G}^{\circ} \rangle + r^{\circ}(\mathbf{A}') - r^{\circ}(\mathbf{A}),$$

with $|r^{\circ}(\mathbf{A}')|, |r^{\circ}(\mathbf{A})| = \mathrm{O}_{\mathbb{P}}((n\rho_n)^{-1})$ as in Corollary 22, and $|\langle \mathrm{diag}(\mathbf{A}), \mathbf{G}^{\circ} \rangle| \le \sum_i |\mathbf{G}^{\circ}_{ii}| \le n \cdot C'd/(n\sigma_d) = \mathrm{O}_{\mathbb{P}}((n\rho_n)^{-1})$ by Proposition 21. The last sentence follows as in Proposition 15. □

B CLOSED-FORM DERIVATIONS FOR STOCHASTIC CO-BLOCKMODELS

B.1 THE EFFECTIVE SHARED DIMENSION

In this section we present how to calculate the underlying effective shared dimension (and its degree-free counterpart) for the standard and degree-corrected stochastic co-blockmodel.

The stochastic co-blockmodel.
Let $\mathbf{B} \in [0,1]^{K\times L}$ of rank d admit a compact singular value decomposition $\mathbf{B} = \mathbf{U_B}\mathbf{\Sigma_B}\mathbf{V_B}^\top$ and define $\mathbf{X_B} = \mathbf{U_B}\mathbf{\Sigma_B}^{1/2} \in \mathbb{R}^{K\times d}$ and $\mathbf{Y_B} = \mathbf{V_B}\mathbf{\Sigma_B}^{1/2} \in \mathbb{R}^{L\times d}$. Let $\mathbf{\Pi} \in [0,1]^{K\times L}$ define the joint distribution for sending and receiving community membership, with row sums $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ and column sums $\boldsymbol{\pi}' = (\pi'_1, \dots, \pi'_L)$. Set

$$\boldsymbol{\xi}_k = \mathbf{X_B}^\top \mathbf{e}_k, \qquad \boldsymbol{\eta}_l = \mathbf{Y_B}^\top \mathbf{f}_l$$

for $k \in \{1, \dots, K\}$ and $l \in \{1, \dots, L\}$, where $\mathbf{e}_k$ and $\mathbf{f}_l$ are the kth and lth standard basis vectors of $\mathbb{R}^K$ and $\mathbb{R}^L$ respectively. Then the second moment matrices are given by

$$\Delta_\xi = \sum_{k=1}^{K} \pi_k \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top, \qquad \Delta_\eta = \sum_{l=1}^{L} \pi'_l \boldsymbol{\eta}_l \boldsymbol{\eta}_l^\top, \qquad \Delta_{\xi\eta} = \sum_{k=1}^{K}\sum_{l=1}^{L} \mathbf{\Pi}_{kl}\, \boldsymbol{\xi}_k \boldsymbol{\eta}_l^\top.$$

Then, by Proposition 3(i),

$$\boxed{\Psi_F = \operatorname{trace}\big(\Delta_\xi^{-1}\Delta_{\xi\eta}\Delta_\eta^{-1}\Delta_{\eta\xi}\big).}$$

For the degree-free value, replace $\boldsymbol{\xi}_k$ and $\boldsymbol{\eta}_l$ with the normalised values:

$$\boldsymbol{\xi}_k^\circ = \frac{\boldsymbol{\xi}_k}{(\mathbf{B}\boldsymbol{\pi}')_k}, \qquad \boldsymbol{\eta}_l^\circ = \frac{\boldsymbol{\eta}_l}{(\mathbf{B}^\top\boldsymbol{\pi})_l},$$

recompute the three moment matrices, and apply the same trace formula to obtain

$$\boxed{\Psi_F^\circ = \operatorname{trace}\big(\Delta_{\xi^\circ}^{-1}\Delta_{\xi^\circ\eta^\circ}\Delta_{\eta^\circ}^{-1}\Delta_{\eta^\circ\xi^\circ}\big)}$$

by Proposition 3(ii).

The degree-corrected stochastic co-blockmodel.
Retaining the notation of the previous example, let each node additionally carry degree parameters (θ, ϕ) (assumed independent of the latent positions) with joint distribution H on $[a,1]^2$, so that a node with communities (k,l) has latent positions

$$\mathbf{x} = \theta\, \boldsymbol{\xi}_k, \qquad \mathbf{y} = \phi\, \boldsymbol{\eta}_l.$$

The second moment matrices are then just scaled by the second moments of the degree parameters:

$$\Delta_x = \mathbb{E}[\theta^2]\sum_{k=1}^{K} \boldsymbol{\pi}_k\, \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top, \qquad \Delta_y = \mathbb{E}[\phi^2]\sum_{l=1}^{L} \boldsymbol{\pi}'_l\, \boldsymbol{\eta}_l \boldsymbol{\eta}_l^\top, \qquad \Delta_{xy} = \mathbb{E}[\theta\phi]\sum_{k=1}^{K}\sum_{l=1}^{L} \mathbf{\Pi}_{kl}\, \boldsymbol{\xi}_k \boldsymbol{\eta}_l^\top,$$

and

$$\boxed{\Psi_F^{\mathrm{DC}} = \operatorname{trace}\big(\Delta_x^{-1}\Delta_{xy}\Delta_y^{-1}\Delta_{yx}\big)}$$

by Proposition 3(i).

