Unsupervised Attributed Dynamic Network Embedding with Stability Guarantees

Emma Ceccherini[1] **Ian Gallagher**[2] **Andrew Jones**[3] **Daniel Lawson**[1]

[1]University of Bristol, U.K.
[2]The University of Melbourne, Australia
[3]University of Edinburgh, U.K.

Abstract

Stability for dynamic network embeddings ensures that nodes behaving the same at different times receive the same embedding, allowing comparison of nodes in the network across time. We present attributed unfolded adjacency spectral embedding (AUASE), a stable unsupervised representation learning framework for dynamic networks in which nodes are attributed with time-varying covariate information. To establish stability, we prove uniform convergence to an associated latent position model. We quantify the benefits of our dynamic embedding by comparing with state-of-the-art network representation learning methods on four real attributed networks. To the best of our knowledge, AUASE is the only attributed dynamic embedding that satisfies stability guarantees without the need for ground truth labels, which we demonstrate provides significant improvements for link prediction and node classification.

1 INTRODUCTION

Representation learning on dynamic networks [Goyal et al., 2020, Zuo et al., 2018, Trivedi et al., 2019, Zhou et al., 2018, Ma et al., 2020, Hamilton et al., 2017, Goyal et al., 2018], which learns a low dimensional representation for each node, is a widely explored problem. While most existing network embedding techniques focus solely on the network features, nodes in real-world networks are associated with a rich set of attributes. For example, in a social network, the user's posts are significantly correlated with trust and following relationships, and it has been shown that jointly exploiting both information sources improves learning performance [Tang et al., 2013].

Network embeddings for static attributed networks include frameworks based on matrix factorisation [Yang et al., 2015], or deep learning [Gao and Huang, 2018, Tu et al., 2017, Tan et al., 2023, Sun et al., 2016, Zhang et al., 2018, Li et al., 2021]. Some existing dynamic network embeddings leverage node attributes, but their exploitation of node attributes is rather limited, as they are usually solely used to initialise the first layer [Sankar et al., 2020, Dwivedi et al., 2023, Liu et al., 2021, Xu et al., 2020b,a].

Approaches that purposefully exploit node attributes include frameworks based on matrix factorisation [Liu et al., 2020, Li et al., 2017], deep learning [Tang et al., 2022, Ahmed et al., 2024, Wei et al., 2019], or Bayesian modelling [Luodi et al., 2024]. However, to the best of our knowledge, none of these methods have stability guarantees, which ensure that if two node/time pairs 'behave the same' in the network, their representation is the same up to noise. Stability allows for the comparison of embeddings over time because the embedding space has a consistent interpretation.

Attributed unfolded adjacency spectral embedding (AUASE) is a framework for unsupervised dynamic attributed network embedding with stability guarantees. AUASE builds on unfolded adjacency spectral embedding (UASE) [Jones and Rubin-Delanchy, 2021], an unsupervised spectral method for dynamic network embedding with stability guarantees [Gallagher et al., 2021]. However, UASE is only suited to unattributed networks. To include node attributes, AUASE combines the adjacency matrix and a covariate matrix, with edge weight proportional to covariate values, into an attributed adjacency matrix.

To analyse the statistical properties of AUASE, we define an attributed dynamic latent position model combining a latent position dynamic network model [Gallagher et al., 2021] with a dynamic covariate model. We show that AUASEs converge asymptotically to the noise-free embedding, which is essential to demonstrate AUASE's stability properties.

Figure 1 shows the embeddings of AUASE and two comparable methods - visually explaining the difference between stable and unstable embeddings. The astute reader will be wondering why leading dynamic graph methods could have

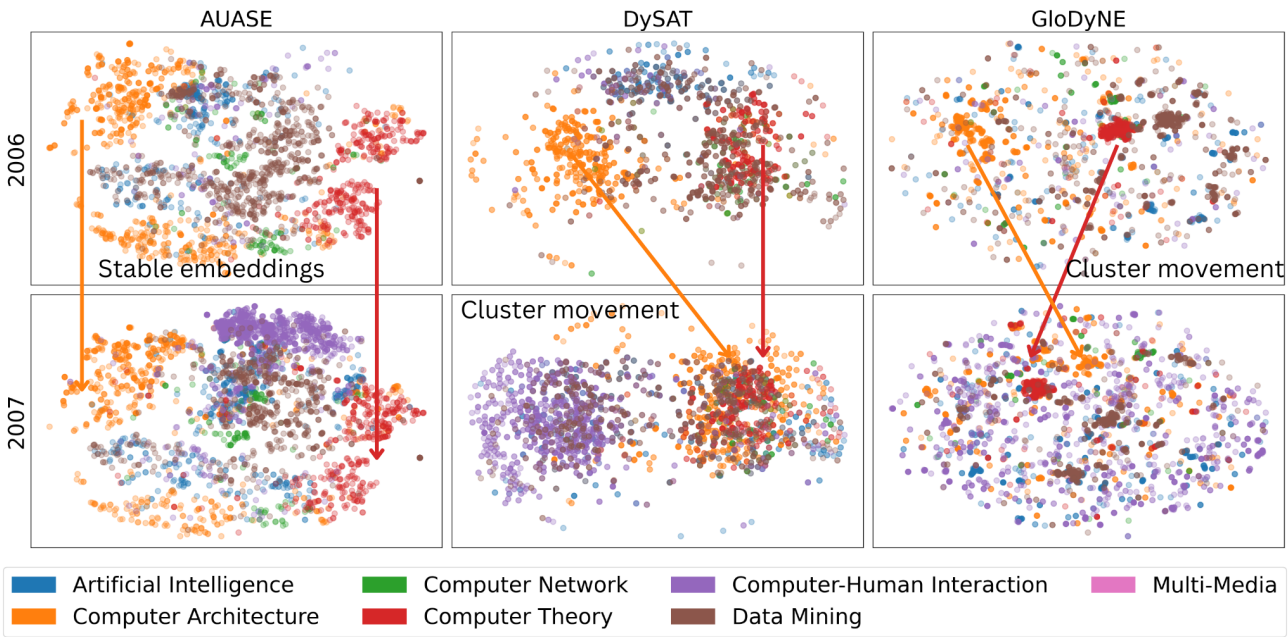

Figure 1: AUSE provides stable embeddings, meaning that nodes that behave the same at different times receive the same embedding illustrated with two-dimensional t-SNE visualisations of node embeddings (in a shared temporal space) of the DBLP datasets. Colours correspond to community membership.

such a visually poor internal representation. There are a wide class of problems - for which graph neural networks (GNNs) were developed - that can be expressed as a supervised learning problem, and the classifier learns a separate map for each time point, for instance, node classification and link prediction [Rossi et al., 2020, Pareja et al., 2020]. However, many problems are not of this form and require *unsupervised stable* embeddings. Examples include but are not limited to: group discovery in purchasing networks for marketing purposes [Palla et al., 2007], anomaly detection in financial or cybersecurity networks [Ranshous et al., 2015] and protein function prediction in dynamic biological networks [Klein et al., 2012, Yue et al., 2020].

We compare AUASE to state-of-the-art unsupervised attributed dynamic embeddings on the tasks of node classification and link prediction as a means to demonstrate AUASE embedding's validity and the value of stability. In extensive experiments on four large real-world datasets AUASE considerably outperforms other methods, which in many cases perform worse than baseline guessing precisely because they lack embedding stability.

2 THEORY & METHODS

Let $\mathcal{G} = (\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(T)})$ represent a dynamic sequence of T node-attributed networks each with nodes $[n] = \{1, \ldots, n\}$. At time $t \in [T]$, we denote $\mathcal{G}^{(t)} = (\mathbf{A}^{(t)}, \mathbf{C}^{(t)})$, where $\mathbf{A}^{(t)} \in \{0,1\}^{n \times n}$ represents the network structure using a binary adjacency matrix and $\mathbf{C}^{(t)} \in \mathbb{R}^{n \times p}$ represents the time-varying node attributes as covariates. The aim is to obtain a low-dimensional embedding $\{\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)}\}_{t \in [T]}$ representing the network and covariate behavior of each node across time.

