

Mixed Membership Estimation in Partial Correlation Network Model*

Endong Wang[†]

Siao Xu[‡]

Summary

Latent group structure is a standard way to model cross-sectional dependence in large panels, yet existing partial-correlation community methods assign each unit to a single group and do not distinguish link formation from link strength. We propose a two-path mixed-membership partial-correlation network (2PMM-PCN): mixed membership governs link formation along an extensive margin, while node-specific sociability governs link strength along an intensive margin. Because the node-specific components enter the population signal only through a common scale, memberships are recoverable by PartialCorr-Mixed-SCORE, for which we establish a convergence rate under weak dependence. Applied to U.S. state employment growth, the method reveals economically meaningful overlaps while recovering known group structure.

Keywords: Concentration matrix; spectral methods; probabilistic graphical model; weighted networks; degree correction. *JEL Classification:* C32; C33; C38; C55; D85

*We are grateful to Jean-Marie Dufour, Juan Carlos Escanciano, Matthew O. Jackson, Evan Munro, Bin Peng, Christopher Rothe, Frank Schorfheide, Carsten Trenkler, Amilcar Velez, and Mengshan Xu for helpful comments and discussions. We also thank participants and organizers of the 2025 (EC)² Conference in Lugano, the 2026 CSW Asia Meeting (Main) in Abu Dhabi, the Mannheim Econometrics Seminar, the 12th HKMetrics Workshop, the 45th International Symposium on Forecasting, the 2025 Harbin International Conference on Econometrics and Statistics, Time Series AI in Macroeconomics and Finance, and the ENTER Jamboree at Universidad Carlos III de Madrid for helpful comments and feedback. All errors are our own.

[†]Department of Economics, University of Mannheim, L7, 3–5, Mannheim, Germany, 68161. TEL: +49 621-181-1879; Email: endong.wang@uni-mannheim.de.

[‡]Graduate School of Economic and Social Sciences, University of Mannheim; siao.xu@students.uni-mannheim.de.

1 Introduction

Latent group structure is widely used to model heterogeneity and cross-sectional dependence in large panels. Two broad approaches are common. One models grouped heterogeneity in panel regressions, allowing coefficients to vary across latent groups but remain homogeneous within groups (Bonhomme & Manresa 2015, Su et al. 2016, Zhang et al. 2019). The other recovers groups from the interdependence of large multivariate systems, often through community or block structures in the implied network (Guðmundsson & Brownlees 2021, Brownlees et al. 2022, Chen et al. 2023, Zhu et al. 2025). Both approaches, however, typically impose pure membership, assigning each unit to a single latent group, whereas overlapping—or equivalently, mixed—membership is common in many applications. Moreover, in the partial-correlation community models most closely related to our setting, the latent network structure primarily governs which links form, while the strength of realized links is not separately parameterized. We address these limitations by developing a mixed-membership model for the conditional-dependence structure of a large panel of time series, allowing units to belong to multiple latent groups while separating link formation from link strength.

Two features of existing partial-correlation community models are particularly restrictive for our purposes. First, they typically assign each unit to a single group. This can be limiting when units carry diversified exposures: regional economies combine several industries, firms participate in overlapping networks, and financial institutions operate across business lines. Dependence between two units may therefore reflect more than one latent channel, which a single group label cannot record. Second, the same latent structure is typically used to characterize the presence of conditional dependence without separately modeling heterogeneity in the strength of existing links. The determinants of these two margins need not coincide: shared exposures may determine whether two units are conditionally dependent, while size, centrality, or openness may affect the strength of that dependence (Gabaix 2011, Acemoglu et al. 2012).

We specify a model with two paths governing link formation and link strength. The first governs the extensive margin. Each unit carries a membership vector over K latent groups, and a link forms with a probability that depends on these vectors, a group-level connectivity matrix, and a node-specific propensity to connect, as in degree-corrected mixed-membership network models (Airoldi et al. 2008, Fan et al. 2022, Jin et al. 2024). The second path governs the intensive margin. Conditional on a link being present, its expected weight depends on the sociabilities of the two units. Embedding the resulting weighted network in the concentration matrix yields the two-path mixed-membership partial-correlation network (2PMM-PCN), in which the sparsity pattern records which conditional dependencies are present and the normalized link weights determine their strength.

The paper makes three contributions. First, we allow overlapping (mixed) membership in a partial-correlation network, nesting pure community models (Brownlees et al. 2022) as the special case in which every unit loads on a single group, so that dependence between two units may reflect several latent channels rather than one. Second, we separate link formation from link strength through two distinct paths, distinguishing the extensive margin of conditional dependence from its intensive margin, in parallel with two-part models for weighted networks (Simpson & Laurienti 2015). Third, we show that the two node-specific paths enter the population signal only through a common scale. This preserves the degree-corrected structure and permits the memberships to be recovered by a single spectral procedure, PartialCorr-Mixed-SCORE, for which we derive a convergence rate for the average membership estimation error under weak dependence. The model is relevant for panels in which units have overlapping exposures and both link incidence and link strength matter, including industry co-movement across firms, dependence among financial institutions operating across business lines, and interdependence among regional economies.

Because the network is latent, the membership structure must be recovered from the concentration matrix estimated from the panel. The node-specific propensity to form links and the node-specific sociability enter the population signal only through their product.

The resulting degree-corrected mixed-membership structure is preserved under spectral normalization: an eigen-ratio transformation removes the effective node-specific scale and maps the observations into a simplex whose vertices represent pure membership in the latent groups and whose barycentric coordinates determine the membership weights. This leads to PartialCorr-Mixed-SCORE, which extends SCORE (Jin 2015) and Mixed-SCORE (Jin et al. 2024) from observed networks to partial-correlation networks estimated from panel data.

We derive population identification and sample recovery results. At the population level, the simplex representation underlying Mixed-SCORE continues to hold under the 2PMM-PCN. At the sample level, the estimation error has two sources: randomness in the realized weighted network and estimation of the concentration matrix from the panel. Under an operator-norm covariance-concentration condition that permits serial dependence, we derive a convergence rate for the average membership estimation error that separates these two sources.

Monte Carlo simulations show that PartialCorr-Mixed-SCORE recovers mixed memberships well, with accuracy improving as the time dimension, block separation, and share of pure nodes increase. We apply the method to quarterly U.S. state payroll employment growth, previously studied by Hamilton & Owyang (2012) and Brownlees et al. (2022). The estimated memberships retain the broad three-group benchmark used in that literature while allowing each state to have continuous membership weights across the three latent groups.

Our analysis relates to several strands of literature. First, we contribute to the literature on latent group structure in panel data. Existing work typically introduces group heterogeneity through regression coefficients, fixed effects, or other unit-specific parameters (Bonhomme & Manresa 2015, Su et al. 2016, Zhang et al. 2019), or recovers groups from cross-sectional dependence in a multivariate system (Guðmundsson & Brownlees 2021, Brownlees et al. 2022, Chen et al. 2023, Zhu et al. 2025). The object we recover differs

from the former and is more general than the latter. We do not group units according to a common regression coefficient. Instead, the latent structure is encoded in the conditional-dependence network of the panel, and each unit is represented by a vector of membership weights rather than by a single group label. Pure community models are nested as the special case in which every row of Π is a canonical basis vector.

Second, the paper links mixed-membership network models ([Airolidi et al. 2008](#), [Zhang et al. 2020](#), [Fan et al. 2022](#), [Jin et al. 2024](#)) to probabilistic graphical models and panel dependence. Unlike the usual mixed-membership network setting, the network is not directly observed. It must be recovered from the concentration matrix of a panel of time series, so estimation is affected by both realization noise in the weighted network and sampling error in the estimated concentration matrix. The concentration matrix connects the model to graphical models because, under conditional Gaussianity, its off-diagonal zeros encode conditional-independence restrictions ([Koller & Friedman 2009](#), [Munro & Ng 2022](#)). Our two-path specification also separates the extensive margin of conditional dependence—whether a link is present, governed by memberships, group connectivity, and node-specific propensity—from its intensive margin—the expected strength of an existing link, governed by sociability—in the spirit of two-part models for weighted networks ([Simpson & Laurienti 2015](#)). In this way, mixed-membership network methods can be used to study cross-sectional dependence in panel data ([Zhu et al. 2019](#), [Chen et al. 2022](#)).

The remainder of the paper is organized as follows. Section 2 presents the model, estimator, and theory. Sections 3 and 4 report the simulation and empirical results, respectively, and Section 5 concludes.

2 Methodology

This section develops the model and the estimation theory.

2.1 The Two-Path Mixed-Membership Partial-Correlation Network

We model the contemporaneous conditional-dependence structure of an N -dimensional, zero-mean, stationary panel time series $\{Y_t\}_{t=1}^T$ that is Gaussian conditional on the latent weighted network W introduced below. Conditional on W , let $\Sigma = \text{Var}(Y_t | W)$ denote the contemporaneous covariance matrix and let $\Lambda = \Sigma^{-1}$ denote the corresponding concentration, or precision, matrix (Dempster 1972). All distributional statements concerning Y_t in this subsection are understood conditional on W . The partial correlation between Y_{it} and Y_{jt} given the remaining components satisfies

$$\rho_{ij} = \text{Corr}(Y_{it}, Y_{jt} | \{Y_{st} : s \neq i, j\}, W) = -\frac{\Lambda_{ij}}{\sqrt{\Lambda_{ii}\Lambda_{jj}}}, \quad i \neq j. \quad (1)$$

Thus Λ and the partial-correlation matrix share the same off-diagonal sparsity pattern. Under conditional Gaussianity, $\Lambda_{ij} = 0$ is equivalent to conditional independence between Y_{it} and Y_{jt} given the remaining components and W (Pourahmadi 2013).

Following the normalized-Laplacian construction of Brownlees et al. (2022), we represent this conditional-dependence structure through a latent weighted graph. Let $W \in \mathbb{R}^{N \times N}$ be its symmetric, nonnegative weighted adjacency matrix, with $W_{ii} = 0$. Let $d_{W,i} = \sum_{j=1}^N W_{ij}$ and $D_W = \text{diag}(d_{W,1}, \dots, d_{W,N})$, and use the convention $(D_W^{-1/2})_{ii} = 0$ whenever $d_{W,i} = 0$. Writing $M_W = D_W^{-1/2} W D_W^{-1/2}$ for the normalized weighted adjacency matrix, the symmetric normalized graph Laplacian is

$$L = I_N - M_W = I_N - D_W^{-1/2} W D_W^{-1/2}. \quad (2)$$

We map the weighted network into the concentration matrix as

$$\Lambda = \frac{1}{\sigma^2} (I_N + \phi L), \quad (3)$$

for $\sigma^2 > 0$, $\phi \geq 0$. Since L is positive semidefinite, Λ is positive definite for $\phi \geq 0$, and

hence defines a valid concentration matrix. Moreover, for $\phi > 0$, Λ and L have the same eigenspaces, so the spectral structure relevant for recovering the latent memberships is preserved under this mapping.

Because $W_{ii} = 0$, the diagonal entries of L equal one. Therefore, $\Lambda_{ii} = (1 + \phi)/\sigma^2$. Let $\mathcal{E}_N = \{\min_{1 \leq i \leq N} d_{W,i} > 0\}$ denote the event that the realized network contains no isolated nodes. On $\mathcal{E}_N$, for $i \neq j$,

$$\Lambda_{ij} = -\frac{\phi W_{ij}}{\sigma^2 \sqrt{d_{W,i} d_{W,j}}}, \quad \rho_{ij} = \frac{\phi}{1 + \phi} \frac{W_{ij}}{\sqrt{d_{W,i} d_{W,j}}} \geq 0. \quad (4)$$

Under the isolated-node convention above, $\Lambda_{ij} = 0$ whenever either node is isolated. Thus, for $\phi > 0$, $\Lambda_{ij} = 0$ if and only if $W_{ij} = 0$. The sparsity pattern of W therefore determines which pairs of variables are conditionally dependent, while its normalized weights determine the magnitudes of the corresponding partial correlations. When $\phi = 0$, the model reduces to conditional independence.

It remains to specify the latent structure that generates W . Building on [Brownlee et al. \(2022\)](#), we extend their network model in two directions: we replace pure community membership with mixed membership, letting each unit belong to several latent groups, and we add node-specific sociability that governs the intensity of conditional dependence separately from its sparsity pattern. This two-path model generates W , which maps into Λ through (2) and (3). Let $\Pi \in \mathbb{R}^{N \times K}$ collect the mixed-membership vectors $\pi_i \in \mathbb{R}^K$ as its rows, where K is the number of latent groups; Π is our main object of interest, since it determines the latent group structure underlying the partial-correlation network.

The first path governs sparsity. Conditional on (Π, B, Θ) , the upper triangular entries $\{A_{ij} : i < j\}$ are independent Bernoulli variables with success probabilities $\Theta_i \pi_i B \pi_j' \Theta_j$, where $B \in \mathbb{R}^{K \times K}$ is the group-level connectivity matrix, whose diagonal and off-diagonal entries govern within- and between-group connectivity, and Θ is diagonal with entries $\Theta_i > 0$ capturing node-specific propensity to form links. The second path governs link intensity:

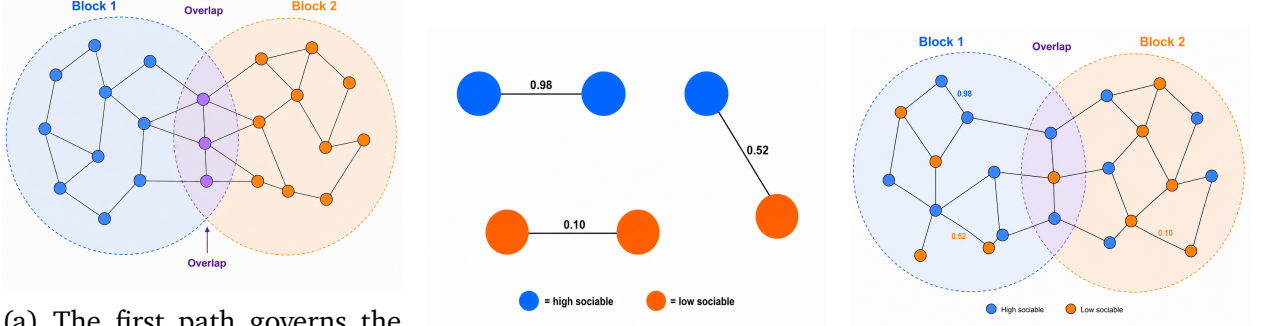
with $U \in \mathbb{R}^{N \times N}$ diagonal and $u_i \in (0, 1)$ the sociability of unit i , the weight $W_{ij} \mid A_{ij} = 1$ follows a Beta distribution with mean $u_i u_j$ and concentration $\kappa_{ij} > 0$, where larger κ_{ij} tightens dispersion around $u_i u_j$. The two paths thus separate the extensive margin, whether two units connect, from the intensive margin, the expected strength given a connection, while retaining the structure required for spectral recovery.