Thus in particular adding degree-correction to a standard stochastic co-blockmodel scales Ψ_F by a factor of

$$\kappa_H = \frac{\mathbb{E}[\theta\phi]^2}{\mathbb{E}[\theta^2]\mathbb{E}[\phi^2]} \in (0,1].$$

For the degree-free value, the expected out-degree of a node is $\rho_n\, \mathbf{x}^\top \boldsymbol{\mu}_y$ with $\boldsymbol{\mu}_y = \mathbb{E}[\mathbf{y}] = \sum_l \mathbb{E}[\phi]\boldsymbol{\pi}'_l\boldsymbol{\eta}_l$, so

$$\mathbf{x}^\circ = \frac{\boldsymbol{\xi}_k}{\mathbb{E}[\phi](\mathbf{B}\boldsymbol{\pi}')_k}, \qquad \mathbf{y}^\circ = \frac{\boldsymbol{\eta}_l}{\mathbb{E}[\theta](\mathbf{B}^\top\boldsymbol{\pi})_l},$$

using $\boldsymbol{\xi}_k^\top \boldsymbol{\eta}_l = \mathbf{B}_{kl}$. As in the previous example, we then compute the second moment matrices and find that

$$\boxed{\Psi_F^{\circ\mathrm{DC}} = \operatorname{trace}\big(\Delta_{x^\circ}^{-1}\Delta_{x^\circ y^\circ}\Delta_{y^\circ}^{-1}\Delta_{y^\circ x^\circ}\big)}$$

by Proposition 3.

B.2 THE VARIANCE TERM

In this section, we present how to calculate the variance terms τ_F^2 and $\tau_F^{\circ 2}$ appearing in Theorems 4 and 5 for stochastic co-blockmodels. We retain the notational conventions of Appendix B.1 throughout.

The stochastic co-blockmodel.
Put

$$\mathbf{T} = \Delta_\xi^{-1}\Delta_{\xi\eta}\Delta_\eta^{-1}, \qquad \mathbf{M}_1 = \mathbf{T}\Delta_{\eta\xi}\Delta_\xi^{-1}, \qquad \mathbf{M}_2 = \mathbf{T}^\top\Delta_{\xi\eta}\Delta_\eta^{-1},$$
$$t_{kl} = \boldsymbol{\xi}_k^\top\mathbf{T}\boldsymbol{\eta}_l, \qquad p_k = \boldsymbol{\xi}_k^\top\mathbf{M}_1\boldsymbol{\xi}_k, \qquad q_l = \boldsymbol{\eta}_l^\top\mathbf{M}_2\boldsymbol{\eta}_l.$$

The function g of Proposition 8, evaluated for a node in communities (k,l), is $g_{kl} = 2t_{kl} - p_k - q_l$. Since t, p and q can all be shown to have mean Ψ_F, we have $\sum_{k,l}\mathbf{\Pi}_{kl}g_{kl} = 0$, and so

$$\boxed{\tau_F^2 = \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbf{\Pi}_{kl}\, g_{kl}^2.}$$

For the degree-free version, use the normalised values $\boldsymbol{\xi}_k^\circ = \boldsymbol{\xi}_k/(\mathbf{B}\boldsymbol{\pi}')_k$ and $\boldsymbol{\eta}_l^\circ = \boldsymbol{\eta}_l/(\mathbf{B}^\top\boldsymbol{\pi})_l$. Form $\mathbf{T}^\circ$, $\mathbf{M}_1^\circ$ and $\mathbf{M}_2^\circ$ from their moment matrices as before, and put

$$u_{kl} = \boldsymbol{\xi}_k^{\circ\top}\mathbf{T}^\circ\boldsymbol{\eta}_l^\circ, \qquad p_k^\circ = \boldsymbol{\xi}_k^{\circ\top}\mathbf{M}_1^\circ\boldsymbol{\xi}_k^\circ, \qquad q_l^\circ = \boldsymbol{\eta}_l^{\circ\top}\mathbf{M}_2^\circ\boldsymbol{\eta}_l^\circ,$$

each with mean Ψ_F°. Applying the previous results to the normalised values gives the first-order term $2u_{kl} - p_k^\circ - q_l^\circ$. For the second-order correction, write $G(\mathbf{a},\mathbf{b})$ for the value of Ψ_F° computed with $\boldsymbol{\xi}_k/(\boldsymbol{\xi}_k^\top\mathbf{b})$ and $\boldsymbol{\eta}_l/(\mathbf{a}^\top\boldsymbol{\eta}_l)$ in place of the normalised values. Its gradients at $(\mathbf{a}_0,\mathbf{b}_0)$ are the d-vectors

$$\mathbf{c}_b = \partial_{\mathbf{b}}G = -2\sum_{k=1}^{K}\sum_{l=1}^{L}\mathbf{\Pi}_{kl}\,(u_{kl} - p_k^\circ)\,\boldsymbol{\xi}_k^\circ, \qquad \mathbf{c}_a = \partial_{\mathbf{a}}G = -2\sum_{k=1}^{K}\sum_{l=1}^{L}\mathbf{\Pi}_{kl}\,(u_{kl} - q_l^\circ)\,\boldsymbol{\eta}_l^\circ.$$

