Check for updates

Received: 23 February 2021 | Revised: 28 March 2021 | Accepted: 11 April 2021

DOI: 10.1002/sta4.381

SPECIAL ISSUE PAPER

WILEY

Spectral analysis of networks with latent space dynamics and signs

Joshua Cape

Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence
Joshua Cape, Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.
Email: joshua.cape@pitt.edu

Funding information
National Science Foundation, Grant/Award Numbers: DMS 1902755, SES 1951005

Abstract

We pursue the problem of modelling and analysing latent space dynamics in collections of networks. Towards this end, we pose and study latent space generative models for signed networks that are amenable to inference via spectral methods. Permitting signs, rather than restricting to unsigned networks, enables richer latent space structure and permissible dynamic mechanisms that can be provably inferred via low rank truncations of observed adjacency matrices. Our treatment of and ability to recover latent space dynamics holds across different levels of granularity, namely, at the overall graph level, for communities of nodes, and even at the individual node level. We provide synthetic and real data examples to illustrate the effectiveness of methodologies and to corroborate accompanying theory. The contributions set forth in this paper complement an emerging statistical paradigm for random graph inference encompassing random dot product graphs and generalizations thereof.

KEYWORDS
dimensionality reduction, dynamics, graphs, latent space models, networks, spectral methods

1 | INTRODUCTION

The analysis and modelling of network data has arisen as an active research area within the statistics community and throughout the sciences (Crane, 2018; Kolaczyk, 2009). Networks (i.e., graphs), which convey information about interactions between entities, serve as useful data objects for representing and studying diverse processes and phenomena, concomitantly finding widespread use in biology (Barabasi & Oltvai, 2004; Horvath, 2011), economics (Jackson, 2011; Mele, 2017), and neuroscience (Bullmore & Bassett, 2011; Bullmore & Sporns, 2009), among other disciplines. An overarching aim shared by a growing number of researchers is to place the analysis and modelling of network data on a firm statistical footing. Contemporary efforts seek to develop analogues of classical statistical concepts for network-centric data settings, particularly for problems involving multiple networks. The purpose of the present paper is to further contribute to such efforts.

In this paper, we observe as data a collection of graphs (i.e., networks) indexed by time. At each discrete time step, the observed graph is posited to arise in a manner governed by an unobserved (i.e., latent) matrix that is also indexed by time. Given an observed collection of graphs, we seek to make inferences about the unobserved collection of underlying latent matrices. When the latent matrices evolve over time according to some (possibly unknown, possibly time-dependent) mechanism, we further seek to make inferences about the generative mechanism using only the observed graphs. Here, generative mechanisms amount to prescribing latent space dynamics, which we model and estimate using spectral methods.

The term *spectral methods* broadly refers to the versatile techniques rooted in matrix analysis that involve (partial) matrix eigendecompositions and singular value decompositions. Spectral methods underlie principal component analysis and are at the heart of multivariate statistical analysis, particularly dimensionality reduction. Of note, truncated matrix decompositions and factorizations are tractable, competitive

Abbreviation: RDPG, random dot product graph.

 © 2021 John Wiley & Sons, Ltd.

approaches for obtaining provable estimation and clustering guarantees in conjunction with concentration inequalities and matrix perturbation analysis. Spectral methods can be easily implemented at relatively large scale using modern statistical software, benefiting in turn from advances in numerical linear algebra routines. Collectively, these advantages motivate our focus on spectral methods.

The contributions set forth in this paper complement an emerging statistical paradigm for random graph inference encompassing random dot product graphs (RDPGs) and generalizations thereof. This paper augments existing modelling, theory, and methods to encompass signed networks and to illustrate the recovery of latent space dynamics and general geometric structure. One insight is that modelling the presence or absence of signed edges can be succinctly achieved by altering the distributional support of existing dot product generative approaches for unsigned graphs, without requiring modifications directly to the generative mechanisms themselves or sacrificing interpretability. Geometrically, this allows one to model latent space as (an unrestricted subset of) the Euclidean unit ball rather than as a collection of assorted vectors required to satisfy stringent positive inner product constraints, hence permitting more diverse dynamics and connectivity phenomena in latent space and in observed networks.

2 | BACKGROUND AND CONTEXT

A growing body of literature in statistics addresses questions of modelling and inference for multiple networks (graphs). Research themes include the development of modelling capabilities, methodologies, and representations for joint, simultaneous dimensionality reduction (Arroyo et al., 2020; Levin et al., 2017; Pantazis et al., 2020), as well as the analysis of time series of networks for the purpose of detecting change-points or anomalous behaviour, at the individual node level or at the overall graph level (Chen et al., 2020; Padilla et al., 2019; Wang et al., 2013; Xu & Hero, 2014). The statistical treatment of multiple networks as data largely builds on historical developments that focus on single-graph modelling, analysis, and inference. A common approach consists of using Erdős–Rényi random graphs (Erdős & Rényi, 1959,1960) and stochastic block models (Holland et al., 1983), together with their variants, as a starting point for studying changes in network structure that can be broadly construed as *dynamics*. Examples of network dynamics include, for example, changes in the overall number of communities, in relative community sizes, or in the connectivity patterns of specific nodes. Pertinent recent works include Chiang et al. (2014), Crane (2015), Han et al. (2015), Kim et al. (2018), Kunegis et al. (2010), Liu et al. (2018), Matias and Robin (2014), Matias and Miele (2017), Rastelli et al. (2018), Sarkar and Moore (2006), Sarkar et al. (2014), Sewell and Chen (2015, 2017), and Zhu et al. (2017), many of which adopt latent space modelling approaches (Hoff et al., 2002).

Spectral methods, in particular spectral clustering approaches (Ng et al., 2002; von Luxburg, 2007), have received considerable attention in the single-network context, enjoying widespread success in community detection (i.e., clustering) (Lei & Rinaldo, 2015; Rohe et al., 2011) as well as more broadly in the discovery of structure within network-valued data. In contemporary research on multiple networks, spectral methods exhibit similar advantages in terms of their methodological flexibility, tractability for developing model-based statistical theory, and computational advantages that leverage concurrent advances in eigen-solvers and parallel scientific computing. There and here, truncated eigen-decompositions are treated as a method of data denoising and dimensionality reduction, insofar as (stochastic) adjacency matrices are often viewed as noisy versions of their (deterministic) population-level counterpart, an expectation matrix (possibly low rank) of Bernoulli edge probabilities. A flurry of recent activity in the literature establishes a high-resolution theoretical understanding of denoising at the level of individual, node-specific low-dimensional representations (i.e., *embeddings*) in large networks (Cape et al., 2019; Fan et al., 2019; Lei, 2019; Lyzinski et al., 2014). Results at this refined level of granularity involve entry-wise and row-wise perturbations of (concatenated) high-dimensional eigenvectors, which together with the aforementioned references collectively form a backdrop for the present paper.

A core theme of this paper is modelling and recovering geometric structure present in latent space that produces connectivity behaviour in observed networks. Such a focus accords with the broad study of low-dimensional Euclidean latent space geometry, encompassing cluster centroids in stochastic block models (Holland et al., 1983), line segments in degree-corrected block models (Karrer & Newman, 2011), simplicial regions in mixed-membership block models (Airoldi et al., 2008), and beyond (Athreya et al., 2021). More recently, investigations into hyperbolic indefinite latent space geometry consider network analysis absent distance-based latent Euclidean geometry (Rubin-Delanchy et al., 2017; Smith et al., 2019).

Notation. For any positive integer n, let $[[n]] \equiv \{1,2,\ldots,n\}$. Write $X \in \mathbb{R}^d$ to denote a real-valued d-dimensional random vector. Deterministic (column) vectors are written in terms of their elements as $\mathbf{x} \equiv (x_1, \ldots, x_d)^\top$, where superscript $\top$ denotes the transpose operation. Matrices are written element-wise as $\mathbf{A} \equiv (a_{ij}) \in \mathbb{R}^{m \times n}$ and indexed in the manner $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}$, and so forth. The symbol $\circ$ denotes element-wise multiplication of matrices, that is, $(\mathbf{A} \circ \mathbf{B})_{ij} \equiv a_{ij} b_{ij}$. Entry-wise absolute value of a matrix is written as $|\mathbf{A}| = (|a_{ij}|) \in \mathbb{R}^{m \times n}$. The standard sign function is denoted by $\mathrm{sgn}(\cdot) \in \{-1,0,1\}$ and acts element-wise when applied to matrices. The Euclidean inner (dot) product between vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_i x_i y_i$ with the corresponding Euclidean norm $\|\mathbf{x}\|_2 \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. The Frobenius norm of a matrix is $\|\mathbf{A}\|_{\mathrm{F}} := \sqrt{\sum_{i,j} |a_{ij}|^2}$, the matrix operator norm is $\|\mathbf{A}\|_{\mathrm{op}} := \sup_{\|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2$, and the maximum absolute entry norm is $\|\mathbf{A}\|_{\max} := \max_{i,j} |a_{ij}|$. The set of d by d real orthogonal matrices is denoted by $\mathbb{O}_d$. Matrices with the prime character, $'$, serve as intermediaries and are not of central importance. Overscript tilde, $\sim$, is used to denote normalized vectors with unit Euclidean norm, that is, $\tilde{\mathbf{x}} := \mathbf{x} \|\mathbf{x}\|_2^{-1}$, while for matrices, the use of overscript tilde depends upon context. Finally, overscript hat notation, $\widehat{\mu}$, is used together with $\doteq$ to denote the (rounded) numerical value of a sample quantity, for example, $\widehat{\mu} \doteq 1.23$.