2.1 ATTRIBUTED UNFOLDED ADJACENCY SPECTRAL EMBEDDING

We present our attributed embedding procedure, which extends unfolded adjacency spectral embedding (UASE).

For each time period $t \in [T]$, we incorporate the covariates $\mathbf{C}^{(t)}$ into the adjacency matrix $\mathbf{A}^{(t)}$ by including them as p attribute nodes in the network. Nodes are connected to these p attribute nodes with edge weight proportional to the corresponding covariate value, producing the augmented adjacency matrix,

$$\mathbf{A}_C^{(t)} = \begin{bmatrix} (1-\alpha)\mathbf{A}^{(t)} & \alpha\mathbf{C}^{(t)} \\ \alpha\mathbf{C}^{(t)\top} & \mathbf{0}_{p \times p} \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)},$$

where the hyperparameter $\alpha \in [0,1]$ balances the relative contributions of $\mathbf{A}^{(t)}$ and $\mathbf{C}^{(t)}$, fixed for all $t \in [T]$.

Given the augmented adjacency matrices $\mathbf{A}_C^{(t)}$, the procedure continues analogously to UASE. Hence, for $\alpha = 0$, AUASE reduces to UASE. The unfolded attributed adjacency matrix $\mathbf{A}_C$ is constructed by concatenating the attributed adjacency matrices, and the dynamic spectral embed-

ding is obtained using the output of the d-truncated SVD. This procedure is detailed in Algorithm 1.

Algorithm 1 Attributed unfolded adjacency spectral embedding (AUASE).

Input: Attributed dynamic network $\mathcal{G}$, embedding dimension d, hyperparameter $\alpha \in [0, 1]$.

1: Construct the attributed adjacency matrix $\mathbf{A}_C^{(t)}$.

2: Construct the unfolded attributed adjacency matrix

$$\mathbf{A}_C = (\mathbf{A}_C^{(1)} \mid \cdots \mid \mathbf{A}_C^{(T)}).$$

3: Compute the d-truncated SVD of

$$\mathbf{A}_C \approx \mathbf{U}_\mathbf{A} \mathbf{\Sigma}_\mathbf{A} \mathbf{V}_\mathbf{A}^\top.$$

4: Divide $\mathbf{V}_\mathbf{A}^\top$ into T blocks

$$\mathbf{V}_\mathbf{A} = (\mathbf{V}_\mathbf{A}^{(1)} \mid \cdots \mid \mathbf{V}_\mathbf{A}^{(T)}).$$

5: Define the attributed unfolded adjacency spectral embeddings (AUASE) as

$$\hat{\mathbf{Y}}_\mathbf{A}^{(t)} = (\mathbf{Y}_\mathbf{A}^{(t)})_{1:n} = (\mathbf{V}_\mathbf{A}^{(t)} \mathbf{\Sigma}_\mathbf{A}^{1/2})_{1:n}^\top,$$

Output: Node embeddings $\{\hat{\mathbf{Y}}_\mathbf{A}^{(t)}\}_{t \in [T]}$.

We focus exclusively on the rows of $\mathbf{Y}_\mathbf{A}^{(t)}$ corresponding to nodes in the dynamic network $\mathcal{G}$. While AUASE produces dynamic embeddings for the p attribute nodes, we do not consider their asymptotic properties as we assume a fixed number of attributes.

2.2 ATTRIBUTED DYNAMIC NETWORK MODEL

To analyse the asymptotic properties of the embeddings, we propose a latent position model for dynamic attributed networks. The use of dynamic latent position models to study dynamic networks is well established [Jones and Rubin-Delanchy, 2021, Sewell and Chen, 2015, Lee and Priebe, 2011, Kim et al., 2018]. To incorporate node attributes, we combine a non-attributed dynamic network model [Gallagher et al., 2021] with a dynamic covariate model using the same dynamic latent positions.

Definition 1 (Dynamic network model)**.** *Conditional on latent positions $Z_i^{(t)} \in \mathcal{Z}$, the sequence of symmetric matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(T)} \in \mathbb{R}^{n \times n}$ is distributed as a dynamic latent position network model, if, for all $i < j$,*

$$\mathbf{A}_{ij}^{(t)} \mid Z_i^{(t)}, Z_j^{(t)} \overset{ind}{\sim} H(Z_i^{(t)}, Z_j^{(t)}),$$

where H is a symmetric real-valued distribution function, $H(Z_1, Z_2) = H(Z_2, Z_1)$, for all $Z_1, Z_2 \in \mathcal{Z}$.

Definition 2 (Dynamic covariate model)**.** *Conditional on latent positions $Z_i^{(t)} \in \mathcal{Z}$, the sequence of covariate matrices $\mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(T)} \in \mathbb{R}^{n \times p}$ is distributed as a dynamic latent position covariate model, if, for all $i \in [n]$, $\ell \in [p]$,*

$$\mathbf{C}_{i\ell}^{(t)} \mid Z_i^{(t)} \overset{ind}{\sim} f_\ell(Z_i^{(t)}),$$

where $f_\ell(Z)$ is a real-valued distribution for all $Z \in \mathcal{Z}$.

Our goal is to describe the attributed adjacency matrices $\mathbf{A}_C^{(t)}$ as a special type of dynamic network model (Definition 1). To include node attributes, we add fixed latent positions $Z_{n+\ell}^{(t)} = i_\ell$ representing an index for each covariate such that $\mathcal{I} = \{i_\ell : \ell \in [p]\}$ is disjoint from the possible node latent positions $\mathcal{Z}$. For convenience, we denote $f_{i_\ell}(Z) = f_\ell(Z)$.

Definition 3 (Attributed dynamic latent position model)**.** *Conditional on latent positions $Z_i^{(t)} \in \mathcal{Z}$ and fixed latent positions $Z_{n+\ell}^{(t)} = i_\ell$, the sequence of symmetric matrices $\mathbf{A}_C^{(1)}, \ldots, \mathbf{A}_C^{(T)} \in \mathbb{R}^{(n+p)\times(n+p)}$ is distributed as an attributed dynamic latent position network model with sparsity parameter $\rho > 0$, if,*

for all $i < j$,

$$\mathbf{A}_{ij}^{(t)} \mid Z_i^{(t)}, Z_j^{(t)} \overset{ind}{\sim} H(Z_i^{(t)}, Z_j^{(t)}),$$

where H is a symmetric real-valued distribution function satisfying

$$H(Z_i, Z_j) = \begin{cases} (1-\alpha)\text{Bernoulli}(\rho f(Z_i, Z_j)) & Z_i, Z_j \in \mathcal{Z}, \\ \alpha \rho^{1/2} f_{Z_j}(Z_i) & Z_i \in \mathcal{Z}, Z_j \in \mathcal{I}, \\ \delta_0 & Z_i, Z_j \in \mathcal{I}, \end{cases}$$

where δ_0 represents the Dirac delta function with mass at zero, and $\mathbf{X} \sim \alpha f$ is shorthand for $\mathbf{X}/\alpha \sim f$.

Definition 3 includes the scenario where the covariates do not change over time, for all nodes, $\mathbf{C}_{i\ell}^{(t)} = \mathbf{C}_{i\ell}$ for all $t \in [T]$. To show this, we construct latent positions which includes the fixed covariate, $Z_i' = (Z_i, \mathbf{C}_{i\ell})$. The dynamic covariate model is then a deterministic function of this new latent position, $f_\ell(Z_i') = \mathbf{C}_{i\ell}$. Multiple covariates can be included this way.

2.3 THEORETICAL RESULTS

In this section, we present theoretical results that describe the asymptotic behaviour of the dynamic embedding $\hat{\mathbf{Y}}^{(t)}$ given the following model assumptions:

Assumption 1 (Low-rank expectation)**.** *There exist maps $\phi : \mathcal{Z}^T \cup \mathcal{I}^T \to \mathbb{R}^d$ and $\phi_t : \mathcal{Z} \cup \mathcal{I} \to \mathbb{R}^d$ for all $t \in [T]$ such that*

$$\mathbb{E}[(\mathbf{A}_C^{(t)})_{ij} \mid Z_i, Z_j] = \phi(Z_i)\phi_t(Z_j^{(t)})^\top,$$

where $Z_i = \{Z_i^{(1)}, \ldots, Z_i^{(T)}\}$.

