

# Convergence guarantees for response prediction for latent structure network time series

Aranyak Acharyya, Francesco Sanna Passino, Michael W. Trosset, Carey E. Priebe

**Abstract**—In this article, we propose a technique to predict the response associated with an unlabeled time series of networks in a semisupervised setting. Our model involves a collection of time series of random networks of growing size, where some of the time series are associated with responses. Assuming that the collection of time series admits an unknown lower dimensional structure, our method exploits the underlying structure to consistently predict responses at the unlabeled time series of networks. Each time series represents a multilayer network on a common set of nodes, and raw stress embedding, a popular dimensionality reduction tool, is used for capturing the unknown latent low dimensional structure. Apart from establishing theoretical convergence guarantees and supporting them with numerical results, we demonstrate the use of our method in the analysis of real-world biological learning circuits of larval *Drosophila*.

**Index Terms**—multilayer networks, doubly unfolded adjacency spectral embedding, raw stress embedding

## 1 INTRODUCTION

In recent times, tools for statistical analysis and inference on random graphs have gained popularity owing to their applicability in extracting information from network data arising from various domains of real life, including neuroscience (Vogelstein et al., 2011), biology and social studies (Holland et al., 1983). Erdős–Rényi random graphs (Erdős and Rényi, 1984) comprise the simplest model of random graphs where the probability of edge formation between any pair of nodes is equal. Stochastic blockmodels (Holland et al., 1983) are graphs where each node is assigned a community membership and the probability of edge formation between any pair of nodes depends only upon the corresponding community memberships. Random dot product graphs (Athreya et al., 2018; Young and Scheinerman, 2007) represent a generalization of stochastic blockmodels, in which every node is assigned a feature vector, also known as its latent positions, and the probability of formation of an edge between any two nodes is the inner product between the corresponding latent positions. The notion of generalized random dot product graphs (Rubin-Delanchy et al., 2022) offer a further generalization to random dot product graphs, where the probability of edge formation between a given pair of nodes is the indefinite inner product between the corresponding latent positions.

While single random graphs have been largely explored in the recent years, the field of studying multiple networks is still emerging. In most cases, the study of multiple networks

entails analysis of a multilayer network, which amounts to a collection of graphs on the same set of nodes. In Jones and Rubin-Delanchy (2020), a popular model, named the *multilayer random dot product graph*, has been proposed to capture the behaviour of a multilayer graph, and a method to obtain node-level embeddings is also proposed. Results in Gallagher et al. (2021) show that *unfolded adjacency spectral embedding*, the multiple network embedding procedure proposed in Jones and Rubin-Delanchy (2020), offers certain desirable stability guarantees, that is, if two nodes behave similarly then they are assigned similar embeddings up to noise.

A popular conjecture suggests that in majority of real-life datasets, the dimension of the datapoints is only artificially high, and in essence the high-dimensional datapoints lie on or cluster around some low-dimensional manifold (Whiteley et al., 2022). This provides the motivation behind manifold learning. Multidimensional scaling (Borg and Groenen, 2005) comprise a class of procedures meant to learn the underlying low-dimensional structure that given high-dimensional datapoints correspond to. Works in Rubin-Delanchy (2020) show that the adjacency spectral embeddings of a latent position random graph with high-dimensional latent positions will be close to a low-dimensional manifold. Results in Trosset et al. (2020) and Trosset and Priebe (2024) establish that manifold learning can be carried out consistently from noisy datapoints sufficiently close to a low-dimensional manifold in a high-dimensional ambient space. Based on these results, Acharyya et al. (2023) and Acharyya et al. (2024) respectively show that node-level and graph-level responses can be predicted in a semisupervised setting from observations on single and multiple graphs corresponding to datapoints on a low-dimensional manifold in a high-dimensional ambient space. This work, where every time series of networks corresponds to a high dimensional datapoint on a low dimensional manifold, can be regarded as an extension to Acharyya et al. (2023) and Acharyya et al. (2024).

- Aranyak Acharyya is with the Mathematical Institute for Data Science, Johns Hopkins University, Baltimore, MD, 21218. E-mail: aachary6@jhu.edu
- Francesco Sanna Passino is with the Department of Mathematics, Imperial College London, London, United Kingdom. E-mail: f.sannapassino@imperial.ac.uk
- Michael W. Trosset is with the Department of Statistics, Indiana University Bloomington, Bloomington, IN 47405. E-mail: mtrosset@iu.edu
- Carey E. Priebe is with the Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218. E-mail: cep@jhu.edu

In this paper, our model involves a collection of time series of networks, each time series corresponding to a point on a one-dimensional manifold in a high-dimensional ambient space. Some of the time series are assumed to be associated with responses linked to the corresponding scalar pre-images via a linear regression model. We propose a technique based on the works in [Baum et al. \(2024\)](#) to predict the response at an unlabeled time series, by exploiting the presence of the scalar pre-images. We establish convergence guarantees of our algorithm and demonstrate its performance guarantees numerically. Besides, we demonstrate the use of our method to analyze the learning circuit of a collection of *Drosophila* larvae.

We organize the manuscript in the following manner. [Section 2](#) introduces the reader to preliminaries of topics like multiplex random graphs, *Doubly Unfolded Adjacency Spectral Embedding* (DUASE; [Baum et al., 2024](#)) and raw stress embedding ([Borg and Groenen, 2005](#); [Trosset and Priebe, 2024](#)). Then [Section 3](#) describes our model, mentions our goal and states our proposed algorithm, and is followed by [Section 4](#) which states our theoretical results. In [Section 5](#), the numerical results are shown in support of the theoretical findings. An illustration of the use of our method in analyzing biological learning circuits of *Drosophila* is presented in [Section 6](#), followed by a conclusion in [Section 7](#). The proofs of our theoretical results are given in [Section A](#).

## 2 IMPORTANT DEFINITIONS, NOTATIONS AND TERMINOLOGIES

Here, we denote the set  $\{1, 2, \dots, n\}$  by  $[n]$ . Also, for any  $k, n \in \mathbb{N}$ , the set  $\{(k-1)n+1, (k-1)n+2, \dots, kn\}$  is denoted by  $S_n^k$ . In this paper, every vector will be represented by a bold lower case letter such as  $\mathbf{v}$ . Any vector by default is a column vector. Matrices will be denoted by bold upper case letters such as  $\mathbf{A}$ . For a matrix  $\mathbf{A}$ , the  $(i, j)$ -th entry will be given by  $\mathbf{A}_{i,j}$ , the  $i$ -th row (written as a column vector) will be given by  $\mathbf{A}_{i,*}$  and the  $j$ -th column will be given by  $\mathbf{A}_{*,j}$ . For any matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  with  $\text{rank}(\mathbf{A}) = r$ , the singular values in descending order will be given by  $\sigma_1(\mathbf{A}) \geq \dots \geq \sigma_r(\mathbf{A})$ , the corresponding left singular vectors will be given by  $\mathbf{u}_1(\mathbf{A}), \dots, \mathbf{u}_r(\mathbf{A})$  and the corresponding right singular vectors will be given by  $\mathbf{v}_1(\mathbf{A}), \dots, \mathbf{v}_r(\mathbf{A})$ . The  $n \times n$  centering matrix will be denoted by  $\mathbf{H}_n = \mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n \mathbf{1}_n^T)$  where  $\mathbf{I}_n$  is the  $n \times n$  identity matrix and  $\mathbf{1}_n$  is the  $n$ -dimensional vector of all ones. For a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{A}_{[S_1, S_2]}$  (where  $S_1 \subseteq [m]$ ,  $S_2 \subseteq [n]$ ) denotes the matrix obtained by retaining the rows with indices in  $S_1$  and the columns with indices in  $S_2$ , and  $\mathbf{A}_{[S_1, \cdot]}$  denotes the matrix obtained by retaining the rows with indices in  $S_1$  and all the columns of  $\mathbf{A}$ , and  $\mathbf{A}_{[\cdot, S_2]}$  denotes the matrix obtained by retaining the columns with indices in  $S_2$  and all the rows of  $\mathbf{A}$ .

Discussed below are some important definitions and notions that we will frequently utilize in this paper.