The distinction between the two paths is economically meaningful even though our primary estimation target is the membership matrix Π . Consider two units with similar membership compositions. A high value of Θ_i means that unit i has a relatively high propensity to form conditional-dependence links across the network, whereas a high value of u_i means that, conditional on a link being present, its interactions tend to be stronger. These are different margins. A unit may be connected to many other units but only weakly so, or it may have relatively few links that are individually strong. For regional economies, for example, common industrial composition may determine which regions are exposed to similar shocks, while openness, scale, or the intensity of interregional interaction may affect the strength with which those shocks are transmitted. For firms, membership may summarize overlapping industry or production-network exposures, while node-specific characteristics may determine how strongly an existing relationship contributes to conditional dependence.

This interpretation also clarifies what the model does and does not identify. The two paths are distinct in the generative law of the weighted network, but their node-specific components enter its first moment through the product $U\Theta$. Consequently, the first-moment spectral procedure developed below does not attempt to disentangle u_i from Θ_i . The purpose of the two-path formulation is instead to distinguish the economic mechanisms generating link incidence and link intensity while retaining a low-rank mixed-membership structure from which Π can be recovered. Separate estimation of the two node-specific channels would require additional information or restrictions beyond those used in the present paper.

The two-path mixed-membership network model is defined formally as follows.

Definition 1 (Two-Path Mixed-Membership Network Model). *Let $B \in [0, 1]^{K \times K}$ be symmet-*



(a) The first path governs the sparsity pattern. Mixed memberships generate overlapping latent groups.

(b) The second path governs connection intensity through node-specific sociability.

(c) Together, the two paths determine the weighted adjacency matrix W .

Figure 1: The two-path mixed-membership network construction. The resulting weighted adjacency matrix is mapped into the concentration matrix through (2) and (3).

ric and positive definite with $B_{kk} = 1$ for $k = 1, \dots, K$, let $\Theta_i > 0$, let $u_i \in (0, 1)$, and let the rows of Π lie on the simplex, $\pi_i \geq 0$ and $\pi_i \mathbf{1} = 1$. Assume $0 \leq \Theta_i \pi_i B \pi_j' \Theta_j \leq 1$ for all $i \neq j$ and $0 < \underline{\kappa} \leq \kappa_{ij} \leq \bar{\kappa} < \infty$. For $1 \leq i < j \leq N$, the weighted network is generated sequentially by

$$A_{ij} \mid \pi_i, \pi_j, B, \Theta_i, \Theta_j \sim \text{Bernoulli}(\Theta_i \pi_i B \pi_j' \Theta_j),$$

$$W_{ij} \mid A_{ij}, \kappa_{ij}, u_i, u_j = \begin{cases} \text{Beta}(u_i u_j \kappa_{ij}, (1 - u_i u_j) \kappa_{ij}), & A_{ij} = 1, \\ 0, & A_{ij} = 0. \end{cases}$$

We impose $A_{ji} = A_{ij}$, $W_{ji} = W_{ij}$, $A_{ii} = W_{ii} = 0$, and $\kappa_{ij} = \kappa_{ji}$. Conditional on the latent parameters $(\Pi, B, \Theta, U, \kappa)$, the unordered pairs $\{(A_{ij}, W_{ij}) : 1 \leq i < j \leq N\}$ are mutually independent.

The normalization $B_{kk} = 1$ fixes the group-specific diagonal rescaling inherent in degree-corrected mixed-membership models (Jin et al. 2024). Our identification target is Π . Since U and Θ enter the population signal only through $U\Theta$, they are not separately identified without further restrictions. Under the pure-node condition, however, the rank- K population signal identifies Π up to a permutation of its columns. Thus estimation of Π uses only the first-moment factorization of the weighted network; the two paths are retained to

distinguish economically between link formation and link intensity. Formal identification is discussed in Section B.1 of the supplementary material.

The sequential construction in Definition 1, together with conditional independence across unordered pairs, implies

$$\begin{aligned} \mathbb{P}(W, A \mid B, \Theta, \Pi, U, \kappa) &= \prod_{1 \leq i < j \leq N} \mathbb{P}(W_{ij} \mid A_{ij}, \kappa_{ij}, u_i, u_j) \\ &\times \prod_{1 \leq i < j \leq N} \mathbb{P}(A_{ij} \mid B, \pi_i, \pi_j, \Theta_i, \Theta_j). \end{aligned} \quad (5)$$

The factorization in (5) makes the sequential interpretation of the model explicit. Conditional on (A, U, κ) , the distribution of link weights does not depend further on (Π, B, Θ) , whereas conditional on (Π, B, Θ) , the distribution of link formation does not involve (U, κ) . The first path therefore determines where conditional-dependence links can occur, and the second determines the distribution of their weights once they occur. Conditional independence across unordered pairs then permits the joint law to be written as the product in (5).

Under Definition 1,

$$\mathbb{E}(W_{ij} \mid A_{ij} = 1) = u_i u_j, \quad \text{Var}(W_{ij} \mid A_{ij} = 1) = \frac{u_i u_j (1 - u_i u_j)}{1 + \kappa_{ij}}, \quad (6)$$

so κ_{ij} affects dispersion but not the conditional mean. The resulting population signal is given next.

Theorem 1. *Consider the 2PMM of Definition 1 and treat the latent parameters as fixed. Define $\Omega_W = U\Theta\Pi B\Pi'\Theta U$. Then, for every $i \neq j$, $\mathbb{E}(W_{ij}) = (\Omega_W)_{ij} = \Theta_i \pi_i B \pi_j' \Theta_j u_i u_j$, whereas $\mathbb{E}(W_{ii}) = 0$.*

Theorem 1 shows how the two paths combine in the first moment of the weighted network. For $i \neq j$, $\mathbb{E}(W_{ij}) = \mathbb{P}(A_{ij} = 1) \mathbb{E}(W_{ij} \mid A_{ij} = 1)$, so link incidence contributes $\Theta_i \pi_i B \pi_j' \Theta_j$ and conditional link intensity contributes $u_i u_j$. Their product yields the low-

rank factorization $\Omega_W = U\Theta\Pi\Pi'\Theta U$. Hence the two node-specific channels enter the population spectral signal through the effective scale $U\Theta$, whereas the mixed-membership matrix Π continues to determine its low-rank cross-sectional geometry.

Because $W_{ii} = 0$, $\mathbb{E}(W) = \Omega_W - \text{diag}(\Omega_W)$ rather than exactly Ω_W . We nevertheless use the full matrix Ω_W as the ideal population signal. Retaining its diagonal preserves the exact low-rank mixed-membership factorization that generates the simplex representation below, while the diagonal discrepancy is of smaller order after degree normalization and is absorbed into the network perturbation in Lemma 3. This distinction is useful conceptually: Ω_W is not an additional observed network, but the deterministic low-rank signal around which the realized weighted network fluctuates.

Combining the two-path network model with the normalized-Laplacian map yields the partial-correlation network model.

Definition 2 (Two-Path Mixed-Membership Partial-Correlation Network). *Let W be generated according to Definition 1. Let L and Λ be defined by (2) and (3), using the isolated-node convention for $D_W^{-1/2}$ stated above. Conditional on W , the process $\{Y_t\}_{t=1}^T$ is Gaussian with covariance matrix $\text{Var}(Y_t | W) = \Sigma = \Lambda^{-1}$.*

Definition 2 formally combines the two layers of the model. The latent parameters $(\Pi, B, \Theta, U, \kappa)$ generate the weighted network W , while W determines the concentration matrix and hence the conditional-dependence structure of Y_t . Under this convention, the symmetric normalized Laplacian is positive semidefinite, so $\Lambda = \sigma^{-2}(I_N + \phi L)$ is positive definite for every $\sigma^2 > 0$ and $\phi \geq 0$. Moreover, under Assumption 2(iv), the expected weighted degree diverges faster than $\log N$, and a standard concentration argument gives $\mathbb{P}(\mathcal{E}_N) \rightarrow 1$. Hence the isolated-node convention is asymptotically inactive.

The non-negativity of the partial correlations follows from the restriction $W_{ij} \geq 0$. We retain this restriction because Barigozzi & Brownlees (2019) and Brownlees et al. (2018) document that negative partial correlations are rare in network models for economic panels. Allowing signed link weights would require a corresponding modification of the normaliza-

tion and is outside the scope of the present model.

2.2 Algorithm

We estimate the mixed-membership structure by adapting Mixed-SCORE (Jin et al. 2024) to the partial-correlation network. Given the panel $\{Y_t\}_{t=1}^T$ and a value $\hat{K}$ for the number of groups, let $\mathbf{Y} = (Y_1, \dots, Y_T)' \in \mathbb{R}^{T \times N}$ and compute the sample covariance matrix $\hat{\Sigma} = \mathbf{Y}'\mathbf{Y}/T$. Throughout, we assume $T > N$ and that $\hat{\Sigma}$ is nonsingular, and set $\hat{\Lambda} = \hat{\Sigma}^{-1}$. A heuristic for choosing $\hat{K}$ in empirical implementation is described in Appendix B.4.

It is useful to view the estimator as three transformations. First, the panel is mapped into an estimated conditional-dependence network through $\hat{\Sigma}$ and $\hat{\Lambda}$. Second, the leading eigenspace of the sign-reversed off-diagonal concentration matrix is transformed by SCORE ratios, which remove the effective node-specific scale and expose a simplex in $\mathbb{R}^{K-1}$. Third, the vertices of this simplex are estimated and each observation is expressed in barycentric coordinates; the Mixed-SCORE normalization then converts these coordinates into membership weights. The first transformation is specific to the present panel setting, while the latter two inherit the geometry of Mixed-SCORE.

For $\phi > 0$, the sign-reversed off-diagonal part of Λ is a positive scalar multiple of the normalized weighted adjacency matrix $M_W = D_W^{-1/2} W D_W^{-1/2}$ of the realized graph. We therefore define

$$M := -\text{offdiag}(\Lambda). \quad (7)$$

By Definition 2, $M = (\phi/\sigma^2)M_W$, so M , M_W , and L share the same eigenvectors and the positive factor ϕ/σ^2 does not affect the SCORE ratios. Because M_W already incorporates degree normalization, M retains the degree-corrected mixed-membership structure required by Mixed-SCORE. We therefore apply Mixed-SCORE directly to $\hat{M} = -\text{offdiag}(\hat{\Lambda})$, its sample counterpart, without additional degree normalization; the error from replacing Λ by $\hat{\Lambda}$ is handled by the spectral perturbation analysis below.

Under the conditions imposed below, the leading population eigenvector has strictly positive entries, so the population SCORE ratios are well defined. To stabilize sample ratios when an estimated denominator is small while preserving the rotational formulation used when the non-leading eigenvectors are not individually identified, Step 4 uses Euclidean projection onto a ball. This projection is nonexpansive and orthogonally equivariant. Let $\tau_N > 0$ be a deterministic sequence satisfying $\tau_N \rightarrow 0$. The floor τ_N in Step 7 guarantees that the finite-sample estimator is real valued and is asymptotically inactive because the corresponding population quadratic forms are bounded away from zero after rescaling. The remaining zero-denominator, singularity, and normalization conventions in the algorithm are likewise asymptotically inactive.

Algorithm 1 PartialCorr-Mixed-SCORE

Input: Panel data $\{Y_t\}_{t=1}^T$ and estimated number of groups $\hat{K}$.

- 1: Compute $\hat{\Sigma} = \mathbf{Y}'\mathbf{Y}/T$ and $\hat{\Lambda} = \hat{\Sigma}^{-1}$.
- 2: Form the sign-reversed off-diagonal matrix $\hat{M}$ whose diagonal entries are zero and whose (i, j) -th entry is $-\hat{\Lambda}_{ij}$ for $i \neq j$.
- 3: Let $\hat{\lambda}_1$ be the largest eigenvalue of $\hat{M}$ and let $\hat{x}_1$ be its corresponding eigenvector. Let $\hat{\lambda}_2, \dots, \hat{\lambda}_{\hat{K}}$ be the remaining $\hat{K} - 1$ eigenvalues largest in magnitude, ordered such that $|\hat{\lambda}_2| \geq \dots \geq |\hat{\lambda}_{\hat{K}}|$, with corresponding eigenvectors $\hat{x}_2, \dots, \hat{x}_{\hat{K}}$. Choose the sign of $\hat{x}_1$ such that $\mathbf{1}'\hat{x}_1 > 0$, and set $\hat{X} = [\hat{x}_1, \dots, \hat{x}_{\hat{K}}]$.
- 4: Set $S = \sqrt{\hat{K}} \log(N)$. For each i , form the untruncated ratio vector $\tilde{r}_i \in \mathbb{R}^{\hat{K}-1}$ with entries $\tilde{r}_i(k) = \hat{X}_{i,k+1}/\hat{X}_{i,1}$ for $1 \leq k \leq \hat{K} - 1$, using the convention $\tilde{r}_i = 0$ if $\hat{X}_{i,1} = 0$. Set $\hat{r}_i = \tilde{r}_i$ if $\|\tilde{r}_i\|_2 \leq S$ and $\hat{r}_i = S\tilde{r}_i/\|\tilde{r}_i\|_2$ otherwise, and let $\hat{R} = (\hat{r}_1, \dots, \hat{r}_N)'$.
- 5: Apply a vertex-search procedure to the rows of $\hat{R}$ and obtain the estimated vertices $\{\hat{v}_1, \dots, \hat{v}_{\hat{K}}\} \subset \mathbb{R}^{\hat{K}-1}$.
- 6: For each i , solve $(1, \hat{R}_i')' = \begin{pmatrix} 1 & \dots & 1 \\ \hat{v}_1 & \dots & \hat{v}_{\hat{K}} \end{pmatrix} \hat{w}_i$; if the augmented vertex matrix is singular, take the minimum-norm solution through its Moore–Penrose inverse.
- 7: For $1 \leq k \leq \hat{K}$, compute $\hat{b}_{1k} = [\max\{\hat{\lambda}_1 + \hat{v}_k' \text{diag}(\hat{\lambda}_2, \dots, \hat{\lambda}_{\hat{K}})\hat{v}_k, \tau_N\}]^{-1/2}$.
- 8: For $1 \leq i \leq N$ and $1 \leq k \leq \hat{K}$, set $\hat{\pi}_{ik}^* = \max\{0, \hat{w}_{ik}/\hat{b}_{1k}\}$.
- 9: If $\|\hat{\pi}_i^*\|_1 > 0$, set $\hat{\pi}_i = \hat{\pi}_i^*/\|\hat{\pi}_i^*\|_1$; otherwise set $\hat{\pi}_i = \hat{K}^{-1}\mathbf{1}$, for $i = 1, \dots, N$.