Assumption 2 (Singular values of $\mathbf{P}_C$)**.** *The* d *non-zero singular values of the mean unfolded adjacency matrix* $\mathbf{P}_C = \mathbb{E}(\mathbf{A}_C)$ *satisfy*

$$\sigma_i(\mathbf{P}_C) = \Theta(T^{1/2}\rho n)$$

for all $i \in [d]$ *with high probability as* $n \to \infty$.

Assumption 3 (Finite number of covariates)**.** $p = \mathrm{O}(1)$.

Assumption 4 (Network sparsity)**.** *The sparsity factor* ρ *satisfies*

$$\rho = \omega(n^{-1}\log^k(n))$$

for some constant k.

Assumption 5 (Subexponential tails)**.** *There exists a constant* $\beta > 0$ *such that*

$$\mathbb{P}\left(|f_l(Z_i^{(t)}) - \mathbb{E}[f_l(Z_i^{(t)})]| > x \mid Z_i^{(t)}\right) \leq \exp\left(-\beta x\right)$$

Assumption 1 allows us to define a canonical choice for the maps ϕ based on the mean attributed unfolded adjacency matrix $\mathbf{P}_C = \mathbb{E}[\mathbf{A}_C]$. The low-rank assumption states that the d-truncated SVD $\mathbf{P}_C = \mathbf{U_P}\mathbf{\Sigma_P}\mathbf{V_P}$ is exact and we define canonical choice for the maps,

$$\phi(Z_i) = (\mathbf{U_P}\mathbf{\Sigma_P}^{1/2})_i \in \mathbb{R}^d,$$
$$\phi_t(Z_i^{(t)}) = (\mathbf{V_P}^{(t)}\mathbf{\Sigma_P}^{1/2})_i = (\mathbf{Y_P}^{(t)})_i \in \mathbb{R}^d.$$

In particular, we will demonstrate how the AUASE output $\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)}$ is a good approximation of the noise-free embedding $\hat{\mathbf{Y}}_{\mathbf{P}}^{(t)} = (\mathbf{Y_P}^{(t)})_{1:n}$ defined by the map ϕ_t.

Assumption 2 is a technical condition which ensures that the growth of the singular values of $\mathbf{P}_C$ is regulated. Assumption 3 assumes a fixed number of covariates which alongside Assumption 4 ensures that the networks are dense enough to recover the latent positions. Versions of these assumptions are required to analyse the multilayer random dot product graph [Jones and Rubin-Delanchy, 2021]. These assumptions are not limiting as for finite n (i.e., any practical application) scaling α would mitigate any imbalance between sparsity in the graph and the covariates.

Assumption 5 prevents large attribute values, which would otherwise ruin the structure in an embedding. It is satisfied trivially for bounded distributions like the Bernoulli and Beta distributions and many other common distributions like the exponential and Gaussian distributions. An equivalent condition is established in the weighted generalised random dot product graph model [Gallagher et al., 2024].

Our main result shows that $\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)}$ is a good approximation of some invertible linear transformation of $\hat{\mathbf{Y}}_{\mathbf{P}}^{(t)}$, as the size of the largest error tends to zero as the number of nodes grows. Let $\|\cdot\|_{2\to\infty}$ denote the two-to-infinity norm [Cape et al., 2019], the maximum row-wise Euclidean norm of a matrix.

Theorem 1 (Uniform consistency)**.** *Under Assumptions 1-5 there exists a sequence of orthogonal matrices* $\mathbf{W} = \mathbf{W}_n \in \mathbb{O}(d)$ *such that, for all* $t \in [T]$,

$$\|\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)} - \hat{\mathbf{Y}}_{\mathbf{P}}^{(t)}\mathbf{W}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1, 1/\beta)}{T^{1/4}\rho^{1/2}n^{1/2}}\right),$$

where for $x, y \in \mathbb{R}$, *we define* $r_\alpha(x, y) = (1-\alpha)x + \alpha y$, *and* $X_n = \mathrm{O}_{\mathbb{P}}(a_n)$ *means that* X_n/a_n *is bounded in probability Janson [2011].*

Theorem 1 has important methodological implications similar to the equivalent theorem in Jones and Rubin-Delanchy [2021]. Uniform consistency in the two-to-infinity norm implies that subsequent statistical analysis is consistent up to rotation.

2.4 STABILITY PROPERTIES

Xue et al. [2022] argues that dynamic network embeddings are desirable to preserve the network's local structure and the long-term dynamic evolution. The survey observes that most dynamic network embeddings possess only one of these properties, but not both. In this section, we show that AUASE captures both local and global network structures, which we refer to as spatial and temporal stability, respectively.

We say that node/time pairs (i, s) and (j, t) are exchangeable if they 'behave the same' within the dynamic network. For the attributed dynamic latent position model (Definition 3) this is equivalent to stating, for all $Z \in \mathcal{Z} \cup \mathcal{I}$,

$$H(Z_i^{(s)}, Z) = H(Z_j^{(t)}, Z).$$

It is desirable for exchangeable node/time pairs (i, s) and (j, t) to have equal embeddings $\hat{\mathbf{Y}}_i^{(s)}$ and $\hat{\mathbf{Y}}_j^{(t)}$ up to noise. If two exchangeable nodes behave consistently at a fixed time ($i \neq j, s = t$), we say the embedding has spatial stability. If a node behaves consistently between two exchangeable time points ($i = j, s \neq t$), we say the embedding has temporal stability.

The following lemma shows that AUASE has both spatial and temporal stability.

Lemma 1 (AUASE stability)**.** *Given exchangeable node/time pairs* (i, s) *and* (j, t), *the embeddings* $(\hat{\mathbf{Y}}_{\mathbf{A}}^{(s)})_i$ *and* $(\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)})_j$ *are asymptotically equal,*

$$\|(\hat{\mathbf{Y}}_{\mathbf{A}}^{(s)})_i - (\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)})_j\| = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1, 1/\beta)}{T^{1/4}\rho^{1/2}n^{1/2}}\right).$$

Spatial and temporal stability is crucial for downstream machine-learning tasks. Training a predictive model using embedding data is only meaningful if the embedding space

has a consistent interpretation. Sections 3.3, 3.5, and 3.6 demonstrate problems when using dynamic embedding techniques that do not have spatial or temporal stability. To the best of our knowledge, AUASE is the only existing attributed dynamic embedding which satisfies both these desirable stability properties.

2.5 PARAMETER SELECTION

The AUASE procedure detailed in Algorithm 1 requires two additional input parameters; the embedding dimension d and the attributed adjacency matrix weight α. This section outlines approaches for selecting these parameters.

In the theory it is assumed that the embedding dimension d is known, however in practice it will need to be estimated. For a given α, the truncated SVD of $\mathbf{A}_C$ is computed up to some maximum embedding dimension and d is determined by a singular value threshold, for example, using profile likelihood Zhu and Ghodsi [2006] or ScreeNOT Donoho et al. [2023].

The hyperparameter α balances the contributions of the network and covariates in the AUASE procedure. The truncated SVD gives the best low-rank approximation of $\mathbf{A}_C$ with respect to the Frobenius norm. By increasing α, the truncated SVD is incentivised to minimise the squared error terms corresponding to the scaled covariate matrices in $\mathbf{A}_C$.

For a downstream task, such as node classification, cross-validation can be used to choose α. For unsupervised learning, one can compare embedding quality metrics [Tsitsulin et al., 2023]. We show that practical performance is robust to the choice of α in Section 3.7.

3 EXPERIMENTS

3.1 METHOD COMPARISON

We compare AUASE to the following baselines:

- DRLAN [Liu et al., 2020]: Efficient framework incorporating an offline and an online network embedding model. Hyperparameter β weights network features and attribute contribution.
- DySAT [Sankar et al., 2020]: Learns node embeddings through self-attention layers and temporal dynamics.
- GloDyNE [Hou et al., 2020]: Dynamic but not attributed, and updates node embeddings for a selected subset of nodes that accumulate the largest topological changes over subsequent network snapshots.
- DyRep [Trivedi et al., 2019]: Inductive GNN-based dynamic graph embedding method, it supports (non-time-varying) node attributes.
- CONN [Tan et al., 2023]: Attributed but static GNN that explicitly exploits node attributes.