### 2.1 Preliminaries on multilayer graphs and DUASE

A graph is an ordered pair  $(V, E)$  where  $V$  denotes the set of vertices and  $E \subseteq V \times V$  denotes the collection of edges. An adjacency matrix  $\mathbf{A}$  of a graph is defined in the following

manner:  $\mathbf{A}_{i,j} = 1$  if  $(i, j) \in E$ , and  $\mathbf{A}_{i,j} = 0$  otherwise. Here, we deal with directed graphs, hence  $\mathbf{A}$  has a positive probability of being asymmetric. Latent position random graphs are those graphs where each node is associated with a vector that is called its latent position. The latent position of the  $i$ -th node is denoted by  $\mathbf{x}_i \in \mathbb{R}^d$  for some natural number  $d$ . First, we state the definition of random graphs.

**Definition 1** (Random graph; [Holland et al. \(1983\)](#)). Suppose  $G$  is a directed random graph with  $n$  nodes, such that the probability of an edge from the  $i$ -th node to the  $j$ -th node is given by  $p_{ij}$ . Then, the probability matrix of outward edges for the graph  $G$  will be given by  $\mathbf{P} = (p_{ij})_{i,j=1}^n$ , henceforth referred to as outward edge formation probability matrix (and sometimes we will drop ‘outward’ for sake of convenience). The adjacency matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  satisfies  $\mathbf{A}_{i,j} \sim^{ind} \text{Bernoulli}(p_{ij})$ , for all  $i, j \in [n]$ ,  $i \neq j$ , and  $\mathbf{A}_{i,i} = 0$ ,  $i \in [n]$ .

Secondly, we state the formal definition of multilayer directed random graphs.

**Definition 2** (Multilayer random graph; [Jones and Rubin-Delanchy \(2020\)](#)). A multilayer random graph is a collection of graphs with a common set of nodes, but varying probability of edge from one node to another, for any pair of nodes. Suppose  $G_1, \dots, G_M$  denote a multilayer graph with a common set  $V$  of  $n$  nodes and  $M$  layers. The probability of an edge from the  $i$ -th node to the  $j$ -th node is given by  $p_{ij}^{(l)}$  for the graph  $G_l$ , for all  $l \in [M]$ . The adjacency matrix  $\mathbf{A}^{(l)} \in \mathbb{R}^{n \times n}$  of  $G_l$  satisfies  $\mathbf{A}_{i,j}^{(l)} \sim^{ind} \text{Bernoulli}(p_{ij}^{(l)})$ , for all  $i, j \in [n]$ ,  $i \neq j$ ,  $l \in [M]$ , and  $\mathbf{A}_{i,i}^{(l)} = 0$ ,  $i \in [n]$ ,  $l \in [M]$ .

In this paper, we will deal with a collection of multilayer random graphs, where the number  $n$  of nodes remains the same across all observations. In total, there are  $N$  multilayers of networks and there are  $M$  graphs in each multilayer. In our notation,  $G^{(k,l)}$  denotes the  $l$ -th network in the  $k$ -th layer, and the corresponding adjacency matrix is given by  $\mathbf{A}^{(k,l)} \in \mathbb{R}^{n \times n}$  for all  $k \in [N], l \in [M]$ . In our paper, we shall be dealing with a collection of  $N$  time series of graphs (in which each series contains  $M$  graphs on a common set of  $n$  nodes) where each time series can be regarded as a multilayer graph. For each  $l \in [M]$ , the  $l$ -th graph  $G^{(k,l)}$  corresponds to a timepoint  $\tau_l \in [0, \tau^*]$ ,  $\tau^* \in \mathbb{R}_+$ , for all  $k \in [N]$ , for some values  $0 < \tau_1 < \tau_2 < \dots < \tau_{N-1} < \tau_N < \tau^*$ . For the kind of collection of multilayer graphs defined above, a reliable method of embedding for subsequent inference is *Doubly Unfolded Adjacency Spectral Embedding* (DUASE; [Baum et al., 2024](#)). It offers the stability guarantee that if two nodes (with the possibility that they belong to different graphs in different layers) behave similarly, then they will be assigned similar embedding. The algorithmic pseudocode for DUASE is described in [Algorithm 1](#).

**Remark 1.** In the original paper on *Doubly Unfolded Adjacency Spectral Embedding* ([Baum et al., 2024](#)), the DUASE algorithm technically returns both the left embeddings  $\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\}$  and the right embeddings  $\{\mathbf{Y}_{\mathcal{A}}^{(l)} : l \in [M]\}$ . However, since our goal in this paper needs only the left embeddings, we state the algorithm as in [Algorithm 1](#), returning only  $\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\}$ .

---

**Algorithm 1** DUASE( $\{\mathbf{A}^{(k,l)}\}_{k \in [N], l \in [M]}; d$ )

---

1: Construct the block matrix

$$\mathcal{A} = \left( \mathbf{A}^{(k,l)} \right)_{k \in [N], l \in [M]} \in \mathbb{R}^{nN \times nM}.$$

2: Define  $\mathbf{U}_{\mathcal{A}} = [\mathbf{u}_1(\mathcal{A}) \dots \mathbf{u}_d(\mathcal{A})] \in \mathbb{R}^{nN \times d}$  to be the matrix of the top  $d$  left singular vectors,  $\mathbf{V}_{\mathcal{A}} = [\mathbf{v}_1(\mathcal{A}) \dots \mathbf{v}_d(\mathcal{A})] \in \mathbb{R}^{nM \times d}$  to be the matrix of the top  $d$  right singular vectors, and  $\mathbf{\Sigma}_{\mathcal{A}} = \text{diag}\{\sigma_1(\mathcal{A}), \dots, \sigma_d(\mathcal{A})\} \in \mathbb{R}^{d \times d}$  to be the diagonal matrix of the top  $d$  singular values of  $\mathcal{A}$ .

3: Compute the left embedding  $\mathbf{X}_{\mathcal{A}} = \mathbf{U}_{\mathcal{A}} \mathbf{\Sigma}_{\mathcal{A}}^{\frac{1}{2}} \in \mathbb{R}^{nN \times d}$  and the right embedding  $\mathbf{Y}_{\mathcal{A}} = \mathbf{V}_{\mathcal{A}} \mathbf{\Sigma}_{\mathcal{A}}^{\frac{1}{2}} \in \mathbb{R}^{nM \times d}$ .

4: The information of the  $k$ -th layer across all timepoints is contained in  $\mathbf{X}_{\mathcal{A}}^{(k)} = (\mathbf{X}_{\mathcal{A}})_{[(k-1)n+1:kn, \cdot]} \in \mathbb{R}^{n \times d}$  for all  $k \in [N]$  and the information of the  $l$ -th timepoint across all layers is contained in  $\mathbf{Y}_{\mathcal{A}}^{(l)} = (\mathbf{Y}_{\mathcal{A}})_{[(l-1)n+1:ln, \cdot]} \in \mathbb{R}^{n \times d}$  for all  $l \in [M]$ .

5: **return**  $\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\}$ .

---

## 2.2 Raw stress embedding

### 2.2.1 Finite sample size

Raw stress embedding is a popular method for nonlinear dimensionality reduction. Given dissimilarities  $\{\Delta_{i,j}\}_{i,j \in [N]}$  for some finite  $N \in \mathbb{N}$ , the goal is to find vectors  $\mathbf{z}_1, \dots, \mathbf{z}_N \in \mathbb{R}^c$  for some predetermined target embedding dimension  $c$ , such that the interpoint Euclidean distances between the  $\mathbf{z}_i$  approximate the corresponding dissimilarities, that is,  $\|\mathbf{z}_i - \mathbf{z}_j\| \approx \Delta_{i,j}$  for all  $i, j \in [N]$ . The full algorithm is given below.

---

**Algorithm 2** RSEmb( $\{\Delta_{i,j}\}_{i,j \in [N]}; c$ )

---

1: Define the raw stress function to be  $\sigma(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N) = \sum_{i,j=1}^N w_{i,j} (\|\mathbf{z}_i - \mathbf{z}_j\| - \Delta_{i,j})^2$  where  $\mathbf{z}_i \in \mathbb{R}^c$  for all  $i \in [N]$ .

2: Obtain  $(\hat{\mathbf{z}}_1, \dots, \hat{\mathbf{z}}_N) = \arg \min \sigma(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$

3: **return**  $(\hat{\mathbf{z}}_1, \dots, \hat{\mathbf{z}}_N)$ .

---

For this article, we restrict our attention to the regime of  $c = 1$ . Moreover, we set  $w_{i,j} = 1$  for all pairs  $(i, j)$ .

**Remark 2.** Iterative majorization (for details, see Chapter 8 of [Borg and Groenen, 2005](#)) is used to minimize the raw stress function. In order to avoid getting trapped in local minima, classical multidimensional scaling outputs are typically used for initialization. In our paper, we assume the global minima is reached for theoretical reasons.