3 | SETUP

We begin by stating a random graph model in which a fixed, given latent matrix $\mathbf{X}$ (of row vectors) gives rise to a random graph $\mathcal{G}$ which is represented by its corresponding (signed, stochastic) adjacency matrix $\mathbf{A} \equiv \mathbf{A}_{\mathcal{G}}$.

Definition 1 Signed dot product graph [sgn-DPG]. Let $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^d$ be a collection of vectors satisfying $\|\mathbf{x}_i\|_2 \le 1$; arrange them row-wise in the matrix $\mathbf{X} := [\mathbf{x}_1|\mathbf{x}_2|...|\mathbf{x}_n]^\top \in \mathbb{R}^{n\times d}$. Let $\mathbf{A} \in \{-1,0,1\}^{n\times n}$ be a symmetric matrix defined in the manner

$$\mathbf{A} := \operatorname{sgn}(\mathbf{X}\mathbf{X}^\top) \circ \mathbf{A}',$$

where the random matrix $\mathbf{A}' \in \{0,1\}^{n\times n}$ satisfies $a'_{ij} \sim \text{Bernoulli}(|\langle \mathbf{x}_i, \mathbf{x}_j\rangle|)$ independently for $i \le j$ and $a'_{ji} = a'_{ij}$. Matrices $\mathbf{A}$ are called signed adjacency matrices of signed dot product graphs (sgn-DPGs) with underlying latent vectors $\{\mathbf{x}_i\}_{i=1}^n$.

The occurrence of $\mathbf{X}\mathbf{X}^\top$ in Definition 1 amounts to the choice of the standard dot product link function (ignoring signs), namely, $f(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j\rangle$ for all rows $\mathbf{x}_i, \mathbf{x}_j$ of $\mathbf{X}$, a choice with a history of use in social network analysis (e.g., Young & Scheinerman, 2007). This choice can be viewed as a first-order approximation of more intricate, nonlinear link functions, in the spirit of a truncated Taylor series expansion. Furthermore, using the standard inner product permits one to apply the full arsenal of matrix analysis for the purpose of analysing matrix factorizations and their perturbations, in a manner reminiscent of factor modelling.

Given an observation $\mathbf{A}$ via Definition 1, one wishes to make inferences about the unobserved latent matrix $\mathbf{X}$. Note that dot products between rows of $\mathbf{X}$ give rise to Bernoulli edge probabilities, namely, the vector $\mathbf{x}_i$ encodes connectivity information corresponding to node i in $\mathbf{A}$ and can therefore be viewed as a parameter of interest. Correspondingly, for any specified latent matrix $\mathbf{X}$, the expectation of $\mathbf{A}$, denoted by $\mathbb{E}_{\mathbf{X}}[\mathbf{A}]$, is given by $\mathbf{X}\mathbf{X}^\top$. Here, the expectation $\mathbb{E}_{\mathbf{X}}[\mathbf{A}]$ can be interpreted as the population-level (signed) adjacency matrix, while observations $\mathbf{A}$ can be interpreted as noisy versions thereof. Importantly, in general, the entries of the expectation matrix take values outside the set $\{-1,0,1\}$.

One source of nonidentifiability to keep in mind is that $\mathbf{X}\mathbf{X}^\top \equiv (\mathbf{X}\mathbf{W})(\mathbf{X}\mathbf{W})^\top$ for any $d \times d$ orthogonal matrix $\mathbf{W}$. In what follows, it will therefore be necessary to either (i) consider estimation and inference modulo orthogonal transformation, (ii) to obtain an orthogonal alignment via additional information, or (iii) to work directly in terms of (estimated) vector-vector products.[1]

The low rank symmetric inner product structure in the above definition leads to that the order $n \times d$ entries of $\mathbf{X}$ constitute statistical quantities of interest. Absent such structure, $\mathbf{A}$ with an arbitrary expectation matrix would have order n^2 possibly different values that would need to be estimated. In what follows, it will be shown to be particularly useful to prescribe an underlying sampling process for the rows of $\mathbf{X}$, which can be viewed as prescribing a prior distribution on latent space.

Definition 2 Signed RDPG [sgn-RDPG]. Let $X_1, \dots, X_n$ be independent and identically distributed vectors with distribution function F where $\operatorname{supp} F \equiv \mathcal{X} \subset \mathbb{R}^d$ with $\|\mathbf{x}\|_2 \le 1$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbb{E}[X_1 X_1^\top] \in \mathbb{R}^{d\times d}$ is full rank. Write $\mathbf{X} := [X_1|X_2|...|X_n]^\top \in \mathbb{R}^{n\times d}$. Let $\mathbf{A} \in \{-1,0,1\}^{n\times n}$ be a symmetric matrix defined in the conditional manner

$$\mathbf{A}|\mathbf{X} := \operatorname{sgn}(\mathbf{X}\mathbf{X}^\top) \circ \mathbf{A}',$$

where $\mathbf{A}' \in \{0,1\}^{n\times n}$ satisfies $a'_{ij}|X_i, X_j \sim \text{Bernoulli}(|\langle X_i, X_j\rangle|)$ independently for $i \le j$ and $a'_{ji} = a'_{ij}$. Matrices $\mathbf{A}$ are called signed adjacency matrices of signed random dot product graphs (sgn-RDPGs).

An adjacency matrix $\mathbf{A}$ arising via Definition 2 can be viewed as a noisy version of its conditional expectation matrix of edge probabilities by observing that $\mathbb{E}[\mathbf{A}|\mathbf{X}] = \mathbf{X}\mathbf{X}^\top$. The important distinction between Definitions 1 and 2 is whether or not the latent matrix $\mathbf{X}$ is given a priori in an arbitrary manner, or whether its rows are homogeneously sampled in an i.i.d. manner. Situations wherein a few rows of $\mathbf{X}$ are not i.i.d. can at times be treated as small, asymptotically negligible perturbations of the above *model*. Here, Definition 2 includes the classical RDPG model as a special case (Athreya et al., 2018).[2] Corresponding mathematical tools and an existing theory can be carried over to the present setting, as discussed in the following section.

[1]We remark that matrix perturbation analysis, in concert with probabilistic concentration bounds, reveals that the recoverability of $\mathbf{X}$ from $\mathbf{A}$ depends crucially on the structure of $\mathbf{X}$. Moreover, the concept of *recovery* itself runs the gamut, covering both weak and strong forms of consistency and various rates of convergence depending on the choice of matrix norm. Under Definition 1, loosely speaking, too much localized relative heterogeneity in the row or column structure of $\mathbf{X}$ precludes uniform entry-wise estimation and residual analysis.

[2]Existing works on RDPGs vary in terms of whether or not self-loops are permitted or forbidden (i.e., whether $\mathbf{A}$ must be hollow or not). Elsewhere and in the present paper, this difference is often immaterial (though strictly speaking not always).

3.1 | Embedding strategies and methodology

The embedding strategies considered in this section are stated without reference to an underlying generative mechanism. When the underlying generative mechanism is as in Definition 2, these strategies can be viewed as estimation procedures, motivating the notational usage of hats. Our discussion of embedding strategies is congruent with the literature on RDPGs, which establishes strong forms of consistent estimation, asymptotic normality, semiparametric and nonparametric hypothesis testing, and efficiency results for spectral-based methods (Athreya et al., 2016; Sussman et al., 2013; Tang, Athreya, Sussman, Lyzinski, Park, & Priebe, 2017; Tang, Athreya, Sussman, Lyzinski, & Priebe, 2017; Tang, Cape, & Priebe, 2017).

The first embedding approach amounts to obtaining a truncated eigen-decomposition of a symmetric (adjacency) input matrix. Retaining the leading eigenvectors is based on the desire to retain cluster-level information and statistical signal, while retaining the leading eigenvalues is based on the desire that the embedding serves as a faithful low-rank approximation to the original input matrix. The following approach, stated here in the form of an algorithm, is widely considered throughout the literature (e.g., see the survey paper Athreya et al., 2018, and the references therein).

Algorithm 1: Adjacency spectral embedding (ASE)

Input: symmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and positive integer $d \in [[n]]$.

1. Truncated spectral decomposition: compute the d largest-in-magnitude eigenvalues of $\mathbf{A}$, arranged in the matrix $\widehat{\boldsymbol{\Lambda}} = \mathrm{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_d)$, with $|\widehat{\lambda}_1| \geq \ldots \geq |\widehat{\lambda}_d|$, and compute d corresponding leading eigenvectors of $\mathbf{A}$, denoted $\{\widehat{\mathbf{u}}_i\}_{i=1}^d$, arranged as ordered orthonormal columns in the matrix $\widehat{\mathbf{U}} \in \mathbb{R}^{n \times d}$.

Output: (unsupervised) adjacency spectral embedding $\widehat{\mathbf{X}} := \widehat{\mathbf{U}} |\widehat{\boldsymbol{\Lambda}}|^{1/2} \in \mathbb{R}^{n \times d}$.