Output: Estimated mixed-membership matrix $\hat{\Pi} \in \mathbb{R}^{N \times \hat{K}}$.

Relative to SCORE and Mixed-SCORE, which operate on an observed adjacency matrix, PartialCorr-Mixed-SCORE starts from panel observations, estimates the concentration

matrix, and uses its sign-reversed off-diagonal component to recover the normalized latent network; the subsequent SCORE-ratio, vertex-search, and membership-reconstruction steps follow the Mixed-SCORE construction. We use the Sketched Vertex Search of [Jin et al. \(2024\)](#) in the implementation, while the theory applies to any vertex-search procedure satisfying Assumption 6.

2.3 Theory

This section establishes theoretical guarantees for PartialCorr-Mixed-SCORE. The analysis reflects the two layers of randomness in the model. The first arises from the generation of the weighted network W ; the second arises from observing the panel $\{Y_t\}_{t=1}^T$ conditional on W . We first establish the population simplex geometry associated with the ideal low-rank signal, then control the deviations introduced by the realized network and by estimation of the concentration matrix.

Throughout this section, K may depend on N and T and is allowed to diverge. The analysis concerns the unregularized regime $T > N$ with $N/T \rightarrow 0$; regularized precision-matrix estimation when $N \geq T$ is outside the scope of the present theory. All constants appearing in the assumptions below are independent of N , T , and K . The results are stated for the true number of groups. If K is estimated, we assume that the selection procedure satisfies $\mathbb{P}(\hat{K} = K) \rightarrow 1$. Because mixed-membership labels are not intrinsically ordered, estimation error is evaluated up to a permutation of the columns of Π . Model identification is discussed in Section B.1 of the supplementary material.

Define the population expected degrees by $d_{\Omega,i} = \sum_{j \neq i} (\Omega_W)_{ij}$ and let $\mathcal{D}_\Omega = \text{diag}(d_{\Omega,1}, \dots, d_{\Omega,N})$. The population normalized adjacency signal, i.e. the population counterpart of M_W obtained by replacing W and D_W by Ω_W and $\mathcal{D}_\Omega$, is

$$\mathcal{M}_W = \mathcal{D}_\Omega^{-1/2} \Omega_W \mathcal{D}_\Omega^{-1/2} = (\mathcal{D}_\Omega^{-1/2} U \Theta \Pi) B (\Pi' \Theta U \mathcal{D}_\Omega^{-1/2}). \quad (8)$$

We retain the full matrix Ω_W in (8), including its diagonal, because this preserves the rank- K mixed-membership factorization. The realized weighted adjacency matrix W has a zero diagonal; the resulting diagonal discrepancy, together with the sampling variation of W , is treated as part of the network perturbation below.

Assumption 1 (Pure-node representation). *For each latent group $k \in \{1, \dots, K\}$, define $\mathcal{I}_k = \{i \in \{1, \dots, N\} : \pi_i = e'_k\}$, where e_k is the k -th canonical basis vector in $\mathbb{R}^K$. There exists a constant $c_0 > 0$ such that $|\mathcal{I}_k| \geq c_0 N/K$ for $k = 1, \dots, K$.*

Assumption 1 is a separability condition. For fixed K , it reduces to the usual positive-fraction condition. When K increases, each group is required to contain a positive fraction of its natural size N/K in pure nodes. Thus all population simplex vertices remain represented without imposing a fixed positive fraction of the entire cross section for each group.

Assumption 2 (Population regularity). *Let $G = \Pi' \Theta U \mathcal{D}_\Omega^{-1} U \Theta \Pi$. There exist constants $0 < c_B < C_B < \infty$, $0 < c_G < C_G < \infty$, $c_\lambda > 0$, $0 < C_d < \infty$, $0 < c_\phi < C_\phi < \infty$, and $c_\sigma > 0$ such that:*

1. $c_B \leq \lambda_{\min}(B) \leq \lambda_{\max}(B) \leq C_B$ and $c_G \leq \lambda_{\min}(G) \leq \lambda_{\max}(G) \leq C_G$.
2. Ω_W is irreducible.
3. Let $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_K > 0$ denote the K nonzero eigenvalues of $\mathcal{M}_W$. Then $\lambda_1 - \lambda_2 \geq c_\lambda$.
4. With $d_{\min} = \min_i d_{\Omega,i}$ and $d_{\max} = \max_i d_{\Omega,i}$, $d_{\min}/\log N \rightarrow \infty$, $d_{\max}/d_{\min} \leq C_d$ and $\max_i (\Omega_W)_{ii} \leq C_d$.
5. $c_\phi \leq \phi/\sigma^2 \leq C_\phi$ and $\sigma^2 \geq c_\sigma$.

Assumption 2(i) keeps the rank- K signal uniformly nondegenerate, since the nonzero eigenvalues of $\mathcal{M}_W$ are those of $B^{1/2}GB^{1/2}$ and $\mathcal{D}_\Omega^{-1/2}U\Theta\Pi$ has full column rank. Parts (ii)–(iii) identify the Perron eigenvector while allowing multiplicities among the remaining

signal eigenvalues, so the non-leading eigenvectors are identified only as an eigenspace. Part (iv) provides the degree concentration required below while permitting bounded node heterogeneity, and part (v) rules out a vanishing partial-correlation signal.

The bounded-degree-ratio restriction in part (iv) is worth distinguishing from the degree correction in the model itself. The parameters Θ_i and u_i allow units to differ systematically in their propensity to connect and in the strength of their interactions. For the present asymptotic theory, however, we restrict this heterogeneity to remain of comparable order across units. Thus the framework accommodates economically meaningful cross-sectional heterogeneity but does not cover sequences in which a vanishing fraction of nodes becomes asymptotically dominant in weighted degree. Allowing $d_{\max}/d_{\min}$ to diverge would require a correspondingly sharper row-specific spectral analysis and is not needed for the membership result developed here.

2.3.1 Population geometry

Let $\mathcal{X} = [x_1, \dots, x_K] \in \mathbb{R}^{N \times K}$ collect an orthonormal basis of the eigenspace associated with the nonzero eigenvalues of $\mathcal{M}_W$, with x_1 denoting the leading eigenvector.

Lemma 1. *Consider the 2PMM of Definition 1 satisfying Assumption 2. For any fixed orthonormal basis $\mathcal{X}$ defined above, there exists a unique nonsingular real matrix $P \in \mathbb{R}^{K \times K}$ such that $\mathcal{X} = (\mathcal{D}_\Omega^{-1/2} U \Theta \Pi) P$.*

For unit i , let $\mathcal{X}_{i,1}$ denote the first entry of the i -th row of $\mathcal{X}$ and let $\mathcal{X}_{i,-1} \in \mathbb{R}^{1 \times (K-1)}$ contain its remaining entries. Define the population SCORE-ratio vector by $r_i = \mathcal{X}'_{i,-1} / \mathcal{X}_{i,1} \in \mathbb{R}^{K-1}$ and let $R = (r_1, \dots, r_N)'$.

Lemma 2. *Consider the 2PMM of Definition 1 satisfying Assumptions 1 and 2. Then:*

1. *The leading eigenvector x_1 can be chosen to have strictly positive entries, so every population SCORE ratio r_i is well defined.*

2. *There exists a nondegenerate K -vertex simplex $\mathcal{S} \subset \mathbb{R}^{K-1}$ with vertices $v_1, \dots, v_K \in \mathbb{R}^{K-1}$.*
3. *Every r_i lies in $\mathcal{S}$, and $r_i = v_k$ whenever $\pi_i = e'_k$.*
4. *Let $w_i \in \mathbb{R}^K$ denote the barycentric coordinate vector satisfying $r_i = \sum_{k=1}^K w_i(k)v_k$. Then $w_i = (\pi'_i \circ b_1) / \|\pi'_i \circ b_1\|_1$, where $\circ$ denotes entry-wise multiplication and $b_{1,k} = [\lambda_1 + v'_k \text{diag}(\lambda_2, \dots, \lambda_K)v_k]^{-1/2}$. Consequently, $\pi'_i = (w_i \oslash b_1) / \|w_i \oslash b_1\|_1$, where $\oslash$ denotes entry-wise division.*

Lemma 2 is the population simplex representation underlying Mixed-SCORE. The effective node scale $\mathcal{D}_\Omega^{-1/2}U\Theta$ is diagonal, so the standard SCORE normalization of Jin (2015) removes it. The formulation is invariant to orthogonal rotations of the non-leading signal eigenspace and therefore permits multiplicities among $\lambda_2, \dots, \lambda_K$.

2.3.2 Two-layer spectral perturbation

The estimator acts on $\widehat{M} = -\text{offdiag}(\widehat{\Lambda})$, the sample counterpart of $M = -\text{offdiag}(\Lambda) = (\phi/\sigma^2)M_W$, whereas the population geometry of Lemma 2 is defined by the ideal signal $\mathcal{M}_W$. The total perturbation can therefore be decomposed as

$$\widehat{M} - \frac{\phi}{\sigma^2}\mathcal{M}_W = (\widehat{M} - M) + \frac{\phi}{\sigma^2}(M_W - \mathcal{M}_W). \quad (9)$$

The first term is the time-series error from estimating the concentration matrix conditional on the realized network; the second is the network error from replacing the ideal normalized signal by the realized weighted network.

Assumption 3 (Time-series covariance concentration). *Conditional on W , $\{Y_t\}$ is a zero-mean stationary process with covariance matrix $\Sigma = \text{Var}(Y_t | W)$. Assume $T > N$, $N/T \rightarrow 0$,*

and, under the joint law of $(W, Y_1, \dots, Y_T)$,

$$\left\| \frac{1}{T} \sum_{t=1}^T \Sigma^{-1/2} Y_t Y_t' \Sigma^{-1/2} - I_N \right\| = O_p \left(\sqrt{\frac{N}{T}} \right). \quad (10)$$

Assumption 3 directly imposes the operator-norm covariance concentration used in the spectral analysis below. The restriction $T > N$, together with the imposed nonsingularity of $\widehat{\Sigma}$, places the analysis in the unregularized regime, while $N/T \rightarrow 0$ ensures that the covariance-estimation error vanishes in operator norm. The condition holds, for example, in the conditionally i.i.d. Gaussian case when $N/T \rightarrow 0$. More generally, serial dependence is permitted whenever the standardized sample covariance satisfies the stated rate. We impose this concentration condition directly rather than additional primitive mixing and tail restrictions, since the latter are not otherwise used in the proofs.

Define $\delta_{N,T} = \sqrt{\log N / d_{\min}} + \sqrt{N/T}$.

Lemma 3 (Two-layer matrix deviation). *Consider the 2PMM-PCN of Definition 2 satisfying Assumption 2, with panel $\{Y_t\}$ satisfying Assumption 3. Then*

$$\|M_W - \mathcal{M}_W\| = O_p \left(\sqrt{\frac{\log N}{d_{\min}}} \right), \quad (11)$$

$$\|\widehat{M} - M\| = O_p \left(\sqrt{\frac{N}{T}} \right), \quad (12)$$

$$\left\| \widehat{M} - \frac{\phi}{\sigma^2} \mathcal{M}_W \right\| = O_p(\delta_{N,T}). \quad (13)$$

Equation (11) absorbs both the sampling variation of the weighted edges and the discrepancy between the zero diagonal of W and the diagonal of the low-rank signal Ω_W . The latter is of order $\max_i (\Omega_W)_{ii} / d_{\min} = O(d_{\min}^{-1})$ and is dominated by the rate in (11). Equation (12) follows from Assumption 3, perturbation of the matrix inverse, and $-\text{offdiag}(\Lambda) = M$; here the perturbation constant is controlled by $\|\Lambda\| \leq \sigma^{-2} + 2\phi/\sigma^2 \leq c_\sigma^{-1} + 2C_\phi = O(1)$ under Assumption 2(v), since the eigenvalues of L lie in $[0, 2]$. The constants are uniform

in K under Assumption 2.

Let $\hat{X}_{-1} = [\hat{x}_2, \dots, \hat{x}_K]$ and $\mathcal{X}_{-1} = [x_2, \dots, x_K]$.

Theorem 2. *Consider the 2PMM-PCN of Definition 2 satisfying Assumption 2, with panel $\{Y_t\}$ satisfying Assumption 3. Let $\hat{\lambda}_1, \dots, \hat{\lambda}_K$ and $\hat{X} = [\hat{x}_1, \dots, \hat{x}_K]$ be selected as in Algorithm 1. Then $\delta_{N,T} \rightarrow 0$, and with probability tending to one Algorithm 1 selects the K eigenvalues converging to $(\phi/\sigma^2)\lambda_1, \dots, (\phi/\sigma^2)\lambda_K$; in particular, all selected signal eigenvalues are positive. There exists a sign $s \in \{-1, 1\}$ such that*

$$\left| \hat{\lambda}_1 - \frac{\phi}{\sigma^2} \lambda_1 \right| + \|s\hat{x}_1 - x_1\|_2 = O_p(\delta_{N,T}). \quad (14)$$

Moreover, there exists an orthogonal matrix $\mathbf{H} \in \mathbb{R}^{(K-1) \times (K-1)}$ such that

$$\|\hat{X}_{-1}\mathbf{H}' - \mathcal{X}_{-1}\|_F = O_p(\sqrt{K} \delta_{N,T}), \quad (15)$$

$$\left\| \mathbf{H} \text{diag}(\hat{\lambda}_2, \dots, \hat{\lambda}_K) \mathbf{H}' - \frac{\phi}{\sigma^2} \text{diag}(\lambda_2, \dots, \lambda_K) \right\| = O_p(\delta_{N,T}). \quad (16)$$

Theorem 2 treats the leading eigenvector separately from the remaining signal eigenspace. The leading eigenvector is individually identified by the gap in Assumption 2(iii), whereas the remaining $K - 1$ eigenvectors are aligned through $\mathbf{H}$ and need not be individually identified. This avoids pairwise eigengaps among $\lambda_2, \dots, \lambda_K$. The $\sqrt{K}$ factor in (15) is the conversion from an operator-norm subspace perturbation to a Frobenius bound for a $(K - 1)$ -dimensional basis. Under Assumption 4, the sign convention in Step 3 of Algorithm 1 selects the sign s in Theorem 2 with probability tending to one.