### 2.2.2 Infinite sample size

Suppose  $\mathcal{M}$  is a compact Riemannian manifold of innate dimension  $c$ , and let  $\Delta : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$  be a Borel-measurable dissimilarity function. Assuming  $g : \mathcal{M} \rightarrow \mathbb{R}^c$  is a Borel-measurable embedding function, define the raw stress function by

$$\begin{aligned} & \sigma((\Delta, \mathcal{P}), g) \\ &= \int_{\mathcal{M}} \int_{\mathcal{M}} (\|g(m') - g(m'')\| - \Delta(m', m''))^2 \mathcal{P}(dm') \mathcal{P}(dm''). \end{aligned}$$

Defining  $\mathbf{D}(m', m'') = \|g(m') - g(m'')\|$ , we redefine the raw stress function as

$$\begin{aligned} & \sigma((\Delta, \mathcal{P}), \mathbf{D}) \\ &= \int_{\mathcal{M}} \int_{\mathcal{M}} (\|g(m') - g(m'')\| - \Delta(m', m''))^2 \mathcal{P}(dm') \mathcal{P}(dm'') \\ &= \|\mathbf{D} - \Delta\|_{\mathcal{P}}^2. \end{aligned}$$

Let  $\mathcal{Y}$  be the cone of all Euclidean pseudometrics, then we define the raw stress minimizer for dissimilarity  $\Delta$  with respect to probability function  $\mathcal{P}$  as

$$\text{Min}(\Delta, \mathcal{P}) = \left\{ \mathbf{D} \in \mathcal{Y} : \sigma((\Delta, \mathcal{P}), \mathbf{D}) \leq \inf_{\mathbf{D} \in \mathcal{Y}} \sigma((\Delta, \mathcal{P}), \mathbf{D}) \right\}.$$

## 3 MODEL AND METHODOLOGY

Our model involves a set of time series of graphs. Each time series can be regarded as a multilayer random directed graph. This is so because a multilayer random graph is a collection of graphs on a common set of nodes, while a time series of graphs in practice involves a collection of realizations of a single graph over multiple timepoints (for instance, sequence of snapshots of a network of neurons in the brain of an organism). It is assumed that some of the time series are associated with a scalar response. It is also assumed that each time series corresponds to a scalar pre-image. A simple linear regression model links the responses to the scalar pre-images of the time series of graphs.

There are  $N$  time series in total, and each time series has  $M$  graphs, where each graph has  $n$  nodes. The adjacency matrix of the  $l$ -th graph in the  $k$ -th time series is denoted by  $\mathbf{A}^{(k,l)}$ , and let the corresponding probability matrix be  $\mathbf{P}^{(k,l)}$ . We define the grand probability matrix to be  $\mathcal{P} = (\mathbf{P}^{(k,l)})_{k \in [N], l \in [M]}$ . We denote

$$\{\mathbf{X}_{\mathcal{P}}^{(k)} : k \in [N]\} = \text{DUASE} \left( \left\{ \mathbf{P}^{(k,l)} \right\}_{k \in [N], l \in [M]} \right),$$

where we recall that for every  $k \in [N]$ ,  $\mathbf{X}_{\mathcal{P}}^{(k)} \in \mathbb{R}^{n \times d}$ . Suppose there exist scalars  $t_i$  such that for all  $k_1, k_2$ ,

$$\|\mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)}\|_{2,\infty} = |t_{k_1} - t_{k_2}|.$$

Suppose  $s \ll K$  is a fixed natural number and for  $k \in [s]$ , response  $y_k$  is associated with the  $k$ -th time series. We further assume that the responses  $y_k$  are linked to the scalar pre-images  $t_k$  via a simple linear regression model, that is,

$$y_k = \alpha + \beta t_k + \epsilon_k$$

where  $\epsilon_k \sim^{iid} N(0, \sigma_{\epsilon}^2)$ .

Our goal is to predict  $y_{s+1}$  for the unlabeled  $(s+1)$ -th time series. To do that, we first estimate  $\mathbf{X}_{\mathcal{P}}^{(k)}$  with  $\mathbf{X}_{\mathcal{A}}^{(k)}$ , where

$$\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\} = \text{DUASE} \left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]} \right).$$

The matrices  $\mathbf{X}_{\mathcal{A}}^{(k)}$  estimate their population counterpart matrices  $\mathbf{X}_{\mathcal{P}}^{(k)}$  consistently up to an orthogonal transformation, and hence the pairwise distances between  $\mathbf{X}_{\mathcal{A}}^{(k)}$  can consistently estimate the corresponding pairwise distances  $\mathbf{X}_{\mathcal{P}}^{(k)}$ . Hence the population (involving probability matrices  $\mathbf{P}^{(k,l)}$ ) dissimilarity matrix  $\Delta$  is estimated by the sample

(involving adjacency matrices  $\mathbf{A}^{(k,l)}$ ) dissimilarity matrix  $\hat{\Delta}$ , where

$$\Delta = \left( \left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\| \right)_{k_1, k_2=1}^N = (|t_{k_1} - t_{k_2}|)_{k_1, k_2=1}^N,$$

$$\hat{\Delta} = \left( \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\| \right)_{k_1, k_2=1}^N.$$

Then, we apply raw stress minimization algorithm on  $\hat{\Delta}$  into  $\mathbb{R}$ , and obtain scalars  $\{\hat{z}_k\}_{k=1}^N$ . Treating the embeddings  $\hat{z}_k$  as proxy regressors, we predict the response  $y_{s+1}$  with  $\tilde{y}_{s+1}$ . The full procedure is described in *Algorithm 3*.

---

**Algorithm 3** PredTSGResp  $\left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]}; d; r \right)$

---

- 1: Compute the DUASE of  $\{\mathbf{A}^{(k,l)}\}_{k \in [N], l \in [M]}$ :

$$\{\mathbf{X}_{\mathcal{A}}^{(k)}\}_{k=1}^N = \text{DUASE} \left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]} \right).$$

- 2: Obtain the estimated pairwise distance matrix

$$\hat{\Delta} = \left( \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} \right)_{k_1, k_2=1}^N.$$

- 3: Use raw stress embedding to obtain

$$\{\hat{z}_k\}_{k=1}^N = \text{RSEmb}(\hat{\Delta}; 1).$$

- 4: Compute the estimated regression coefficients:

$$\hat{b} = \frac{\sum_{i=1}^s (y_i - \bar{y})(\hat{z}_i - \bar{\hat{z}})}{\sum_{i=1}^s (\hat{z}_i - \bar{\hat{z}})^2}, \quad \hat{a} = \bar{y} - \hat{b}\bar{\hat{z}},$$

where  $\bar{\hat{z}} = s^{-1} \sum_{i=1}^s \hat{z}_i$ .

- 5: **return** Predicted responses  $\tilde{y}_r = \hat{a} + \hat{b}\hat{z}_r$ .
- 

From now onwards, we shall be dealing with scenarios where  $N, M, n$  vary all together. Specifically, there exists a sequence  $(N_K, M_K, n_K)$  such that  $N_K \rightarrow \infty$ ,  $M_K \rightarrow \infty$ ,  $n_K \rightarrow \infty$  as  $K \rightarrow \infty$  in a manner that the *Theorem 1* holds. If it is clear from the context that the  $K$ -th instant is being spoken of, we shall omit the subscript  $K$  and replace  $N_K$  with  $N$  (and  $M_K$  with  $M$ ,  $n_K$  with  $n$ ). Stated below are our model assumptions.

**Assumption 1.** The grand probability matrix, defined as  $\mathcal{P} = (\mathbf{P}^{(k,l)})_{k \in [N_K], l \in [M_K]}$ , satisfies  $\text{rank}(\mathcal{P}) = d$  for all sufficiently large  $K$ .

**Assumption 2.** The number of graphs per multilayer and the number of multilayers both must grow slower than the number of nodes in each graph, that is,  $M_K = o(n_K)$ ,  $N_K = o(n_K)$ .

**Assumption 3.** The scalars  $t_k \sim_{iid} \mathcal{P}$  for all  $k$ , where  $\mathcal{L} = \text{support}(\mathcal{P}) \subset \mathbb{R}$  is closed and bounded.

Having discussed our setting in this section, we move on to the next section to state the results establishing asymptotic properties of our proposed PredTSGResp algorithm.

## 4 THEORETICAL RESULTS

In this section, we present our theoretical results. The results are primarily based on two results from the literature, one

that establishes consistency of DUASE embeddings (*Theorem 1*) and another that establishes continuity of raw stress embeddings (*Theorem 2*).