Optional step: normalize the rows of $\widehat{\mathbf{X}}$ *to each have unit norm.*

The notation $\widehat{\mathbf{X}}$ is intended to suggest that $\widehat{\mathbf{X}} \approx \mathbf{X}$ for $\mathbf{X}$ as in Definitions 1 and 2. Some care must be exercised, for the notion of approximately, $\approx$, necessarily takes into account aligning $\widehat{\mathbf{X}}$ and $\mathbf{X}$ via orthogonal transformation, together with the fact that $\widehat{\mathbf{U}}$ is computed non-uniquely up to a choice of column-wise signs.[3]

One approach for aligning embeddings is to incorporate information about *seed nodes*, considered, for example, in the problem of vertex nomination, which seeks to leverage known information about nodes of interest in one network in order to discover nodes of interest in another related network (Agterberg et al., 2020). Additional outside knowledge about a set of seed nodes and their corresponding known latent representations allows one to obtain *supervised* orthogonal transformations that orient embeddings.

Algorithm 2: Supervised ASE (sup-ASE)

Input: symmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, positive integer $d \in [[n]]$, seed node latent vectors $\mathbf{X}_{(1)} \in \mathbb{R}^{s \times d}$ with $s \geq d$, $\mathrm{rank}(\mathbf{X}_{(1)}) = d$, and with known correspondence in $[[n]]$.

1. Adjacency spectral embedding: obtain $\widehat{\mathbf{X}} \in \mathbb{R}^{n \times d}$ via Algorithm 1.
2. Supervised alignment: compute $\mathbf{W}_{\mathrm{S}} \in \underset{\mathbf{W} \in \mathbb{O}_d}{\arg\inf} \|\widehat{\mathbf{X}}_{(1)}\mathbf{W} - \mathbf{X}_{(1)}\|_{\mathrm{F}}^2$, where $\widehat{\mathbf{X}}_{(1)}$ is a submatrix of $\widehat{\mathbf{X}}$ using the known correspondence to $\mathbf{X}_{(1)}$ in $[[n]]$. In particular, write the singular value decomposition of $\widehat{\mathbf{X}}_{(1)}^\top \mathbf{X}_{(1)}$ as $\mathbf{W}'\boldsymbol{\Sigma}\mathbf{W}''$ and set $\mathbf{W}_{\mathrm{S}} := \mathbf{W}'\mathbf{W}''$.[4]

Output: supervised adjacency spectral embedding $\widehat{\mathbf{X}}_{\mathrm{S}} := \widehat{\mathbf{X}}\mathbf{W}_{\mathrm{S}} \in \mathbb{R}^{n \times d}$.

Optional step: normalize the rows of $\widehat{\mathbf{X}}_{\mathrm{S}}$ *to each have unit norm.*

The methodology in Algorithm 2 remains applicable for vertex-aligned multiple graph settings without seed nodes. Specifically, one may first compute $\widehat{\mathbf{X}}^{(1)}, \widehat{\mathbf{X}}^{(2)}$ from $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}$, after which compute $\mathbf{W}_{\mathrm{S}}$ using both embeddings to obtain $\widehat{\mathbf{X}}_{\mathrm{S}}^{(2)} := \widehat{\mathbf{X}}^{(1)}\mathbf{W}_{\mathrm{S}}$.

3.2 | Theory

The formulation of sgn-RDPGs in Definition 2 is stated for general distributions F on a suitable support set $\mathcal{X} \subset \mathbb{R}^d$, in accordance with the standard description of RDPGs. This paper treats the uniform distribution on the unit ball $\mathcal{B}_d = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\|_2 \leq 1\}$ and by association the uniform distribution on its surface, the unit sphere $\mathcal{S}_d = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\|_2 = 1\}$, as two natural choices of *null distribution*, noting that $\mathcal{B}_d \equiv \mathcal{X} \subset \mathbb{R}^d$ corresponds to the maximal d-dimensional ambient space such that $\mathrm{supp}\, F \subseteq \mathcal{B}_d$ ensures valid probabilities $|\langle \mathbf{x}_i, \mathbf{x}_j \rangle| \in [0, 1]$. Absent these choices, both $\mathcal{B}_d$ and $\mathcal{S}_d$ remain canonical latent space domains containing the support of distributions F.

[3]For the sake of concreteness, one may without loss of generality take the first entry in each column of $\widehat{\mathbf{U}}$ to be nonnegative. When $\mathbf{A}$ exhibits large negative eigenvalues, it is sometimes useful in practice (and for organizing theory) to index the leading eigenvalues alternatively in the manner $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ and $-\lambda_{p+1} \geq -\lambda_{p+2} \geq \ldots \geq -\lambda_d > 0$.

[4]In low-dimensional examples one could alternatively minimize with respect to the maximum absolute entry norm or the maximum Euclidean row norm, as highlighted in Cape (2020). Such an approach, necessarily requiring numerical optimization in the absence of a general closed-form analytic solution, would be particularly suitable when certain parametrized, structured orthogonal transformations are of application-specific interest.

3.2.1 | Uniform row-wise consistency

Fix F to be a (nondegenerate) distribution on (a subset of) $\mathcal{B}_d$, including but not limited to Uniform$(\mathcal{S}_d)$ and Uniform$(\mathcal{B}_d)$. Then, there exists a sequence of orthogonal matrices $\mathbf{W}^{(n)}$ such that with high probability for large n ,

$$\max_{1 \le i \le n} \|\mathbf{W}^{(n)}\widehat{X}_i - X_i\|_2 \le C\sqrt{\frac{d\log^c(n)}{n}} \tag{1}$$

for some constants $C, c > 0$ depending possibly on F and where the dependence of $\widehat{X}_i$ on n is suppressed. Existing results of this form share a common proof architecture, modulo the treatment of logarithmic factors, network sparsity considerations, and the specification of "high probability" (Athreya et al., 2016; Pantazis et al., 2020; Rubin-Delanchy et al., 2017).

The (theoretical) orthogonal matrices $\mathbf{W}^{(n)}$, which depend on both $\mathbf{X}$ and $\widehat{\mathbf{X}}$, can be computed in simulation examples but are unavailable in practice since $\mathbf{X}$ is unknown. Figure 1 illustrates that, when available, supervised spectral embeddings are capable of yielding similar estimation accuracy. Elsewhere, it is desirable to obtain uniform perturbation bounds to establish consistency absent $\mathbf{W}^{(n)}$, as discussed next.

3.2.2 | Low-rank truncation and uniform entry-wise consistency

Equipped with Equation (1), a boundedness argument together with an application of the triangle inequality yields that $\mathbf{A}_{[d]} := \widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top$, the best rank d approximation of $\mathbf{A}$ in Frobenius norm, closely approximates the conditional expectation $\mathbf{X}\mathbf{X}^\top$. In particular, as before, with high probability as $n \to \infty$,

$$\|\mathbf{A}_{[d]} - \mathbf{X}\mathbf{X}^\top\|_{\max} \le C\sqrt{\frac{d\log^c(n)}{n}}. \tag{2}$$

Without the use of the generative latent vector distribution F, matrix perturbation theory can be employed to show that the bound in Equation (2) still holds upon replacing each X_i with $\mathbf{x}_i$ for suitable deterministic collections of (sequences of) vectors $\{\mathbf{x}_i\}_{i=1}^n$ under additional coherence-type assumptions.

When the latent vectors lie on the unit sphere, or on a sphere of known radius $r \le 1$, a natural normalization step is to correspondingly project the estimates $\widehat{X}_i$ in hopes of improving estimation accuracy. Towards this end, write $\tilde{X}_i := \widehat{X}_i\|\widehat{X}_i\|_2^{-1}$ and similarly define $\tilde{\mathbf{X}} := \left[\tilde{X}_1|\ldots|\tilde{X}_n\right]^\top$ so that $\tilde{\mathbf{A}}_{[d]} := \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top$. Under minor modifications, it follows that for d fixed and n large, with high probability $\left\|\tilde{\mathbf{A}}_{[d]} - \mathbf{X}\mathbf{X}^\top\right\|_{\max}$ satisfies the same

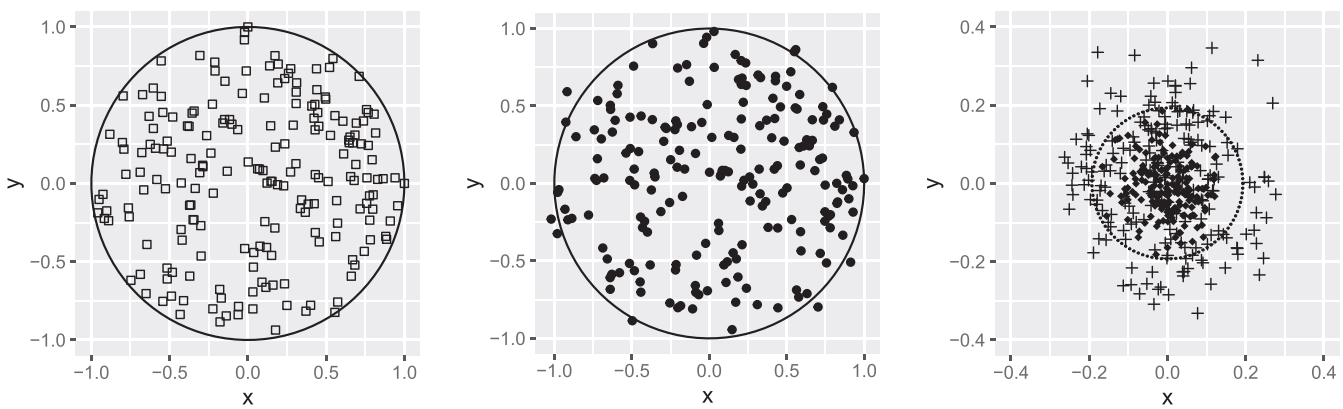

FIGURE 1 A spectral embedding example with $n = 200$ vertices having latent vectors in dimension $d = 2$. Left plot: hollow squares denote latent vectors generated uniformly inside the unit circle. Middle plot: solid circles denote estimated latent vectors obtained via a supervised spectral embedding of the observed network. Right plot: crosses depict raw vector–vector differences ignoring supervised orthogonal alignment, solid diamonds depict vector–vector differences between supervised spectral embedding and corresponding latent vectors (i.e., left plot minus middle plot); dotted circle has a radius equal to the largest Euclidean residual row norm after aligning by the theoretical orthogonal transformation in Equation (1). Note the zoomed-in scale (i.e., see the axes) for the residuals shown in the rightmost plot

high-probability bound as in Equation (2), though this bound can presumably be improved via more delicate perturbation analysis that leverages uniformity on the sphere, together with a reduction of variance argument.