2.3.3 SCORE ratios and membership recovery

Assumption 4 (Population delocalization). *There exist constants $0 < c_x < C_x < \infty$ such that $c_x/\sqrt{N} \leq \mathcal{X}_{i,1} \leq C_x/\sqrt{N}$ and $\max_{2 \leq k \leq K} |\mathcal{X}_{i,k}| \leq C_x/\sqrt{N}$ for $i = 1, \dots, N$.*

Assumption 4 is a coordinate-wise delocalization condition. In particular, it permits

$\|\mathcal{X}_{i,\cdot}\|_2$ to be of order $\sqrt{K/N}$ and therefore does not implicitly bound K . It also implies that every population SCORE-ratio coordinate is uniformly bounded.

The estimator applies no rotation; $\mathbf{H}$ only aligns the sample and population coordinate systems, and the Euclidean projection commutes with it, so the SCORE perturbation can be analyzed through $\|\mathbf{H}\widehat{r}_i - r_i\|_2$.

Theorem 3. *Consider the 2PMM-PCN of Definition 2 satisfying Assumptions 1, 2, and 4, with panel $\{Y_t\}$ satisfying Assumption 3. If $K \log^2 N \delta_{N,T}^2 \rightarrow 0$, then*

$$\begin{aligned} \frac{1}{N} \|\widehat{R}\mathbf{H}' - R\|_F^2 &= O_p(K \log^2 N \delta_{N,T}^2) \\ &= O_p\left[K \log^2 N \left(\frac{\log N}{d_{\min}} + \frac{N}{T}\right)\right]. \end{aligned} \quad (17)$$

The factor K in Theorem 3 is the dimensional cost of estimating the K -dimensional signal eigenspace. The logarithmic factor arises from the Euclidean truncation radius $S = \sqrt{K} \log N$ used to control rows with small estimated leading coordinates, together with the use of a global Frobenius eigenspace bound. No inverse pairwise-eigengap factor appears because the non-leading eigenvectors are aligned as a single eigenspace.

Let $\mathcal{V} = \begin{pmatrix} 1 & \cdots & 1 \\ v_1 & \cdots & v_K \end{pmatrix} \in \mathbb{R}^{K \times K}$ collect the augmented vertices as its columns.

Assumption 5 (Well-conditioned simplex). *There exist constants $0 < c_V < C_V < \infty$ such that $c_V \sqrt{K} \leq \sigma_{\min}(\mathcal{V})$ and $\|\mathcal{V}\| \leq C_V \sqrt{K}$.*

Assumption 5 is the growing-dimensional analogue of the usual simplex-conditioning requirement. It implies $\|\mathcal{V}^{-1}\| = O(K^{-1/2})$. It also gives the uniform vertex scaling needed below. Indeed, for each k , $\sigma_{\min}(\mathcal{V}) \leq \|\mathcal{V}e_k\|_2 = (1 + \|v_k\|_2^2)^{1/2} \leq \|\mathcal{V}\|$, and hence $\|v_k\|_2 \asymp \sqrt{K}$ uniformly in k . Combining this with Assumption 2(i) and the expression for $b_{1,k}$ in Lemma 2 gives $b_{1,k} \asymp K^{-1/2}$ uniformly in k . Thus these uniform properties are consequences of Assumption 5, rather than additional assumptions.

Let $\mathfrak{P}_K$ denote the set of $K \times K$ permutation matrices.

Assumption 6 (Strongly efficient vertex search). *Let $\hat{v}_1, \dots, \hat{v}_K$ be the vertices returned by the vertex-search step. For the same orthogonal matrix $\mathbf{H}$ as in Theorem 2, there exists a random $\mathcal{P} \in \mathfrak{P}_K$ such that, uniformly over the admissible sequence of K ,*

$$\max_{1 \leq k \leq K} \|\mathbf{H}\hat{v}_{\mathcal{P}(k)} - v_k\|_2^2 = O_p\left(\frac{1}{N} \|\hat{R}\mathbf{H}' - R\|_F^2\right). \quad (18)$$

Assumption 6 is a high-level geometric stability condition. Its form is the growing- K analogue of the strong-efficiency criterion used for Mixed-SCORE: the same orthogonal transformation that aligns the SCORE coordinates also aligns the estimated simplex vertices. We impose this property directly and uniformly over the admissible sequence of K . Existing strong-efficiency results for SVS are established under additional membership configuration conditions and, in the results used here as motivation, fixed K . Thus the growing- K requirement in Assumption 6 should be viewed as an explicit high-level condition rather than as a consequence of the existing SVS theory. No separate orthogonal-equivariance assumption on the vertex-search map is needed because the required alignment is stated directly in (18).

Combining Theorem 3 and Assumption 6 gives $\max_k \|\mathbf{H}\hat{v}_{\mathcal{P}(k)} - v_k\|_2 = O_p(\sqrt{K} \log N \delta_{N,T})$.

Because the selected eigenvalues estimate $(\phi/\sigma^2)\lambda_k$, the sample scaling vector $\hat{b}_1$ estimates b_1 up to the common factor $(\phi/\sigma^2)^{-1/2}$. The common factor cancels under the final ℓ_1 normalization.

Lemma 4 (Stable scaling vector). *Under the conditions of Theorem 3 and Assumptions 5 and 6, with $\mathcal{P}$ from Assumption 6, $\max_{1 \leq k \leq K} |\hat{b}_{1,\mathcal{P}(k)} / \{(\phi/\sigma^2)^{-1/2} b_{1,k}\} - 1| = O_p(\log N \delta_{N,T})$. In particular, the quantities inside the square roots defining $\hat{b}_{1,k}$ are strictly positive with probability tending to one, so the floor τ_N in Algorithm 1 is asymptotically inactive.*

To see why no additional factor of K appears in Lemma 4, write the sample quadratic form after alignment as $(\mathbf{H}\hat{v}_k)' \{\mathbf{H} \text{diag}(\hat{\lambda}_2, \dots, \hat{\lambda}_K) \mathbf{H}'\} (\mathbf{H}\hat{v}_k)$. This is exactly equal to the quadratic form used by the algorithm before rotation. By Assumption 5, $\|v_k\|_2^2 \asymp K$ and

$b_{1,k}^{-2} \asymp K$. The perturbation of the quadratic form is therefore of order $K \log N \delta_{N,T}$, while the population quadratic form itself is of order K , yielding the relative rate in Lemma 4.

Lemma 5 (Stable barycentric coordinates). *Under the conditions of Theorem 3 and Assumptions 5 and 6, with $\mathcal{P}$ from Assumption 6,*

$$\frac{1}{N} \sum_{i=1}^N \|\mathcal{P}' \hat{w}_i - w_i\|_2^2 = O_p(\log^2 N \delta_{N,T}^2). \quad (19)$$

Lemma 5 follows from $\|\mathcal{V}^{-1}\| = O(K^{-1/2})$. The SCORE-coordinate error in Theorem 3 is therefore reduced by a factor K^{-1} after squaring. The vertex perturbation has the same rate: since w_i is a barycentric coordinate vector, $w_i \geq 0$ and $\|w_i\|_1 = 1$, the perturbation generated by the estimated vertices is a convex combination of the individual vertex errors and is controlled by their maximum rather than by their sum.

The membership reconstruction map combines the barycentric coordinates and the scaling vector. Since $b_{1,k} \asymp K^{-1/2}$ uniformly in k , the normalization in Lemma 2 implies that the contribution of the barycentric-coordinate error is at most of order K times its average squared ℓ_2 error, while a relative perturbation of b_1 enters without an additional factor of K . Consequently, Lemmas 4 and 5 yield

$$\frac{1}{N} \sum_{i=1}^N \|\mathcal{P}' \hat{\pi}'_i - \pi'_i\|_2^2 = O_p(K \log^2 N \delta_{N,T}^2). \quad (20)$$

This gives the main membership result.

Theorem 4. *Consider the 2PMM-PCN of Definition 2 satisfying Assumptions 1, 2, 4, 5, and 6, with panel $\{Y_t\}$ satisfying Assumption 3. If*

$$K \log^2 N \left(\frac{\log N}{d_{\min}} + \frac{N}{T} \right) \rightarrow 0, \quad (21)$$

then

$$\min_{\mathcal{P} \in \mathfrak{P}_K} \frac{1}{N} \|\widehat{\Pi} \mathcal{P} - \Pi\|_F^2 = O_p \left[K \log^2 N \left(\frac{\log N}{d_{\min}} + \frac{N}{T} \right) \right]. \quad (22)$$

Theorem 4 establishes consistency of the membership estimator up to a permutation of the latent-group labels while allowing K to diverge, provided its growth is slow relative to the network density and the time-series sample size. The two components of the rate correspond respectively to realization of the weighted network and estimation of the concentration matrix from the panel. The result does not require pairwise separation among the non-leading signal eigenvalues $\lambda_2, \dots, \lambda_K$, which may be arbitrarily close or have multiplicities; instead it relies on the high-level growing- K vertex-stability condition in Assumption 6. The time-series layer permits serial dependence whenever the covariance concentration condition in Assumption 3 holds; no restriction of the form $Y_t = \epsilon_t$ is imposed. The logarithmic loss reflects the use of truncated SCORE ratios together with a global eigenspace perturbation bound.

2.4 Discussion of the rate and scope

The rate in Theorem 4 separates the two statistical problems that arise because the network is latent rather than directly observed. The term $\log N/d_{\min}$ is associated with realization of the weighted network: larger expected degree makes the realized normalized network more tightly concentrated around its low-rank population signal. The term N/T is specific to observing that network indirectly through the panel and reflects estimation of the N -dimensional concentration matrix from T time-series observations. Consequently, increasing T can eliminate the second source of error but not the first; even with an arbitrarily long panel, the realized network remains a random draw from the 2PMM. Conversely, a dense network with a strong low-rank signal does not by itself eliminate the error from estimating its concentration structure when the time dimension is short.

The dependence on K has a different source. It is the dimensional cost of recovering a growing signal eigenspace and converting the resulting SCORE coordinates into membership vectors. Importantly, the result does not require pairwise eigengaps among $\lambda_2, \dots, \lambda_K$. These eigenvalues may coincide, in which case the associated eigenvectors are not individually identified. The analysis instead aligns the entire non-leading eigenspace by the orthogonal matrix $\mathbf{H}$. Since SCORE ratios, simplex vertices, and the corresponding barycentric coordinates rotate jointly, the membership representation remains well defined. What is required is separation of the rank- K signal from the null space and separate identification of the leading Perron eigenvector.

The restrictions $T > N$ and $N/T \rightarrow 0$ also play different roles. The former is a computational restriction associated with the unregularized estimator $\hat{\Lambda} = \hat{\Sigma}^{-1}$: it permits direct inversion of the sample covariance matrix. The latter is the asymptotic rate restriction that makes concentration-matrix estimation negligible in operator norm under Assumption 3. Thus $T > N$ should not be interpreted as an intrinsic requirement of mixed-membership recovery from a partial-correlation network. When $N \geq T$, the same conceptual procedure could be combined with a regularized precision-matrix estimator, but its first-stage error would have to replace the $\sqrt{N/T}$ term in the present analysis. We leave that high-dimensional extension outside the scope of the current paper.

Finally, Theorem 4 deliberately separates the panel-specific spectral analysis from the geometric requirements of the vertex-search step. Assumption 6 is imposed as a high-level stability condition: it requires the vertex estimator to transform the average SCORE perturbation into uniformly accurate estimates of the simplex vertices along the admissible growing- K sequence. Existing strong-efficiency arguments for vertex search motivate this condition, but the growing- K property is not derived here from primitive assumptions for a particular implementation. This formulation makes clear which part of the result comes from estimating a latent network from panel data and which part is inherited from mixed-membership simplex recovery. The $\log^2 N$ factor is likewise conservative and results

from the truncated SCORE ratios together with a global eigenspace perturbation argument; sharper row-wise analysis may reduce this loss under stronger conditions.

3 Simulation

We evaluate the finite-sample behavior of PartialCorr-Mixed-SCORE in a Monte Carlo experiment with $K = 3$ communities. To respect the normalization $B_{kk} = 1$, we set the diagonal entries of B equal to one and its common off-diagonal entry equal to $b_{\text{out}}/b_{\text{in}}$. We study three designs indexed by $(b_{\text{in}}, b_{\text{out}}) \in \{(0.25, 0.01), (0.50, 0.01), (0.25, 0.05)\}$, with within-to-between ratios of 25, 50, and 5.

Table 1: Estimation errors of PartialCorr-Mixed-SCORE.

$N = 120$					
N_0	$b_{\text{in}}/b_{\text{out}}$	$T = 300$	$T = 500$	$T = 800$	$T = 1000$
30	0.25/0.01	0.168	0.158	0.150	0.145
	0.50/0.01	0.150	0.122	0.104	0.100
	0.25/0.05	0.172	0.169	0.164	0.162
50	0.25/0.01	0.187	0.170	0.154	0.142
	0.50/0.01	0.151	0.102	0.068	0.056
	0.25/0.05	0.195	0.189	0.182	0.177
80	0.25/0.01	0.195	0.154	0.103	0.090
	0.50/0.01	0.115	0.045	0.026	0.022
	0.25/0.05	0.221	0.203	0.181	0.170
100	0.25/0.01	0.200	0.124	0.069	0.056
	0.50/0.01	0.082	0.031	0.019	0.016
	0.25/0.05	0.235	0.202	0.165	0.147
120	0.25/0.01	0.186	0.089	0.048	0.040
	0.50/0.01	0.057	0.024	0.016	0.013
	0.25/0.05	0.242	0.188	0.128	0.108

Notes: Each entry reports the average, over 500 simulation replications, of the minimum MSE over all permutations of the group labels.