To the best of our knowledge, DRLAN and DySAT are the only unsupervised time-varying attributed dynamic embeddings with available source code, hence directly comparable to AUASE. The motivations for the exclusion of other comparable methods are in Appendix D.2.

For the sake of completeness, we also include a dynamic but not attributed method (GloDyNE), a dynamic embedding with non-time-varying attributes (DyRep) and a static method (CONN). However, these methods are not directly comparable with AUASE and, therefore, are not suited for tasks which require a stable dynamic attributed embedding on data with time-varying attributes.

3.2 IMPLEMENTATION AND EFFICIENCY

Both AUASE and UASE can be easily implemented with the *pyemb* Python package[1]. The code for the reproducing experiments is available at this GitHub repository [2].

The computational times to compute the dynamic embeddings of four large real-data networks for each baseline are reported in Table 2. The runtime of AUASE implemented with the *pyemb* Python package is orders of magnitude faster than competing methods, showing that AUASE is highly efficient for large graphs.

We do not provide a comprehensive time and space complexity analysis as AUASE is essentially a sparse truncated SVD whose space and time complexity has been extensively studied. The space complexity in the sparsest regime we consider is $O(Tn\log^k(n))$.

The truncated SVD of $\mathbf{A}_C$ can be computed using the Augmented Implicitly Restarted Lanczos Bidiagonalization algorithm [Baglama and Reichel, 2005] implemented in the *irlba* package in R, or the *irlbpy* in Python. Although the exact time complexity of this algorithm has not been studied theoretically, it is known that the time complexity of more general Lanczos-type algorithms for computing the rank truncated SVD is $O(Nd)$ [Tomás et al., 2023], in the sparsest regime we consider, this is $O(Tn\log^k(n)d)$. The author of the *irlba* package demonstrates its speed by performing a simulated experiment in which they compute the first 2 singular vectors of a sparse 10M x 10M matrix with 1M non-zero entries which takes approximately 6 seconds on a computer with two Intel Xeon CPU E5-2650 processors (16 physical CPU cores) running at 2 GHz equipped with 128 GB of ECC DDR3 RAM [3].

[1] https://pyemb.github.io/pyemb/html/index.html

[2] https://github.com/emmaceccherini/AUASE

[3] https://bwlewis.github.io/irlba/comparison.html

3.3 SIMULATED EXAMPLE

To illustrate the properties of AUASE we simulate a sequence of attributed networks using a dynamic stochastic block model with three communities with time-varying node attributes.

We assume three possible behaviours denoted $Z_i^{(t)} \in \{1, 2, 3\}$ and define the attributed dynamic latent position model in Definition 3,

$$\mathbf{A}_{ij}^{(t)} \mid Z_i^{(t)}, Z_j^{(t)} \overset{\text{ind}}{\sim} \text{Bernoulli}(\mathbf{B}_{Z_i^{(t)}, Z_j^{(t)}}),$$
$$\mathbf{C}_{i\ell}^{(t)} \mid Z_i^{(t)} \overset{\text{ind}}{\sim} \text{Normal}(\mathbf{D}_{Z_i^{(t)}, \ell}, \mathbf{I}),$$

where the latent position indexes into the community edge probability matrix $\mathbf{B} \in \mathbb{R}^{3\times 3}$ and the mean attribute matrix $\mathbf{D} \in \mathbb{R}^{3\times p}$,

$$\mathbf{B} = \begin{pmatrix} p_1 & p_0 & p_0 \\ p_0 & p_0 & p_0 \\ p_0 & p_0 & p_0 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}.$$

Note that in this experiment it is impossible to distinguish $Z_i^{(t)} = 2$ and $Z_i^{(t)} = 3$ using the adjacency matrices alone. Likewise, it is impossible to distinguish between $Z_i^{(t)} = 1$ and $Z_i^{(t)} = 2$ using the covariate matrices alone. Both are required to determine all three types of behaviour.

We construct the attributed dynamic latent position model with $n = 1000$ nodes and $p = 150$ covariates over $T = 10$ time points. Nodes are assigned with equal probability to three communities, depicted by the colours in Figure 2a, corresponding to a trajectory of local behaviours $(Z_i^{(1)}, \ldots, Z_i^{(T)})$, represented by the shapes in Figure 2a. The edge probabilities are $p_0 = 0.5$, $p_1 = 0.3$ and mean attributes $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^p$ are detailed in Appendix C.

Figure 2b demonstrates the behaviour of the three communities we would like the dynamic embedding techniques to recover, shown through the noise-free embedding of $\mathbf{P}_C = \mathbb{E}(\mathbf{A}_C)$. For instance, for $t \in \{4, 5\}$, all three communities have $Z_i^{(t)} = 2$ (star), so the embeddings are the same due to spatial stability. After $t = 7$ community $Z_i = 1$ (blue) switches from latent position $Z_i^{(t)} = 2$ (star) to $Z_i^{(t)} = 1$ (circle), so the embedding returns to its previous position due to temporal stability.

We compute the dynamic embedding using UASE, AUASE with $\alpha = 0.2$, DRLAN and DySAT into $d = 3$ dimensions, the known number of communities. Further details and experiments regarding the choice of these parameters and specific algorithm hyperparameters can be found in Appendix C.

Figures 2c-2f show the dynamic network embedding for these four techniques. For visualisation, we reduce the three-dimensional embeddings to one dimension with UMAP McInnes et al. [2020]. The solid line shows the mean embedding for each community with a 90% confidence interval.

Only AUASE is able to fully recover the underlying structure of the attributed network, displaying both spatial and temporal stability as predicted by the theory. UASE cannot detect the differences between communities 2 and 3, as we expected, as these only differ in their covariates, which is not considered by the technique.

DRLAN captures the dynamic patterns from both edge and attribute information, although it is unable to produce a sensible embedding at time $t = 0$ and has wide confidence intervals. The stability is also very sensitive to the choice of hyperparameter (see Appendix C.3). DySAT is clearly unstable, for instance, the switching of embeddings between times $t = 7$ and $t = 8$. Moreover, the embeddings of communities 2 and 3 appear merged for the whole time frame as observed with UASE. This occurs because DySAT, like most existing neural network methods, only incorporates node attributes as input to the first layer [Dwivedi et al., 2022].

3.4 DATASETS

As discussed in Section 2.4, stability is a desirable intrinsic property of a dynamic embedding essential for any downstream analysis. To provide an objective measure of its value, we evaluate the learned node representations on the downstream tasks node classification and link prediction for the following four attributed dynamic networks. Further details about the construction of the adjacency matrices $\mathbf{A}^{(t)}$ and covariate matrices $\mathbf{C}^{(t)}$ are given in Appendix D.