3.2.3 | Asymptotic normality

Aside from (strong) consistency, row-wise asymptotic normality results also hold for the transformed residuals of the estimated latent vectors. A direct adaptation of the proof architecture in Pantazis et al. (2020) to the setting of Definition 2 yields that there exists a sequence of $d \times d$ orthogonal matrices $\mathbf{W}^{(n)}$ such that for any fixed (row) index i, for any fixed choice $\mathbf{z} \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \mathbb{P}\left[n^{1/2}(\mathbf{W}^{(n)}\widehat{X}_i - X_i) \le \mathbf{z}\right] = \int_{\operatorname{supp} F} \Phi(\mathbf{z}, \boldsymbol{\Sigma}(\mathbf{x}))\mathrm{d}F(\mathbf{x}), \tag{3}$$

where $\Phi(\cdot, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian cumulative distribution function with mean zero and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}(\mathbf{x}) := \boldsymbol{\Delta}^{-1}\mathbb{E}\left[(|\langle \mathbf{x}, X_1\rangle| - |\langle \mathbf{x}, X_1\rangle|^2)X_1X_1^\top\right]\boldsymbol{\Delta}^{-1} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{\Delta} := \mathbb{E}[X_1X_1^\top]$. By convention, X_1 is written here since $X_i \sim F$ for each i. A side-by-side comparison with the established literature reveals the role of supp$F$ and signed inner products for lifting RDPGs to sgn-RDPGs.

In what follows, the theoretical observations collected above motivate and are illustrated in the successful estimation of latent matrix dynamics, both at the individual node level and *uniformly* over all nodes in certain large networks.

3.3 | Multiple graphs

At time $t \in [[T]] \equiv \{1, 2, \ldots, T\}$, let $\mathcal{G}^{(t)} \equiv (\mathcal{V}^{(t)}, \mathcal{E}^{(t)})$ denote an observed graph on n_t vertices with corresponding observed signed adjacency matrix $\mathbf{A}^{(t)}$ depending on the unobserved latent matrix $\mathbf{X}^{(t)} \in \mathbb{R}^{n_t \times d_t}$. Let $\mathcal{V}_{\mathrm{S}}^{(t)}$ be a collection of seed nodes of cardinality s_t with known vertex correspondence across all graphs. Our goal is to obtain $\{\widehat{\mathbf{X}}_{\mathrm{S}}^{(t)}\}_{t\in[[T]]}$ from $\{\mathbf{A}^{(t)}\}_{t\in[[T]]}$ with the help of (latent vectors for) the seed nodes $\{\mathcal{V}_{\mathrm{S}}^{(t)}\}_{t\in[[T]]}$, in order to recover $\{\mathbf{X}^{(t)}\}_{t\in[[T]]}$ and hence the to-be-specified mechanism(s) governing $\mathbf{X}^{(t)} \mapsto \mathbf{X}^{(t+1)}$.

The presence of a few seed nodes relative to a large node set can be viewed as a small perturbation of the model setup in Section 3, leading to computable orthogonal transformations $\mathbf{W}_{\mathrm{S}}$ that approximate the unknown orthogonal transformations in Section 3.2. For simplicity, we focus on the setting where one has $s_t \equiv d$ *pure seed nodes*, that is, nodes whose latent vectors are standard basis vectors. This choice, while arguably canonical, is strictly speaking not required and can be relaxed. We further consider $\mathbb{R}^{d_t}$ with $d_t \equiv d$ at all time points t, since tackling the general problem of unequal latent dimensions $\{d_t\}_{t\in[[T]]}$ is substantially more complicated and not well understood for spectral methods.

4 | AUTOREGRESSIVE PROCESSES IN LATENT SPACE

Vector autoregressive models are a staple of time series analysis and arise naturally in the statistical treatment of dynamics. Here, we show that autoregressive modelling and inference can be carried out within inner product latent space. A key difference in the present network-centric context, unlike in classical statistics, is that autoregressive behaviour arises implicitly in the observed random vectors $\widehat{X}_i$ when prescribing dynamics for the corresponding latent vectors $\mathbf{x}_i$.

The traditional vector autoregressive model of order p, denoted VAR(p), can be expressed at time t in the manner

$$Y^{(t)} = \boldsymbol{\mu}_Y + \mathbf{R}^{(1)}Y^{(t-1)} + \ldots + \mathbf{R}^{(p)}Y^{(t-p)} + \boldsymbol{\epsilon}^{(t)}, \tag{4}$$

where $Y^{(t)}$ is an observed d-dimensional random vector, the (unobserved) matrices $\mathbf{R}^{(j)} \in \mathbb{R}^{d\times d}$ are fixed, $\boldsymbol{\mu}_Y \in \mathbb{R}^d$ is a fixed (unobserved) intercept vector, and $\epsilon^{(t)}$ is a d-dimensional random (noise) vector satisfying $\mathbb{E}[\epsilon^{(t)}] = \mathbf{0}_d$, $\mathbb{E}[\epsilon^{(t)}(\epsilon^{(t)})^\top] = \boldsymbol{\Sigma}_\epsilon \in \mathbb{R}^{d\times d}$, and $\mathbb{E}[\epsilon^{(t)}(\epsilon^{(t')})^\top] = \mathbf{0}_{d\times d}$ for $t \neq t'$

[5]We do not prescribe the intercept vector $\boldsymbol{\mu}_Y$ nor do we explicitly prescribe the noise vector $\epsilon^{(t)}$. In practice and in simulations, some care must be taken to ensure identifiability and estimability of model parameters.

(Lütkepohl, 2005). Given vector-valued observations $\{Y^{(t)}\}_{t \in [[T]]}$, the objective is to make inferences about the matrices $\{\mathbf{R}^{(j)}\}_{j \in [[p]]}$ and vector $\boldsymbol{\mu}_Y$. The multiple graphs paradigm with latent dynamics inspires the following generative autoregressive setup and subsequent examples.[5]

Generative setup:

- Fix positive integers $1 \le d, p \ll n$ and a subset $\mathcal{I} \subset [[n]]$ of nonseed nodes of interest.
  - For $i \in \mathcal{I}$, choose $\{\mathbf{R}_i^{(j)}\}_{j \in [[p]]} \subset \mathbb{R}^{d \times d}$ such that $\sum_{j \le p} \|\mathbf{R}_i^{(j)}\|_{\mathrm{op}} \le 1$.
  - Obtain "pre-vectors" $X_i^{(-p+1)}, \ldots, X_i^{(0)} \in \mathcal{B}_d$, say i.i.d. Uniform$(\mathcal{B}_d)$.
  - For $t \ge 1$, assign $\mathbf{x}_i^{(t)} := \mathbf{R}_i^{(1)} \mathbf{x}_i^{(t-1)} + \ldots + \mathbf{R}_i^{(p)} \mathbf{x}_i^{(t-p)}$.
- For all remaining nonseed nodes, $i' \in [[n]]$, $i' \notin \mathcal{I}$, and $t \ge 1$, let $X_{i'}^{(t)} \overset{\text{i.i.d.}}{\sim} F$ on $\mathcal{B}_d$, say $F \equiv$ Uniform$(\mathcal{B}_d)$.
- For each $t \ge 1$, generate $\mathbf{A}^{(t)}$ according to Definition 1 given $\mathbf{X}^{(t)}$.

4.1 | Estimation of single node dynamics

For VAR(1) with $d = 2$, consider autoregressive processes at the node-specific level that exhibit periodicity via orthogonal rotation matrices. For $\theta \in [0, 2\pi)$ radians, the 2×2 rotation matrix about the origin through an angle of θ in the counterclockwise direction is given by

$$\mathbf{R}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The specification of such latent vector dynamics may be associated with, say, seasonal preferences or cyclic connectivity changes in latent "social space." See Sewell and Chen (2015) for another example involving rotations in the context of network dynamics and latent space models.

For concreteness, set $\theta_i := 2\pi i / \mathrm{card}(\mathcal{I})$ with $i = 1, 2, \ldots, \mathrm{card}(\mathcal{I}) = 10$. By the matrix perturbation analysis underlying the row-wise bound in Equation (1) and the asymptotic normality in Equation (3), which mirrors discussion in Rubin-Delanchy et al. (2017), the estimated latent vectors behave in the manner

$$\widehat{X}_{\mathrm{S},i}^{(t)} \approx \boldsymbol{\mu}_i + \mathbf{R}_{\theta_i}^{(1)} \widehat{X}_{\mathrm{S},i}^{(t-1)} + \widehat{\epsilon}_i^{(t)}, \tag{5}$$

where $\widehat{\epsilon}_i^{(t)}$ reflects a shrinking, approximately Gaussian matrix perturbation residual noise term. Here, $\boldsymbol{\mu}_i \approx \mathbf{0}_d$, and so we do not compute relative errors for the intercept estimates presented below.