These follow in one line from $\partial\boldsymbol{\xi}^\circ/\partial b_j = -\xi_j^\circ\boldsymbol{\xi}^\circ$, which gives $\partial\Delta_{\xi^\circ}/\partial b_j = -2\mathbb{E}[\xi_j^\circ\boldsymbol{\xi}^\circ\boldsymbol{\xi}^{\circ\top}]$ and $\partial\Delta_{\xi^\circ\eta^\circ}/\partial b_j = -\mathbb{E}[\xi_j^\circ\boldsymbol{\xi}^\circ\boldsymbol{\eta}^{\circ\top}]$, contracted against the derivatives $2\mathbf{T}^\circ$ and $-\mathbf{M}_1^\circ$ of h. Then

$$g_{kl}^\circ = 2u_{kl} - p_k^\circ - q_l^\circ + \mathbf{c}_a^\top(\boldsymbol{\xi}_k - \mathbf{a}_0) + \mathbf{c}_b^\top(\boldsymbol{\eta}_l - \mathbf{b}_0),$$

and

$$\boxed{\tau_F^{\circ 2} = \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbf{\Pi}_{kl}\,(g_{kl}^\circ)^2,}$$

the correction terms having mean zero.

The degree-corrected stochastic co-blockmodel
A node with communities (k,l) and degree parameters (θ,ϕ) has positions $\theta\boldsymbol{\xi}_k$ and $\phi\boldsymbol{\eta}_l$. The second moment matrices are those of the co-blockmodel scaled by $\mathbb{E}[\theta^2]$, $\mathbb{E}[\phi^2]$ and $\mathbb{E}[\theta\phi]$, so the function g is given by:

$$g = \kappa_H\left(2\,\frac{\theta\phi}{\mathbb{E}[\theta\phi]}\,t_{kl} - \frac{\theta^2}{\mathbb{E}[\theta^2]}\,p_k - \frac{\phi^2}{\mathbb{E}[\phi^2]}\,q_l\right), \qquad \kappa_H = \frac{\mathbb{E}[\theta\phi]^2}{\mathbb{E}[\theta^2]\,\mathbb{E}[\phi^2]},$$

with t, p and q as before. Thus with the matrices

$$\mathbf{S} = \sum_{k,l}\mathbf{\Pi}_{kl}\,\mathbf{v}_{kl}\mathbf{v}_{kl}^\top, \quad \mathbf{v}_{kl} = (2t_{kl},\,-p_k,\,-q_l)^\top,$$

and

$$\mathbf{R} = \mathbb{E}_H\big[\mathbf{w}\mathbf{w}^\top\big], \quad \mathbf{w} = \left(\frac{\theta\phi}{\mathbb{E}[\theta\phi]},\,\frac{\theta^2}{\mathbb{E}[\theta^2]},\,\frac{\phi^2}{\mathbb{E}[\phi^2]}\right)^\top,$$

we have

$$\boxed{\tau_F^{2,\mathrm{DC}} = \kappa_H^2\ \mathrm{trace}(\mathbf{RS}).}$$

For the degree-free version, the normalised positions do not depend on the node's own degree parameters, and rescaling the atoms by constants leaves u, p° and q° unchanged, so the first-order part of the variance is exactly that of the standard stochastic co-blockmodel. The correction terms, however, now acquire the relative degrees $\tilde{\theta} = \theta/\mathbb{E}[\theta]$ and $\tilde{\phi} = \phi/\mathbb{E}[\phi]$:

$$g_{kl}^{\circ,\mathrm{DC}} = g_{kl}^{\circ} + (\tilde{\theta} - 1)\, \mathbf{c}_a^\top \boldsymbol{\xi}_k + (\tilde{\phi} - 1)\, \mathbf{c}_b^\top \boldsymbol{\eta}_l,$$

with g_{kl}°, $\mathbf{c}_a$ and $\mathbf{c}_b$ defined as before. Since $\tilde{\theta} - 1$ and $\tilde{\phi} - 1$ have mean zero and are independent of the communities,

$$\boxed{\tau_F^{\circ 2,\mathrm{DC}} = \tau_F^{\circ 2} + \frac{\mathrm{Var}(\theta)}{\mathbb{E}[\theta]^2} \sum_k \pi_k (\mathbf{c}_a^\top \boldsymbol{\xi}_k)^2 + \frac{\mathrm{Var}(\phi)}{\mathbb{E}[\phi]^2} \sum_l \pi'_l (\mathbf{c}_b^\top \boldsymbol{\eta}_l)^2 + 2\, \frac{\mathrm{Cov}(\theta,\phi)}{\mathbb{E}\theta\, \mathbb{E}\phi} \sum_{k,l} \mathbf{\Pi}_{kl}\, (\mathbf{c}_a^\top \boldsymbol{\xi}_k)(\mathbf{c}_b^\top \boldsymbol{\eta}_l).}$$