**Theorem 1.** (*Baum et al., 2024*) Suppose there are  $N$  time series of networks, each consisting of  $M$  directed random latent position graphs where each graph has  $n$  nodes. Each time series can be regarded as a multiplex of graphs. Let  $\mathbf{P}^{(k,l)}$  denote the outward edge formation probability matrix and  $\mathbf{A}^{(k,l)}$  denote the adjacency matrix of the  $l$ -th graph in the  $k$ -th series, for all  $k \in [N]$ ,  $l \in [M]$ . Further, assume that *Assumptions 1* and *2* hold. Denoting

$$\{\mathbf{X}_{\mathcal{P}}^{(k)} : k \in [N]\} = \text{DUASE} \left( \left\{ \mathbf{P}^{(k,l)} \right\}_{k \in [N], l \in [M]}; d \right)$$

and

$$\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\} = \text{DUASE} \left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]}; d \right),$$

there exists  $\mathbf{Q} \in \mathcal{O}(d)$  such that for each  $k \in \mathbb{N}$ ,

$$\left\| \mathbf{X}_{\mathcal{A}}^{(k)} - \mathbf{X}_{\mathcal{P}}^{(k)} \mathbf{Q} \right\|_{2,\infty} \rightarrow^P 0$$

as  $K \rightarrow \infty$ .

The above theorem states that as the number of time series, number of graphs and number of nodes increase simultaneously, the sample DUASE embedding for every time series converges to its population counterpart up to an orthogonal transformation. Since the maximum row norm of a matrix is invariant to orthogonal transformations, we can conclude that the distance between two sample DUASE embeddings approach the distance between the two corresponding population DUASE embeddings.

**Proposition 1.** In the setting of *Theorem 1*, define the matrices

$$\Delta^{(K)} = \left( \left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\|_{2,\infty} \right)_{k_1, k_2 \in [N]},$$

$$\hat{\Delta}^{(K)} = \left( \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} \right)_{k_1, k_2 \in [N]}.$$

Then, for each  $k_1, k_2 \in \mathbb{N}$ ,

$$\left| \hat{\Delta}_{k_1, k_2}^{(K)} - \Delta_{k_1, k_2}^{(K)} \right| \rightarrow^P 0$$

as  $K \rightarrow \infty$ .

Observe that as the number of time series, number of graphs and number of nodes grow simultaneously (under our model assumptions, one of which demands that the number of multilayers and number of graphs per multilayer must grow slower than the the number of nodes per graph), each entry of the population pairwise distance matrix  $\Delta$  converges entrywise to the pairwise distance matrix  $\mathbf{D}$  between the scalar pre-images  $t_i$ . Our next result from *Trosset and Priebe (2024)* states that the globally minimizing EDM-1 matrices of a sequence of dissimilarity matrices converge to the globally minimizing EDM-1 matrix of the limit of the dissimilarity matrices.

**Theorem 2.** (*Trosset and Priebe, 2024*) Suppose  $\hat{\Delta}^{(K)}$  is uniformly bounded and converges to the dissimilarity function  $\hat{\Delta}^{(\infty)}$  in the topology of pointwise convergence. Assume that the sequence of empirical probability measures  $\{\hat{\mathcal{P}}_K\}$  weakly converges



to the probability measure  $\mathcal{P}_\infty$ . Then every sequence  $\{\hat{\mathbf{D}}^{(K)}\}$  where  $\hat{\mathbf{D}}^{(K)} \in \text{Min}(\hat{\Delta}^{(K)}, \hat{\mathcal{P}}_K)$ , will have an accumulation point, and if  $\hat{\mathbf{D}}^{(\infty)}$  is an accumulation point of  $\{\hat{\mathbf{D}}^{(K)}\}$ , then it will satisfy  $\hat{\mathbf{D}}^{(\infty)} \in \text{Min}(\hat{\Delta}^{(\infty)}, \mathcal{P}_\infty)$ .

The above theorem establishes consistency of dissimilarity minimizers in a growing sample size scenario. It helps us prove the following result which establishes the consistency of raw stress embeddings obtained from the maximum row norm differences between the sample DUASE embeddings.

**Proposition 2.** In the setting of Theorem 1, define

$$\{\hat{z}_i\}_{i=1}^N = \text{RSEmb}(\hat{\Delta}^{(K)}; 1).$$

Then as  $K \rightarrow \infty$ , for every  $k_1, k_2 \in N$ ,

$$(|\hat{z}_{k_1} - \hat{z}_{k_2}| - |t_{k_1} - t_{k_2}|) \rightarrow^P 0.$$

Recall from our assumptions that the pre-images  $t_k$  are the unknown regressors in our regression model for which we wish to predict a response. Proposition 2 tells us that pairwise distances between the raw stress embeddings approach the pairwise distance between the true regressors, thereby helping the raw stress embeddings closely approximate an affine transformation of the regressors, justifying the use of the raw stress embeddings as proxy regressors in a linear regression model.

**Theorem 3.** Suppose there are  $N$  time series of random directed graphs where each series has  $M$  graphs and each graph has  $n$  nodes. Define  $\mathcal{P} = (\mathbf{P}^{(k,l)})_{k \in [N], l \in [M]}$  and  $\mathcal{A} = (\mathbf{A}^{(k,l)})_{k \in [N], l \in [M]}$ , where  $\mathbf{P}^{(k,l)}$  and  $\mathbf{A}^{(k,l)}$  denote the probability matrix and the adjacency matrix for the  $l$ -th graph in the  $k$ -th series. Denote the population DUASE embeddings by  $\{\mathbf{X}_{\mathcal{P}}^{(k)} : k \in [N]\} = \text{DUASE}(\mathbf{P}^{(k,l)}; d)$  and the sample DUASE embeddings by  $\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\} = \text{DUASE}(\mathbf{A}^{(k,l)}; d)$  and assume that there exist scalar pre-images  $t_k$  such that for every  $k_1, k_2$ ,

$$\|\mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)}\|_{2,\infty} = |t_{k_1} - t_{k_2}|.$$

Suppose responses  $y_1, \dots, y_s$  are observed corresponding to the first  $s$  time series such that the following model hold

$$y_k = \alpha + \beta t_k + \epsilon_k$$

where  $\epsilon_k \sim^{iid} N(0, \sigma_\epsilon^2)$ ,  $k \in [s]$ . Denote the raw stress embeddings by  $\{\hat{z}_k\}_{k=1}^N = \text{RSEmb}(\hat{\Delta}; 1)$  where

$$\hat{\Delta} = \left( \|\mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)}\|_{2,\infty} \right)_{k_1, k_2=1}^N.$$

Then the predicted response

$$\tilde{y}_r = \text{PredTSGResp} \left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]}; d, r \right)$$

satisfies

$$|\tilde{y}_r - \hat{y}_r| \rightarrow^P 0$$

as  $K \rightarrow \infty$ , where  $\hat{y}_r$  is the predicted response at the  $r$ -th time series based on the true regressors  $t_k$ .

The above theorem tells us that as the number of unlabeled (auxiliary) time series increases (along with number of graphs and size of graphs), the predicted response obtained

from our method approaches the predicted response obtained from the true regressors. In order to test the validity of a simple linear regression model, we deploy an  $F$ -test that uses the observed responses and the predicted responses obtained from the true regressors. In the absence of the true regressors, we can still use predicted responses obtained from our method and hope to mimic the power of the original  $F$ -test, by virtue of Theorem 3.

**Corollary 1.** In the setting of Theorem 3, suppose we want to test  $H_0 : \beta = 0$  against  $H_1 : \beta \neq 0$  at level of significance  $\tilde{\alpha}$ . Define the following test statistics:

$$F^* = (s-2) \frac{\sum_{k=1}^s (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^s (y_k - \bar{y})^2}, \quad \hat{F} = (s-2) \frac{\sum_{k=1}^s (\tilde{y}_k - \bar{y})^2}{\sum_{k=1}^s (y_k - \bar{y})^2}.$$

Suppose  $\pi^*$  is the power of the test carried out by the principle: reject  $H_0$  if  $F^* > c_{\tilde{\alpha}}$ , and let  $\hat{\pi}$  be the power of the test with the principle: reject  $H_0$  if  $\hat{F} > c_{\tilde{\alpha}}$ . Then, for every  $(\alpha, \beta)$ ,  $|\hat{\pi} - \pi^*| \rightarrow 0$  as  $K \rightarrow \infty$ .

The above result paves a way for testing the validity of a proposed linear regression model between the responses and the scalar pre-images in a realistic setting where the regressors are unknown.