- DBLP[4] [Ley, 2009]: Co-authorship network consisting of bibliography data from computer science publications. Nodes represent authors with edges denoting co-authorship. Attributes are derived from an author's abstracts published in each time interval. For each interval an author has one of seven labels based on their most frequent publication area.
- ACM[4]: Co-authorship network similar to DBLP.
- Epinions[5] [Tang et al., 2013]: Trust network among users of a social network for sharing reviews. Nodes represent users with edges denoting a mutual trust relationship between users. Attributes are derived from users' reviews in each time interval. For each interval, a user has one of 22 labels based on their frequent review category.
- ogbn-mag [6] [Wang et al., 2020]: Co-authorship network from the Microsoft Academic Graph. We construct 10 adjacency matrices and attributes matrices

[4] https://www.aminer.cn/citation
[5] http://www.epinions.com
[6] https://ogb.stanford.edu/docs/nodeprop

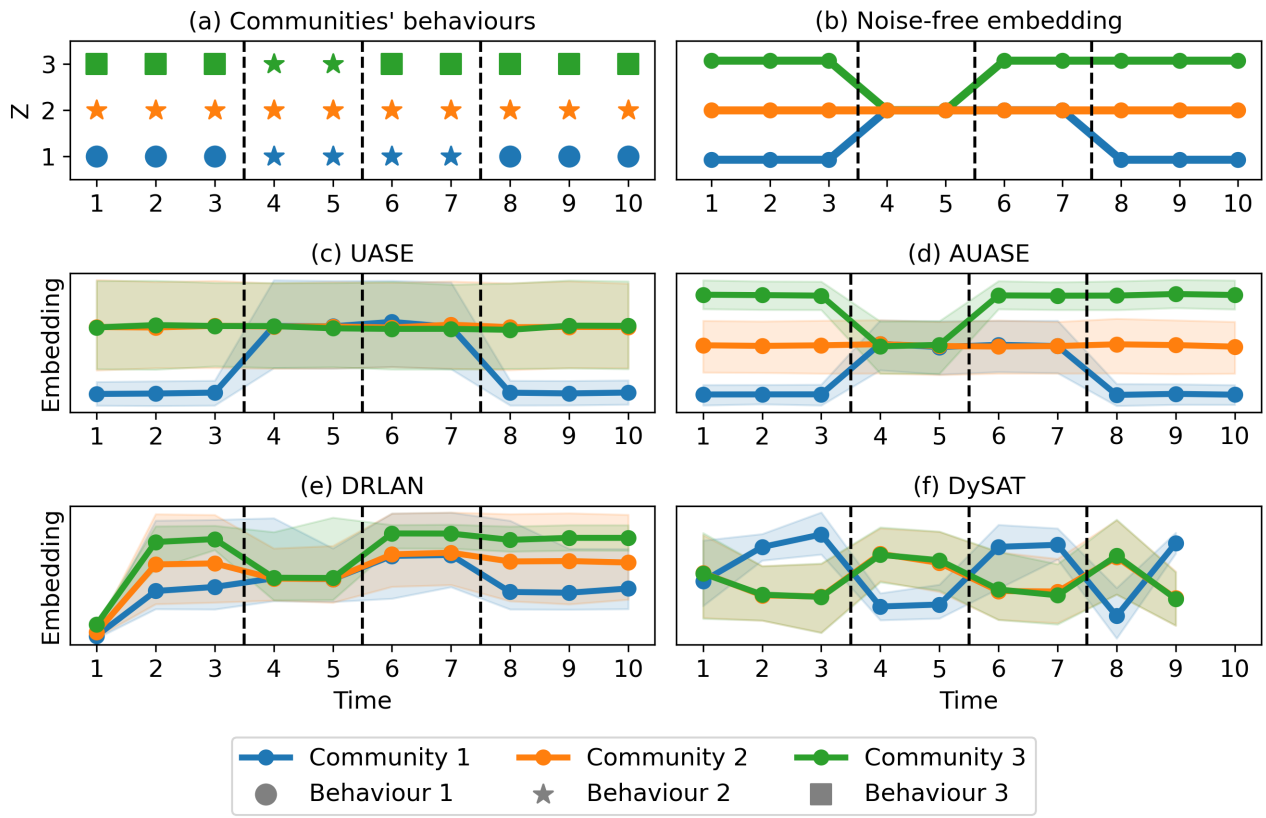

Figure 2: (a) Representation of global community (colour) and local behaviours (shape) of the synthetic attributed dynamic network. (b)-(f) One-dimensional UMAP visualisation of different node embedding techniques. The coloured lines show the mean embedding for each community with a 90% confidence interval. The y-axis should only be interpreted as an illustration of community separation, rotated to best align the ground truth.

where authors are the nodes. The tasks we perform are at author-level not paper-level (usual for ogbn-mag) which creates much harder tasks (see Appendix D for more details).

Dataset statistics are shown in Table 1.

Table 1: Attributed dynamic network statistics for the datasets showing the number of nodes n, the number of covariates p, the number of time intervals T, and the number of node labels K.

Datasets	Nodes - n	Attributes - p	Intervals - T	Labels - K
DBLP	10092	4291	9	7
ACM	34339	6489	15	2
Epinions	15851	7726	11	22
ogbn-mag	479979	128	10	349

The embedding dimension is selected using ScreeNOT [Zhu and Ghodsi, 2006] giving $d = 15$, $d = 29$ and $d = 22$ for DBLP, ACM and Epinions, respectively. The one exception is DySAT on the ACM dataset where $d = 32$. For ogbn-mag we choose $d = 50$ based on the scree plot. We use degree correction for AUASE and UASE embeddings [Passino et al., 2022].

The number of nodes in ogbn-mag is an order of magnitude larger than in the other datasets, therefore only UASE, AUASE and DRLAN are computationally feasible. More details about the computational barriers of DySAT, GloDyNE, DyRep and CONN are given in Appendix D.

Figure 1 shows the dynamic embedding using AUASE, DySAT and GloDyNE for the DBLP dataset for two consecutive time intervals. We reduce each embedding into two-dimensions using t-SNE Van der Maaten and Hinton [2008] jointly over all time intervals to preserve stability. The DySAT and GloDyNE embeddings find community structures, but they appear to have moved over time showing a lack of temporal stability. Conversely, AUASE shows both types of stability, for instance, authors in Computer Architecture and Computer Theory staying approximately in the same position over time.

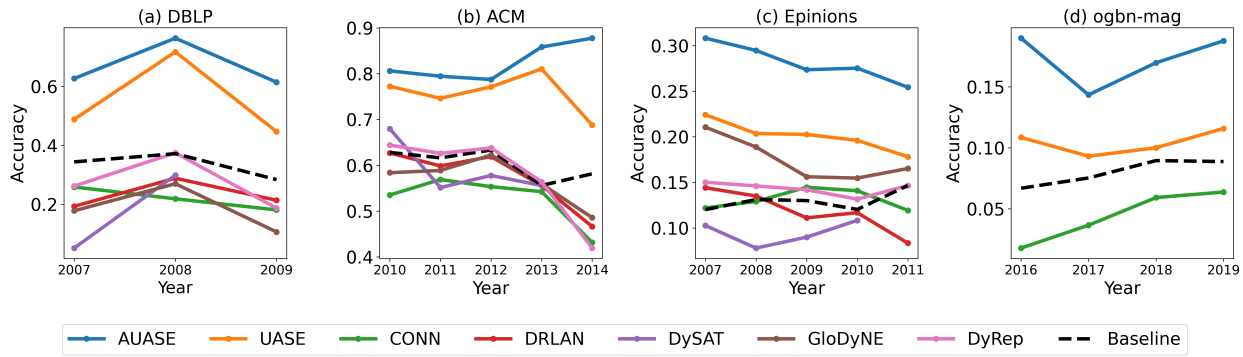

Figure 3: Node label classification accuracies for DBLP, ACM, Epinions and ogbn-mag datasets for each time interval in the corresponding test datasets. The dynamic embedding techniques outlined in Section 3.1 are solid lines, and different colors represents different methods. A basic baseline predicting the most common label is dashed.

3.5 NODE CLASSIFICATION

Node classification aims to predict label categories for current nodes using the historical node embedding data. We train a classifier using XGBoost [Chen and Guestrin, 2016] on the first 65% of time intervals for all the datasets and test on the remaining data. We reserve 10% of the training data for validation to select the weight hyperparameters via cross-validation for AUASE, DRLAN and CONN. Other hyperparameters are set to the default values provided by the original authors.

Figure 3 reports the classification accuracy for each test year, while Table 2 reports the mean accuracy over all test data. Other metrics were considered and gave equivalent results (see Appendix D, Tables 4, 5 and 6). Because label prediction requires predicting current embeddings, the stability illustrated in Figure 1 allows AUASE to dramatically outperform baselines in terms of accuracy - 40% over DyRep on DBLP, 23% over DySAT on ACM, 11% over GloDyNE on Epinions and 12% over DRLAN on ogbn-mag. AUASE also performs noticeably better than UASE, suggesting that the inclusion of covariate information is improving the quality of the dynamic embedding. Stability is clearly driving this performance increase, as Figure 3 shows that no unstable dynamic method consistently outperforms the baseline classifier of predicting the most common class.