C AN EXPLICIT FORM FOR THE QUADRATIC BIAS

In this appendix we provide the explicit form of the quadratic bias term $Q(\mathbf{P})$ from Section 6, which we recall is defined as

$$Q(\mathbf{P}) = \frac{1}{2} \sum_{i \neq j} \mathbf{P}_{ij}(1 - \mathbf{P}_{ij})\, \frac{\partial^2 f}{\partial \mathbf{P}_{ij}^2}(\mathbf{P}).$$

We retain the notation of previous sections, and for each pair (i, j) define

$$c_{ij}(\mathbf{P}) = \frac{1}{2}\, \frac{\partial^2 f}{\partial \mathbf{P}_{ij}^2}(\mathbf{P}).$$

For a node i, $\mathbf{u}_i = \mathbf{U}^\top \mathbf{e}_i$ and $\mathbf{v}_i = \mathbf{V}^\top \mathbf{e}_i$ are the ith rows of $\mathbf{U}$ and $\mathbf{V}$, as vectors in $\mathbb{R}^d$. Two further vectors in $\mathbb{R}^d$ appear repeatedly:

$$\mathbf{V}^\top \mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{e}_i = \mathbf{v}_i - \mathbf{V}^\top \mathbf{U} \mathbf{u}_i, \qquad \mathbf{U}^\top \mathbf{\Pi}_{\mathbf{V}}^{\perp} \mathbf{e}_j = \mathbf{u}_j - \mathbf{U}^\top \mathbf{V} \mathbf{v}_j,$$

the receiving coordinates of the part of $\mathbf{e}_i$ outside the sending subspace, and the sending coordinates of the part of $\mathbf{e}_j$ outside the receiving subspace. Note $(\mathbf{\Pi}_{\mathbf{U}}^{\perp})_{ii} = 1 - \|\mathbf{u}_i\|^2$, $(\mathbf{\Pi}_{\mathbf{V}}^{\perp})_{jj} = 1 - \|\mathbf{v}_j\|^2$, $(\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^\top)_{ij} = \mathbf{u}_i^\top \mathbf{\Sigma}^{-1} \mathbf{v}_j$ and, for $i \neq j$, $(\mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{\Pi}_{\mathbf{V}}^{\perp})_{ij} = -\mathbf{u}_i^\top \mathbf{u}_j - \mathbf{v}_i^\top \mathbf{v}_j + \mathbf{u}_i^\top \mathbf{U}^\top \mathbf{V} \mathbf{v}_j$.

Proposition 24. *For every i, j,*

$$c_{ij}(\mathbf{P}) = c_{ij}^{\mathrm{send}} + c_{ij}^{\mathrm{rec}} + c_{ij}^{\mathrm{joint}},$$

where

$$\begin{aligned} c_{ij}^{\mathrm{send}} = {} & 2\big(1 - \|\mathbf{v}_j\|^2\big)(\mathbf{V}^\top \mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{e}_i)^\top \mathbf{V}^\top \mathbf{U} \mathbf{\Sigma}^{-2} \mathbf{u}_i - 2\big(\mathbf{u}_i^\top \mathbf{\Sigma}^{-1} \mathbf{v}_j\big)(\mathbf{V}^\top \mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{e}_i)^\top \mathbf{V}^\top \mathbf{U} \mathbf{\Sigma}^{-1} \mathbf{v}_j \\ & + \|\mathbf{V}^\top \mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{e}_i\|^2\, \|\mathbf{\Sigma}^{-1} \mathbf{v}_j\|^2 - \big(1 - \|\mathbf{u}_i\|^2\big) \|\mathbf{V}^\top \mathbf{U} \mathbf{\Sigma}^{-1} \mathbf{v}_j\|^2, \end{aligned}$$

$$\begin{aligned} c_{ij}^{\mathrm{rec}} = {} & 2\big(1 - \|\mathbf{u}_i\|^2\big)(\mathbf{U}^\top \mathbf{\Pi}_{\mathbf{V}}^{\perp} \mathbf{e}_j)^\top \mathbf{U}^\top \mathbf{V} \mathbf{\Sigma}^{-2} \mathbf{v}_j - 2\big(\mathbf{u}_i^\top \mathbf{\Sigma}^{-1} \mathbf{v}_j\big)(\mathbf{U}^\top \mathbf{\Pi}_{\mathbf{V}}^{\perp} \mathbf{e}_j)^\top \mathbf{U}^\top \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{u}_i \\ & + \|\mathbf{U}^\top \mathbf{\Pi}_{\mathbf{V}}^{\perp} \mathbf{e}_j\|^2\, \|\mathbf{\Sigma}^{-1} \mathbf{u}_i\|^2 - \big(1 - \|\mathbf{v}_j\|^2\big) \|\mathbf{U}^\top \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{u}_i\|^2, \end{aligned}$$

and

$$c_{ij}^{\mathrm{joint}} = 2\big(\mathbf{u}_i^\top \mathbf{\Sigma}^{-1} \mathbf{V}^\top \mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{e}_i\big)\big(\mathbf{v}_j^\top \mathbf{\Sigma}^{-1} \mathbf{U}^\top \mathbf{\Pi}_{\mathbf{V}}^{\perp} \mathbf{e}_j\big) + 2\, (\mathbf{\Pi}_{\mathbf{U}}^{\perp} \mathbf{\Pi}_{\mathbf{V}}^{\perp})_{ij}\, \mathbf{u}_i^\top \mathbf{\Sigma}^{-1} \mathbf{V}^\top \mathbf{U} \mathbf{\Sigma}^{-1} \mathbf{v}_j.$$

