When Post-Hoc Rotation Cannot Replace Non-Negativity: A Boundary for Low-Rank Factor Recovery Through a Nonlinear Link

Anonymous Authors[1]

Abstract

Rohe & Zeng (2023) showed that an unconstrained spectral fit followed by a Varimax rotation recovers the latent factors of a broad class of linear low-rank models—including the stochastic blockmodel and latent Dirichlet allocation—when the factors are leptokurtic. We ask whether this recipe survives a nonlinear link between latent factors and observation, using two rank-K factor models for binary matrices that share an inner-product core but differ in link and constraint: the latent space model (signed, logistic link) of Hoff et al. (2002) and the edge partition model (non-negative, Bernoulli–Poisson link) of Zhou (2015). Post-hoc Varimax aligns the LSM fit with the correct axes in moderately dense networks containing anchor nodes, but the resulting embedding remains partly negative, only weakly sparse, and Gaussian-like in shape at every density we test, never matching the non-negative, sparse, heavy-tailed loadings of the constrained fit. Although the LSM-MLE Gram is empirically close to a nonlinear transformation of $\Theta\Theta^\top$ (Gram-vs-Gram correlation 0.82–0.88), the column-wise leptokurtosis of Θ on which the identifying assumption depends is not inherited by the LSM-MLE fit; Varimax cannot manufacture leptokurtic axes when the column space it operates on does not contain any. This gives a boundary for the spectral-plus-rotation recipe and a structural reason why constraint-based factorizations cannot be fully replaced by post-hoc rotation of an unconstrained fit.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1. Introduction

Identifiability of low-rank factor models. Recovering the latent factors of a low-rank model is, in general, possible only up to an orthogonal rotation. Two complementary routes around this ambiguity have been studied. The non-negative matrix factorization literature imposes constraints on the factor matrices—non-negativity together with separability—and recovers them up to permutation and rescaling (Donoho & Stodden, 2003; Arora et al., 2013; Mao et al., 2017). Rohe & Zeng (2023) take a different route: under the semi-parametric factor model $\mathbb{E}(A \mid Z, Y) = ZBY^\top$ with leptokurtic latent factors, PCA followed by a Varimax rotation consistently estimates the factor matrices in $\|\cdot\|_{2\to\infty}$ (Rohe & Zeng, 2023, Theorem 7.1). Their result is a *unification*: the same fast spectral procedure recovers the latent structure targeted by the stochastic blockmodel (Holland et al., 1983), its degree-corrected and mixed-membership variants, and a natural variant of latent Dirichlet allocation. The methodological implication is that, when the latent factors carry the right distributional shape, an unconstrained spectral fit followed by a free orthogonal rotation can substitute for fitting a constrained generative model directly. Constraints like non-negativity, simplex membership, or hard clustering need not be enforced during fitting; they emerge post-hoc from a Varimax rotation of generic principal components.

A natural question. The Rohe–Zeng result is established for linear factor models: the data matrix A is a linear function of the latent factors Z, Y plus noise, so PCA on a centred A inherits the column-space and the distributional shape of Z. Many latent-factor models that practitioners care about, however, place a *nonlinear link* between the latent factors and the observed data. This is true throughout network analysis: latent-space and latent-feature models for binary adjacency matrices apply a logit, probit, or Bernoulli–Poisson link to an inner-product score. A natural question is whether the spectral-plus-rotation recipe survives such a link. If yes, the Rohe–Zeng program extends well beyond the linear setting; if no, the boundary tells us something about what the constraint-based generative models are doing that an unconstrained spectral fit cannot.

(LSM, EPM) as a clean test case. The latent space model (LSM) and the edge partition model (EPM) share the same inner-product latent geometry $s_{ij} = \boldsymbol{\theta}_i^\top \boldsymbol{\theta}_j$ but differ exactly in the two ingredients we want to isolate: constraint and link. LSM places $\boldsymbol{\theta}_i \in \mathbb{R}^K$ and uses a logistic link. Its likelihood is $O(K)$-invariant, so it identifies only the column space of Θ. EPM places $\boldsymbol{\theta}_i \in \mathbb{R}_+^K$ and uses a Bernoulli–Poisson link. Non-negativity breaks rotation invariance, and under separability (Donoho & Stodden, 2003; Arora et al., 2013) EPM is identifiable up to column permutation. EPM's non-negative loadings are widely used because they are read off directly as community memberships—the model bakes interpretability into the constraint set. Asking whether LSM + Varimax recovers EPM's structure is therefore the network analogue of the Rohe–Zeng question, with the added difficulty of a nonlinear link.

What we find. We study LSM + Varimax versus EPM on synthetic networks generated from EPM with known Θ. The findings, stated honestly:

1. **Column directions transfer (in moderate-density regimes).** In the moderate-density, separable regime, LSM + Varimax recovers EPM's column directions: per-column cosine similarity to Θ is ≈ 0.95 at $K = 2$. We state this as a conjecture rather than a theorem; a finite-sample directional theorem would require concentration of the LSM-MLE Gram and a Varimax identifiability result for non-negative mixed-membership column directions, neither of which is in hand.
2. **Distributional shape does not transfer.** Across all densities we test, ∼25–30% of LSM + Varimax entries remain negative, row-Hoyer sparsity sits ∼0.2 below EPM's, and column kurtosis remains sub-Gaussian ($\kappa \approx 2$) where the EPM truth Θ is leptokurtic ($\kappa = 5$–30). The leptokurtic identifiability condition of Rohe & Zeng (2023) therefore holds for Θ but fails for the LSM-MLE fit $\hat{Z}$.
3. **Why the recipe stops at directions.** At the population level, an LSM that matches the EPM probabilities satisfies $ZZ^\top = \Lambda_{\mathrm{EPM}}(\Theta\Theta^\top) - \alpha \mathbf{1}\mathbf{1}^\top$ entrywise, with $\Lambda_{\mathrm{EPM}}(s) = \log(e^s - 1)$ (Section 3). Empirically, the finite-sample LSM-MLE Gram tracks this closely enough that the rank-K structure is approximately preserved (Gram-vs-Gram correlation $r = 0.82$–0.88), but the column-wise leptokurtosis of Θ is not inherited by $\hat{Z}$. Varimax can only return leptokurtic axes when leptokurtic directions exist in the column space; on the LSM-MLE column space they do not. We do not claim to have isolated a mechanism; we report what we observe and the implication that follows for Varimax.
4. **Side observation: complementary failure regimes.** LSM + Varimax fails in the sparse regime (where the link distortion is largest) and EPM fails in the dense regime (where Bernoulli–Poisson saturation makes loadings weakly identified). At $K = 2$ the crossover sits near density 0.4. The sparse-side advantage of EPM is robust at $K \in \{2, 3, 5\}$; the dense-side crossover is observed at $K \in \{2, 3\}$ but the densities reachable at $N = 300, K = 5$ do not enter EPM's saturation regime.

What this means. The result is a boundary for the Rohe–Zeng recipe through a nonlinear link. The geometric half—column-space recovery—generalises; the distributional half does not. **Concretely, this says that the non-negativity constraint in models like EPM is not redundant decoration on top of a generative model whose structure could be recovered post-hoc.** It encodes distributional information (sparsity, leptokurtosis) that the column space of an unconstrained inner-product fit, viewed through a nonlinear link, does not contain. Constrained models earn their keep precisely in the part of the structure that post-hoc rotation cannot reach. The complementary-failure-modes finding is a downstream practical corollary: where EPM's sampler also struggles (the dense regime), LSM + Varimax can substitute for EPM's directional structure; where the link distorts (the sparse regime), it cannot.

Related work. Hoff et al. (2002); Hoff (2005; 2008) introduced and analysed LSM. EPM (Zhou, 2015) factorises binary adjacencies via a hierarchical gamma process and the Bernoulli–Poisson link; Caron & Fox (2017); Schein et al. (2015); Acharya et al. (2015); Leisen et al. (2021) are related Poisson/gamma-factorization models. Nakis et al. (2023) place a latent *distance* model on the simplex for NMF-style identifiability; our object of comparison is the inner-product LSM. Identifiability of mixed memberships under separability traces to Donoho & Stodden (2003); Arora et al. (2013); Mao et al. (2017); Airoldi et al. (2008). Our anchor is Rohe & Zeng (2023), which we take as the linear-link benchmark and ask how much survives the LSM–EPM link.

2. Setup

We observe a symmetric binary adjacency matrix $Y \in \{0,1\}^{N \times N}$ with $Y_{ii} = 0$, generated from a model with latent factors $\boldsymbol{\theta}_i \in \mathbb{R}^K$ (or $\mathbb{R}_+^K$) and edge probability depending only on the inner product $s_{ij} \coloneqq \boldsymbol{\theta}_i^\top \boldsymbol{\theta}_j$.

LSM. With intercept $\alpha \in \mathbb{R}$ absorbing density,

$$P(Y_{ij} = 1 \mid \boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \alpha) = \sigma(\alpha + \boldsymbol{\theta}_i^\top \boldsymbol{\theta}_j), \qquad (1)$$

with σ logistic. The likelihood is $O(K)$-invariant under $\boldsymbol{\theta}_i \mapsto Q\boldsymbol{\theta}_i$, so LSM identifies only the column space of Θ.

EPM. With prior $\theta_{ik} \sim \text{Gamma}(\alpha_0, 1)$ and $\boldsymbol{\theta}_i \in \mathbb{R}_+^K$, EPM models edges as a sum of latent per-community Poisson counts, $M_{ijk} \sim \text{Poisson}(\theta_{ik}\theta_{jk})$, with the binary observation an indicator that any community contributed: $Y_{ij} = \mathbb{1}\big[\sum_k M_{ijk} \geq 1\big]$. Marginalising the counts gives the Bernoulli–Poisson link

$$P(Y_{ij} = 1 \mid \boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = 1 - \exp(-\boldsymbol{\theta}_i^\top \boldsymbol{\theta}_j). \tag{2}$$

Non-negativity breaks rotation invariance; under anchor-row separability, EPM is identifiable up to column permutation.

Varimax. Given a fitted $Z \in \mathbb{R}^{N\times K}$, Varimax (Kaiser, 1958) returns $Z_{\text{vmax}} = ZQ^\star$ where

$$Q^\star = \arg\max_{Q\in O(K)} \sum_{k=1}^{K} \left[\frac{1}{N}\sum_i (ZQ)_{ik}^4 - \left(\frac{1}{N}\sum_i (ZQ)_{ik}^2 \right)^2 \right] \tag{3}$$

We sign-fix each column to have non-negative sum. Main analysis at $K = 2$; $K \in \{2, 3, 5\}$ in Section 4.7.