3.6 LINK PREDICTION

For link prediction, the goal is to predict whether two nodes form an edge in the current network, given historical embeddings. Training data consists of two node embeddings from the same time interval,

$$\left((\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)})_i, (\hat{\mathbf{Y}}_{\mathbf{A}}^{(t)})_j \right).$$

These are assigned a positive label if there exists an edge between the two nodes at time $t+1$, namely, $\mathbf{A}_{ij}^{(t+1)} = 1$, otherwise, they are assigned a negative label, balanced so the classes have the same size. We train a binary classifier using XGBoost [Chen and Guestrin, 2016] on the first $T-2$ intervals for all the datasets, test on the remaining data and repeat the sampling 10 times.

Table 2 reports the area under the receiver operating characteristic curve which lies in the interval $[0, 1]$ with higher values indicating better prediction. Stability allows using more history in the classifier, which helps AUASE outperform nearly all of its competitors. However, there is still strong information about edge probabilities in unstable embeddings, allowing GloDyNE to perform marginally better for the ACM dataset. There is less distinction between AUASE and UASE which suggests that covariate information is not as beneficial for this task on these data.

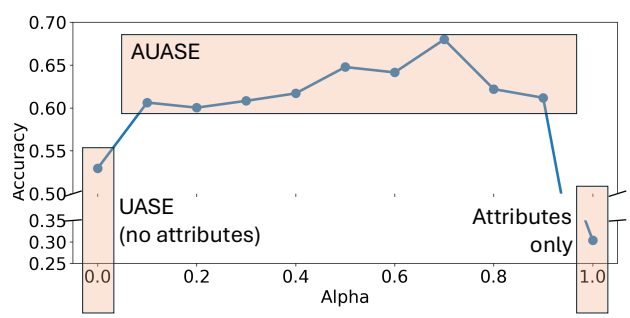

Figure 4: AUASE is not sensitive to the attribute weight hyperparameter α as shown by accuracy on DBLP dataset.

3.7 PARAMETER SENSITIVITY

In an unsupervised setting, there are many proposed heuristics for hyperparameter estimation, e.g., Tsitsulin et al. [2023], which might result in different choices of α in AUASE. Importantly, the *information* contained in the em-

Table 2: AUC for link prediction, average accuracy for node classification and time to compute the dynamic embedding for each baseline. Errors show the 90% CI, the best performance is **bold**, second underlined.

Method	Task	DBLP	ACM	Epinions	ogbn-mag
CONN	Link Pred.	0.741 ± 0.007	0.771 ± 0.008	0.602 ± 0.072	-
	Node Class.	0.219	0.526	0.131	-
	Time	$\approx$1h	$\approx$1h	$\approx$1h	-
GloDyNE	Link Pred.	0.883 ± 0.004	$\mathbf{0.907} \pm 0.005$	0.729 ± 0.016	-
	Node Class.	0.185	0.568	0.175	-
	Time	$\approx$10m	$\approx$1h	$\approx$1h	-
DRLAN	Link Pred.	0.575 ± 0.018	0.641 ± 0.015	0.626 ± 0.081	$\underline{0.939} \pm 0.009$
	Node Class.	0.232	0.573	0.118	0.044
	Time	$\approx$1s	$\approx$1s	$\approx$1s	$\approx$ 10m
DyRep	Link Pred.	0.511 ± 0.014	0.552 ± 0.010	0.548 ± 0.022	-
	Node Class.	0.272	0.578	0.143	-
	Time	$\approx$5h	$\approx$40h	$\approx$48h	-
DySAT	Link Pred.	0.802 ± 0.018	0.758 ± 0.023	0.596 ± 0.107	-
	Node Class.	0.175	0.591	0.095	-
	Time	$\approx$1h	$\approx$5h	$\approx$3h	-
UASE	Link Pred.	$\underline{0.907} \pm 0.003$	$\underline{0.896} \pm 0.003$	$\underline{0.806} \pm 0.009$	0.911 ± 0.002
	Node Class.	<u>0.550</u>	<u>0.758</u>	<u>0.201</u>	<u>0.140</u>
	Time	$\approx$1s	$\approx$5s	$\approx$1s	$\approx$30m
AUASE	Link Pred.	$\mathbf{0.915} \pm 0.004$	$\underline{0.896} \pm 0.005$	$\mathbf{0.809} \pm 0.009$	$\mathbf{0.954} \pm 0.003$
	Node Class.	**0.668**	**0.825**	**0.281**	**0.173**
	Time	$\approx$5s	$\approx$5s	$\approx$5s	$\approx$30m

bedding is not sensitive to this choice. To demonstrate this, Figure 4 presents the prediction accuracy for node classification on AUASE embeddings of the DBLP dataset.

AUASE performs better than competing methods shown in Figure 3a for any $\alpha \in [0.1, 0.9]$ for which attributes and network are jointly embedded, and fine-tuning increases accuracy by 15% compared to UASE. Since AUASE is computationally efficient, a fast exploration of different hyperparameter choices is possible. An equivalent analysis on link prediction gives similar results (see Appendix D, Table 3).

4 CONCLUSION

By combining provable stability over time with attributed embedding, we have shown that AUASE learns intrinsically 'better' dynamic attributed embeddings that solve problems that all previous (i.e., unstable) methods cannot. This empirically improves predictive performance for node classification and link prediction problems, it is expected to help with unsupervised tasks and aids interpretability.

A limitation of our method is the choice of embedding dimension and of the weight hyperparameter to balance the contribution of network features and attributes. While these are common challenges for unsupervised methods, we have a solution for supervised tasks and demonstrated AUASE robustness to the choice of α.

Future research includes exploring non-linear solutions to attributed dynamic embeddings exploiting AUASE stability properties [Davis et al., 2023]. Extending AUASE to deal with continuous-time network data [Modell et al., 2024] could also be valuable.

Acknowledgements

Emma Ceccherini gratefully acknowledges support by the Centre for Doctoral Training in Computational Statistics and Data Science (Compass, EPSRC Grant number EP/S023569/1). This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - `http://www.bris.ac.uk/acrc/`. Andrew Jones gratefully acknowledges support by the NeST EPSRC programme grant EP/X002195/1.

References

Nourhan Ahmed, Ahmed Rashed, and Lars Schmidt-Thieme. Learning attentive attribute-aware node embeddings in dynamic environments. *International Journal of Data Science and Analytics*, 17(2):189–201, 2024.

James Baglama and Lothar Reichel. Augmented implicitly restarted lanczos bidiagonalization methods. *SIAM Journal on Scientific Computing*, 27(1):19–42, 2005.

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

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

Ed Davis, Ian Gallagher, Daniel John Lawson, and Patrick Rubin-Delanchy. A simple and powerful framework for stable dynamic network embedding, 2023. URL `https://arxiv.org/abs/2311.09251`.

David Donoho, Matan Gavish, and Elad Romanov. Screenot: Exact mse-optimal singular value thresholding in correlated noise. *The Annals of Statistics*, 51(1):122–148, 2023.

Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks, 2022. URL `https://arxiv.org/abs/2003.00982`.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023.

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.

Ian Gallagher, Andrew Jones, Anna Bertiger, Carey E Priebe, and Patrick Rubin-Delanchy. Spectral embedding of weighted graphs. *Journal of the American Statistical Association*, 119(547):1923–1932, 2024.

Hongchang Gao and Heng Huang. Deep attributed network embedding. In *Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI))*, 2018.

Palash Goyal, Nitin Kamra, Xinran He, and Yan Liu. DynGEM: Deep embedding method for dynamic graphs. *arXiv preprint arXiv:1805.11273*, 2018.

Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 187:104816, 2020.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

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

Chengbin Hou, Han Zhang, Shan He, and Ke Tang. GloDyNE: Global topology preserving dynamic network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 2020. doi: 10.1109/TKDE.2020.3046511.

Svante Janson. Probability asymptotics: notes on notation. *arXiv preprint arXiv:1108.3924*, 2011.

Andrew Jones and Patrick Rubin-Delanchy. The multilayer random dot product graph, 2021. URL `https://arxiv.org/abs/2007.10455`.