## 5 SIMULATIONS

In this section, we describe the simulation experiments. We carry out two simulation experiments that provide numerical support for the theoretical results Theorem 3 and Corollary 1.

First, we describe the simulation that numerically shows the predicted response obtained from Algorithm 3 approaches the predicted response obtained from the true regressors. The number of labeled datapoints is fixed at  $s = 5$ . Denoting the common index that controls the growth of  $n$  (number of nodes per graph),  $M$  (number of graphs per multilayer) and  $N$  (number of multilayers) by  $K$ , we set  $n_K = 15 + \lfloor (K-1)^{1.5} \rfloor$ ,  $N_K = 10 + (K-1)$  and  $M_K = 8 + (K-1)$ , while the common index  $K$  varies in the range  $\{1, 2, 3, \dots, 30\}$ . We repeat the following task for each  $K$  on each of 100 Monte Carlo samples. At first, we obtain the regressors  $t_1, \dots, t_s \sim^{iid} U(0, 1)$  associated with observed responses  $y_1, \dots, y_s$  where  $y_k = \alpha + \beta t_k + \epsilon_k$ ,  $\epsilon_k \sim^{iid} N(0, \sigma_\epsilon^2)$  with regression parameters  $\alpha = 2.0$ ,  $\beta = 8.0$ ,  $\sigma_\epsilon = 0.01$ . We generate the pre-images for the auxiliary points as  $t_{s+1}, \dots, t_N \sim^{iid} U(0, 1)$ . We define two matrices  $\tilde{\mathbf{X}} \in \mathbb{R}^{nN \times d}$  and  $\tilde{\mathbf{Y}} \in \mathbb{R}^{nM \times d}$ , where we choose the embedding dimension  $d = 2$ . Now,

$$\tilde{\mathbf{X}}_{[(i-1)n+1]:in, \cdot} = \frac{t_i}{\sqrt{d}} \mathbf{J}_{(n,d)}$$

where  $\mathbf{J}_{(n,d)}$  is the  $n \times d$  matrix of all ones. Also, for every pair  $(i, j) \in [nM] \times [d]$ ,  $\tilde{\mathbf{Y}}_{ij} \sim^{iid} U(0.2, 0.5)$ . We define the grand probability matrix  $\mathcal{P} = \tilde{\mathbf{X}}\tilde{\mathbf{Y}}^T$ . For any  $i_1, i_2 \in [nN]$ ,

$$\|\tilde{\mathbf{X}}_{i_1, \cdot} - \tilde{\mathbf{X}}_{i_2, \cdot}\| = \left\| \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2} + \eta} \right)_{i_1, \cdot} - \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2} + \eta} \right)_{i_2, \cdot} \right\|.$$

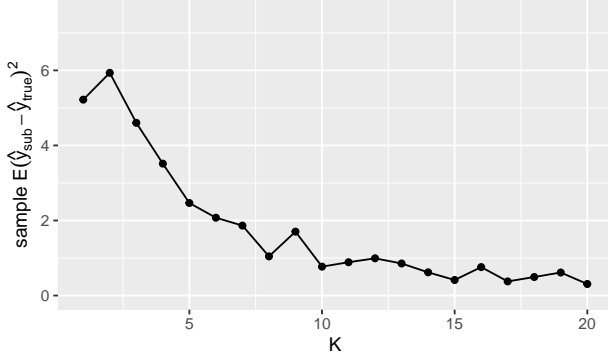


Figure 1: Plot showing that average squared difference between the predicted response based on the true regressors and the predicted response obtained from our method approaches zero, as the number of multilayers, number of graphs per multilayer and the number of nodes per graph increases in a suitable manner.

Hence, for any  $k_1, k_2 \in [N]$ ,

$$\begin{aligned} |t_{k_1} - t_{k_2}| &= \left\| \tilde{\mathbf{X}}_{[s_{n^1},]} - \tilde{\mathbf{X}}_{[s_{n^2},]} \right\|_{2,\infty} \\ &= \left\| \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2}+\eta} \right)_{[s_{n^1},]} - \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2}+\eta} \right)_{[s_{n^2},]} \right\|_{2,\infty} \\ &= \left\| \left\{ \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2}} \right)_{[s_{n^1},]} - \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2}} \right)_{[s_{n^2},]} \right\} \mathbf{S}_{\mathcal{P}}^{\eta} \right\|_{2,\infty} \end{aligned}$$

where  $\eta \in [-\frac{1}{2}, \frac{1}{2}]$ . Clearly, if  $d = 1$ , then we could scale  $t_k$  by  $\sigma_1(\mathcal{P})^\eta$  for all  $k$ , and reformulated our regression model. But even if  $d > 1$ , it can be numerically shown that the second largest singular value of  $\mathcal{P}$  is much smaller than its largest singular value, and hence the scaling of the regressors and the reformulation of the regression model still holds approximately, enough to ensure numerical results are satisfactory. We provide numerical evidence in [Appendix A.2](#).

Next, we generate the adjacency matrix  $\mathbf{A}^{(k,l)}$  for the  $l$ -th graph in the  $k$ -th multilayer following  $\mathbf{A}_{i,j}^{(k,l)} \sim \text{ind Bernoulli}(\mathbf{P}_{i,j}^{(k,l)})$  for all  $i \neq j$ , where  $\mathbf{P}^{(k,l)}$  is the probability matrix for the  $l$ -th graph in the  $k$ -th multilayer. We obtain the DUASE embedding

$$\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\} = \text{DUASE} \left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]} ; d \right)$$

and thereby compute the raw stress embeddings

$$\{\hat{z}_i\}_{i=1}^N = \text{RSEmb}(\{\hat{\Delta}_{i,j}\}_{i,j \in [N]}; 1).$$

Using a simple linear regression model on  $(y_i, t_i)_{i=1}^s$ , the response at the  $(s+1)$ -th multilayer is predicted with  $\hat{y}_{s+1}$ . Similarly, using a simple linear regression model on  $(y_i, \hat{z}_i)_{i=1}^s$ , we predict the response at the  $(s+1)$ -th multilayer by  $\tilde{y}_{s+1}$ . The mean of the values of the squared difference  $(\hat{y}_{s+1} - \tilde{y}_{s+1})^2$  is computed over all the 100 Monte Carlo samples, and plotted against  $K$ , and the resulting plot is given in [Figure 1](#). It is seen that the sample average squared distance between the predicted response from the true regressors and the predicted response obtained from our method approaches zero as  $K$  goes to infinity, thus supporting [Theorem 3](#).

We next present simulation results in support of [Corollary 1](#). For testing  $H_0 : \beta = 0$  versus  $H_1 : \beta \neq 0$ ,

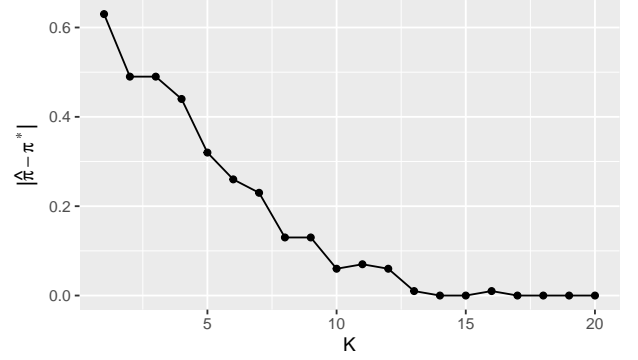


Figure 2: Plot showing the difference between estimated powers of the tests based on the true and the substitute  $F$ -statistics approaching zero as the number of multilayers, number of graphs per multilayer and number of nodes per graph increase at suitable rates. A set of 100 Monte Carlo samples of a collection of multilayer directed random graphs are generated, and one-dimensional embeddings are obtained by raw-stress minimization on the Double Unfolded Adjacency Spectral Embeddings. Responses are regressed against these one-dimensional raw-stress embeddings to obtain a substitute  $F$ -statistic.