The link, as background. The two log-odds curves agree at moderate-to-large inner products and diverge near zero. For LSM with intercept, $\Lambda_{\text{LSM}}(s) = \alpha + s$. For EPM with $P = 1 - e^{-s}$, $\Lambda_{\text{EPM}}(s) = \log(e^s - 1)$, defined for $s > 0$. A short calculation gives:

Proposition 2.1 (Shape of the log-odds)**.** (i) $\Lambda_{\text{EPM}}(s) = s + \log(1 - e^{-s}) \to s$ *as* $s \to \infty$*; in particular* $|\Lambda_{\text{EPM}}(s) - s| < 0.14$ *for* $s \geq 2$. (ii) $\Lambda_{\text{EPM}}(s) \sim \log s$ *as* $s \to 0^+$*, with* $\Lambda'_{\text{EPM}}(s) \to \infty$. (iii) *An affine LSM log-odds with slope one cannot match this curvature regardless of* α.

Proof. (i) $\Lambda_{\text{EPM}}(s) = \log(e^s(1-e^{-s})) = s + \log(1-e^{-s})$ and $|\log(1-e^{-s})| \leq e^{-s}/(1-e^{-s})$. (ii) Differentiate; near zero $e^s - 1 \approx s$. (iii) Immediate. □

We use this only as background: it identifies the sparse regime as the place where the link distortion is largest, and so where directional recovery is least likely to hold. It is not in itself a contribution. See Figure 1.

3. The question and a directional conjecture

The Rohe–Zeng recipe takes a fitted unconstrained embedding $\hat{U}$ (from PCA), applies Varimax, and returns a rotated $\hat{U}Q^\star$ that recovers the latent factors Z in $\|\cdot\|_{2\to\infty}$ under the leptokurtic identifying assumption (Rohe & Zeng, 2023, Theorem 7.1). Translating this recipe to the network setting gives: fit LSM by maximum likelihood, obtain $\hat{Z}$, apply Varimax, ask how close $\hat{Z}_{\text{vmax}}$ is to the EPM truth Θ.

Why the directional half should plausibly work. Three ingredients support directional recovery. First, when s_{ij} is bounded away from zero, the LSM and EPM log-odds are approximately equal (Proposition 2.1), so the Gram of the LSM MLE is close to the Gram of Θ in this regime, and the LSM column space is close to the column space of Θ. Second, LSM is $O(K)$-invariant, so the LSM solution is determined only up to a rotation; the question becomes which rotation to pick. Third, separable non-negative loadings have a distinctive fourth-moment signature—each column has zeros from anchors of other communities and large values from anchors of itself, making column-squared entries spiky. Varimax (Equation 3) maximises exactly this fourth-moment quantity, and so among all rotations of a column space spanned by separable non-negative columns, it is plausibly the one that returns to the non-negative basis up to permutation and sign.

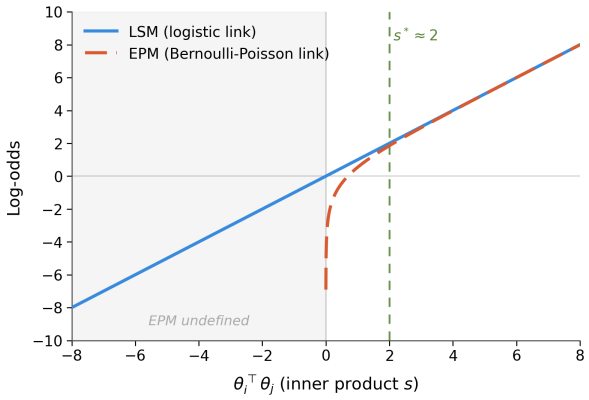

Figure 1. LSM ($\alpha + s$, blue, $\alpha = 0$) and EPM ($\log(e^s - 1)$, dashed orange) log-odds. The intercept can shift LSM vertically but cannot reproduce EPM's curvature near zero. The two curves agree for moderate-to-large s.

Why the distributional half should fail. The Rohe–Zeng theorem uses leptokurtosis of the latent factors as its identifying assumption (Rohe & Zeng, 2023, Theorem 5.1, Assumption 1). In the linear setting, PCA on a centred data matrix preserves the leptokurtic structure of Z because the data matrix is linearly related to Z. In our setting the relationship is no longer linear: an LSM that matches the EPM probabilities at the population level satisfies $\sigma(\alpha + z_i^\top z_j) = 1 - \exp(-\theta_i^\top \theta_j)$, which inverts entrywise to

$$ZZ^\top = \Lambda_{\text{EPM}}\big(\Theta\Theta^\top\big) - \alpha\,\mathbf{1}\mathbf{1}^\top, \tag{4}$$

with $\Lambda_{\text{EPM}}(s) = \log(e^s - 1)$ applied entrywise. This is a population statement; the finite-sample LSM-MLE will deviate from it. There is no general reason for the entrywise nonlinear map Λ_{EPM} to preserve the column-wise distributional shape of Θ, even when it preserves the column space approximately. We therefore expect—and confirm in Section 4.3—that the leptokurtic identifying assumption

fails on $\hat{Z}$ even when it holds on Θ. This bars distributional recovery.