For the choice θ_1 with $n = 200$ and $T = 50$, we first obtain a collection of aligned, time-indexed, estimated latent matrices using the supervised ASE procedure in Algorithm 2. Thereafter, applying the method of ordinary least squares (OLS) for VAR(p) estimation (Lütkepohl, 2005) yields the node-specific estimates

$$\widehat{\boldsymbol{\mu}}_{1,\mathrm{OLS}} \doteq \begin{bmatrix} -0.00083 \\ -0.00135 \end{bmatrix}, \widehat{\mathbf{R}}_{\theta_1,\mathrm{OLS}}^{(1)} \doteq \begin{bmatrix} 0.766 & -0.578 \\ 0.581 & -0.820 \end{bmatrix}, \text{ where } \mathbf{R}_{\theta_1}^{(1)} \doteq \begin{bmatrix} 0.809 & -0.587 \\ 0.587 & -0.809 \end{bmatrix}, \tag{6}$$

and the latter estimate has a relative Frobenius norm error of 0.0322.

The above estimation procedure, which already yields a small Frobenius norm error, is agnostic to the fact that the coefficient matrix is an orthogonal matrix. We can further project the OLS estimate to the nearest orthogonal matrix in the manner

$$\widehat{\mathbf{R}}_{\theta_i,\mathrm{ORTH}}^{(1)} \in \underset{\mathbf{W} \in \mathbb{O}_d}{\arg\inf} \|\widehat{\mathbf{R}}_{\theta_i,\mathrm{OLS}}^{(1)} - \mathbf{W}\|_{\mathrm{F}}^2, \tag{7}$$

which produces the improved estimate

$$\widehat{\mathbf{R}}_{\theta_i,\mathrm{ORTH}}^{(1)} \doteq \begin{bmatrix} 0.807 & -0.590 \\ 0.590 & -0.807 \end{bmatrix}, \text{ rel. Frob. err.} \doteq 0.00286. \tag{8}$$

Simulations with larger choices of n (i.e., node "sample size") produce improved estimates and relative errors as expected, as do larger values of T, up to a point.

4.2 | Estimation under choice of VAR lag order

Any VAR(p) model can be expressed by recursion as a VAR(p') model where $p' \geq p$. It is therefore still possible to estimate $\mathbf{R}_{\theta_i}^{(1)}$ under the (mis) specification VAR(2), albeit with inflated error. Consider that Equation (5) can be written for any value $\gamma \in [0, 1]$ as

$$\widehat{X}_{\mathrm{S},i}^{(t)} \approx \boldsymbol{\mu}_i + \gamma \mathbf{R}_{\theta_i}^{(1)} \widehat{X}_{\mathrm{S},i}^{(t-1)} + (1-\gamma) \mathbf{R}_{2\theta_i}^{(2)} \widehat{X}_{\mathrm{S},i}^{(t-2)} + \widehat{\epsilon}_i^{(t)}, \tag{9}$$

since $\mathbf{R}_{2\theta_i}^{(2)} \equiv \left(\mathbf{R}_{\theta_i}^{(1)}\right)^2$. The presence of γ introduces an apparent nonidentifiability at the estimation step. Nevertheless, performing the orthogonal projection update after solving the VAR(2) ordinary least squares problem yields the reasonable estimates and corresponding errors

$$\widehat{\boldsymbol{\mu}}_{1,\mathrm{OLS}}^{\mathrm{VAR}(2)} \doteq \begin{bmatrix} -0.00491 \\ -0.00058 \end{bmatrix}, \widehat{\mathbf{R}}_{\theta_1,\mathrm{ORTH}}^{(1),\mathrm{VAR}(2)} \doteq \begin{bmatrix} 0.726 & -0.687 \\ 0.687 & -0.726 \end{bmatrix}, \widehat{\mathbf{R}}_{2\theta_1,\mathrm{ORTH}}^{(2),\mathrm{VAR}(2)} \doteq \begin{bmatrix} 0.389 & -0.920 \\ 0.920 & -0.389 \end{bmatrix}, \text{rel. Frob. errs.} \doteq (0.129, 0.086). \tag{10}$$

4.3 | Uniform recovery of full network dynamics

Now consider graphs with $n = 200$ nodes across $T = 200$ time steps, where all nonseed nodes simultaneously exhibit their own periodic latent space dynamics governed by $\{\theta_i\}$ as above. Does it remain possible to recover individual nodes' latent space dynamics, and if so, uniformly for the entire network? Figure 2 provides an illustration in the affirmative. In the left panel, each point is identified with an individual node, with the right panel displaying the corresponding norm of its initial latent (pre)vector, generated uniformly in $\mathcal{B}_2$. Nodes' latent dynamics given by $\mathbf{R}_{\theta_i}^{(1)}$ are estimated with small Frobenius norm relative errors (see x-axis) and with even smaller errors after orthogonally projecting the estimates (see y-axis). Note the order of magnitude improvement and concentration of points near the origin (i.e., bottom-left region in the plots).

5 | NONLINEAR DYNAMICS IN LATENT SPACE

This section demonstrates that nonlinear dynamics and their long-run behaviour can also be modelled and estimated using the spectral methods considered in this paper. Here, we turn to degree-corrected block models and their characteristic *radial* latent space geometric structure as a point of departure. Towards this end, consider the logistic difference equation, originally from population biology (May, 1976; Verhulst, 1845), in the canonical form

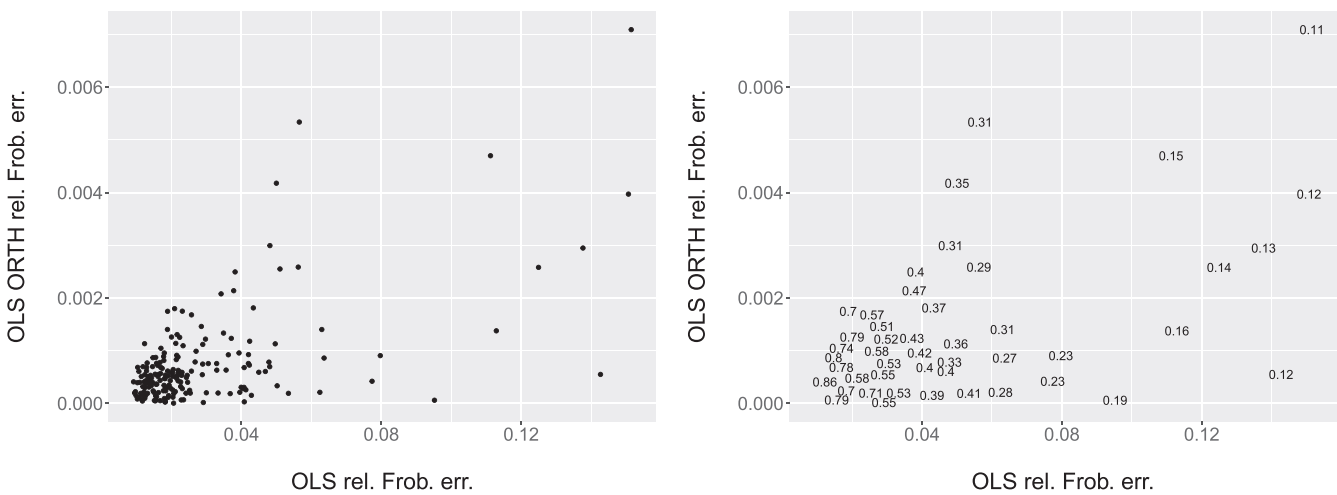

FIGURE 2 Residuals illustrating an example of uniformly accurate estimation of autoregressive latent space dynamics for all nodes in a network time series. Some points are not shown in the right panel due to crowding out of text size. Details are in Section 4

$$R^{(t+1)} = \mu \times R^{(t)} \times (1 - R^{(t)}), 0 \le R^{(t)} \le 1, 0 \le \mu \le 4. \tag{11}$$

Here, $R^{(t)}$ represents a proportion of a population at time t, and μ is a parameter that governs the nonlinear dynamical behaviour arising from the above recurrence relation as $t \to \infty$ (Braun & Golubitsky, 1983, Chapter 1). Different values of μ lead to different observed behaviours for almost all initial values of the population.[6]

The one-dimensional dynamics in Equation (11) can be lifted to higher dimensions by associating $R^{(t)}$ to line segments extending from the origin to the unit sphere. We arrive at the following generative procedure for multiple graphs.