we choose level of significance  $\tilde{\alpha}$ . The setting is same as before, except this time we take  $n_K = 15 + \lfloor (K-1)^{1.5} \rfloor$ ,  $M_K = 8 + (K-1)$  and  $N_K = 10 + (K-1)$ . For every  $K$  in the range  $\{1, 2, \dots, 30\}$ , on each of 100 Monte Carlo samples, we proceed as before to obtain the raw-stress embeddings  $\{\hat{z}_i\}_{i=1}^N$ . Using a linear regression model on the bivariate data  $(y_i, t_i)_{i=1}^s$  we compute the predicted responses  $\{\hat{y}_i\}_{i=1}^s$  and thus obtain the true  $F$ -statistic  $F^*$ , and using a linear regression model on the bivariate data  $(y_i, \hat{z}_i)_{i=1}^s$ , we obtain the predicted responses  $\{\tilde{y}_i\}_{i=1}^s$  and thus obtain the substitute  $F$ -statistic  $\hat{F}$ , where

$$F^* = (s-2) \frac{\sum_{k=1}^s (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^s (y_k - \hat{y}_k)^2}, \quad \hat{F} = (s-2) \frac{\sum_{k=1}^s (\tilde{y}_k - \bar{y})^2}{\sum_{k=1}^s (y_k - \tilde{y}_k)^2}.$$

The test based on  $F^*$  (and equivalently, also the test based on  $\hat{F}$ ) rejects  $H_0$  at significance level  $\tilde{\alpha}$  if  $F^* > c_{\tilde{\alpha}}$  for pre-specified threshold  $c_{\tilde{\alpha}}$ . For each statistic amongst  $F^*$  and  $\hat{F}$ , we estimate the power of the test based on that statistic by computing the proportion of times the test based on that statistic rejects  $H_0$  at level  $\tilde{\alpha}$ . We calculate the absolute difference between the estimated powers of the tests based on  $F^*$  and  $\hat{F}$  and plot them against  $K$ , and the resulting plot is given in [Figure 2](#). We observe that the difference between the estimated powers of the tests approaches zero as  $K$  increases.

## 6 REAL DATA ANALYSIS

In this section, we demonstrate the use of our methodology in the analysis of biological learning networks of larval *Drosophila*. The wiring diagram, also known as connectome, of the larval *Drosophila* has been recently completed ([Windling et al., 2023](#)), which enables simulation of biologically realistic models of the circuits of neurons based on known anatomical connectivity ([Eschbach et al., 2020](#)). There have been recent works on studying the learning networks (circuit of neurons responsible for learning in an organism) by training connectome-constrained models to perform associative learning in simulations where a given stimulus is delivered, eliciting a certain network output in the animal

(for instance, when an odour is coupled with pain, the odour loses its attraction to the organism).

To be more specific, in our case, the network models are trained to perform extinction learning. In this phenomenon, an association between a conditioned stimulus (e.g an odour) and reinforcement (e.g pain) is initially learnt, and then weakened by exposure to the same conditioned stimulus in absence of the reinforcement. The behaviour of the network over different timepoints is simulated, comprising a time series of networks corresponding to a single extinction learning trial. A total of 143 such trials are performed corresponding to 11 different replications for each of 13 different models, each model being a result of removal of a particular synapse from the parent network. In each trial, at first a conditioned stimulus, followed by reinforcement (pain or reward), is delivered, and then after significant gaps the stimulus is delivered again twice, without being coupled with the reinforcement. A learning score is recorded for every extinction learning trial. The learning score is defined as the ratio of the network output at the third conditioned stimulus to that at the second conditioned stimulus, where the network output at a particular time is defined as the ratio of degree of aversion to the degree of attraction to the conditioned stimulus at that time.

We thus have  $M = 143$  time series of networks, each associated with a learning score. Each time series has  $N = 160$  networks and each network has  $n = 140$  nodes. We convert each network into a binary one by choosing to record the entry of the adjacency matrix as one if its modulus exceeds a particular threshold, and as zero otherwise. The threshold is taken to be the 25-th percentile of the absolute values of the original edge weights. We perform Double Unfolded Adjacency Spectral Embedding on the collection of these time series of networks (with embedding dimension  $d = 3$ ), and thus obtain a matrix representation for every time series. We obtain a dissimilarity matrix of the pairwise differences (measured in two-to-infinity norm) of the matrix representations of the time series, and by subsequent raw-stress minimization we obtain one-dimensional embeddings for all the time series. A linear regression model is assumed to link the responses with the one-dimensional raw-stress embeddings, and an  $F$ -test with  $p$ -value = 0.026 justifies it (at level of significance 0.05). The scatterplot of the responses against the one-dimensional raw-stress embeddings is given in Figure 3, along with the fitted regression line.

## 7 DISCUSSION

In this paper, we propose a method to predict a response corresponding to an unlabeled time series of networks, in a semisupervised setting. We assume that each time series of networks correspond to a scalar pre-image such that a suitable measure of pairwise distances between the time series is captured by the interpoint distances between the scalar pre-images. Assuming a linear regression model links the responses to the scalar pre-images, we propose to predict the responses by using raw-stress minimization to find proxies for the true regressors. We provide theoretical and numerical justification for our method in Sections 4 and 5 respectively.

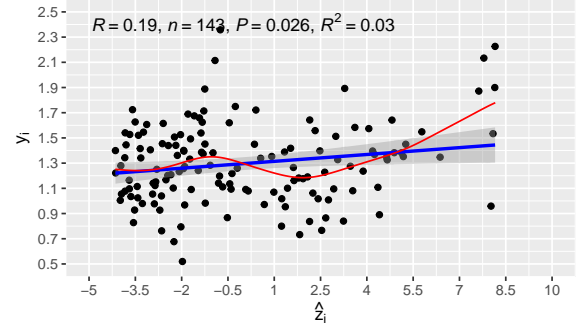


Figure 3: Scatterplot of the responses  $y_i$  against the one-dimensional raw-stress embeddings  $\hat{z}_i$ , along with the fitted linear regression line in blue. An  $F$ -test is performed to check the validity of the simple regression model  $y_i = a + b\hat{z}_i + \eta_i$  and  $p = 0.026$  is obtained, justifying the use of the simple linear regression model. The estimated model parameters are  $\hat{a} = 1.296$  and  $\hat{b} = 0.018$ . The curve in red denotes the fitted nonparametric regression curve, by the method of local linear regression, which gives an R-squared value of 0.1667.

We also demonstrate the use of our method in analysis of learning networks in larval *Drosophila*. A collection of time series of networks, each representing the behaviour of the learning circuit in larval *Drosophila* over snapshots of time in a learning trial, is observed. Each time series is associated with a response dubbed learning score. Our method obtains one-dimensional embeddings corresponding to all the time series, such that the responses can be linked to the embeddings via a linear regression model at level 0.05 (an  $F$ -test results in a  $p$ -value of 0.026).

The desirable asymptotic properties of our method are reliant on the guarantee of vanishing uniform bound on the errors in estimating the regressors. Such guarantee can help extend our result to the regime where a nonparametric regression is used to link the responses with the regressors. We provide an example of predicting the responses via a nonparametric regression model, from our real dataset (details in Section 6). The setting is same as the one described in Section 6, where we have 143 time series of networks, each associated with a response. A one-dimensional embedding is obtained corresponding to every time series by raw stress minimization upon the DUASE embeddings, and a local linear regression model is used to predict the responses. The plot of the responses against the regressors, along with the fitted regression curve, is given in Figure 3.

The results in this paper involve establishing asymptotic convergence guarantees for the output of the proposed algorithm. Finding the rate of convergence of the predicted response obtained from our method to the predicted response obtained from the true regressors is an interesting open problem in this area. Solving that problem would give us an idea of how large the set of auxiliary time series needs to be in order to achieve a given level of accuracy. Moreover, making finite-sample improvements to the algorithmic output taking the measurement error into account, comprises a potentially intriguing problem solving which will likely be beneficial to the practitioners.

## ACKNOWLEDGMENTS

Francesco Sanna Passino acknowledges funding from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/Y002113/1.