The conjecture. With these caveats, we state:

Conjecture 3.1 (Directional recovery)**.** *Suppose* $Y \sim \mathrm{EPM}(\Theta)$ *with* Θ *satisfying anchor-row separability and full-rank* $\mathbb{E}[\theta\theta^\top]$*. In the moderate-density regime where* $\langle\theta_i, \theta_j\rangle$ *is bounded away from* 0*, the LSM MLE after Varimax rotation and sign-fixing recovers the column* directions *of* Θ*: there exists a permutation* π *and signs* $\varepsilon_k \in \{-1, +1\}$ *such that for each* k*, the per-column cosine similarity* $|\cos(\hat{Z}_{\mathrm{vmax},\cdot k}, \Theta_{\cdot\pi(k)})|$ *approaches one.*

The conjecture is supported by the empirics in Section 4.2, but a formal proof is open: it would require concentration of the LSM-MLE Gram around $\Theta\Theta^\top$ under separability and a Varimax identifiability theorem for non-negative mixed-membership column directions in a non-leptokurtic column space. The conjecture deliberately stops at directions and does not claim entry-wise recovery (Section 4.3).

4. Empirical analysis

4.1. Setup and metrics

We generate networks from EPM with known $\Theta_{\mathrm{true}} \in \mathbb{R}_+^{N\times K}$. Each row is either an anchor row (all mass on one coordinate; ensures separability) or a Dirichlet-mixed row. A scalar `scale` multiplies Θ_{true} to control density. We fit LSM with intercept by maximum likelihood, apply Varimax with sign-fix, and fit EPM by Gibbs sampling. $N = 300$, $K = 2$, five replicates per setting.

Two metrics for two questions. The cross-model question—do LSM and EPM recover similar column spaces of Θ?—needs a metric that quotients out rotation, since LSM is rotationally non-identifiable while EPM is identifiable up to permutation. We use similarity Procrustes (Schönemann, 1966),

$$d_{\mathrm{sim}}(M, T) = \min_{\substack{Q\in O(K),\, c>0,\\ \Pi \text{ signed perm.}}} \frac{\|cMQ\Pi - T\|_F}{\|T\|_F}, \qquad (5)$$

which is rotation-invariant and so scores LSM raw and LSM + Varimax identically.

The Varimax-evaluation question—did Varimax pick the correct basis within the LSM column space?—is exactly the question for which rotation invariance is wrong. Following the spirit of Rohe & Zeng's $\|\cdot\|_{2\to\infty}$ over signed permutations, we use per-column scaled Procrustes

$$d_{\mathrm{pcs}}(M, T) = \min_{\Pi,\, c_1,\ldots,c_K>0} \frac{\left\|\sum_k c_k (M\Pi)_{\cdot k} e_k^\top - T\right\|_F}{\|T\|_F}, \qquad (6)$$

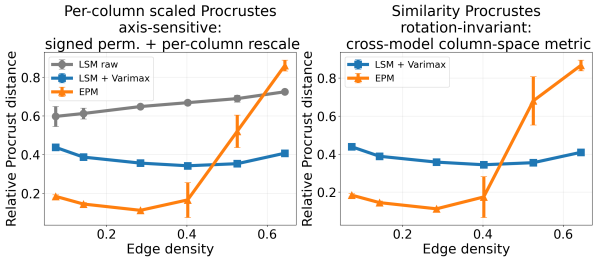

Figure 2. Density sweep, two metrics. *Left (d_{pcs}, axis-sensitive):* LSM raw and LSM + Varimax separate clearly, with the Varimax effect narrowing the relative distance from ∼0.6–0.7 to ∼0.34–0.44 across the entire range. *Right (d_{sim}, rotation-invariant):* LSM raw and LSM + Varimax coincide (the metric optimises over rotations). EPM's d_{sim} rises with density; LSM + Varimax stays nearly flat. Crossover near $\rho \approx 0.4$.

which optimises over a signed permutation and one positive rescale per column, but no rotation. The per-column rescaling c_k here is a metric convention, not a model invariance: as noted in Section 2, symmetric EPM does not have a per-column scale freedom, but Gibbs runs return $\hat{\Theta}$ at slightly different overall column scales across replicates (and prior shape/rate parameters need not match the truth's), so we allow $c_k > 0$ in the alignment to avoid penalising that gap. d_{pcs} is bounded below by $\sin\angle(\hat{Z}_{\cdot k}, \Theta_{\cdot k})$ per column, so a small value implies the directional recovery in Conjecture 3.1. The point that a rotation-invariant metric cannot evaluate a rotation method, and that the two questions need distinct metrics, is itself a small methodological clarification.

Distributional diagnostics. We report column-averaged Pearson kurtosis κ on the raw fitted embedding, alongside fraction of negative entries, row-averaged Gini of $|M|$, and row-Hoyer sparsity (Hoyer, 2004).

4.2. Directional recovery (testing the conjecture)

We sweep `scale` $\in \{0.2, 0.3, 0.5, 0.7, 1.0, 1.5\}$ holding anchor rows fixed, producing densities $\rho \approx 0.07$ to 0.64. Five replicates per scale. Figure 2 reports both Procrustes metrics.

The axis-sensitive panel (d_{pcs}, left) is the test of Conjecture 3.1. LSM raw sits at $d_{\mathrm{pcs}} \approx 0.6$–$0.7$ across the range; LSM + Varimax sits at 0.34–0.44, below LSM raw at every density. The Varimax rotation moves the embedding measurably toward the EPM axes, with per-column cosine similarity to truth ≈ 0.95 in the moderate-to-dense window. This supports directional recovery in the regime the conjecture targets.

The rotation-invariant panel (d_{sim}, right) does not evaluate the rotation but speaks to cross-model column-space agreement. LSM + Varimax stays nearly flat at 0.34–0.44. EPM's

d_{sim} rises from 0.18 at $\rho = 0.07$ to 0.87 at $\rho = 0.64$, a steep degradation under Bernoulli–Poisson saturation. The two methods cross near $\rho \approx 0.4$, with EPM dominating below and LSM + Varimax above. Section 4.4 returns to this.