The panel has $N = 120$ series, of which N_0 are pure and the remaining $N - N_0$ are mixed; we let $N_0 \in \{30, 50, 80, 100, 120\}$. Each pure series takes a canonical membership vector $\pi_i \in$

$\{[1, 0, 0], [0, 1, 0], [0, 0, 1]\}$, and the mixed series are split evenly between $\pi_i = [0.5, 0.5, 0]$ and $\pi_i = [0, 0.5, 0.5]$. The pure series are allocated as evenly as possible across the three groups. The sociabilities u_i are i.i.d. $\text{Uniform}[0.3, 1.0]$, and the degree-heterogeneity parameters Θ_i are equally spaced on $[0.9\sqrt{b_{\text{in}}}, \sqrt{b_{\text{in}}}]$. This parameterization gives exactly the same link-formation probabilities as using diagonal entries b_{in} , off-diagonal entries b_{out} , and degree-heterogeneity parameters equally spaced on $[0.9, 1.0]$, while satisfying the normalization $B_{kk} = 1$ imposed in Definition 1. We set $\kappa_{ij} = 1$ for all $i \neq j$ and set $\phi = \sigma^2 = 1$.

We draw the weighted network from the 2PMM of Definition 1, map it to the concentration matrix Λ through Definition 2, and set $\Sigma = \Lambda^{-1}$. The panel is generated as $Y_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$, $t = 1, \dots, T$. PartialCorr-Mixed-SCORE forms $\hat{\Sigma} = \mathbf{Y}'\mathbf{Y}/T$ and $\hat{\Lambda} = \hat{\Sigma}^{-1}$ and estimates each π_i . Over 500 replications, we report the minimum MSE over all permutations of the group labels, $\min_{\mathcal{P} \in \mathfrak{P}_K} N^{-1} \|\hat{\Pi}\mathcal{P} - \Pi\|_F^2$.

Table 1 collects the estimation results. The clearest pattern is the role of the time dimension. Holding the network design fixed, the error declines monotonically with T along every row, in line with the N/T component of Theorem 4. The improvement can be substantial when the latent groups are well separated. For example, under $(0.50, 0.01)$ and $N_0 = 120$, the MSE falls from 0.057 at $T = 300$ to 0.013 at $T = 1000$. This pattern is particularly relevant in the present setting because, unlike Mixed-SCORE applied to an observed network, PartialCorr-Mixed-SCORE must first learn the relevant network object from the time-series sample.

Block separation is equally important in finite samples. Recovery is most accurate under $(0.50, 0.01)$ and least accurate under $(0.25, 0.05)$. At $N_0 = 120$ and $T = 1000$, for example, the corresponding MSEs are 0.013 and 0.108, with $(0.25, 0.01)$ between them at 0.040. The effect of the number of pure nodes is more nuanced. Pure nodes identify and stabilize the simplex vertices, and their benefit becomes especially visible once the time-series noise is sufficiently small: under $(0.50, 0.01)$ and $T = 1000$, the MSE falls from 0.100 at $N_0 = 30$ to 0.013 at $N_0 = 120$. At shorter samples the pattern need not be monotone, because changing

N_0 affects the geometry of the SCORE cloud while concentration-matrix estimation remains noisy. This distinction also helps reconcile the simulation with the theory: Assumption 1 guarantees representation of the simplex vertices, whereas the strength of block separation is absorbed into the population spectral and simplex-conditioning constants rather than appearing as a separate term in the displayed asymptotic rate.

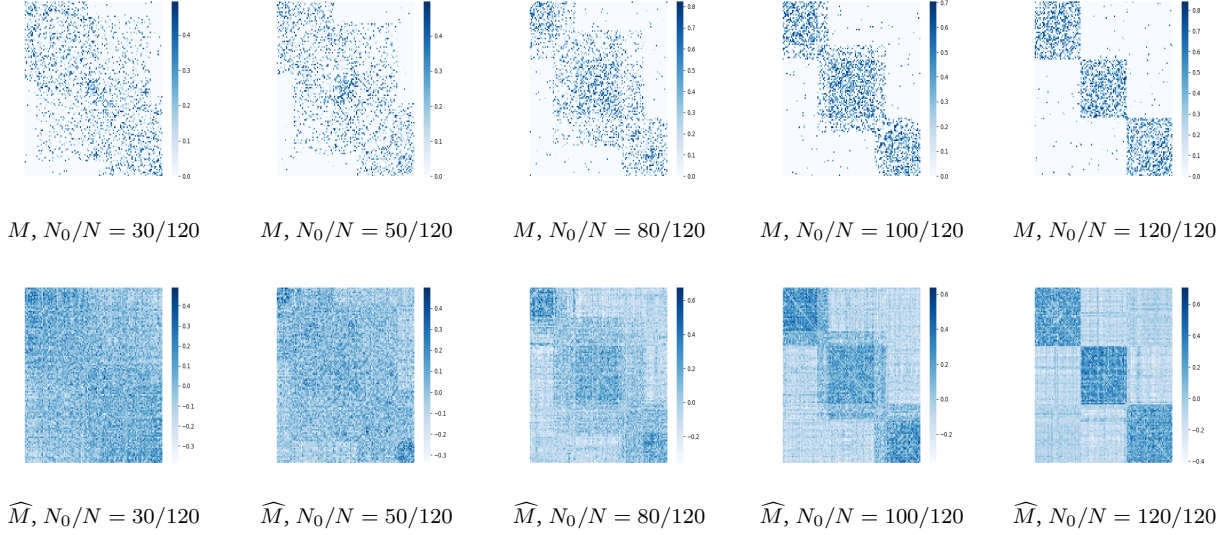


Figure 2: Simulated and estimated network structures as the share of pure nodes varies. The upper panels show the sign-reversed off-diagonal concentration matrix $M = -\text{offdiag}(\Lambda)$, and the lower panels show its sample counterpart $\widehat{M} = -\text{offdiag}(\widehat{\Lambda})$. Since $\phi = \sigma^2 = 1$ in the simulation, $M = (\phi/\sigma^2)M_W = M_W$ coincides with the normalized weighted network. As the share of pure nodes increases (left to right), the overlapping structure gradually gives way to a discrete block structure that remains visible in the estimated network.

Figure 2 provides a graphical counterpart to these results by comparing the simulated normalized weighted network M_W with the estimated sign-reversed off-diagonal concentration matrix $\widehat{M}$. As the share of pure nodes increases, the overlapping pattern sharpens into a discrete block structure. The corresponding structure remains visible in $\widehat{M}$, illustrating that the estimated partial-correlation network preserves the main features of the latent mixed-membership network.

4 Application: Learning the Mixed Memberships of U.S. States

We apply PartialCorr-Mixed-SCORE to state-level payroll employment in the United States. The series are seasonally adjusted, annualized quarter-to-quarter growth rates in payroll employment from 1956 to 2007, taken from the U.S. Bureau of Labor Statistics. Following [Hamilton & Owyang \(2012\)](#) and [Brownlees et al. \(2022\)](#), we drop Alaska and Hawaii, leaving $N = 48$ states and $T = 207$ quarterly observations. We estimate the latent group structure across states and compare it with the groupings reported in those two studies.

The two studies provide different benchmarks. [Hamilton & Owyang \(2012\)](#) specify a Markov-switching model in which states cluster by the timing of their business cycles, estimated by Markov chain Monte Carlo; cross-validation selects three idiosyncratic regional clusters. Their clusters are not mutually exclusive—a state may be affiliated with more than one—but affiliation is recorded through binary indicators rather than a continuous composition. [Brownlees et al. \(2022\)](#) instead model the partial-correlation matrix as a stochastic block model and recover a hard partition by spectral clustering. We keep the same panel and the same three-group benchmark ($K = 3$) and represent each state by a continuous membership vector on the simplex, allowing its employment dynamics to exhibit varying degrees of affiliation with several latent groups. This matters here: a state whose labor market responds to both energy prices and national manufacturing demand has no natural place in a hard partition, and forcing it into one discards information about its affinities with multiple latent dependence patterns.

Table 2 reports estimated memberships for a set of representative states; the full set of 48 appears in Section B.2 of the supplement. Using near-pure states as anchors, we label the three groups broadly as financial and services, oil-producing and agricultural, and manufacturing, corresponding to the finance, oil, and manufacturing characteristics that [Hamilton & Owyang \(2012\)](#) use to explain cluster affiliation. These labels summarize dom-

Table 2: Mixed-membership estimates for selected U.S. states.

State	Financial and Services	Oil-Producing and Agricultural	Manufacturing
Massachusetts	0.736	0.000	0.264
New Jersey	0.777	0.000	0.223
New York	0.524	0.000	0.476
California	0.848	0.032	0.120
Texas	0.062	0.600	0.339
Florida	0.112	0.732	0.155
Missouri	0.129	0.871	0.000
Maine	0.046	0.954	0.000
Louisiana	0.247	0.480	0.273
Michigan	0.602	0.000	0.398
Indiana	0.340	0.000	0.660

inant comovement patterns rather than sectoral accounting identities, and the weights are coordinates in the estimated partial-correlation simplex rather than employment shares: a manufacturing loading, for instance, may reflect industrial composition, common shocks, or input–output linkages that generate similar conditional dynamics. Because the groups are dependence coordinates rather than business-cycle-timing clusters, state-by-state affiliations need not coincide with the earlier studies, though the three recovered dimensions relate to those they highlight.

The weights also differ in kind from the posterior quantities in [Hamilton & Owyang \(2012\)](#), who report, for each state and cluster, the posterior probability of a binary affiliation $p(h_{nk} = 1 \mid Y)$ and allow a state to be affiliated with several clusters at once. Our entries are instead continuous weights on the simplex that sum to one across groups, so a vector such as (0.52, 0, 0.48) does not report the probability that a state belongs to each group; it places the state between the two corresponding population vertices.

Several service-oriented economies place their largest weight on the first group: New Jersey (0.777), Massachusetts (0.736), New York (0.524), and California (0.848). The second group collects states whose dynamics are more closely tied to energy and agriculture, such as Texas (0.600) and Missouri (0.871), while Louisiana spreads its weight across all three (0.247, 0.480, 0.273). The mixed weights are most informative away from the vertices:

Indiana remains manufacturing-oriented (0.660) but carries a sizable financial-and-services weight (0.340), and Michigan (0.602, 0.398) and New York (0.524, 0.476) load materially on two groups at once—affiliations a hard partition cannot represent. For states with one weight near one the method reproduces a hard partition; its added value lies precisely in these intermediate states, whose positions are retained rather than forced to the nearest vertex.

5 Conclusion

This paper introduced the 2PMM-PCN, a partial-correlation network model that extends pure-membership community networks in two directions: latent groups may overlap, and the incidence of a conditional-dependence link is separated from its intensity. We proposed PartialCorr-Mixed-SCORE to recover the membership matrix from the panel’s estimated concentration matrix, and showed that the average membership error is governed by two sources specific to this setting—realization of the latent weighted network and estimation of the concentration matrix from a finite time series. The rate permits the number of latent groups to grow and does not require pairwise separation among the non-leading signal eigenvalues. Simulations and an application to U.S. state payroll-employment growth, in which states occupy intermediate positions between latent groups, illustrate the method.

Several extensions remain. The theory relies on an unregularized concentration-matrix estimator and so covers the regime $T > N$ with $N/T \rightarrow 0$; the case $N \geq T$ would require pairing the spectral analysis with regularized precision-matrix estimation. The growing- K result also treats the stability of the vertex-search step as a high-level condition, and inference for individual membership vectors and a formal criterion for selecting K would complement the consistency results developed here.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data analyzed in this article are openly available at <https://econweb.ucsd.edu/~jhamilto/software.htm>.

References

- Acemoglu, D., Carvalho, V. M., Ozdaglar, A. & Tahbaz-Salehi, A. (2012), ‘The network origins of aggregate fluctuations’, *Econometrica* **80**(5), 1977–2016.
- Airoldi, E. M., Blei, D. M., Fienberg, S. E. & Xing, E. P. (2008), ‘Mixed membership stochastic blockmodels’, *Journal of Machine Learning Research* **9**(65), 1981–2014.
- Barigozzi, M. & Brownlees, C. (2019), ‘NETS: Network estimation for time series’, *Journal of Applied Econometrics* **34**(3), 347–364.
- Bonhomme, S. & Manresa, E. (2015), ‘Grouped patterns of heterogeneity in panel data’, *Econometrica* **83**(3), 1147–1184.
- Brownlees, C., Guðmundsson, G. S. & Lugosi, G. (2022), ‘Community detection in partial correlation network models’, *Journal of Business & Economic Statistics* **40**(1), 216–226.
- Brownlees, C., Nualart, E. & Sun, Y. (2018), ‘Realized networks’, *Journal of Applied Econometrics* **33**(7), 986–1006.
- Chen, C. Y.-H., Härdle, W. K. & Klochkov, Y. (2022), ‘SONIC: Social network analysis with influencers and communities’, *Journal of Econometrics* **228**(2), 177–220.

- Chen, E. Y., Fan, J. & Zhu, X. (2023), ‘Community network auto-regression for high-dimensional time series’, *Journal of Econometrics* **235**(2), 1239–1256.
- Dempster, A. P. (1972), ‘Covariance selection’, *Biometrics* pp. 157–175.
- Fan, J., Fan, Y., Han, X. & Lv, J. (2022), ‘SIMPLE: Statistical inference on membership profiles in large networks’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84**(2), 630–653.
- Gabaix, X. (2011), ‘The granular origins of aggregate fluctuations’, *Econometrica* **79**(3), 733–772.
- Guðmundsson, G. S. & Brownlees, C. (2021), ‘Detecting groups in large vector autoregressions’, *Journal of Econometrics* **225**(1), 2–26.
- Hamilton, J. D. & Owyang, M. T. (2012), ‘The propagation of regional recessions’, *The Review of Economics and Statistics* **94**(4), 935–947.
- Jin, J. (2015), ‘Fast community detection by SCORE’, *The Annals of Statistics* **43**(1), 57–89.
- Jin, J., Ke, Z. T. & Luo, S. (2024), ‘Mixed membership estimation for social networks’, *Journal of Econometrics* **239**(2), 105369.
- Koller, D. & Friedman, N. (2009), *Probabilistic graphical models: principles and techniques*, MIT Press.
- Munro, E. & Ng, S. (2022), ‘Latent Dirichlet analysis of categorical survey responses’, *Journal of Business & Economic Statistics* **40**(1), 256–271.
- Pourahmadi, M. (2013), *High-dimensional covariance estimation: with high-dimensional data*, John Wiley & Sons.
- Simpson, S. L. & Laurienti, P. J. (2015), ‘A two-part mixed-effects modeling framework for analyzing whole-brain network data’, *NeuroImage* **113**, 310–319.