Generative setup:

- Fix positive integers $1 \le d \ll n$ and a subset $\mathcal{I} \subset [[n]]$ of nonseed nodes of interest.
  - For $i \in \mathcal{I}$, let $X_i^{(0)} \sim F$ on $\mathcal{B}_d$, say $F \equiv \text{Uniform}(\mathcal{B}_d)$.
  - Given $X_i^{(0)} = \mathbf{x}_i^{(0)}$ set $R_i^{(0)} := \|\mathbf{x}_i^{(0)}\|_2^{-1}$, hence $X_i^{(0)} = R_i^{(0)} \tilde{\mathbf{x}}_i^{(0)}$.
  - Choose $\mu_i \in [0, 4]$ and denote the corresponding dynamic behaviour by $c(i)$ from among C1, ..., C6 (see footnote).
  - For $t \ge 1$, let $R_i^{(t)}$ evolve according to Equation (11) and assign $\mathbf{x}_i^{(t)} := R_i^{(t)} \tilde{\mathbf{x}}_i^{(0)}$.
- For all remaining nonseed nodes i', and $t \ge 1$, let $X_{i'}^{(t)} \overset{\text{i.i.d.}}{\sim} F$ on $\mathcal{B}_d$, say $F \equiv \text{Uniform}(\mathcal{B}_d)$.
- For each time $t \ge 1$, generate $\mathbf{A}^{(t)}$ given $\mathbf{X}^{(t)}$ according to Definition 1.

In words, for a given value of μ_i and initial vector $\mathbf{x}_i^{(0)}$, the above procedure prescribes a sequence of latent vectors that are scaled versions of the unit norm vector $\tilde{\mathbf{x}}_i^{(0)}$. Such latent geometry and its variants can be interpreted in the context of social networks, where individuals' directions in social space may be fixed across time, while their relative level of involvement in different social dimensions may vary over time. For example, hypothetical Alice might generally enjoy, say, music (x-axis) and politics (y-axis), whereas hypothetical Bob might generally enjoy music but dislike politics, in which case their latent vectors in social space (the xy–plane) directionally lie in the first and fourth quadrants, with changing magnitudes based on feeling or level of involvement over time.

5.1 | Recovering idiosyncratic node dynamics

Consider $T = 100$ networks each with $n = 1000$ nodes having associated latent vectors of dimension $d = 2$, evolving according to the nonlinear dynamics described above. Take two nodes as being seed nodes with standard basis latent vectors, and take 977 nodes as having latent vectors that evolve i.i.d. in the null manner $\text{Uniform}(\mathcal{B}_2)$. For the remaining 21 nodes, associate to each a distinct value $\mu \in$ {0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 3.60, 3.70, 3.80, 3.90, 3.925, 3.95, 3.975, 3.99} so that each node's latent space trajectory evolves according to the condition corresponding to its value of μ. For each of these 21 nodes of interest, the goal is to recover the corresponding latent vector long-term behaviour. Here, we focus on the values $\mu = 3.5$ (C5), 3.25 (C4), and 2.25 (C3) corresponding to Figures 3–5, respectively. Each figure consists of three panels depicting the dynamics of a single node with the corresponding μ value, across all T networks. In each figure, the leftmost panel shows the estimated latent vectors as a scatter plot, with each point representing a single node embedding computed from an observed network, with the unit circle shown for reference. The middle panel shows the zoomed-in estimated latent vectors displayed according to the time index of each network embedding in the time series. The rightmost panel shows Gaussian mixture modelling (via `mclust` in `R`) applied to the estimated latent vectors; solid lines depict latent vector radial line segments (true domain), while the dashed line represents a regression-based estimate thereof; crosshairs depict empirical Gaussian cluster centroids, while solid black squares correspond to the true theoretical asymptotic latent space points of convergence or oscillation.

Figures 3 and 4 illustrate that Gaussian clustering applied to the estimated latent vectors accurately recovers the true latent locations of eventual permanent oscillations. Furthermore, fitting a least squares regression line to the estimated latent vectors accurately recovers the radial domain of the underlying latent vector process. Accurate estimation occurs in spite of the majority of nodes evolving randomly (inside the unit ball) and in spite of other nodes behaving according to their own, differing, nonlinear latent space dynamics.

[6]See Chapter 2 in Sternberg (2010) for additional details concerning the following behaviours.

C1. For μ between 0 and 1, the population dies out, that is, $R^{(t)} \to 0$ as $t \to \infty$.

C2. For μ between 1 and 2, the population proportion approaches a fixed point, that is, $R^{(t)} \to \frac{\mu-1}{\mu}$ as $t \to \infty$.

C3. For μ between 2 and 3, the population proportion approaches the fixed point $\frac{\mu-1}{\mu}$ in an oscillatory, spiralling fashion, that is, $R^{(t)} \to \frac{\mu-1}{\mu}$ as $t \to \infty$ but more slowly than under the previous condition.

C4. For μ between 3 and $1 + \sqrt{6} \approx 3.4494$, the population proportion approaches permanent oscillations between the two values $1/2 + 1/(2\mu) \pm (1/(2\mu))\sqrt{(\mu+1)(\mu-3)}$.

C5. For μ between $1 + \sqrt{6} \approx 3.4494$ and approximately 3.5440, the population proportion approaches permanent oscillations between four values that are functions of μ.

C6. For μ near 4, the population exhibits increasingly chaotic behaviour for large t, though with occasional islands of stability.

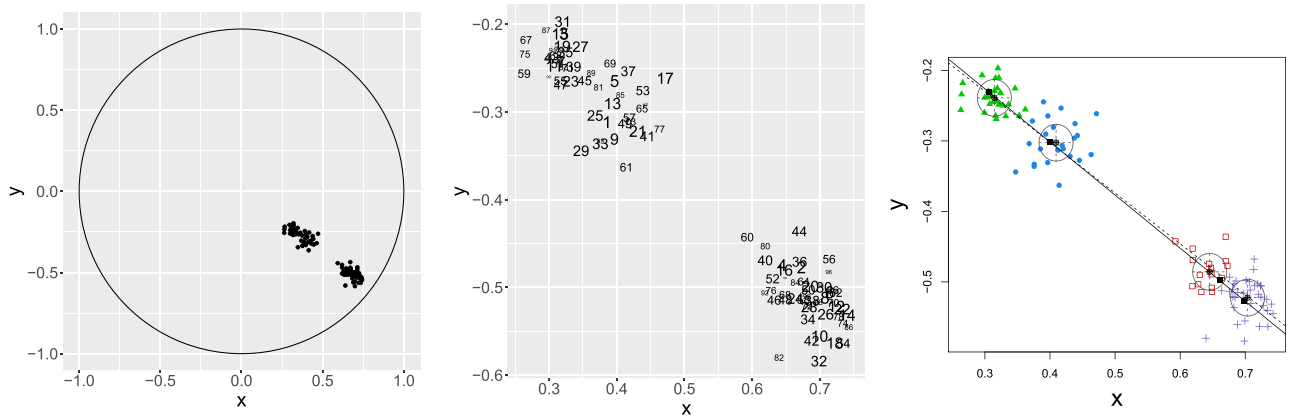

FIGURE 3 Estimated latent dynamics for a node with latent vectors evolving according to $\mu = 3.5$ (C5). Details in Section 5.1

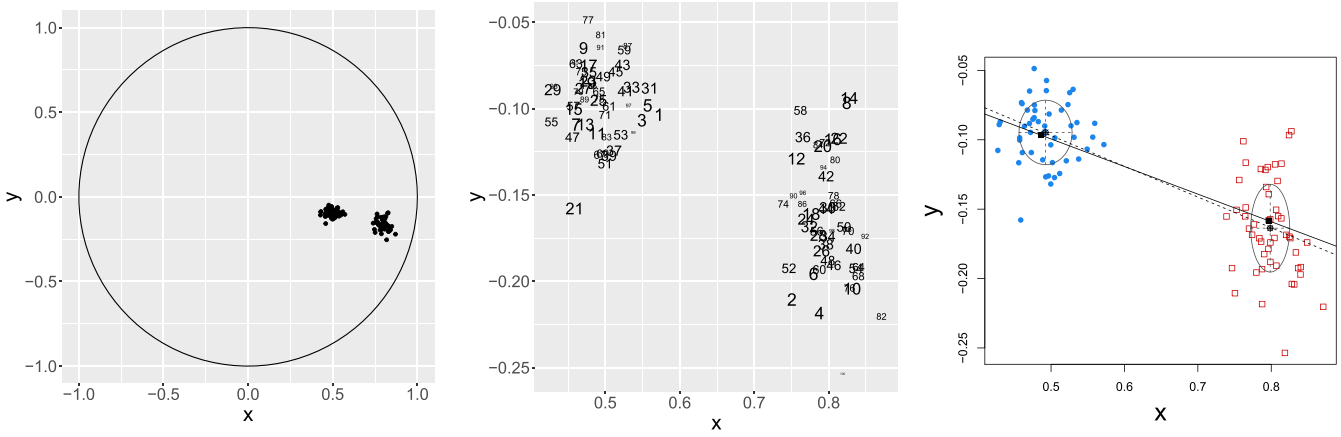

FIGURE 4 Estimated latent dynamics for a node with latent vectors evolving according to $\mu = 3.25$ (C4). Details in Section 5.1