Corollary 25 (Matrix form)**.** *Let $\mathbf{s}^2$ be the $n \times n$ matrix with entries $\mathbf{P}_{ij}(1 - \mathbf{P}_{ij})$ off the diagonal and zero on it,* $\mathrm{diag}(\cdot)$ *the vector of diagonal entries, $\circ$ the entrywise product and $\langle \cdot, \cdot \rangle$ the entrywise*

inner product. Then

$$\begin{aligned}
Q(\mathbf{P}) \;=\; & 2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{V}\mathbf{V}^{\top}\mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^{\top})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{V}}^{\perp}) \\
& + 2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{U}}^{\perp})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{U}\mathbf{U}^{\top}\mathbf{V}\mathbf{\Sigma}^{-2}\mathbf{V}^{\top}) \\
& - \mathrm{diag}(\mathbf{\Pi}_{\mathbf{U}}^{\perp})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\top}\mathbf{V}\mathbf{V}^{\top}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}) \\
& - \mathrm{diag}(\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}\mathbf{U}\mathbf{U}^{\top}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\top})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{V}}^{\perp}) \\
& + \mathrm{diag}(\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{\Pi}_{\mathbf{V}}\mathbf{\Pi}_{\mathbf{U}}^{\perp})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{V}\mathbf{\Sigma}^{-2}\mathbf{V}^{\top}) \\
& + \mathrm{diag}(\mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^{\top})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{\Pi}_{\mathbf{U}}\mathbf{\Pi}_{\mathbf{V}}^{\perp}) \\
& + 2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\top})^{\top}\mathbf{s}^2\,\mathrm{diag}(\mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}) \\
& + 2\big\langle \mathbf{s}^2, (\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{\Pi}_{\mathbf{V}}^{\perp}) \circ (\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top})\big\rangle \\
& - 2\big\langle \mathbf{s}^2, (\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}) \circ (\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{V}\mathbf{V}^{\top}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top})\big\rangle \\
& - 2\big\langle \mathbf{s}^2, (\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}) \circ (\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}\mathbf{U}\mathbf{U}^{\top}\mathbf{\Pi}_{\mathbf{V}}^{\perp})\big\rangle.
\end{aligned}$$

Proving these requires a couple of intermediate results. For each pair (i, j) define the $n \times d$ matrices:

$$\begin{aligned}
\mathbf{U}_{ij}^{(1)} &= \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i\mathbf{v}_j^{\top}\mathbf{\Sigma}^{-1}, \\
\mathbf{U}_{ij}^{(2)} &= -(\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-1}\mathbf{v}_j)\,\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i\mathbf{v}_j^{\top}\mathbf{\Sigma}^{-1} + (1 - \|\mathbf{v}_j\|^2)\,\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-2} \\
&\quad - (1 - \|\mathbf{u}_i\|^2)\,\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{v}_j\mathbf{v}_j^{\top}\mathbf{\Sigma}^{-1}, \\
\mathbf{V}_{ij}^{(1)} &= \mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{e}_j\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-1}, \\
\mathbf{V}_{ij}^{(2)} &= -(\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-1}\mathbf{v}_j)\,\mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{e}_j\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-1} + (1 - \|\mathbf{u}_i\|^2)\,\mathbf{\Pi}_{\mathbf{V}}^{\perp}\mathbf{e}_j\mathbf{v}_j^{\top}\mathbf{\Sigma}^{-2} \\
&\quad - (1 - \|\mathbf{v}_j\|^2)\,\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{u}_i\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-1}.
\end{aligned}$$

Proposition 26. *Let $\widetilde{\mathbf{P}} = \mathbf{P} + t\,\mathbf{e}_i\mathbf{e}_j^{\top}$, with leading d-dimensional left and right singular subspaces $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$. For t small enough, $\mathbf{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{U}$ and $\mathbf{\Pi}_{\widetilde{\mathbf{V}}}\mathbf{V}$ are bases of these subspaces, and*

$$\mathbf{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{U} = \mathbf{U} + t\,\mathbf{U}_{ij}^{(1)} + t^2\,\mathbf{U}_{ij}^{(2)} + O(t^3), \qquad \mathbf{\Pi}_{\widetilde{\mathbf{V}}}\mathbf{V} = \mathbf{V} + t\,\mathbf{V}_{ij}^{(1)} + t^2\,\mathbf{V}_{ij}^{(2)} + O(t^3).$$