- Su, L., Shi, Z. & Phillips, P. C. (2016), ‘Identifying latent structures in panel data’, *Econometrica* **84**(6), 2215–2264.
- Zhang, Y., Levina, E. & Zhu, J. (2020), ‘Detecting overlapping communities in networks using spectral methods’, *SIAM Journal on Mathematics of Data Science* **2**(2), 265–283.
- Zhang, Y., Wang, H. J. & Zhu, Z. (2019), ‘Quantile-regression-based clustering for panel data’, *Journal of Econometrics* **213**(1), 54–67.
- Zhu, X., Chang, X., Li, R. & Wang, H. (2019), ‘Portal nodes screening for large scale social networks’, *Journal of Econometrics* **209**(2), 145–157.
- Zhu, X., Xu, G. & Fan, J. (2025), ‘Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes’, *Journal of Econometrics* **249**, 105564.

Supporting Information for “Mixed Membership Estimation in Partial Correlation Network Model”

Endong Wang

Siao Xu

University of Mannheim

University of Mannheim

September 8, 2026

Endong Wang. Department of Economics, University of Mannheim, L7, 3–5, Mannheim, Germany, 68161. TEL: +49 621-181-1879; Email: endong.wang@uni-mannheim.de.

Siao Xu. Graduate School of Economic and Social Sciences, University of Mannheim; siao.xu@students.uni-mannheim.de.

SUPPLEMENTARY MATERIAL

2PMM-PCN Supplementary Material: This supplementary material contains all theoretical proofs for the paper, “Mixed Membership Estimation in Partial Correlation Network Model.” In addition, it discusses model identification, reports supplementary results for the empirical application, provides a graphical representation of the model, and describes the heuristic used to select the number of groups.

U.S. state-level employment growth rate dataset: The dataset is obtained from the U.S. Bureau of Labor Statistics.

Deidentified data have been made available at the following URL:
<https://econweb.ucsd.edu/~jhamilto/software.htm>.

1 Notation and graph conventions

Let N denote the number of series and T the number of observations. For a matrix C , let $C_{i,\cdot}$, $C_{\cdot,j}$, and C_{ij} denote its i -th row, j -th column, and (i,j) -th entry, respectively. The operator norm and Frobenius norm are denoted by $\|C\|$ and $\|C\|_F = \{\text{tr}(C'C)\}^{1/2}$. For a symmetric matrix C , $\lambda_{\max}(C)$ and $\lambda_{\min}(C)$ denote its largest and smallest eigenvalues. The notation $\text{diag}(C)$ denotes the diagonal matrix formed from the diagonal of C , and $\text{offdiag}(C) = C - \text{diag}(C)$.

A weighted undirected graph is a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, where $\mathcal{V} = \{1, \dots, N\}$ is the vertex set, $\mathcal{E}$ is the set of unordered edges, and W is the symmetric weighted adjacency matrix. The binary adjacency matrix A satisfies $A_{ij} = 1$ if $\{i, j\} \in \mathcal{E}$ and $A_{ij} = 0$ otherwise. Throughout, $A_{ii} = W_{ii} = 0$.

Appendix A. Proofs

Proof of Theorem 1. For every $i \neq j$, iterated expectations give

$$\begin{aligned}\mathbb{E}(W_{ij}) &= \mathbb{E}[\mathbb{E}(W_{ij} \mid A_{ij})] \\ &= \mathbb{P}(A_{ij} = 1)\mathbb{E}(W_{ij} \mid A_{ij} = 1).\end{aligned}\tag{1}$$

Under Definition 1,

$$\mathbb{P}(A_{ij} = 1) = \Theta_i \pi_i B \pi_j' \Theta_j, \quad \mathbb{E}(W_{ij} \mid A_{ij} = 1) = u_i u_j.\tag{2}$$

Consequently,

$$\mathbb{E}(W_{ij}) = \Theta_i \pi_i B \pi_j' \Theta_j u_i u_j = (\Omega_W)_{ij}.\tag{3}$$

Since $W_{ii} = 0$, $\mathbb{E}(W_{ii}) = 0$. □

Positive definiteness of the concentration matrix. The normalized graph Laplacian L is symmetric. Under the isolated-node convention used in Definition 2, for every $x \in \mathbb{R}^N$,

$$x' L x = \frac{1}{2} \sum_{i,j: d_{W,i} d_{W,j} > 0} W_{ij} \left(\frac{x_i}{\sqrt{d_{W,i}}} - \frac{x_j}{\sqrt{d_{W,j}}} \right)^2 + \sum_{i: d_{W,i} = 0} x_i^2 \geq 0. \quad (4)$$

Hence L is positive semidefinite. It follows that, for every nonzero $x \in \mathbb{R}^N$,

$$x' \Lambda x = \frac{1}{\sigma^2} (\|x\|_2^2 + \phi x' L x) \geq \frac{1}{\sigma^2} \|x\|_2^2 > 0. \quad (5)$$

Therefore, Λ is symmetric and positive definite for every $\sigma^2 > 0$ and $\phi \geq 0$. $\square$

Proof of Lemma 1. Let $F = \mathcal{D}_\Omega^{-1/2} U \Theta \Pi$. Under Assumption 2(i), $G = F' F$ is positive definite, so F has full column rank. The same assumption implies that B is positive definite. Therefore,

$$\mathcal{M}_W = F B F' \quad (6)$$

has rank K and $\text{col}(\mathcal{M}_W) = \text{col}(F)$. The columns of $\mathcal{X}$ form an orthonormal basis for this column space, so there exists a nonsingular $P \in \mathbb{R}^{K \times K}$ such that $\mathcal{X} = F P$. If $F P_1 = F P_2$, then $F(P_1 - P_2) = 0$; the full column rank of F implies $P_1 = P_2$. Thus P is unique for the chosen basis $\mathcal{X}$. $\square$

Proof of Lemma 2. Let $F = \mathcal{D}_\Omega^{-1/2} U \Theta \Pi$ and write $\mathcal{X} = F P$, where P is the nonsingular matrix in Lemma 1. Since Ω_W is irreducible, $\mathcal{M}_W = F B F'$ is nonnegative and irreducible. The Perron–Frobenius theorem implies that its largest eigenvalue is simple and that the associated eigenvector x_1 can be chosen to have strictly positive entries.

Write $P_{k,1}$ for the first entry of the k -th row of P and $P_{k,-1}$ for the remaining entries. If $i \in \mathcal{I}_k$ is a pure node, then $\mathcal{X}_{i,1} = d_{\Omega,i}^{-1/2} u_i \Theta_i P_{k,1}$. Since $\mathcal{X}_{i,1} > 0$, $P_{k,1} > 0$ for every k . Define

$v_k = P'_{k,-1}/P_{k,1}$. For an arbitrary node i ,

$$r_i = \frac{\mathcal{X}'_{i,-1}}{\mathcal{X}_{i,1}} = \sum_{k=1}^K \frac{\pi_{ik} P_{k,1}}{\sum_{\ell=1}^K \pi_{i\ell} P_{\ell,1}} v_k. \quad (7)$$

Thus

$$w_i(k) = \frac{\pi_{ik} P_{k,1}}{\sum_{\ell=1}^K \pi_{i\ell} P_{\ell,1}}, \quad k = 1, \dots, K. \quad (8)$$

The coordinates are nonnegative and sum to one. If $\pi_i = e'_k$, then $w_i = e_k$ and $r_i = v_k$.

The simplex is nondegenerate. Indeed,

$$P = \text{diag}(P_{1,1}, \dots, P_{K,1}) \mathcal{V}'. \quad (9)$$

Since P and the diagonal matrix on the right are nonsingular, $\mathcal{V}$ is nonsingular.

It remains to identify $P_{k,1}$. Orthonormality of $\mathcal{X}$ gives

$$P' F' F P = I_K. \quad (10)$$

Moreover, $\mathcal{M}_W \mathcal{X} = \mathcal{X} \text{diag}(\lambda_1, \dots, \lambda_K)$. Substituting $\mathcal{M}_W = F B F'$ and $\mathcal{X} = F P$ and using the full column rank of F gives

$$B(F' F) P = P \text{diag}(\lambda_1, \dots, \lambda_K). \quad (11)$$

Equation (10) implies $(F' F) P = P'^{-1}$, and hence

$$B = P \text{diag}(\lambda_1, \dots, \lambda_K) P'. \quad (12)$$

Taking the k -th diagonal entry and using $B_{kk} = 1$,

$$1 = P_{k,1}^2 [\lambda_1 + v'_k \text{diag}(\lambda_2, \dots, \lambda_K) v_k]. \quad (13)$$

Since $P_{k,1} > 0$,

$$P_{k,1} = [\lambda_1 + v'_k \text{diag}(\lambda_2, \dots, \lambda_K) v_k]^{-1/2} = b_{1,k}. \quad (14)$$

Substitution into (8) gives $w_i = (\pi'_i \circ b_1) / \|\pi'_i \circ b_1\|_1$, and therefore

$$\pi'_i = \frac{w_i \oslash b_1}{\|w_i \oslash b_1\|_1}. \quad (15)$$

□

Proof of Lemma 3. Let

$$\overline{\Omega}_W = \mathbb{E}(W) = \Omega_W - \text{diag}(\Omega_W), \quad (16)$$

$$\overline{\mathcal{M}}_W = \mathcal{D}_\Omega^{-1/2} \overline{\Omega}_W \mathcal{D}_\Omega^{-1/2}. \quad (17)$$

The degree matrix of $\overline{\Omega}_W$ is $\mathcal{D}_\Omega$ because $d_{\Omega,i} = \sum_{j \neq i} (\Omega_W)_{ij}$. Conditional on the latent parameters, the upper-triangular entries $\{W_{ij} : i < j\}$ are mutually independent, take values in $[0, 1]$, and satisfy

$$\text{Var}(W_{ij}) \leq \mathbb{E}(W_{ij}^2) \leq \mathbb{E}(W_{ij}) = (\overline{\Omega}_W)_{ij}. \quad (18)$$

Scalar Bernstein's inequality applied to each weighted degree, followed by a union bound, gives

$$\max_{1 \leq i \leq N} \frac{|d_{W,i} - d_{\Omega,i}|}{d_{\Omega,i}} = O_p \left(\sqrt{\frac{\log N}{d_{\min}}} + \frac{\log N}{d_{\min}} \right) = O_p \left(\sqrt{\frac{\log N}{d_{\min}}} \right). \quad (19)$$

The last equality uses $d_{\min}/\log N \rightarrow \infty$. In particular, $\mathbb{P}(\min_i d_{W,i} > 0) \rightarrow 1$.

Consider

$$\begin{aligned} & \mathcal{D}_\Omega^{-1/2}(W - \bar{\Omega}_W)\mathcal{D}_\Omega^{-1/2} \\ &= \sum_{i < j} \frac{W_{ij} - (\bar{\Omega}_W)_{ij}}{\sqrt{d_{\Omega,i}d_{\Omega,j}}}(e_i e'_j + e_j e'_i). \end{aligned} \quad (20)$$

Each summand has operator norm at most $d_{\min}^{-1}$. By (18), the matrix variance is bounded by $d_{\min}^{-1}$. Matrix Bernstein's inequality therefore yields

$$\left\| \mathcal{D}_\Omega^{-1/2}(W - \bar{\Omega}_W)\mathcal{D}_\Omega^{-1/2} \right\| = O_p\left(\sqrt{\frac{\log N}{d_{\min}}}\right). \quad (21)$$

On the event in (19),

$$\left\| D_W^{-1/2}\mathcal{D}_\Omega^{1/2} - I_N \right\| = O_p\left(\sqrt{\frac{\log N}{d_{\min}}}\right), \quad (22)$$

and $\|D_W^{-1/2}\mathcal{D}_\Omega^{1/2}\| = O_p(1)$. Moreover,

$$M_W = D_W^{-1/2}\mathcal{D}_\Omega^{1/2} \left[\bar{\mathcal{M}}_W + \mathcal{D}_\Omega^{-1/2}(W - \bar{\Omega}_W)\mathcal{D}_\Omega^{-1/2} \right] \mathcal{D}_\Omega^{1/2} D_W^{-1/2}. \quad (23)$$

Since $\bar{\mathcal{M}}_W$ is the normalized adjacency matrix associated with $\bar{\Omega}_W$, $\|\bar{\mathcal{M}}_W\| \leq 1$. Combining (21) and (22),

$$\|M_W - \bar{\mathcal{M}}_W\| = O_p\left(\sqrt{\frac{\log N}{d_{\min}}}\right). \quad (24)$$

The difference between $\bar{\mathcal{M}}_W$ and the rank- K signal is purely diagonal:

$$\mathcal{M}_W - \bar{\mathcal{M}}_W = \mathcal{D}_\Omega^{-1/2} \text{diag}(\Omega_W) \mathcal{D}_\Omega^{-1/2}, \quad (25)$$

so

$$\|\mathcal{M}_W - \overline{\mathcal{M}}_W\| = \max_i \frac{(\Omega_W)_{ii}}{d_{\Omega,i}} \leq \frac{\max_i (\Omega_W)_{ii}}{d_{\min}} = O(d_{\min}^{-1}). \quad (26)$$

Assumption 2(iv) makes this term negligible relative to $\sqrt{\log N/d_{\min}}$. Therefore,

$$\|M_W - \mathcal{M}_W\| = O_p\left(\sqrt{\frac{\log N}{d_{\min}}}\right). \quad (27)$$

For the time-series layer, let $E_T = \Sigma^{-1/2}\widehat{\Sigma}\Sigma^{-1/2} - I_N$. Then

$$\widehat{\Sigma} = \Sigma^{1/2}(I_N + E_T)\Sigma^{1/2}, \quad \widehat{\Lambda} = \Lambda^{1/2}(I_N + E_T)^{-1}\Lambda^{1/2}. \quad (28)$$

Assumption 3 implies $\|E_T\| = O_p(\sqrt{N/T}) = o_p(1)$. On $\{\|E_T\| < 1\}$,

$$\|\widehat{\Lambda} - \Lambda\| \leq \|\Lambda\| \frac{\|E_T\|}{1 - \|E_T\|}. \quad (29)$$

The eigenvalues of a symmetric normalized graph Laplacian lie in $[0, 2]$, including under the isolated-node convention. Hence

$$\|\Lambda\| \leq \frac{1 + 2\phi}{\sigma^2} \leq c_\sigma^{-1} + 2C_\phi \quad (30)$$

under Assumption 2(v), and therefore $\|\widehat{\Lambda} - \Lambda\| = O_p(\sqrt{N/T})$. For every symmetric matrix C , $\|\text{offdiag}(C)\| \leq 2\|C\|$. Since $-\text{offdiag}(\Lambda) = (\phi/\sigma^2)M_W$,

$$\|\widehat{M} - M\| = \left\| \widehat{M} - \frac{\phi}{\sigma^2} M_W \right\| = O_p\left(\sqrt{\frac{N}{T}}\right). \quad (31)$$

The final claim follows from (9) and the triangle inequality. $\square$

Proof of Theorem 2. Assumptions 2(iv) and 3 imply $\delta_{N,T} \rightarrow 0$. By Assumption 2(i), the K

nonzero eigenvalues of $\mathcal{M}_W$ are the eigenvalues of $B^{1/2}GB^{1/2}$ and hence satisfy

$$c_B c_G \leq \lambda_K \leq \lambda_1 \leq C_B C_G. \quad (32)$$

Together with Assumption 2(iii) and (v), the leading eigenvalue is separated from the remaining signal spectrum by a positive constant and the whole signal spectrum is separated from zero by a positive constant, uniformly in K .