4.3. The leptokurtic limit (the centrepiece)

We now turn to the prediction in Section 3: distributional recovery should fail because the link destroys leptokurtosis. Three diagnostics make this concrete.

Non-negativity and sparsity. EPM is non-negative by construction; LSM raw is roughly half-negative; LSM + Varimax cuts this to ∼25–30% but does not push it below. EPM's row-Hoyer sparsity is consistently 0.73–0.84 while LSM + Varimax sits at 0.47–0.62, with the gap largest in sparse regimes (Figure 3).

Kurtosis. The EPM truth Θ has column kurtosis 5–30 across regimes (Figure 4, orange), well into the leptokurtic regime $\kappa > 3$ that Rohe & Zeng's identifiability requires. LSM raw has $\kappa \approx 2.5$, and LSM + Varimax remains at $\kappa \approx 2$, indistinguishable from LSM raw up to rotation noise. The leptokurtic condition fails for $\hat{Z}$.

What this implies for Varimax. Rohe & Zeng's argument requires the embedding column space to inherit the distributional shape of the latent factors. PCA on a centred data matrix achieves this because the data matrix is linearly related to the factors. In our setting the population analogue is Eq. (4): $ZZ^\top = \Lambda_{\text{EPM}}(\Theta\Theta^\top) - \alpha\mathbf{1}\mathbf{1}^\top$, with Λ_{EPM} applied entrywise. Empirically, the finite-sample LSM-MLE Gram tracks this population picture closely enough that the rank-K structure is approximately preserved—Gram-vs-Gram entries (between $\hat{Z}\hat{Z}^\top$ and $\Theta\Theta^\top$) are correlated $r = 0.82$–0.88 across the regimes we test—but the column-wise leptokurtosis of Θ is not inherited by $\hat{Z}$. Varimax can only return leptokurtic axes when leptokurtic directions exist in the column space; on the LSM-MLE column space they do not, so Varimax cannot manufacture them. The rotation does its job within the column space available to it; the column space itself is the limit.

This is consistent with Conjecture 3.1 stopping at directional recovery: post-hoc rotation cannot reach distributional structure that the unconstrained fit's column space does not contain. We emphasise that the chain—population relation Eq. (4), finite-sample MLE, non-leptokurtic $\hat{Z}$—is suggestive rather than proven; what we establish is the empirical pattern and its consequence for Varimax, not a derivation of the pattern from the link.

4.4. Side observation: complementary failure regimes

The cross-model panel of Figure 2 also delivers a finding that is useful as practical guidance, though it is not the headline of the paper. EPM dominates LSM + Varimax in column-space recovery below $\rho \approx 0.4$; LSM + Varimax dominates above.

LSM + Varimax fails in the sparse regime, where Proposition 2.1 establishes that the link distortion is largest—small-s curvature is exactly what an affine LSM log-odds cannot reproduce.

EPM degrades in the dense regime, where the picture is more entangled. Two non-exclusive explanations are consistent with the empirics. *First*, Bernoulli–Poisson saturation makes $P_{ij} \approx 1$ when $\boldsymbol{\theta}_i^\top \boldsymbol{\theta}_j$ is large, so binary observations carry little information about the magnitude of the inner product once that product is large. The likelihood is weakly informative in the magnitude direction, and any inference procedure would struggle. *Second*, the Gibbs sampler may converge slowly in this regime. We do not cleanly separate the two: in the `heavy_overlap_no_anchors` setting (Section 4.5), a 5× longer burn-in left EPM output essentially unchanged (d_{sim} std 0.02), which points away from slow convergence *in that case*, but we lack multi-chain $\hat{R}$ diagnostics, effective-sample-size estimates, and have not tested alternative samplers (e.g. HMC). What we can say cleanly: at $\rho = 1.0$, EPM's d_{sim} standard deviation across replicates exceeds 0.13, versus LSM + Varimax's < 0.02. EPM is unstable in the dense regime for reasons we partly but not fully isolate.

4.5. Three illustrative regimes

To make the column-space versus distributional gap concrete, we inspect three regimes that bracket the conjecture's assumptions: `dense_separable` (anchor rows, $\rho \approx 0.52$), `sparse` (anchor rows, $\rho \approx 0.13$), and `heavy_overlap_no_anchors` (Dirichlet mix only, $\rho \approx 0.62$). Figure 5 plots raw embeddings; Table 1 reports diagnostics.

In `dense_separable`, LSM + Varimax reaches the EPM truth's column space ($d_{\text{sim}} = 0.34$) while EPM saturates ($d_{\text{sim}} = 0.69$ with std 0.28). In `sparse`, the picture is reversed: EPM recovers the L cleanly ($d_{\text{sim}} = 0.14$) while LSM + Varimax cannot. The most informative regime is `heavy_overlap_no_anchors`, where the truth has no axis structure: even with a 5× longer EPM burn-in, EPM produces a sharp axis-aligned L (Hoyer 0.89, $\kappa = 24.7$) far from the truth; LSM + Varimax sits much closer to the truth's column space at the cost of less concentrated row sparsity (Hoyer 0.35, $\kappa = 3.2$). The high replicate variance of EPM in saturation (d_{sim} std 0.28) is consistent across runs.