Proof. $\mathbf{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{U}$ tends to $\mathbf{U}$, so it has full column rank for small t, and its column space is then that of $\widetilde{\mathbf{U}}$. For a general perturbation $\widetilde{\mathbf{P}} = \mathbf{P} + \mathbf{E}$, the second-order expansion of the projector $\mathbf{\Pi}_{\widetilde{\mathbf{U}}}$ (Xia, 2021, Theorem 1) applied to $\mathbf{U}$ has first-order term $\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{E}\mathbf{V}\mathbf{\Sigma}^{-1}$ and the second-order term

$$\begin{aligned}
& - \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{E}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\top}\mathbf{E}\mathbf{V}\mathbf{\Sigma}^{-1} - \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{E}\mathbf{V}\mathbf{V}^{\top}\mathbf{E}^{\top}\mathbf{U}\mathbf{\Sigma}^{-2} \\
& \quad - \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\top}\mathbf{E}^{\top}\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{E}\mathbf{V}\mathbf{\Sigma}^{-1} + \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{E}\mathbf{E}^{\top}\mathbf{U}\mathbf{\Sigma}^{-2},
\end{aligned}$$

Setting $\mathbf{E} = t\,\mathbf{e}_i\mathbf{e}_j^{\top}$ and using the facts

$$\mathbf{e}_j^{\top}\mathbf{V} = \mathbf{v}_j^{\top}, \quad \mathbf{U}^{\top}\mathbf{e}_i = \mathbf{u}_i, \quad \mathbf{e}_i^{\top}\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i = 1 - \|\mathbf{u}_i\|^2, \quad \mathbf{e}_j^{\top}\mathbf{e}_j = 1$$

we find that the first-order term is $t\,\mathbf{U}_{ij}^{(1)}$, while the four second-order terms become

$$\begin{aligned}
& -(\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-1}\mathbf{v}_j)\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i\mathbf{v}_j^{\top}\mathbf{\Sigma}^{-1}, \quad -\|\mathbf{v}_j\|^2\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-2}, \\
& -(1 - \|\mathbf{u}_i\|^2)\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{v}_j\mathbf{v}_j^{\top}\mathbf{\Sigma}^{-1}, \quad \mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i\mathbf{u}_i^{\top}\mathbf{\Sigma}^{-2},
\end{aligned}$$

each multiplied by t^2. Summing over these gives $t^2\,\mathbf{U}_{ij}^{(2)}$, and interchanging $\mathbf{U}$ and $\mathbf{V}$, $\mathbf{E}$ and $\mathbf{E}^{\top}$ and i and j gives the corresponding result for $\mathbf{V}$. □

Proposition 27. *Let $\mathbf{B}(t) = \mathbf{U} + t\,\mathbf{B}_1 + t^2\mathbf{B}_2 + O(t^3)$ and $\mathbf{C}(t) = \mathbf{V} + t\,\mathbf{C}_1 + t^2\mathbf{C}_2 + O(t^3)$ be $n \times d$ matrices with $\mathbf{U}^{\top}\mathbf{B}_1 = \mathbf{V}^{\top}\mathbf{C}_1 = \mathbf{0}$. Then the coefficient of t^2 in $\mathrm{tr}(\mathbf{\Pi}_{\mathbf{B}(t)}\mathbf{\Pi}_{\mathbf{C}(t)})$ is*

$$\mathcal{T}(\mathbf{B}; \mathbf{C}) + \mathcal{T}(\mathbf{C}; \mathbf{B}) + 2\mathrm{tr}(\mathbf{B}_1\mathbf{U}^{\top}\mathbf{C}_1\mathbf{V}^{\top}) + 2\mathrm{tr}(\mathbf{B}_1\mathbf{U}^{\top}\mathbf{V}\mathbf{C}_1^{\top}),$$

where

$$\mathcal{T}(\mathbf{B}; \mathbf{C}) = 2\mathrm{tr}(\mathbf{U}^{\top}\mathbf{\Pi}_{\mathbf{V}}\mathbf{\Pi}_{\mathbf{U}}^{\perp}\mathbf{B}_2) + \mathrm{tr}(\mathbf{B}_1^{\top}\mathbf{\Pi}_{\mathbf{V}}\mathbf{B}_1) - \mathrm{tr}(\mathbf{B}_1^{\top}\mathbf{B}_1\,\mathbf{U}^{\top}\mathbf{\Pi}_{\mathbf{V}}\mathbf{U}),$$

and $\mathcal{T}(\mathbf{C}; \mathbf{B})$ is the same with $(\mathbf{U}, \mathbf{B}_1, \mathbf{B}_2)$ and $(\mathbf{V}, \mathbf{C}_1, \mathbf{C}_2)$ exchanged.