Lemma 3 gives

$$\left\| \widehat{M} - \frac{\phi}{\sigma^2} \mathcal{M}_W \right\| = O_p(\delta_{N,T}) = o_p(1). \quad (33)$$

Weyl's inequality and (32) imply that, with probability tending to one, the K signal eigenvalues of $\widehat{M}$ are positive and uniformly separated from the remaining eigenvalues, whose magnitudes are $O_p(\delta_{N,T})$. Hence Step 3 selects exactly these K eigenvalues. Weyl's inequality also gives

$$\max_{1 \leq k \leq K} \left| \hat{\lambda}_k - \frac{\phi}{\sigma^2} \lambda_k \right| = O_p(\delta_{N,T}), \quad (34)$$

where both signal spectra are ordered decreasingly on this event.

The Davis–Kahan theorem applied to the leading eigenvector yields a sign $s \in \{-1, 1\}$ such that

$$\|s\hat{x}_1 - x_1\|_2 = O_p(\delta_{N,T}). \quad (35)$$

For the remaining signal subspace, let $\widehat{\mathcal{P}}_{-1} = \widehat{X}_{-1}\widehat{X}'_{-1}$ and $\mathcal{P}_{-1} = \mathcal{X}_{-1}\mathcal{X}'_{-1}$. The spectral separation established above and Davis–Kahan give $\|\widehat{\mathcal{P}}_{-1} - \mathcal{P}_{-1}\| = O_p(\delta_{N,T})$. The orthogonal

Procrustes alignment therefore yields an orthogonal $\mathbf{H} \in \mathbb{R}^{(K-1) \times (K-1)}$ satisfying

$$\|\widehat{X}_{-1}\mathbf{H}' - \mathcal{X}_{-1}\| = O_p(\delta_{N,T}), \quad \|\widehat{X}_{-1}\mathbf{H}' - \mathcal{X}_{-1}\|_F = O_p\left(\sqrt{K} \delta_{N,T}\right). \quad (36)$$

This proves (15).

It remains to prove (16). Since the columns of $\widehat{X}_{-1}$ are eigenvectors of $\widehat{M}$,

$$\widehat{\mathcal{P}}_{-1}\widehat{M}\widehat{\mathcal{P}}_{-1} = \widehat{X}_{-1} \text{diag}(\hat{\lambda}_2, \dots, \hat{\lambda}_K) \widehat{X}_{-1}'. \quad (37)$$

Similarly,

$$\mathcal{P}_{-1} \frac{\phi}{\sigma^2} \mathcal{M}_W \mathcal{P}_{-1} = \frac{\phi}{\sigma^2} \mathcal{X}_{-1} \text{diag}(\lambda_2, \dots, \lambda_K) \mathcal{X}_{-1}'. \quad (38)$$

Using (33), the projector bound above, and the uniform boundedness of $\|\mathcal{M}_W\|$, the difference between these two operators is $O_p(\delta_{N,T})$ in operator norm. Premultiplying and postmultiplying by the aligned orthonormal bases and using the operator-norm bound in (36) gives

$$\left\| \mathbf{H} \text{diag}(\hat{\lambda}_2, \dots, \hat{\lambda}_K) \mathbf{H}' - \frac{\phi}{\sigma^2} \text{diag}(\lambda_2, \dots, \lambda_K) \right\| = O_p(\delta_{N,T}). \quad (39)$$

No separation among $\lambda_2, \dots, \lambda_K$ is used. □

Proof of Theorem 3. Under Assumption 4, $\mathbf{1}'x_1 \geq c_x \sqrt{N}$. Let s be the sign in Theorem 2. Then

$$\mathbf{1}'(s\hat{x}_1) \geq c_x \sqrt{N} - \sqrt{N} \|s\hat{x}_1 - x_1\|_2 > 0 \quad (40)$$

with probability tending to one. Hence the sign convention in Step 3 selects this orientation. We use this orientation below.

Let $\widetilde{r}_i$ denote the untruncated ratio vector in Step 4. Define the set of rows whose

estimated leading coordinate is well separated from zero by

$$\mathcal{G}_N = \left\{ 1 \leq i \leq N : |\hat{x}_1(i) - x_1(i)| \leq \frac{c_x}{2\sqrt{N}} \right\}. \quad (41)$$

For $i \in \mathcal{G}_N$, $|\hat{x}_1(i)| \geq c_x/(2\sqrt{N})$. Moreover, Assumption 4 gives $|x_1(i)| \geq c_x/\sqrt{N}$ and $\|\mathcal{X}_{i,-1}\|_2 \leq C_x \sqrt{K/N}$. We have

$$\mathbf{H}\tilde{r}_i - r_i = \frac{\mathbf{H}\hat{X}'_{i,-1} - \mathcal{X}'_{i,-1}}{\hat{x}_1(i)} + \mathcal{X}'_{i,-1} \left(\frac{1}{\hat{x}_1(i)} - \frac{1}{x_1(i)} \right). \quad (42)$$

Therefore,

$$\|\mathbf{H}\tilde{r}_i - r_i\|_2 \leq C\sqrt{N} \|(\hat{X}_{-1}\mathbf{H}' - \mathcal{X}_{-1})_{i,\cdot}\|_2 + C\sqrt{KN} |\hat{x}_1(i) - x_1(i)|. \quad (43)$$

The Euclidean projection used in Step 4 is nonexpansive and commutes with orthogonal transformations. Also, $\|r_i\|_2 \leq (C_x/c_x)\sqrt{K} < S = \sqrt{K} \log N$ for all sufficiently large N . Hence, for $i \in \mathcal{G}_N$,

$$\|\mathbf{H}\hat{r}_i - r_i\|_2 \leq \|\mathbf{H}\tilde{r}_i - r_i\|_2. \quad (44)$$

Squaring (43), summing over $\mathcal{G}_N$, and using Theorem 2,

$$\begin{aligned} \frac{1}{N} \sum_{i \in \mathcal{G}_N} \|\mathbf{H}\hat{r}_i - r_i\|_2^2 &\leq C \|\hat{X}_{-1}\mathbf{H}' - \mathcal{X}_{-1}\|_F^2 + CK \|\hat{x}_1 - x_1\|_2^2 \\ &= O_p(K\delta_{N,T}^2). \end{aligned} \quad (45)$$

For $i \notin \mathcal{G}_N$, $|\hat{x}_1(i) - x_1(i)|^2 > c_x^2/(4N)$. Therefore,

$$\frac{|\mathcal{G}_N^c|}{N} \leq \frac{4}{c_x^2} \|\hat{x}_1 - x_1\|_2^2 = O_p(\delta_{N,T}^2). \quad (46)$$

By construction, $\|\widehat{r}_i\|_2 \leq S$, while $\|r_i\|_2 \leq C\sqrt{K} \leq CS$. Thus

$$\frac{1}{N} \sum_{i \in \mathcal{G}_N^c} \|\mathbf{H}\widehat{r}_i - r_i\|_2^2 \leq CS^2 \frac{|\mathcal{G}_N^c|}{N} = O_p(K \log^2 N \delta_{N,T}^2). \quad (47)$$

Combining (45) and (47) gives

$$\frac{1}{N} \|\widehat{R}\mathbf{H}' - R\|_F^2 = O_p(K \log^2 N \delta_{N,T}^2). \quad (48)$$

Finally, $\delta_{N,T}^2 \leq 2\{\log N/d_{\min} + N/T\}$, which yields (17). $\square$

Proof of Lemma 4. Relabel the estimated vertices according to the permutation $\mathcal{P}$ in Assumption 6. Theorem 3 and Assumption 6 give

$$\max_{1 \leq k \leq K} \|\mathbf{H}\widehat{v}_{\mathcal{P}(k)} - v_k\|_2 = O_p\left(\sqrt{K} \log N \delta_{N,T}\right) = o_p(1), \quad (49)$$

where the last equality follows from $K \log^2 N \delta_{N,T}^2 \rightarrow 0$. By Assumption 5, $\|v_k\|_2 \asymp \sqrt{K}$ uniformly in k , and hence $\max_k \|\mathbf{H}\widehat{v}_{\mathcal{P}(k)}\|_2 = O_p(\sqrt{K})$.

Because $\mathbf{H}$ is orthogonal, the quadratic form used by the algorithm can be written exactly as

$$\widehat{\lambda}_1 + (\mathbf{H}\widehat{v}_{\mathcal{P}(k)})' \left\{ \mathbf{H} \text{diag}(\widehat{\lambda}_2, \dots, \widehat{\lambda}_K) \mathbf{H}' \right\} (\mathbf{H}\widehat{v}_{\mathcal{P}(k)}). \quad (50)$$

Theorem 2, (49), and Assumption 2(i) imply, uniformly in k ,

$$\left| \widehat{\lambda}_1 + \widehat{v}_{\mathcal{P}(k)}' \text{diag}(\widehat{\lambda}_2, \dots, \widehat{\lambda}_K) \widehat{v}_{\mathcal{P}(k)} - \frac{\phi}{\sigma^2} \{ \lambda_1 + v_k' \text{diag}(\lambda_2, \dots, \lambda_K) v_k \} \right| = O_p(K \log N \delta_{N,T}). \quad (51)$$

Indeed, the rotated spectral-operator error contributes $O_p(K \delta_{N,T})$, while the vertex perturbation contributes $O_p(K \log N \delta_{N,T})$.

By Assumptions 2(i) and 5,

$$\lambda_1 + v_k' \text{diag}(\lambda_2, \dots, \lambda_K) v_k \asymp K \quad (52)$$

uniformly in k . Dividing (51) by this quantity gives a uniform relative error $O_p(\log N \delta_{N,T}) = o_p(1)$. The sample quadratic forms are therefore positive and of order K with probability tending to one, so the floor $\tau_N \rightarrow 0$ is inactive. Applying the mean-value theorem to $x^{-1/2}$ on the corresponding relative neighborhood yields

$$\max_{1 \leq k \leq K} \left| \frac{\hat{b}_{1, \mathcal{P}(k)}}{(\phi/\sigma^2)^{-1/2} b_{1,k}} - 1 \right| = O_p(\log N \delta_{N,T}). \quad (53)$$

□

Proof of Lemma 5. Let

$$\tilde{\mathcal{V}} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ \hat{v}_1 & \cdots & \hat{v}_K \end{pmatrix} \mathcal{P}. \quad (54)$$

Its k -th column is $(1, (\mathbf{H} \hat{v}_{\mathcal{P}(k)})')'$. By Assumption 6 and Theorem 3,

$$\begin{aligned} \|\tilde{\mathcal{V}} - \mathcal{V}\| &\leq \|\tilde{\mathcal{V}} - \mathcal{V}\|_F \leq \sqrt{K} \max_k \|\mathbf{H} \hat{v}_{\mathcal{P}(k)} - v_k\|_2 \\ &= O_p(K \log N \delta_{N,T}) = o_p(\sqrt{K}). \end{aligned} \quad (55)$$

The last equality follows from $K \log^2 N \delta_{N,T}^2 \rightarrow 0$. Assumption 5 and Weyl's inequality therefore imply

$$\sigma_{\min}(\tilde{\mathcal{V}}) \geq \frac{c_{\mathcal{V}}}{2} \sqrt{K}, \quad \|\tilde{\mathcal{V}}^{-1}\| = O_p(K^{-1/2}). \quad (56)$$

The sample barycentric equation in Step 6, after relabeling its columns and premulti-

plying by $\text{diag}(1, \mathbf{H})$, is

$$\tilde{\mathcal{V}}(\mathcal{P}'\hat{w}_i) = \begin{pmatrix} 1 \\ \mathbf{H}\hat{r}_i \end{pmatrix}. \quad (57)$$

Subtracting $\mathcal{V}w_i = (1, r'_i)'$ gives

$$\mathcal{P}'\hat{w}_i - w_i = \tilde{\mathcal{V}}^{-1} \left[\begin{pmatrix} 0 \\ \mathbf{H}\hat{r}_i - r_i \end{pmatrix} - (\tilde{\mathcal{V}} - \mathcal{V})w_i \right]. \quad (58)$$

Since $w_i \geq 0$ and $\|w_i\|_1 = 1$,

$$\|(\tilde{\mathcal{V}} - \mathcal{V})w_i\|_2 \leq \max_k \|\mathbf{H}\hat{v}_{\mathcal{P}(k)} - v_k\|_2. \quad (59)$$

Squaring (58), averaging over i , and using (56), Theorem 3, and Assumption 6,

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \|\mathcal{P}'\hat{w}_i - w_i\|_2^2 &\leq \frac{C}{K} \left\{ \frac{1}{N} \|\hat{R}\mathbf{H}' - R\|_F^2 + \max_k \|\mathbf{H}\hat{v}_{\mathcal{P}(k)} - v_k\|_2^2 \right\} \\ &= O_p(\log^2 N \delta_{N,T}^2). \end{aligned} \quad (60)$$

□

Proof of Theorem 4. Let $\mathcal{P}$ be the permutation in Assumption 6 and define

$$a_i = \left(\frac{\phi}{\sigma^2} \right)^{1/2} (w_i \odot b_1), \quad (61)$$

$$\hat{a}_i = \left\{ (\mathcal{P}'\hat{w}_i) \odot (\mathcal{P}'\hat{b}_1) \right\}_+. \quad (62)$$

The common positive factor in a_i does not affect normalization, so $\pi'_i = a_i / \|a_i\|_1$. Whenever $\hat{a}_i \neq 0$, the corresponding row of $\hat{\Pi}\mathcal{P}$ equals $\hat{a}'_i / \|\hat{a}_i\|_1$.