Bomin Kim, Kevin Lee, Lingzhou Xue, and Xiaoyue Niu. A review of dynamic network models with latent variables, 2018. URL `https://arxiv.org/abs/1711.10421`.

Cecilia Klein, Andrea Marino, Marie-France Sagot, Paulo Vieira Milreu, and Matteo Brilli. Structural and dynamical analysis of biological networks. *Briefings in functional genomics*, 11(6):420–433, 2012.

Nam H Lee and Carey E Priebe. A latent process model for time series of attributed random graphs. *Statistical inference for stochastic processes*, 14:231–253, 2011.

Michael Ley. DBLP: some lessons learned. *Proceedings of the VLDB Endowment*, 2(2):1493–1500, 2009.

Jundong Li, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu. Attributed network embedding for learning in a dynamic environment. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 387–396, 2017.

Zhao Li, Xin Wang, Jianxin Li, and Qingpeng Zhang. Deep attributed network representation learning of complex coupling and interaction. *Knowledge-Based Systems*, 212:106618, 2021.

Zhijun Liu, Chao Huang, Yanwei Yu, Peng Song, Baode Fan, and Junyu Dong. Dynamic representation learning for large-scale attributed networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1005–1014, 2020.

Zhijun Liu, Chao Huang, Yanwei Yu, and Junyu Dong. Motif-preserving dynamic attributed network embedding. In *Proceedings of the Web Conference 2021*, pages 1629–1638, 2021.

Xie Luodi, Hui Tian, and Hong Shen. Learning dynamic embeddings for temporal attributed networks. *Knowledge-Based Systems*, 286:111308, 2024.

Yao Ma, Ziyi Guo, Zhaocun Ren, Jiliang Tang, and Dawei Yin. Streaming graph neural networks. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 719–728, 2020.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction, 2020. URL `https://arxiv.org/abs/1802.03426`.

Xian Mo, Jun Pang, and Zhiming Liu. Deep autoencoder architecture with outliers for temporal attributed network embedding. *Expert Systems with Applications*, 240: 122596, 2024.

Alexander Modell, Ian Gallagher, Emma Ceccherini, Nick Whiteley, and Patrick Rubin-Delanchy. Intensity profile projection: A framework for continuous-time representation learning for dynamic networks. *Advances in Neural Information Processing Systems*, 36, 2024.

Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446 (7136):664–667, April 2007. ISSN 1476-4687. doi: 10.1038/nature05670. URL `http://dx.doi.org/10.1038/nature05670`.

Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 5363–5370, 2020.

Francesco Sanna Passino, Nicholas A Heard, and Patrick Rubin-Delanchy. Spectral clustering on spherical coordinates under the degree-corrected stochastic blockmodel. *Technometrics*, 64(3):346–357, 2022.

Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, and Nagiza F Samatova. Anomaly detection in dynamic networks: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3): 223–247, 2015.

Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2020.

Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the 13th international conference on web search and data mining*, pages 519–527, 2020.

Daniel K Sewell and Yuguo Chen. Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657, 2015.

Xiaofei Sun, Jiang Guo, Xiao Ding, and Ting Liu. A general framework for content-enhanced network representation learning. *arXiv preprint arXiv:1610.02906*, 2016.

Qiaoyu Tan, Xin Zhang, Xiao Huang, Hao Chen, Jundong Li, and Xia Hu. Collaborative graph neural networks for attributed network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. Exploiting homophily effect for trust prediction. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 53–62, 2013.

Shaowei Tang, Zaiqiao Meng, and Shangsong Liang. Dynamic co-embedding model for temporal attributed networks. *IEEE Transactions on Neural Networks and Learning Systems*, 35(3):3488–3502, 2022.

Andrés E Tomás, Enrique S Quintana-Orti, and Hartwig Anzt. Fast truncated svd of sparse and dense matrices on graphics processors. *The International Journal of High Performance Computing Applications*, 37(3-4):380–393, 2023.

Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. DyRep: Learning representations over dynamic graphs. In *International conference on learning representations*, 2019.

J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

Anton Tsitsulin, Marina Munkhoeva, and Bryan Perozzi. Unsupervised embedding quality evaluation. In *Topological, Algebraic and Geometric Learning Workshops 2023*, pages 169–188. PMLR, 2023.

Cunchao Tu, Han Liu, Zhiyuan Liu, and Maosong Sun. Cane: Context-aware network embedding for relation modeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1722–1731, 2017.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11), 2008.

Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1):396–413, 2020.

Hao Wei, Guyu Hu, Wei Bai, Shiming Xia, and Zhisong Pan. Lifelong representation learning in dynamic attributed networks. *Neurocomputing*, 358:1–9, 2019.

Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs. *arXiv preprint arXiv:2002.07962*, 2020a.

Zenan Xu, Zijing Ou, Qinliang Su, Jianxing Yu, Xiaojun Quan, and Zhenkun Lin. Embedding dynamic attributed networks by modeling the evolution processes. *arXiv preprint arXiv:2010.14047*, 2020b.

Guotong Xue, Ming Zhong, Jianxin Li, Jia Chen, Chengshuai Zhai, and Ruochen Kong. Dynamic network embedding survey. *Neurocomputing*, 472:212–223, 2022.

Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information. In *IJCAI*, volume 2015, pages 2111–2117, 2015.

Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102: 315–323, 2015.

Xiang Yue, Zhen Wang, Jingong Huang, Srinivasan Parthasarathy, Soheil Moosavinasab, Yungui Huang, Simon M Lin, Wen Zhang, Ping Zhang, and Huan Sun. Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics*, 36(4):1241–1251, 2020.

Zhen Zhang, Hongxia Yang, Jiajun Bu, Sheng Zhou, Pinggang Yu, Jianwei Zhang, Martin Ester, and Can Wang. Anrl: attributed network representation learning via deep neural networks. In *Ijcai*, volume 18, pages 3155–3161, 2018.

Lekui Zhou, Yang Yang, Xiang Ren, Fei Wu, and Yueting Zhuang. Dynamic network embedding by modeling triadic closure process. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

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

Yuan Zuo, Guannan Liu, Hao Lin, Jia Guo, Xiaoqian Hu, and Junjie Wu. Embedding temporal network via neighborhood formation. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2857–2866, 2018.

Supplementary Material

Emma Ceccherini[1] **Ian Gallagher**[2] **Andrew Jones**[3] **Daniel Lawson**[1]

[1]University of Bristol, U.K.
[2]The University of Melbourne, Australia
[3]University of Edinburgh, U.K.

A PROOF OF THEOREM 1

Note that row and column permutations of a matrix leave its singular values unchanged, while simply permuting the entries of its left- and right-singular vectors. Thus, for ease of exposition, we may assume without loss of generality that our matrices $\mathbf{A}_C$ and $\mathbf{P}_C$ take the form

$$\mathbf{A}_C = \begin{pmatrix} (1-\alpha)\mathbf{A} & \alpha\mathbf{C} \\ \alpha\mathbf{C}_* & \mathbf{0}_{p\times Tp} \end{pmatrix} \text{ and } \mathbf{P}_C = \begin{pmatrix} (1-\alpha)\mathbf{P} & \alpha\mathbf{D} \\ \alpha\mathbf{D}_* & \mathbf{0}_{p\times Tp} \end{pmatrix},$$

where $\mathbf{C} \in \mathbb{R}^{n\times Tp}$ and $\mathbf{C}_* \in \mathbb{R}^{p\times Tn}$ are defined by

$$\mathbf{C} = \left(\mathbf{C}^{(1)} \mid \ldots \mid \mathbf{C}^{(T)}\right) \text{ and } \mathbf{C}_* = \left(\mathbf{C}^{(1)\top} \mid \ldots \mid \mathbf{C}^{(T)\top}\right),$$

and $\mathbf{D}$ and $\mathbf{D}_*$ are the expectation matrices of $\mathbf{C}$ and $\mathbf{C}_*$ respectively.

Proposition 1. $\|\mathbf{A}_C - \mathbf{P}_C\| = \mathrm{O}_{\mathbb{P}}\left(T^{1/2}\rho^{1/2}n^{1/2}\log^{1/2}(n)\, r_\alpha(1, 1/\beta)\right)$.