Proof. Since $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$ and $\mathbf{U}^\top\mathbf{B}_1 = \mathbf{0}$, we have $\mathbf{B}^\top\mathbf{B} = \mathbf{I} + t^2\mathbf{S} + O(t^3)$ with $\mathbf{S} = \mathbf{B}_1^\top\mathbf{B}_1 + \mathbf{U}^\top\mathbf{B}_2 + \mathbf{B}_2^\top\mathbf{U}$, so $(\mathbf{B}^\top\mathbf{B})^{-1} = \mathbf{I} - t^2\mathbf{S} + O(t^3)$ and

$$\begin{aligned}\mathbf{\Pi}_{\mathbf{B}(t)} &= \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top \\ &= \mathbf{\Pi}_\mathbf{U} + t\,(\mathbf{B}_1\mathbf{U}^\top + \mathbf{U}\mathbf{B}_1^\top) \\ &\quad + t^2\big(\mathbf{B}_1\mathbf{B}_1^\top - \mathbf{U}\mathbf{B}_1^\top\mathbf{B}_1\mathbf{U}^\top + \mathbf{\Pi}_\mathbf{U}^\perp\mathbf{B}_2\mathbf{U}^\top + \mathbf{U}\mathbf{B}_2^\top\mathbf{\Pi}_\mathbf{U}^\perp\big) + O(t^3),\end{aligned}$$

and likewise for $\mathbf{\Pi}_{\mathbf{C}(t)}$. The coefficient of t^2 in $\mathrm{tr}(\mathbf{\Pi}_{\mathbf{B}(t)}\mathbf{\Pi}_{\mathbf{C}(t)})$ is therefore the trace of the t^2-coefficient of $\mathbf{\Pi}_{\mathbf{B}(t)}$ against $\mathbf{\Pi}_\mathbf{V}$, plus the same with the roles exchanged, plus the trace of the product of the two t-coefficients. By the cyclic property of the trace, and since $\mathrm{tr}(\mathbf{U}\mathbf{B}_2^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{\Pi}_\mathbf{V}) = \mathrm{tr}(\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{B}_2\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}) = \mathrm{tr}(\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{B}_2)$, the first is $\mathcal{T}(\mathbf{B};\mathbf{C})$ and the second is $\mathcal{T}(\mathbf{C};\mathbf{B})$. The third is

$$\mathrm{tr}\big((\mathbf{B}_1\mathbf{U}^\top + \mathbf{U}\mathbf{B}_1^\top)(\mathbf{C}_1\mathbf{V}^\top + \mathbf{V}\mathbf{C}_1^\top)\big),$$

whose four terms are equal in pairs, each pair being transposes of one another, giving $2\mathrm{tr}(\mathbf{B}_1\mathbf{U}^\top\mathbf{C}_1\mathbf{V}^\top) + 2\mathrm{tr}(\mathbf{B}_1\mathbf{U}^\top\mathbf{V}\mathbf{C}_1^\top)$. □

Proof of Proposition 24. By Proposition 26, $f(\mathbf{P} + t\,\mathbf{e}_i\mathbf{e}_j^\top) = \mathrm{tr}(\mathbf{\Pi}_{\mathbf{B}(t)}\mathbf{\Pi}_{\mathbf{C}(t)})$ with $\mathbf{B}(t) = \mathbf{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{U}$ and $\mathbf{C}(t) = \mathbf{\Pi}_{\widetilde{\mathbf{V}}}\mathbf{V}$, whose expansions have $\mathbf{B}_k = \mathbf{U}_{ij}^{(k)}$ and $\mathbf{C}_k = \mathbf{V}_{ij}^{(k)}$. Since $c_{ij}(\mathbf{P})$ is the coefficient of t^2 in this expression, and $\mathbf{U}^\top\mathbf{U}_{ij}^{(1)} = \mathbf{0}$ and $\mathbf{V}^\top\mathbf{V}_{ij}^{(1)} = \mathbf{0}$, we may apply Proposition 27. We claim that $\mathcal{T}(\mathbf{B};\mathbf{C}) = c_{ij}^{\mathrm{send}}$. Indeed, recall that $\mathbf{B}_1 = \mathbf{U}_{ij}^{(1)} = \mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\mathbf{v}_j^\top\mathbf{\Sigma}^{-1}$, $\mathbf{B}_2 = \mathbf{U}_{ij}^{(2)}$, and

$$\mathcal{T}(\mathbf{B};\mathbf{C}) = 2\mathrm{tr}(\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{U}_{ij}^{(2)}) + \mathrm{tr}(\mathbf{U}_{ij}^{(1)\top}\mathbf{\Pi}_\mathbf{V}\mathbf{U}_{ij}^{(1)}) - \mathrm{tr}(\mathbf{U}_{ij}^{(1)\top}\mathbf{U}_{ij}^{(1)}\,\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{U}).$$

For the first term, since $\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{\Pi}_\mathbf{U}^\perp = \mathbf{\Pi}_\mathbf{U}^\perp$ and $\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{U} = \mathbf{0}$, applying $\mathbf{\Pi}_\mathbf{U}^\perp$ to the three terms of $\mathbf{U}_{ij}^{(2)}$ leaves the first two unchanged and annihilates the third:

$$\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{U}_{ij}^{(2)} = -(\mathbf{u}_i^\top\mathbf{\Sigma}^{-1}\mathbf{v}_j)\,\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\mathbf{v}_j^\top\mathbf{\Sigma}^{-1} + (1 - \|\mathbf{v}_j\|^2)\,\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\mathbf{u}_i^\top\mathbf{\Sigma}^{-2}.$$