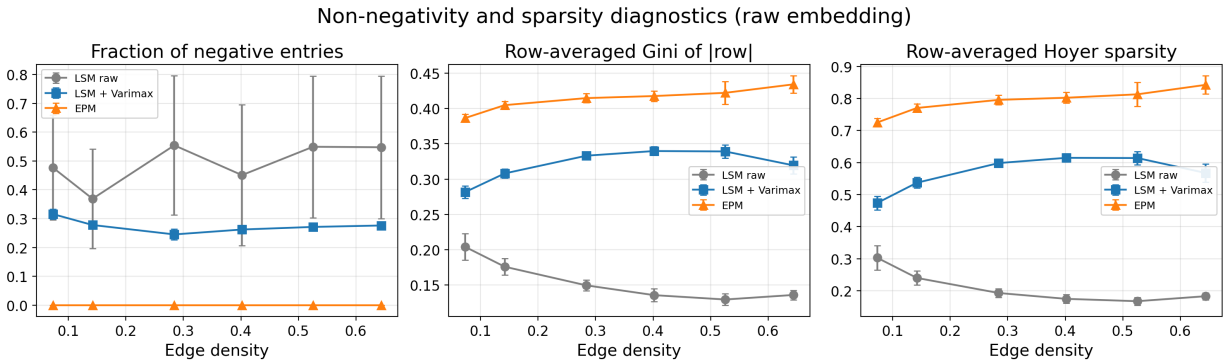

Figure 3. Non-negativity and sparsity diagnostics on the raw embedding versus density. LSM + Varimax retains ∼25–30% negative entries and Hoyer sparsity ∼0.2 below EPM's. Varimax narrows but never closes these gaps.

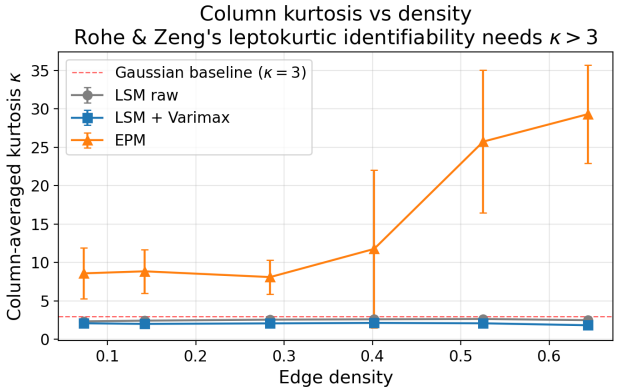

Figure 4. Column-averaged Pearson kurtosis κ versus density. EPM is strongly leptokurtic ($\kappa = 8$–29); LSM raw is sub-Gaussian ($\kappa \approx 2.5$); LSM + Varimax remains sub-Gaussian ($\kappa \approx 2$), matching LSM raw to within rotation noise. The leptokurtic identifiability condition holds for Θ but not for $\hat{Z}$.

4.6. Geometry-vs-interpretability trade-off

Combining the column-space and row-sparsity diagnostics into a single picture clarifies that no method dominates uniformly (Figure 6). EPM occupies the high-Hoyer band uniformly but its highest-density markers veer rightward into poor column-space fidelity. LSM + Varimax sits in a tight middle band: stable Hoyer near 0.5–0.6, stable d_{sim} near 0.34–0.44. LSM raw lies at the bottom: low Hoyer everywhere, similar d_{sim} to LSM + Varimax (same Gram), but without Varimax's row-sparsity bonus. The plot summarises the rest of the empirical analysis: column-space fidelity and row-sparsity are not jointly optimised by any of the three methods across the density range.

4.7. Generalisation to $K > 2$

We repeat the density sweep at $K \in \{2, 3, 5\}$ with $N = 300$ and anchor-row separability fixed (Figure 7).

Table 1. Three regimes ($N = 300$, $K = 2$, mean±std over 5 replicates). EPM in `heavy_overlap_no_anchors` uses 1500 burn-in iterations.

Method	d_{sim}	d_{pcs}	frac<0	Hoyer	κ
dense_separable ($\rho \approx 0.52$; $\Theta\,\kappa \approx 5.1$)					
LSM raw	0.34	0.72	0.35	0.16	2.5
LSM+Varimax	**0.34**	**0.34**	0.27	0.61	2.0
EPM	0.69	0.32	0.00	0.83	17.0
sparse ($\rho \approx 0.14$; $\Theta\,\kappa \approx 5.1$)					
LSM raw	0.38	0.73	0.27	0.21	2.7
LSM+Varimax	0.38	0.38	0.27	0.52	2.0
EPM	**0.14**	**0.14**	0.00	0.77	7.7
heavy_overlap_no_anchors ($\rho \approx 0.67$)					
LSM raw	0.33	0.57	0.49	0.44	3.1
LSM+Varimax	**0.33**	**0.48**	0.10	0.35	3.2
EPM	0.78	0.77	0.00	0.89	24.7

EPM dominates LSM + Varimax in the sparse regime at every K, indicating that the link-function obstruction is not specific to $K = 2$. The dense-side crossover appears at $K = 2, 3$ but not at $K = 5$, where the densities reachable at $N = 300$ ($\rho \lesssim 0.23$) do not enter EPM's saturation regime. Reaching the dense-side crossover at $K = 5$ would require larger N or larger `scale`; we leave this to future work. The complementary-failure-modes finding is therefore K-robust on the sparse side and supported up to $K = 3$ on the dense side. Note that LSM + Varimax's d_{sim} grows with K ($\approx 0.34, 0.41, 0.55$ at $K = 2, 3, 5$), suggesting that rotational identification gets harder at higher K.