## REFERENCES

- Aranyak Acharyya, Joshua Agterberg, Michael W Trosset, Youngser Park, and Carey E Priebe. Semisupervised regression in latent structure networks on unknown manifolds. *Applied Network Science*, 8(1):75, 2023.
- Aranyak Acharyya, Jesús Arroyo Relión, Michael Clayton, Marta Zlatic, Youngser Park, and Carey E Priebe. Consistent response prediction for multilayer networks on unknown manifolds. *arXiv preprint arXiv:2405.03225*, 2024.
- Avanti Athreya, Donniell E. Fishkind, Minh Tang, Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, Keith Levin, Vince Lyzinski, Yichen Qin, and Daniel L Sussman. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.
- Maximilian Baum, Francesco Sanna Passino, and Axel Gandy. Doubly unfolded adjacency spectral embedding of dynamic multiplex graphs. *arXiv preprint arXiv:2410.09810*, 2024.
- Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.
- Paul L. Erdős and Alfréd Rényi. On the evolution of random graphs. *Transactions of the American Mathematical Society*, 286:257–257, 1984.
- Claire Eschbach, Akira Fushiki, Michael Winding, Casey M Schneider-Mizell, Mei Shao, Rebecca Arruda, Katharina Eichler, Javier Valdes-Aleman, Tomoko Ohyama, Andreas S Thum, et al. Recurrent architecture for adaptive regulation of learning in the insect brain. *Nature Neuroscience*, 23(4):544–555, 2020.
- Ian Gallagher, Andrew Jones, and Patrick Rubin-Delanchy. Spectral embedding for dynamic networks with stability guarantees. *Advances in Neural Information Processing Systems*, 34:10158–10170, 2021.
- Paul Holland, Kathryn B. Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5:109–137, 1983.
- Andrew Jones and Patrick Rubin-Delanchy. The multilayer random dot product graph. *arXiv preprint arXiv:2007.10455*, 2020.
- Patrick Rubin-Delanchy. Manifold structure in graph embeddings. *Advances in Neural Information Processing Systems*, 33:11687–11699, 2020.
- Patrick Rubin-Delanchy, Carey E. Priebe, Minh Tang, and Joshua Cape. A statistical interpretation of spectral embedding: The generalised random dot product graph. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84:1446–1473, 2022.
- Michael W Trosset and Carey E Priebe. Continuous multidimensional scaling. *arXiv preprint arXiv:2402.04436*, 2024.
- Michael W Trosset, Mingyue Gao, Minh Tang, and Carey E Priebe. Learning 1-dimensional submanifolds for subsequent inference on random dot product graphs. *arXiv preprint arXiv:2004.07348*, 2020.
- Joshua T. Vogelstein, William R. Gray Roncal, R. Jacob Vogelstein, and Carey E. Priebe. Graph classification using signal-subgraphs: Applications in statistical connectomics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1539–1551, 2011.
- Nick Whiteley, Annie Gray, and Patrick Rubin-Delanchy.

Discovering latent topology and geometry in data: a law of large dimension. *arXiv e-prints*, pages arXiv–2208, 2022.

Michael Winding, Benjamin D Pedigo, Christopher L Barnes, Heather G Patsolic, Youngser Park, Tom Kazimiers, Akira Fushiki, Ingrid V Andrade, Avinash Khandelwal, Javier Valdes-Aleman, et al. The connectome of an insect brain. *Science*, 379(6636):eadd9330, 2023.

Stephen J. Young and Edward R. Scheinerman. Random dot product graph models for social networks. In *Workshop on Algorithms and Models for the Web-Graph*, 2007.

## APPENDIX A

### BACKGROUND JUSTIFICATIONS AND PROOFS

#### A.1 Certain background justifications

**Use of Theorem 2:** From the entrywise convergence of a sequence of dissimilarity matrices of growing size, we use *Theorem 2* to conclude the convergence of the corresponding globally minimizing EDM-1 matrices. However, *Theorem 2* actually rests on the a setting of a sequence of dissimilarity functions  $\{\hat{\Delta}^{(K)}\}$  converging pointwise to another dissimilarity function  $\{\hat{\Delta}^{(\infty)}\}$ , and a sequence of probability distributions  $\{\hat{\mathcal{P}}_K\}$  converging uniformly to a probability distribution  $\mathcal{P}_\infty$ , and it states that any sequence  $\hat{\mathbf{D}}^{(K)} \in \text{Min}(\hat{\Delta}^{(K)}, \hat{\mathcal{P}}_K)$  will have an accumulation point  $\hat{\mathbf{D}}^{(\infty)} \in \text{Min}(\hat{\Delta}^{(\infty)}, \mathcal{P}_\infty)$ .

**Statement of Proposition 2:** While the conclusion in our paper for *Proposition 2* is pointwise convergence of the raw stress embeddings, a much stronger statement holds true in this regard. The mode of convergence is  $L^p$ , that is:

$$\lim_{K \rightarrow \infty} \int_{\mathcal{L}} \int_{\mathcal{L}} (|\hat{z}_{k_1} - \hat{z}_{k_2}| - |t_{k_1} - t_{k_2}|)^p \mathcal{P}(dt_{k_1}) \mathcal{P}(dt_{k_2}) = 0$$

for some  $p \geq 1$ , under *Assumption 3*.

#### A.2 Numerical justification for approximation in higher dimensional ambient spaces

In our simulation in *Section 5*, the relation

$$|t_{k_1} - t_{k_2}| = \left\| \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2}} \right)_{[s_{n^{k_1}}]} - \left( \mathbf{U}_{\mathcal{P}} \mathbf{S}_{\mathcal{P}}^{\frac{1}{2}} \right)_{[s_{n^{k_2}}]} \right\|_{2,\infty}$$

up to rescaling of all the  $t_k$  holds exactly for  $d = 1$  and approximately for  $d > 1$ , because for  $d > 1$  the second largest singular value of  $\mathcal{P}$  is much smaller than its largest singular value. Numerical evidence is shown in *Table 1*.

#### A.3 Proofs of theoretical results

**Proposition 1.** In the setting of *Theorem 1*, define the matrices

$$\Delta^{(K)} = \left( \left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\|_{2,\infty} \right)_{k_1, k_2 \in [N]},$$

$$\hat{\Delta}^{(K)} = \left( \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} \right)_{k_1, k_2 \in [N]}.$$

Then, for each  $k_1, k_2 \in \mathbb{N}$ ,

$$\left| \hat{\Delta}_{k_1, k_2}^{(K)} - \Delta_{k_1, k_2}^{(K)} \right| \rightarrow^P 0$$



$K$	$\sigma_1(\mathcal{P})$	$\sigma_2(\mathcal{P})$
1	38.3913207802614	3.79012067821168e-17
2	46.2467967723536	4.30423371659228e-18
3	53.3483912419807	2.19173601047544e-17
4	70.6204284932543	7.59644834333919e-17
5	87.3256963731673	1.60271756976842e-16
6	105.690843867424	3.75449557243395e-17
7	127.893833191334	5.12974328627644e-16
8	153.452174685367	7.32395180182678e-16
9	181.817642929274	1.08072901300687e-15
10	216.55689743623	2.64045904981925e-15
11	252.293666389214	4.88745113265974e-15
12	295.071278399014	2.1915113364941e-14
13	343.504129675349	4.7492033185641e-14
14	393.134326882093	1.0073712770101e-13
15	458.2653037287	1.97145542975649e-13
16	519.272524528707	2.97958596948405e-13
17	575.884298576391	4.07037061348058e-13
18	650.49260937855	5.36758516418727e-13
19	724.990035514813	6.43326958379969e-13
20	811.462566958402	7.98294314677235e-13

Table 1: Largest and second largest singular values of the graph probability matrix  $\mathcal{P}$  as the number of nodes in each graph, number of graphs in each time series and the number of time series grow simultaneously.

as  $K \rightarrow \infty$ .

*Proof.* Fix  $k_1, k_2 \in \mathbb{N}$ . Observe that by triangle inequality and invariance of Frobenius norm to orthogonal transformation,

$$\begin{aligned} & \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} - \left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\|_{2,\infty} \\ & \leq \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_1)} \mathbf{Q} \right\|_{2,\infty} + \left\| \mathbf{X}_{\mathcal{A}}^{(k_2)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \mathbf{Q} \right\|_{2,\infty}. \end{aligned}$$

Likewise, we can write,

$$\begin{aligned} & \left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\|_{2,\infty} - \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} \\ & \leq \left\| \left( \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_1)} \mathbf{Q} \right) - \left( \mathbf{X}_{\mathcal{A}}^{(k_2)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \mathbf{Q} \right) \right\|_{2,\infty}. \end{aligned}$$

Thus, from the above two equations combined,

$$\begin{aligned} & \left| \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} - \left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\|_{2,\infty} \right| \\ & \leq \left\| \left( \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_1)} \mathbf{Q} \right) - \left( \mathbf{X}_{\mathcal{A}}^{(k_2)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \mathbf{Q} \right) \right\|_{2,\infty}. \end{aligned}$$