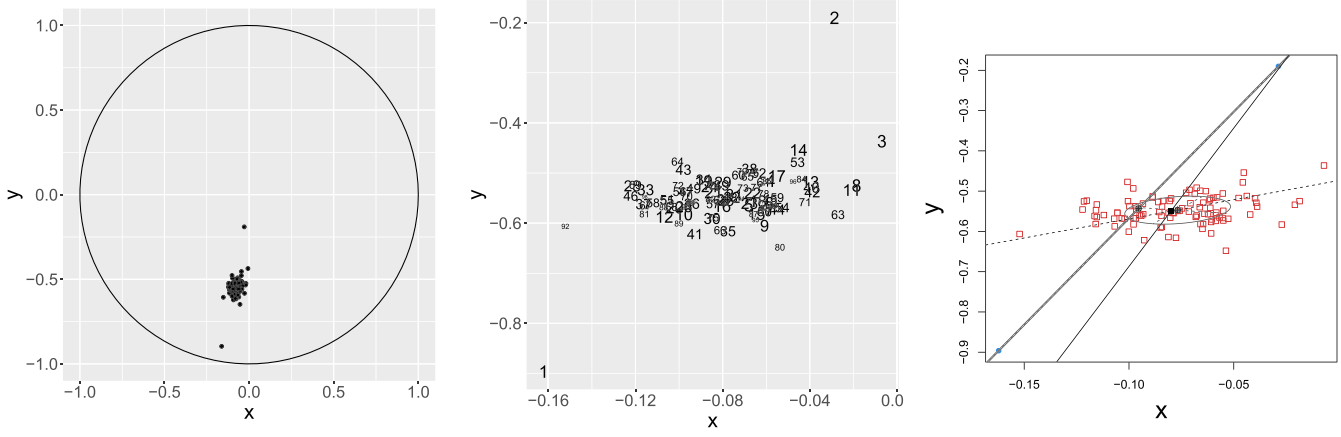

FIGURE 5 Estimated latent dynamics for a node with latent vectors evolving according to $\mu = 2.25$ (C3). Details in Section 5.1

Figure 5 similarly shows that the cluster centroid of the red Gaussian component closely estimates the asymptotic fixed point around which the latent vector process oscillates. A spurious blue cluster, consisting of only two latent vector estimates, arises from keeping the embedding values for the first two networks in the time series. The variability in the eventual oscillations prevents the regression-based

estimate (dashed line) from reasonably estimating the latent space domain (line segment, solid line), while the major axis of the blue component happens to be better suited as an estimate thereof. The latter observations are unsurprising, given that the first two embedded points visibly behave as outliers.

6 | BITCOIN CRYPTOCURRENCY WHO-TRUSTS-WHOM USER NETWORKS

This section provides a brief real data analysis to complement the previous sections and to illustrate bridging theory with practice. We consider publicly available data from the Stanford Network Analysis Project (https://snap.stanford.edu/), specifically two who-trusts-whom networks of users engaged in trading the cryptocurrency Bitcoin. The two networks correspond to the two datasets `soc-sign-bitcoin-otc` and `soc-sign-bitcoin-alpha`, consisting of user–user ratings from the platforms Bitcoin OTC and Bitcoin Alpha, respectively (Kumar et al., 2016, 2018). Both datasets, stored in the form of edge lists, are weighted, signed, directed, temporal networks over the years 2010–2016. Each data point (edge) records the node (user) ID of the rater, the node (user) of the ratee, an integer-valued rating between -10 and 10, and the date/time of the rating. Positive ratings convey trust in user activity and reputation, whereas negative ratings convey perceived distrust. User IDs (node labels) within each network are anonymized; hence, no additional meaning can be attached to the given node ID values. No additional node-level covariates are available, nor are there a priori known node correspondences between the two networks.

Prior to analysis, we apply data preprocessing steps as described below separately to each dataset. First, we exclude edges formed in the years 2010 and 2016, of which there are only several dozen in each dataset; hence, we focus on the 5-year time period 2011–2015. Six graphs are constructed from each dataset, one graph for each individual year as well as an overall graph aggregating edges across all chosen years 2011–2015. Each graph is obtained by symmetrizing the appropriate raw data adjacency matrix derived from the edge list, then binarizing the positive and negative edge weights to take values ±1, and finally keeping the largest connected component. The resulting graphs, which conform to the data setting considered in this paper, are summarized in Table 1.

For each graph derived from each dataset, we compute an adjacency spectral embedding in dimension $\widehat{d} = 2$, chosen by inspecting scree plots and to facilitate low-dimensional visualization. Importantly, the node sets differ across graphs within each dataset, and no further information save the rating time stamps is available to align each of the graph embeddings. Together, these conditions pose a challenge for obtaining a supervised alignment of the separate embeddings. Our approach is to apply *varimax* rotations to each of the embeddings, followed by specifying that the first coordinates in each embedding dimension be positive. Our choice is motivated by the established empirical success of varimax rotations in classical factor analysis (Anderson, 1962; Kaiser, 1958) together with modern advances in the theoretical understanding of varimax from a mathematical statistics perspective (Rohe & Zeng, 2020). The resulting embeddings, displayed in Figure 6, exhibit a form of radial streaking along the axes, amounting to a proxy embedding alignment.

Upon careful inspection, panels (a), (d), (e), and (f) in Figure 6 exhibit similarities in embedding space across the two datasets. The agreement is particularly striking when comparing the panel (a) plots: look for example at the plotted points with x-values exceeding 0.5, or the plotted points with y-values exceeding 0.5. Inspection of the panel (d) plots also suggests that certain individual nodes behave similarly in each dataset. Panels (f) on the left and on the right exhibit shared embedding structure for the aggregate graphs, modulo an orthogonal transformation and the presence of additional radially embedded nodes for Bitcoin OTC (right). These observations suggest the strong possibility of user overlap across the data collected by the platforms or else the presence of different users with conspicuously similar connectivity behaviour across the graphs. Taken together, our embedding approach reveals both similarities and differences across platforms, *globally* at the graph level as well as *locally* at the individual node level. Figure 6 and Table 1 provide complementary views of the graph data in terms of latent space geometry as well as network connectivity structure.

TABLE 1 Summary statistics for postprocessed Bitcoin networks

	2011	2012	2013	2014	2015	Agg.
Alpha						
Nodes	1587	1376	1335	686	146	3772
Edges	4266	4343	3726	1850	304	14077
% Neg. Edg.	3.02	7.41	13.10	20.30	12.20	9.31
Density	0.003	0.005	0.004	0.008	0.029	0.002
OTC						
Nodes	1623	1917	2672	1112	346	5872
Edges	4366	5791	8205	2853	660	21431
% Neg. Edges	3.11	12.50	20.00	21.60	10.60	14.70
Density	0.003	0.003	0.002	0.005	0.011	0.001

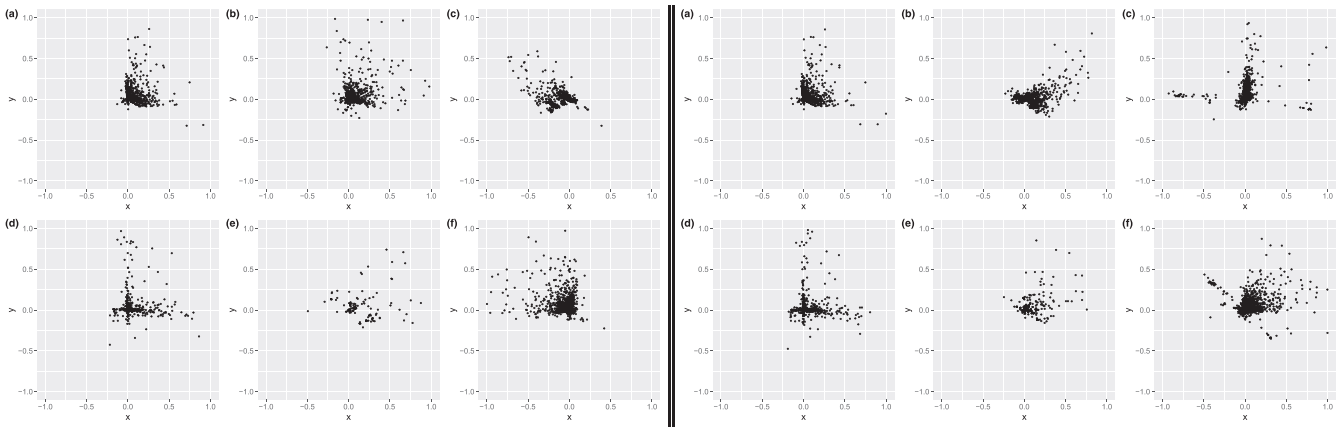

FIGURE 6 Two-dimensional adjacency spectral embeddings for the Bitcoin Alpha data (left) and the Bitcoin OTC data (right). Panels (a) to (e) correspond to years 2011–2015. Panel (f) depicts the embedding of the aggregate graph for 2011–2015. Several outlier points are omitted

Edges in each dataset are time-stamped, so it may be tempting to try to directly align the edges and hence establish node ID correspondences across datasets. Such an approach runs into difficulty, however, because the time unit for each rating in Bitcoin Alpha (resp., Bitcoin OTC) is measured at the granularity of "day" (resp., "second"). Moreover, upon querying edge lists after comparing alike panels, there are apparent temporal discrepancies between the data collected for seemingly similar nodes. A more sophisticated investigation in this direction is beyond the scope of the present paper.

From the discussion above, a deeper analysis could subsequently be undertaken, for example, with the assistance of experienced cryptocurrency traders possessing platform-specific knowledge, to further investigate these data from the early years of Bitcoin. If additional covariate information were available, it would be interesting to track and interpret the similarly behaving nodes in each dataset across different years. It would further be interesting to determine if embedding geometry exhibits quantifiable characteristics reflecting graph structure stemming from trends in cryptocurrency markets and the development of blockchain technologies more generally. Such extensions are left for future work.