5. Discussion

What we have shown. Asking whether LSM + Varimax recovers EPM's structure is the network analogue of the Rohe–Zeng question, with a nonlinear link added. Empirically, the geometric half generalises: in the moderate-density, separable regime, LSM + Varimax recovers EPM's

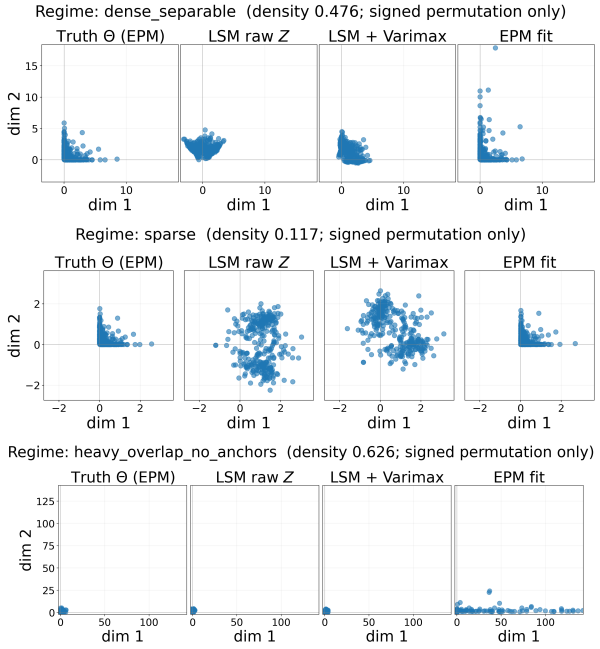

Figure 5. Raw fitted embeddings in three regimes (signed-permutation alignment only, no rotation). Top: `dense_separable`—Varimax pulls the LSM cloud into the upper-right quadrant, with rows directionally aligning with the EPM axes, although the bulk distribution remains a Gaussian-shape blob rather than a sharp L. Middle: `sparse`—LSM positions are diffuse; only EPM recovers the L. Bottom: `heavy_overlap_no_anchors`—truth has no cone; LSM and Varimax sit close to truth while EPM forces an L that is not present.

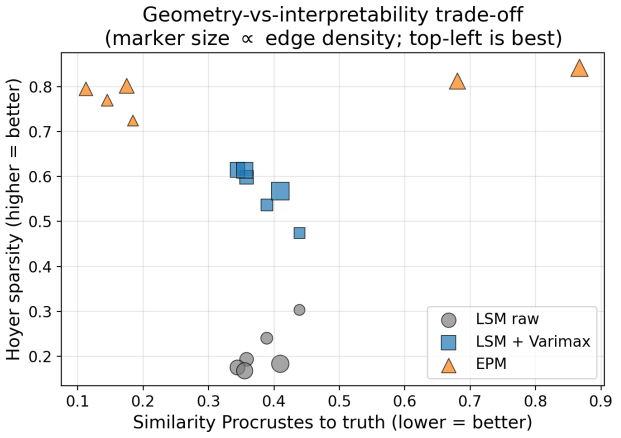

Figure 6. Column-space fidelity (d_{sim}, lower-better) versus row sparsity (Hoyer, higher-better) per method per density. Marker size $\propto$ density. EPM dominates LSM + Varimax in row sparsity everywhere but is dominated in column-space fidelity above $\rho \sim 0.4$. LSM + Varimax has the most stable column-space fidelity across the range.

column directions to per-column cosine similarity ≈ 0.95. The distributional half does not: across all densities tested, the LSM + Varimax embedding remains $\sim$25–30% negative, has Hoyer sparsity $\sim$0.2 below EPM's, and is sub-Gaussian where the EPM truth is leptokurtic. Empirically, the LSM-MLE Gram tracks the population relation in Eq. (4), $ZZ^\top = \Lambda_{\text{EPM}}(\Theta\Theta^\top) - \alpha \mathbf{1}\mathbf{1}^\top$, closely enough that the rank-K column space is approximately preserved (Gram-vs-Gram correlation 0.82–0.88), but the column-wise leptokurtosis of Θ is not inherited by $\hat{Z}$. Whatever combination of the link, the MLE objective, and the optimisation produces this pattern, the consequence for Varimax is the same: the leptokurtic identifying assumption holds for Θ but fails for $\hat{Z}$. Constrained models like EPM are not fully replaced by a post-hoc rotation of an unconstrained fit because the constraint encodes distributional information (sparsity, leptokurtosis) that the unconstrained column space, in this setting, does not contain.

Where this sits relative to Rohe & Zeng. Their result is positive and unifying: a fast spectral plus rotation procedure consistently estimates the latent structure of a broad class of linear factor models, including SBM and LDA. Our result is a boundary on the obvious extension. Through a logistic / Bernoulli–Poisson link, the column-space half of the recipe survives but the distributional half does not. This neither contradicts their theorem nor diminishes its scope; it identifies the part of the program that does not generalise mechanically. A more general statement of the recipe would have to distinguish the column-space identification (which we conjecture survives nonlinear links of this kind, given separability) from the distributional identification (which apparently does not, and would require a different identifying assumption on $\hat{Z}$, not on Θ).

Practical guidance. A side observation, useful for practitioners. LSM + Varimax and EPM fail in opposite density regimes. The crossover at $K = 2$ sits near $\rho \approx 0.4$. Below this, EPM's Bernoulli–Poisson link gives a sparse-regime calibration advantage, and its loadings are well identified. Above this, EPM is unstable under Bernoulli–Poisson saturation—the likelihood is weakly informative in the magnitude direction and inference struggles, with replicate variance climbing accordingly (Section 4.4)—while LSM + Varimax is stable. The two are therefore complementary tools, with LSM + Varimax appropriate exactly where EPM also struggles—provided one is content with directional rather than distributional recovery.