Recall that from [Theorem 1](#), the right hand side goes to zero, hence so does the left hand side. Thus, for each  $k_1, k_2 \in \mathbb{N}$ ,

$$\left| \hat{\Delta}_{k_1, k_2}^{(K)} - \Delta_{k_1, k_2}^{(K)} \right| \rightarrow^P 0$$

as  $K \rightarrow \infty$ .  $\square$

**Proposition 2.** In the setting of [Theorem 1](#), define

$$\{\hat{z}_i\}_{i=1}^N = \text{RSEmb}(\hat{\Delta}^{(K)}; 1).$$

Then as  $K \rightarrow \infty$ , for every  $k_1, k_2 \in \mathbb{N}$ ,

$$(|\hat{z}_{k_1} - \hat{z}_{k_2}| - |t_{k_1} - t_{k_2}|) \rightarrow^P 0.$$

*Proof.* From [Proposition 1](#), for every  $k_1, k_2$ ,  $|\hat{\Delta}_{k_1, k_2} - \Delta_{k_1, k_2}| \rightarrow^P 0$  as  $K \rightarrow \infty$ . From model assumptions, for every  $k_1, k_2$ ,  $\Delta_{k_1, k_2} = |t_{k_1} - t_{k_2}|$  as  $K \rightarrow \infty$ . Hence, for every  $k_1, k_2$ ,  $|\hat{\Delta}_{k_1, k_2} - |t_{k_1} - t_{k_2}|| \rightarrow^P 0$  as  $K \rightarrow \infty$ . By [Theorem 3](#) in [Trosset and Priebe \(2024\)](#), we have that  $||\hat{z}_{k_1} - \hat{z}_{k_2}| - |t_{k_1} - t_{k_2}|| \rightarrow^P 0$ .  $\square$

**Theorem 3.** Suppose there are  $N$  time series of random directed graphs where each series has  $M$  graphs and each graph has  $n$  nodes. Define  $\mathcal{P} = (\mathbf{P}^{(k,l)})_{k \in [N], l \in [M]}$  and  $\mathcal{A} = (\mathbf{A}^{(k,l)})_{k \in [N], l \in [M]}$ , where  $\mathbf{P}^{(k,l)}$  and  $\mathbf{A}^{(k,l)}$  denote the probability matrix and the adjacency matrix for the  $l$ -th graph in the  $k$ -th series. Denote the population DUASE embeddings by  $\{\mathbf{X}_{\mathcal{P}}^{(k)} : k \in [N]\} = \text{DUASE}(\mathcal{P}; d)$  and the sample DUASE embeddings by  $\{\mathbf{X}_{\mathcal{A}}^{(k)} : k \in [N]\} = \text{DUASE}(\mathcal{A}; d)$  and assume that there exist scalar pre-images  $t_k$  such that for every  $k_1, k_2$ ,

$$\left\| \mathbf{X}_{\mathcal{P}}^{(k_1)} - \mathbf{X}_{\mathcal{P}}^{(k_2)} \right\|_{2,\infty} = |t_{k_1} - t_{k_2}|.$$

Suppose responses  $y_1, \dots, y_s$  are observed corresponding to the first  $s$  time series such that the following model hold

$$y_k = \alpha + \beta t_k + \epsilon_k$$

where  $\epsilon_k \sim^{iid} N(0, \sigma_\epsilon^2)$ ,  $k \in [s]$ . Denote the raw stress embeddings by  $\{\hat{z}_k\}_{k=1}^N = \text{RSEmb}(\hat{\Delta}; 1)$  where

$$\hat{\Delta} = \left( \left\| \mathbf{X}_{\mathcal{A}}^{(k_1)} - \mathbf{X}_{\mathcal{A}}^{(k_2)} \right\|_{2,\infty} \right)_{k_1, k_2=1}^N.$$

Then the predicted response

$$\tilde{y}_r = \text{PredTSGResp} \left( \left\{ \mathbf{A}^{(k,l)} \right\}_{k \in [N], l \in [M]}; d, r \right)$$

satisfies

$$|\tilde{y}_r - \hat{y}_r| \rightarrow^P 0$$

as  $K \rightarrow \infty$ , where  $\hat{y}_r$  is the predicted response at the  $r$ -th time series based on the true regressors  $t_k$ .

*Proof.* From [Proposition 2](#) we have,

$$\max_{k_1, k_2 \in [s]} (|\hat{z}_{k_1} - \hat{z}_{k_2}| - |t_{k_1} - t_{k_2}|) \rightarrow^P 0$$

as  $N, M, n \rightarrow \infty$ . As the difference between the interpoint distances between the embeddings  $\{\hat{z}_k\}_{k=1}^s$  and the interpoint distances between the true regressors  $\{t_k\}_{k=1}^s$  approach zero, the raw stress embeddings approach an affine transformation on the true regressors. Since we know that an affine transformation upon the true regressors in a simple linear regression model does not alter a predicted response value, the predicted response  $\tilde{y}_r$  based on the embeddings  $\hat{z}_k$  approach the predicted response  $\hat{y}_r$  based on the true regressors  $t_k$ .  $\square$

**Corollary 1.** In the setting of [Theorem 3](#), suppose we want to test  $H_0 : \beta = 0$  against  $H_1 : \beta \neq 0$  at level of significance  $\tilde{\alpha}$ . Define the following test statistics:

$$F^* = (s-2) \frac{\sum_{k=1}^s (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^s (y_k - \hat{y}_k)^2}, \quad \hat{F} = (s-2) \frac{\sum_{k=1}^s (\tilde{y}_k - \bar{y})^2}{\sum_{k=1}^s (y_k - \tilde{y}_k)^2}.$$

Suppose  $\pi^*$  is the power of the test carried out by the principle: reject  $H_0$  if  $F^* > c_{\tilde{\alpha}}$ , and let  $\hat{\pi}$  be the power of the test with the principle: reject  $H_0$  if  $\hat{F} > c_{\tilde{\alpha}}$ . Then, for every  $(\alpha, \beta)$ ,  $|\hat{\pi} - \pi^*| \rightarrow 0$  as  $K \rightarrow \infty$ .

*Proof.* We know from [Theorem 3](#), for any  $(\alpha, \beta) \in \mathbb{R}^2$ , for all  $r \in [s]$ ,  $|\tilde{y}_r - \hat{y}_r| \rightarrow 0$  as  $K \rightarrow \infty$ . Hence, for all  $(\alpha, \beta) \in \mathbb{R}^2$ ,  $|\hat{F} - F^*| \rightarrow 0$  as  $K \rightarrow \infty$ , and hence for all  $(\alpha, \beta) \in \mathbb{R}^2$ , for any significance level  $\tilde{\alpha}$ ,  $|\hat{\pi} - \pi^*| = |\mathbb{P}_{\alpha, \beta}[\hat{F} > c_{\tilde{\alpha}}] - \mathbb{P}_{\alpha, \beta}[F^* > c_{\tilde{\alpha}}]| \rightarrow 0$  as  $K \rightarrow \infty$ .  $\square$

**Aranyak Acharyya** Aranyak Acharyya received his Bachelors degree in Statistics from Presidency University at Kolkata, India in 2017, and his Masters degree in Statistics from Indian Institute of Technology Kanpur in 2019. From 2019 to 2024, he was a PhD student in the Department of Applied Mathematics and Statistics at Johns Hopkins University, and then he began working as a postdoctoral fellow in the Mathematical Institute for Data Science in Johns Hopkins University. His research interests include statistical inference on networks, manifold learning and artificial intelligence.

**Francesco Sanna Passino** Francesco Sanna Passino received a double BSc degree in Statistics from the universities of Bologna and Glasgow in 2016, a MSc degree in Statistics from Imperial College London in 2017, and a PhD in Statistics from Imperial College London in 2021. Since 2022, he has been a Lecturer (Assistant Professor) in Statistics at Imperial College London. His research interests mostly revolve around statistical analysis of dynamic networks, latent variable models and model-based clustering.

**Michael W. Trosset** Michael W. Trosset received his BA degree in Mathematics from Rice University 1978, and the PhD degree in Statistics from the University of California at Berkeley in 1983. He has held faculty positions at the University of Arizona, the College of William and Mary, and Indiana University where he currently chairs the Department of Statistics. He is the author of an introductory textbook, *An Introduction to Statistical Inference and Its Applications to R*. His research interests include euclidean representations of proximity data, nonlinear dimension reduction, computer experiments and stochastic optimization.

**Carey E. Priebe** Carey E. Priebe received the Bachelors degree in Mathematics from Purdue University in 1984, the Masters degree in Computer Science from San Diego State University in 1988, and the PhD degree in Information Technology from George Mason University in 1993. Since 1994 he has been a professor at the Department of Applied Mathematics and Statistics at Johns Hopkins University. His research interests include statistical analysis of random networks, statistical pattern recognition and statistical inference on high dimensional data. He is a senior member of the IEEE, and Elected Member of the International Statistical Institute, a Fellow of the Institute of Mathematical Statistics, and a Fellow of the American Statistical Association.