Each term has the form $\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\mathbf{w}^\top$ for a vector $\mathbf{w} \in \mathbb{R}^d$, and for such a term

$$\mathrm{tr}(\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\mathbf{w}^\top) = \mathbf{w}^\top\mathbf{U}^\top\mathbf{V}\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i = (\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i)^\top\mathbf{V}^\top\mathbf{U}\,\mathbf{w}.$$

With $\mathbf{w} = \mathbf{\Sigma}^{-1}\mathbf{v}_j$ and $\mathbf{w} = \mathbf{\Sigma}^{-2}\mathbf{u}_i$ respectively,

$$\begin{aligned}2\mathrm{tr}(\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{U}_{ij}^{(2)}) &= 2(1 - \|\mathbf{v}_j\|^2)\,(\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i)^\top\mathbf{V}^\top\mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{u}_i \\ &\quad - 2(\mathbf{u}_i^\top\mathbf{\Sigma}^{-1}\mathbf{v}_j)\,(\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i)^\top\mathbf{V}^\top\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{v}_j,\end{aligned}$$

which are the first two terms of c_{ij}^{send}.

For the second term, since $\mathbf{U}_{ij}^{(1)} = \mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i(\mathbf{\Sigma}^{-1}\mathbf{v}_j)^\top$,

$$\mathbf{U}_{ij}^{(1)\top}\mathbf{\Pi}_\mathbf{V}\mathbf{U}_{ij}^{(1)} = \mathbf{\Sigma}^{-1}\mathbf{v}_j\,\big(\mathbf{e}_i^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{V}\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\big)\,\mathbf{v}_j^\top\mathbf{\Sigma}^{-1} = \|\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\|^2\,\mathbf{\Sigma}^{-1}\mathbf{v}_j\mathbf{v}_j^\top\mathbf{\Sigma}^{-1},$$

whose trace is

$$\mathrm{tr}(\mathbf{U}_{ij}^{(1)\top}\mathbf{\Pi}_\mathbf{V}\mathbf{U}_{ij}^{(1)}) = \|\mathbf{V}^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i\|^2\,\|\mathbf{\Sigma}^{-1}\mathbf{v}_j\|^2,$$

the third term of c_{ij}^{send}.

Finally, in the same manner we see that

$$\mathbf{U}_{ij}^{(1)\top}\mathbf{U}_{ij}^{(1)} = \mathbf{\Sigma}^{-1}\mathbf{v}_j\,(\mathbf{e}_i^\top\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{\Pi}_\mathbf{U}^\perp\mathbf{e}_i)\,\mathbf{v}_j^\top\mathbf{\Sigma}^{-1} = (1 - \|\mathbf{u}_i\|^2)\,\mathbf{\Sigma}^{-1}\mathbf{v}_j\mathbf{v}_j^\top\mathbf{\Sigma}^{-1},$$

and, with $\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{U} = \mathbf{U}^\top\mathbf{V}\mathbf{V}^\top\mathbf{U}$,

$$\begin{aligned}-\mathrm{tr}(\mathbf{U}_{ij}^{(1)\top}\mathbf{U}_{ij}^{(1)}\,\mathbf{U}^\top\mathbf{\Pi}_\mathbf{V}\mathbf{U}) &= -(1 - \|\mathbf{u}_i\|^2)\,\mathbf{v}_j^\top\mathbf{\Sigma}^{-1}\mathbf{U}^\top\mathbf{V}\mathbf{V}^\top\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{v}_j \\ &= -(1 - \|\mathbf{u}_i\|^2)\,\|\mathbf{V}^\top\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{v}_j\|^2,\end{aligned}$$

the fourth term of c_{ij}^{send}. Adding the three gives $\mathcal{T}(\mathbf{B};\mathbf{C}) = c_{ij}^{\mathrm{send}}$.

Similarly, we see that $\mathcal{T}(\mathbf{C};\mathbf{B}) = c_{ij}^{\text{rec}}$. Finally, combining the facts that

$$2\text{tr}(\mathbf{U}_{ij}^{(1)}\mathbf{U}^\top\mathbf{V}_{ij}^{(1)}\mathbf{V}^\top) = 2\,(\mathbf{u}_i^\top\boldsymbol{\Sigma}^{-1}\mathbf{V}^\top\boldsymbol{\Pi}_{\mathbf{U}}^{\perp}\mathbf{e}_i)(\mathbf{v}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{U}^\top\boldsymbol{\Pi}_{\mathbf{V}}^{\perp}\mathbf{e}_j),$$
$$2\text{tr}(\mathbf{U}_{ij}^{(1)}\mathbf{U}^\top\mathbf{V}\mathbf{V}_{ij}^{(1)\top}) = 2\,(\boldsymbol{\Pi}_{\mathbf{U}}^{\perp}\boldsymbol{\Pi}_{\mathbf{V}}^{\perp})_{ij}\;\mathbf{v}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{U}^\top\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{u}_i,$$

and $\mathbf{v}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{U}^\top\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{u}_i = \mathbf{u}_i^\top\boldsymbol{\Sigma}^{-1}\mathbf{V}^\top\mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{v}_j$, gives c_{ij}^{joint}. □