Limitations. Symmetric, undirected, binary networks; synthetic data with known Θ; $N = 300$. Conjecture 3.1 is supported by empirics but unproven. The dense-side crossover finding holds at $K \in \{2, 3\}$; at $K = 5$ the density range with $N = 300$ is too narrow to enter EPM's saturation regime. No real-network experiments.

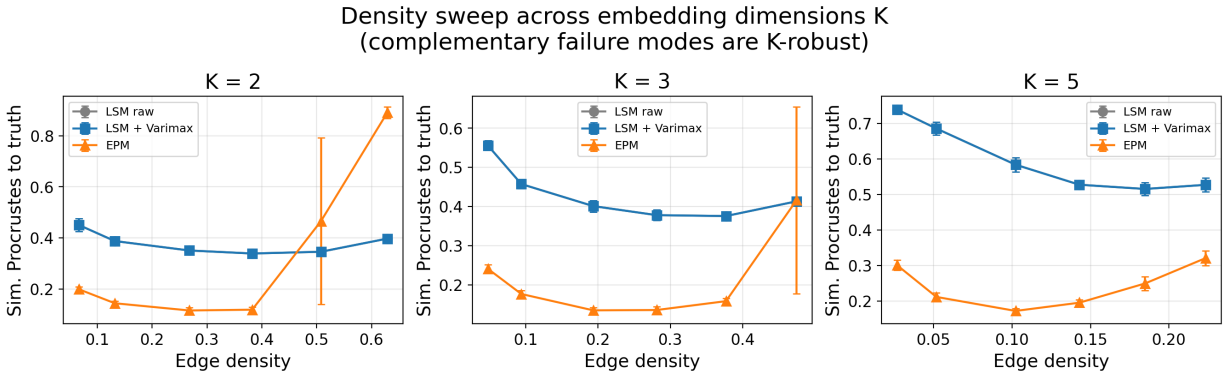

Figure 7. Similarity Procrustes versus density at $K \in \{2, 3, 5\}$. EPM's sparse-regime advantage holds across all K. The dense-side crossover is observed at $K = 2, 3$; at $K = 5$ the densities reachable with $N = 300$ stop short of EPM's saturation regime.

Future work. The most direct theoretical step is a finite-sample directional identifiability theorem matching Conjecture 3.1: concentration of the LSM-MLE Gram around $\Theta\Theta^\top$ under separability, plus a Varimax identifiability statement for non-negative mixed-membership column directions in a column space that is no longer leptokurtic. Empirically: larger N and K for the dense-side crossover; directed networks; real networks with mixed-membership ground truth. More broadly, the Rohe–Zeng program through nonlinear links is largely open territory: which other links and constraint sets admit column-space-only generalisations of the recipe?

Acknowledgements

Do not include acknowledgments in the initial submission.

Impact Statement

This paper presents work whose goal is to advance the theoretical understanding of probabilistic network embeddings within the field of machine learning. Our contributions are methodological and theoretical; no application-specific risks are foreseen.

References

Acharya, A., Ghosh, J., and Zhou, M. Nonparametric Bayesian factor analysis for dynamic count matrices. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 1–9. PMLR, 2015.

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., Wu, Y., and Zhu, M. A practical algorithm for topic modeling with provable guarantees. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28 of *Proceedings of Machine Learning Research*, pp. 280–288. PMLR, 2013. URL https://proceedings.mlr.press/v28/arora13.html.

Caron, F. and Fox, E. B. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B*, 79(5):1295–1366, 2017.

Donoho, D. and Stodden, V. When does non-negative matrix factorization give a correct decomposition into parts? In *Advances in Neural Information Processing Systems*, volume 16, pp. 1141–1148, 2003.

Hoff, P. D. Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469): 286–295, 2005.

Hoff, P. D. Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems*, 20:657–664, 2008.

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

Holland, P. W., Laskey, K. B., and Leinhardt, S. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

Hoyer, P. O. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004. URL https://www.jmlr.org/papers/volume5/hoyer04a/hoyer04a.pdf.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.

Leisen, F., Villa, C., and Walker, S. G. Nonnegative Bayesian nonparametric factor models with completely random measures. *Statistics and Computing*, 31(5):1–14, 2021.

Mao, X., Sarkar, P., and Chakrabarti, D. On mixed memberships and symmetric nonnegative matrix factorizations. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2324–2333. PMLR, 2017.

Nakis, N., Çelikkanat, A., and Mørup, M. HM-LDM: A hybrid-membership latent distance model. In *Complex Networks and Their Applications XI*, volume 1077 of *Studies in Computational Intelligence*, pp. 350–363. Springer, 2023.

Rohe, K. and Zeng, M. Vintage factor analysis with Varimax performs statistical inference. *Journal of the Royal Statistical Society: Series B*, 85(4):1037–1060, 2023.

Schein, A., Paisley, J., Blei, D. M., and Wallach, H. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the 21th ACM SIGKDD International conference on knowledge discovery and data mining*, pp. 1045–1054, 2015.

Schönemann, P. H. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10, 1966. doi: 10.1007/BF02289451.

Zhou, M. Infinite edge partition models for overlapping community detection and link prediction. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 1135–1143. PMLR, 2015.