Lemma 4 and $b_{1,k} \asymp K^{-1/2}$ imply, with probability tending to one,

$$\max_k \frac{1}{(\mathcal{P}'\hat{b}_1)(k)} = O_p(\sqrt{K}), \quad (63)$$

$$\max_k \left| \frac{1}{(\mathcal{P}'\hat{b}_1)(k)} - \left(\frac{\phi}{\sigma^2} \right)^{1/2} \frac{1}{b_{1,k}} \right| = O_p(\sqrt{K} \log N \delta_{N,T}). \quad (64)$$

Before taking the positive part, (63)–(64) give

$$\begin{aligned} & \left\| (\mathcal{P}'\hat{w}_i) \oslash (\mathcal{P}'\hat{b}_1) - a_i \right\|_2^2 \\ & \leq CK \|\mathcal{P}'\hat{w}_i - w_i\|_2^2 + CK \log^2 N \delta_{N,T}^2 \|w_i\|_2^2. \end{aligned} \quad (65)$$

Since $\|w_i\|_2 \leq \|w_i\|_1 = 1$, Lemma 5 yields

$$\frac{1}{N} \sum_{i=1}^N \left\| (\mathcal{P}'\hat{w}_i) \oslash (\mathcal{P}'\hat{b}_1) - a_i \right\|_2^2 = O_p(K \log^2 N \delta_{N,T}^2). \quad (66)$$

Because a_i is componentwise nonnegative, projection onto the nonnegative orthant is non-expansive. Hence the same bound holds with the expression inside the norm replaced by $\hat{a}_i - a_i$.

By Assumptions 2(v) and 5, uniformly in i ,

$$c\sqrt{K} \leq \|a_i\|_1 \leq C\sqrt{K}, \quad \|a_i\|_2 \leq C\sqrt{K} \quad (67)$$

for constants $0 < c < C < \infty$. Consider first rows satisfying $\|\hat{a}_i - a_i\|_2 \leq c/2$. Then $\|\hat{a}_i - a_i\|_1 \leq \sqrt{K}\|\hat{a}_i - a_i\|_2$ and (67) imply $\|\hat{a}_i\|_1 \geq c\sqrt{K}/2$. Therefore,

$$\begin{aligned} \left\| \frac{\hat{a}_i}{\|\hat{a}_i\|_1} - \frac{a_i}{\|a_i\|_1} \right\|_2 & \leq \frac{\|\hat{a}_i - a_i\|_2}{\|\hat{a}_i\|_1} + \|a_i\|_2 \frac{|\|\hat{a}_i\|_1 - \|a_i\|_1|}{\|\hat{a}_i\|_1 \|a_i\|_1} \\ & \leq C\|\hat{a}_i - a_i\|_2. \end{aligned} \quad (68)$$

For the remaining rows, the fraction of indices with $\|\hat{a}_i - a_i\|_2 > c/2$ is $O_p(K \log^2 N \delta_{N,T}^2)$

by (66). The squared Euclidean distance between two probability vectors is at most two. If $\hat{a}_i = 0$, the algorithm assigns the uniform membership vector; such a row is included in the same remainder because (67) implies $\|a_i\|_2 \geq c$.

Combining these observations with (66) gives

$$\frac{1}{N} \|\hat{\Pi} \mathcal{P} - \Pi\|_F^2 = O_p(K \log^2 N \delta_{N,T}^2). \quad (69)$$

Finally,

$$\delta_{N,T}^2 = O\left(\frac{\log N}{d_{\min}} + \frac{N}{T}\right), \quad (70)$$

which proves (22) after minimizing over $\mathfrak{P}_K$. $\square$

Appendix B. Additional Results

Appendix B.1 Identification of Mixed Memberships

The identification result relevant for the estimator concerns the membership matrix Π . The matrices U and Θ have different structural roles in the generative model, but they enter the ideal population signal only through their product. They are therefore not separately identified from that signal without additional restrictions. Likewise, κ_{ij} governs the conditional dispersion of an existing link but does not enter the first-moment factorization used to recover Π ; its role is therefore descriptive for the membership-estimation problem studied here.

Theorem B.1 (Identification from the ideal population signal). *Suppose the conditions of Lemma 2 hold and the ideal population normalized signal $\mathcal{M}_W$ is known. Then Π is identified up to a permutation of its columns.*

Proof. By Assumption 2(ii)–(iii), the leading eigenvalue of $\mathcal{M}_W$ is simple and its eigenvector

x_1 has a unique strictly positive orientation. No simplicity is required for $\lambda_2, \dots, \lambda_K$. Choose any orthonormal eigenbasis for the remaining signal eigenspace and construct the SCORE ratios and simplex as in Lemma 2.

If a non-leading eigenvalue has multiplicity, another admissible eigenbasis is obtained by an orthogonal rotation within the corresponding eigenspace. Under such a change of basis, all SCORE ratios and all simplex vertices are rotated by the same orthogonal transformation. The barycentric coordinates w_i are therefore unchanged. The rotation also commutes with $\text{diag}(\lambda_2, \dots, \lambda_K)$ within each repeated-eigenvalue block, so

$$\lambda_1 + v'_k \text{diag}(\lambda_2, \dots, \lambda_K) v_k \quad (71)$$

is unchanged. Hence b_1 is basis invariant. Lemma 2 then recovers

$$\pi'_i = \frac{w_i \oslash b_1}{\|w_i \oslash b_1\|_1}, \quad i = 1, \dots, N. \quad (72)$$

The simplex vertices are intrinsically unordered, so the only remaining ambiguity is a common permutation of the columns of Π . $\square$

To see why U and Θ are not separately identified, let C be any positive diagonal matrix sufficiently close to I_N that the transformed parameters remain in their admissible parameter spaces. The transformation

$$U^* = UC, \quad \Theta^* = C^{-1}\Theta \quad (73)$$

leaves $U^*\Theta^* = U\Theta$ and therefore leaves $\Omega_W = U\Theta\Pi B\Pi'\Theta U$ unchanged. Separate identification of the two node-specific channels requires restrictions or information beyond the population network signal used for membership recovery.

Appendix B.2 Additional Discussion of the Application

We apply PartialCorr-Mixed-SCORE to a panel of seasonally adjusted, annualized quarter-to-quarter U.S. state payroll-employment growth rates. The sample runs from 1956:Q2 to 2007:Q4 and excludes Alaska and Hawaii, so $N = 48$ and $T = 207$. The data were constructed by [Hamilton & Owyang \(2012\)](#) and subsequently used by [Brownlees et al. \(2022\)](#).

[Hamilton & Owyang \(2012\)](#) study business-cycle synchronization using a Markov-switching model and Bayesian inference. [Brownlees et al. \(2022\)](#) model the partial-correlation network using a stochastic block structure and recover discrete communities by spectral clustering. Our estimator also produces point estimates, but each state is represented by a membership vector on the simplex rather than by a single group label. These entries are structural membership shares under the proposed network model; they should not be interpreted as posterior classification probabilities.

Following the three-group specifications in the earlier studies, we set $K = 3$. Using the same interpretation as in the main text, we describe the groups broadly as financial and services, oil-producing and agricultural, and manufacturing. These labels summarize dominant economic patterns rather than exhaustive sectoral classifications. The mixed-membership formulation allows a state to load on more than one of these dimensions. For example, New York loads on both the financial-and-services and manufacturing groups, while Texas loads on both the oil-producing-and-agricultural and manufacturing groups. The estimates are reported in [Table 1](#).

Table 1: Estimated Mixed-Membership Vectors for U.S. States

State	Group 1	Group 2	Group 3
<i>Alabama</i>	0.000	0.456	0.544
<i>Arizona</i>	0.000	0.000	1.000
<i>Arkansas</i>	0.215	0.281	0.504

State	Group 1	Group 2	Group 3
<i>California</i>	0.848	0.032	0.120
<i>Colorado</i>	0.286	0.247	0.467
<i>Connecticut</i>	0.000	0.797	0.203
<i>Delaware</i>	0.357	0.494	0.150
<i>Florida</i>	0.112	0.732	0.155
<i>Georgia</i>	0.921	0.000	0.079
<i>Idaho</i>	0.000	1.000	0.000
<i>Illinois</i>	0.000	0.000	1.000
<i>Indiana</i>	0.340	0.000	0.660
<i>Iowa</i>	0.806	0.169	0.025
<i>Kansas</i>	0.789	0.000	0.211
<i>Kentucky</i>	0.000	0.000	1.000
<i>Louisiana</i>	0.247	0.480	0.273
<i>Maine</i>	0.046	0.954	0.000
<i>Maryland</i>	0.000	0.762	0.238
<i>Massachusetts</i>	0.736	0.000	0.264
<i>Michigan</i>	0.602	0.000	0.398
<i>Minnesota</i>	0.327	0.368	0.305
<i>Mississippi</i>	0.938	0.062	0.000
<i>Missouri</i>	0.129	0.871	0.000
<i>Montana</i>	0.758	0.174	0.068
<i>Nebraska</i>	0.477	0.000	0.523
<i>Nevada</i>	0.000	0.815	0.185
<i>New Hampshire</i>	0.121	0.879	0.000
<i>New Jersey</i>	0.777	0.000	0.223
<i>New Mexico</i>	0.025	0.000	0.975

State	Group 1	Group 2	Group 3
<i>New York</i>	0.524	0.000	0.476
<i>North Carolina</i>	0.000	0.008	0.992
<i>North Dakota</i>	0.212	0.000	0.788
<i>Ohio</i>	0.000	1.000	0.000
<i>Oklahoma</i>	0.332	0.326	0.342
<i>Oregon</i>	0.011	0.218	0.772
<i>Pennsylvania</i>	1.000	0.000	0.000
<i>Rhode Island</i>	0.000	0.262	0.738
<i>South Carolina</i>	0.400	0.515	0.084
<i>South Dakota</i>	0.088	0.165	0.747
<i>Tennessee</i>	0.775	0.105	0.120
<i>Texas</i>	0.062	0.600	0.339
<i>Utah</i>	0.990	0.010	0.000
<i>Vermont</i>	0.000	0.000	1.000
<i>Virginia</i>	1.000	0.000	0.000
<i>Washington</i>	0.370	0.000	0.630
<i>West Virginia</i>	0.573	0.000	0.427
<i>Wisconsin</i>	0.667	0.306	0.027
<i>Wyoming</i>	0.404	0.000	0.596

Appendix B.3 Graphical Representation of the Model

Figure 1 summarizes the two layers of the model. The latent parameters (Π, B, Θ) generate the binary adjacency matrix A , while (A, U, κ) generate the weighted adjacency matrix W . The normalized Laplacian map then determines the concentration matrix Λ , and the panel $\{Y_t\}_{t=1}^T$ is generated conditionally on W through $\Sigma = \Lambda^{-1}$. The conditional-independence

restrictions associated with the network-generation layer are those in Definition 1 and (5).

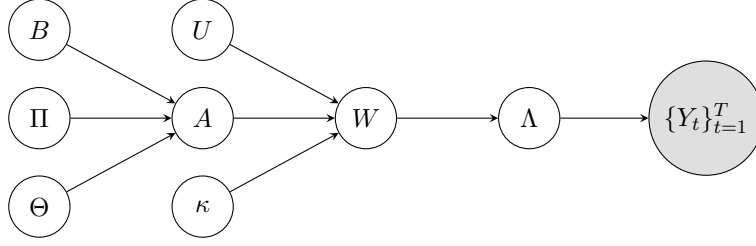


Figure 1: Graphical representation of the two-path mixed-membership partial-correlation network. The map from W to Λ is deterministic, whereas the remaining arrows represent conditional distributions. The shaded node is observed.

Appendix B.4 A Heuristic for Selecting the Number of Groups

Algorithm 1 requires a value of K . The theory in Section 2.3 is stated for the true number of groups and allows an estimated value whenever $\mathbb{P}(\hat{K} = K) \rightarrow 1$. This subsection records a non-backtracking heuristic for empirical implementation. It is not used in the proofs above.

Given $\widehat{M} = -\text{offdiag}(\widehat{\Lambda})$, choose a finite grid of thresholds Γ . For each $\gamma \in \Gamma$, define the symmetric binary matrix

$$A(\gamma)_{ij} = \mathbf{1}\{\widehat{M}_{ij} > \gamma\}, \quad i \neq j, \quad A(\gamma)_{ii} = 0, \quad (74)$$

and let $D(\gamma)$ be its degree matrix. Form the $2N \times 2N$ non-backtracking linearization

$$\mathcal{B}_{\text{nb}}(\gamma) = \begin{pmatrix} 0 & D(\gamma) - I_N \\ -I_N & A(\gamma) \end{pmatrix}. \quad (75)$$

Let $\rho\{\mathcal{B}_{\text{nb}}(\gamma)\}$ denote its spectral radius and count the real eigenvalues satisfying

$$|\lambda_j\{\mathcal{B}_{\text{nb}}(\gamma)\}| > \rho\{\mathcal{B}_{\text{nb}}(\gamma)\}^{1/2}. \quad (76)$$

Denote this count by $\hat{K}(\gamma)$. The threshold may be chosen from a region in which $\hat{K}(\gamma)$ is stable, and the corresponding stable count is used as the estimated number of groups.

This procedure extends the non-backtracking intuition of [Le & Levina \(2022\)](#) and [Li et al. \(2022\)](#) to the estimated partial-correlation network. Its finite-sample behavior depends on the threshold grid, and a formal consistency analysis is left for future work.

References

- Brownlees, C., Guðmundsson, G. S. & Lugosi, G. (2022), ‘Community detection in partial correlation network models’, *Journal of Business & Economic Statistics* **40**(1), 216–226.
- Hamilton, J. D. & Owyang, M. T. (2012), ‘The propagation of regional recessions’, *The Review of Economics and Statistics* **94**(4), 935–947.
- Le, C. M. & Levina, E. (2022), ‘Estimating the number of communities by spectral methods’, *Electronic Journal of Statistics* **16**(1), 3315–3342.
- Li, T., Lei, L., Bhattacharyya, S., Van den Berge, K., Sarkar, P., Bickel, P. J. & Levina, E. (2022), ‘Hierarchical community detection by recursive partitioning’, *Journal of the American Statistical Association* **117**(538), 951–968.