7 | DISCUSSION

This paper demonstrates how a contemporary toolkit of spectral-based theory and methodology can be transferred to the setting of networks exhibiting latent space dynamics and signs. Existing modelling approaches are extended in a manner permitting the specification of inferable generative dynamics in latent space. The examples provided herein are designed to illustrate the versatility of phenomena and behaviours that can be recovered via spectral methods for models based on matrix factorizations.

It is perhaps worth emphasizing the distinction between *latent space dynamics*, as considered in this paper, and the terminology *dynamic networks*. The reader, who might be inclined to prescribe edge-level dynamics when thinking about dynamic networks, will notice that here the (unspoken) concept of edge dynamics is inherited as a consequence of specifying latent space dynamics in a generative network model. Of note, networks with latent space dynamics exhibit variation at different network scales. In contrast, dynamic networks need not even necessarily have an associated notion of latent space, though when they do (at least approximately), one may stand to benefit from parsimonious dimensionality reduction.

Opportunities for future, related work abound. One natural extension is to develop more refined finite-sample and asymptotic theory for structured network settings in high dimensional regimes where $d \equiv d(n)$ grows as the number of vertices n diverges. Such work will foreseeably continue to build on contemporary advances in matrix perturbation analysis, in conjunction with modern developments in high-dimensional probability theory and statistics. Another avenue of pursuit is to undertake a deeper investigation of seed node fidelity and the quality of supervised orthogonal matrix transformations. Advances in this direction are already well underway in the domain of graph matching and vertex nomination. Further still, while the present paper considers spectral analysis of network adjacency matrices, there is ample precedence to investigate properties of different matrix representations, notably (normalized, signed) graph Laplacians.

ACKNOWLEDGEMENTS

This work was supported in part by NSF grant DMS–1902755 and by NSF grant SES–1951005. The author thanks Xianshi Yu for detailed feedback and discussions. The author also thanks Suman Chakraborty for stimulating discussions at the inception of this work.

DATA AVAILABILITY STATEMENT

The datasets used in this study are available at https://snap.stanford.edu/. The code that supports the findings of this study is available at https://github.com/jcape1.

ORCID

Joshua Cape https://orcid.org/0000-0002-1471-1650

REFERENCES

Agterberg, J., Park, Y., Larson, J., White, C., & Priebe, C. E. (2020). Vertex nomination, consistent estimation, and adversarial modification. *Electronic Journal of Statistics*, *14*(2), 3230–3267.

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

Anderson, T. W. (1962). *An introduction to multivariate statistical analysis*. Wiley.

Arroyo, J., Athreya, A., Cape, J., Chen, G., Priebe, C. E., & Vogelstein, J. T. (2020). Inference for multiple heterogeneous networks with a common invariant subspace. arXiv1906.10026.

Athreya, A., Fishkind, D. E., Tang, M., Priebe, C. E., Park, Y., Vogelstein, J. T., & Qin, Y. (2018). Statistical inference on random dot product graphs: A survey. *Journal of Machine Learning Research*, *18*(1), 1–92.

Athreya, A., Priebe, C. E., Tang, M., Lyzinski, V., Marchette, D. J., & Sussman, D. L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, *78*(1), 1–18.

Athreya, A., Tang, M., Park, Y., & Priebe, C. E. (2021). On estimation and inference in latent structure random graphs. *Statistical Science*, *36*(1), 68–88.

Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113.

Braun, M., & Golubitsky, M. (1983). *Differential equations and their applications*: Springer.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Bullmore, E. T., & Bassett, D. S. (2011). Brain graphs: Graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, *7*, 113–140.

Cape, J. (2020). A note on the orthogonal Procrustes problem and norm-dependent optimality. *Electronic Journal of Linear Algebra*, *36*(36), 158–168.

Cape, J., Tang, M., & Priebe, C. E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika*, *106*(1), 243–250.

Chen, G., Arroyo, J., Athreya, A., Cape, J., Vogelstein, J. T., Park, Y., & Priebe, C. E. (2020). Multiple network embedding for anomaly detection in time series of graphs. arXiv2008.10055.

Chiang, K. Y., Hsieh, C. J., Natarajan, N., Dhillon, I. S., & Tewari, A. (2014). Prediction and clustering in signed networks: A local to global perspective. *Journal of Machine Learning Research*, *15*(1), 1177–1213.

Crane, H. (2015). Time-varying network models. *Bernoulli*, *21*(3), 1670–1696.

Crane, H. (2018). *Probabilistic foundations of statistical network analysis*. CRC Press.

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

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

Fan, J., Fan, Y., Han, X., & Lv, J. (2019). SIMPLE: Statistical inference on membership profiles in large networks. arXiv1910.01734.

Han, Q., Xu, K., & Airoldi, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning* (pp. 1511–1520).

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

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

Horvath, S. (2011). *Weighted network analysis: Applications in genomics and systems biology*: Springer Science & Business Media.

Jackson, M. O. (2011). An overview of social networks and economic applications. In *Handbook of Social Economics* (Vol. 1, pp. 511–585). Elsevier.

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

Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, *83*(1), 16107.

Kim, B., Lee, K. H., Xue, L., & Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, *12*, 105–135.

Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. Springer Science & Business Media.

Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., & Subrahmanian, V. (2018). Rev2: Fraudulent user prediction in rating platforms. In *proceedings of the Eleventh ACM International Conference onWeb Search and Data Mining* (pp. 333–341).

Kumar, S., Spezzano, F., Subrahmanian, V., & Faloutsos, C. (2016). Edge weight prediction in weighted signed networks. In *2016 IEEE 16th International Conference on Data Mining (ICDM)* (pp. 221–230).

Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., De Luca, E. W., & Albayrak, S. (2010). Spectral analysis of signed graphs for clustering, prediction and visualization. In *proceedings of the 2010 SIAM International Conference on Data Mining* (pp. 559–570).

Lei, J., & Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, *43*(1), 215–237.

Lei, L. (2019). Unified $\ell_{2\to\infty}$, eigenspace perturbation theory for symmetric random matrices. arXiv1909.04798.

Levin, K., Athreya, A., Tang, M., Lyzinski, V., Park, Y., & Priebe, C. E. (2017). A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference. arXiv1705.09355.

Liu, F., Choi, D., Xie, L., & Roeder, K. (2018). Global spectral clustering in dynamic networks. *Proceedings of the National Academy of Sciences*, *115*(5), 927–932.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Lyzinski, V., Sussman, D. L., Tang, M., Athreya, A., & Priebe, C. E. (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electronic Journal of Statistics*, *8*(2), 2905–2922.

Matias, C., & Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(4), 1119–1141.

Matias, C., & Robin, S. (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. *ESAIM: Proceedings and Surveys*, *47*, 55–74.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *261*(5560), 459–467.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, *85*(3), 825–850.

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* (pp. 849–856).

Padilla, O. H. M., Yu, Y., & Priebe, C. E. (2019). Change point localization in dependent dynamic nonparametric random dot product graphs. arXiv1911.07494.

Pantazis, K., Athreya, A., Frost, W. N., Hill, E. S., & Lyzinski, V. (2020). The importance of being correlated: Implications of dependence in joint spectral inference across multiple networks. arXiv2008.00163.

Rastelli, R., Latouche, P., & Friel, N. (2018). Choosing the number of groups in a latent stochastic blockmodel for dynamic networks. *Network Science*, *6*(4), 469–493.

Rohe, K., Chatterjee, S., & Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, *39*(4), 1878–1915.

Rohe, K., & Zeng, M. (2020). Vintage factor analysis with varimax performs statistical inference. arXiv2004.05387.

Rubin-Delanchy, P., Cape, J., Tang, M., & Priebe, C. E. (2017). A statistical interpretation of spectral embedding: The generalised random dot product graph. arXiv1709.05506.

Sarkar, P., Chakrabarti, D., & Jordan, M. (2014). Nonparametric link prediction in large scale dynamic networks. *Electronic Journal of Statistics*, *8*(2), 2022–2065.

Sarkar, P., & Moore, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*, pp. 1145–1152.

Sewell, D. K., & Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, *110*(512), 1646–1657.

Sewell, D. K., & Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, *12*(2), 351–377.

Smith, A. L., Asta, D. M., & Calder, C. A. (2019). The geometry of continuous latent space models for network data. *Statistical Science*, *34*(3), 428–453.

Sternberg, S. (2010). *Dynamical systems*: Dover Publications.

Sussman, D. L., Tang, M., & Priebe, C. E. (2013). Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(1), 48–57.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., Park, Y., & Priebe, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, *26*(2), 344–354.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., & Priebe, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, *23*(3), 1599–1630.

Tang, M., Cape, J., & Priebe, C. E. (2017). Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral. arXiv1710.10936.

Verhulst, P.-F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences, des Lettres et des Beax-Arts de Belgique*, *18*, 1–38.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416.

Wang, H., Tang, M., Park, Y., & Priebe, C. E. (2013). Locality statistics for anomaly detection in time series of graphs. *IEEE Transactions on Signal Processing*, *62*(3), 703–717.

Xu, K. S., & Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, *8*(4), 552–562.

Young, S. J., & Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for theWeb-Graph* (pp. 138–149).

Zhu, X., Pan, R., Li, G., Liu, Y., & Wang, H. (2017). Network vector autoregression. *Annals of Statistics*, *45*(3), 1096–1123.

How to cite this article: Cape, J. (2021). Spectral analysis of networks with latent space dynamics and signs. *Stat*, *10*(1), e381. https://doi.org/10.1002/sta4.381