Proof. Condition on a choice of latent positions. A standard application of the triangle inequality then tells us that

$$\|\mathbf{A}_C - \mathbf{P}_C\| \leq (1-\alpha)\|\mathbf{A} - \mathbf{P}\| + \alpha \max\left\{\|\mathbf{C} - \mathbf{D}\|, \|\mathbf{C}_* - \mathbf{D}_*\|\right\}.$$

By Proposition 8 of Jones and Rubin-Delanchy [2021] we know that $\|\mathbf{A} - \mathbf{P}\| = \mathrm{O}_{\mathbb{P}}\left(\rho^{1/2}T^{1/2}n^{1/2}\log^{1/2}(n)\right)$, so it suffices to find a bound for the spectral norm of the centred covariate matrices.

We use the following notation:

- Let $\mathbf{E}_{il}^{(t)}$ and $\mathbf{F}_i$ denote the $n \times Tp$ (respectively $(n+Tp)\times(n+Tp)$) matrix with $(i, (t-1)p+l)$th (respectively (i,i)th) entry equal to 1 and all other entries equal to 0.
- For any matrix $\mathbf{M} \in \mathbb{R}^{n\times Tp}$, the symmetric dilation of $\mathbf{M}$ is the matrix $\mathcal{S}(\mathbf{M}) \in \mathbb{R}^{(n+Tp)\times(n+Tp)}$ given by

$$\mathcal{S}(\mathbf{M}) = \begin{pmatrix} \mathbf{0}_{n\times n} & \mathbf{M} \\ \mathbf{M}^\top & \mathbf{0}_{Tp\times Tp} \end{pmatrix}.$$

Note that

$$\mathcal{S}(\mathbf{E}_{il}^{(t)})^k = \begin{cases} \mathbf{F}_i + \mathbf{F}_{n+(t-1)p+l} & k \text{ even} \\ \mathcal{S}(\mathbf{E}_{il}^{(t)}) & k \text{ odd} \end{cases}$$

for any $i \in [n]$ and $l \in [p]$, and that the matrix $\mathbf{F}_i + \mathbf{F}_{n+(t-1)p+l} - \mathcal{S}(\mathbf{E}_{il}^{(t)})$ is positive semi-definite, since for any $\mathbf{x} \in \mathbb{R}^{n+Tp}$ we have

$$\mathbf{x}^\top[\mathbf{F}_i + \mathbf{F}_{n+(t-1)p+l} - \mathcal{S}(\mathbf{E}_{il}^{(t)})]\mathbf{x} = (\mathbf{x}_i - \mathbf{x}_{n+(t-1)p+l})^2 \geq 0.$$

Thus in particular $\mathcal{S}(\mathbf{E}_{il}^{(t)})^k \preccurlyeq \mathbf{F}_i + \mathbf{F}_{n+(t-1)p+l}$ for all k, where $\preccurlyeq$ denotes the standard partial order on positive semi-definite matrices.

Now, note that we can write

$$\mathcal{S}(\mathbf{C}-\mathbf{D}) = \sum_{i=1}^{n}\sum_{l=1}^{p}\left[\sum_{t=1}^{T}\rho^{1/2}(f_l(Z_i^{(t)}) - \mathbb{E}[f_l(Z_i^{(t)})])\mathcal{S}(\mathbf{E}_{il}^{(t)})\right]$$

where each of the bracketed terms in the sum is an independent matrix.

Let

$$\mathbf{M}_{il} = \sum_{t=1}^{T} f_l(Z_i^{(t)}) - \mathbb{E}[f_l(Z_i^{(t)})])\mathcal{S}(\mathbf{E}_{il}^{(t)}).$$

By Assumption 5 and comparison with the exponential distribution we see that

$$\mathbb{E}[\mathbf{M}_{il}^k] \preccurlyeq \frac{2k!}{\beta^k}\left[\sum_{t=1}^{T}\mathcal{S}(\mathbf{E}_{il}^{(t)})^2\right]$$

for any $k \geq 2$ (where we have used the fact that $\mathcal{S}(\mathbf{E}_{il}^{(s)})\mathcal{S}(\mathbf{E}_{il}^{(t)}) = 0$ if $s \neq t$) and thus we may apply a subexponential version of matrix Bernstein (Theorem 6.2 in Tropp [2012]) to find that

$$\mathbb{P}\left(\lambda_{\max}\left(\mathcal{S}(\mathbf{C}-\mathbf{D})\right) \geq t\right) \leq (n+p)\exp\left(\frac{-t^2/2\rho}{4\sigma^2/\beta^2 + t/\beta\rho^{1/2}}\right)$$

where

$$\sigma^2 = \left\|\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\sum_{t=1}^{T}\mathcal{S}(\mathbf{E}_{ij}^{(t)})^2\right]\right\| = \left\|\sum_{i=1}^{n}\sum_{l=1}^{p}\left[T\mathbf{F}_i + \sum_{t=1}^{T}\mathbf{F}_{n+(t-1)p+l}\right]\right\|.$$

We find that the sum over all $i \in [n]$ and $l \in [p]$ is a diagonal matrix whose first n entries are equal to Tp and remaining Tp entries are equal to n, and thus $\sigma^2 = n$, and so we deduce that

$$\lambda_{\max}\left(\mathcal{S}(\mathbf{C}-\mathbf{D})\right) = \mathrm{O}_{\mathbb{P}}\left(\tfrac{1}{\beta}\rho^{1/2}n^{1/2}\log^{1/2}(n)\right).$$

Since the spectral norm of any matrix is equal to the greatest eigenvalue of its symmetric dilation, we therefore conclude that

$$\|\mathbf{C}-\mathbf{D}\| = \mathrm{O}_{\mathbb{P}}\left(\tfrac{1}{\beta}\rho^{1/2}n^{1/2}\log^{1/2}(n)\right).$$

An identical argument shows that $\|\mathbf{C}_* - \mathbf{D}_*\| = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\beta}T^{1/2}\rho^{1/2}n^{1/2}\log^{1/2}(n)\right)$, from which the result follows.

Thus, taking a union bound and integrating over all possible sets of latent positions gives us our desired result. □

Corollary 1. *The non-zero singular values $\sigma_i(\mathbf{A}_C)$ satisfy*

$$\sigma_i(\mathbf{A}_C) = \Theta_{\mathbb{P}}\left(T^{1/2}\rho n\right)$$

for each $i \in [d]$.

Proof. A corollary of Weyl's inequalities (Horn and Johnson [2012], Corollary 7.3.5) states that $|\sigma_i(\mathbf{A}_C) - \sigma_i(\mathbf{P}_C)| \leq \|\mathbf{A}_C - \mathbf{P}_C\|$ for all i, and so in particular

$$\sigma_i(\mathbf{P}_C) - \|\mathbf{A}_C - \mathbf{P}_C\| \leq \sigma_i(\mathbf{A}_C) \leq \sigma_i(\mathbf{P}_C) + \|\mathbf{A}_C - \mathbf{P}_C\|$$

and the result follows directly from Assumption 2 and Proposition 1. □

Proposition 2. *The following bounds hold:*

(i) $\|\mathbf{U_A}\mathbf{U_A}^\top - \mathbf{U_P}\mathbf{U_P}^\top\|$, $\|\mathbf{V_A}\mathbf{V_A}^\top - \mathbf{V_P}\mathbf{V_P}^\top\| = \mathrm{O}_\mathbb{P}\left(\frac{\log^{1/2}(n)\, r_\alpha(1,1/\beta)}{\rho^{1/2} n^{1/2}}\right)$.

(ii) $\|\mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\|_{\mathrm{Frob}} = \mathrm{O}_\mathbb{P}\left(\log^{1/2}(n)\, r_\alpha(1, \rho^{1/2}/\beta)\right)$.

(iii) $\|\mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_A}\|_{\mathrm{Frob}}$, $\|\mathbf{U_A}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\|_{\mathrm{Frob}} = \mathrm{O}_\mathbb{P}\left(T^{1/2}\log(n)\, r_\alpha(1,1/\beta)^2\right)$.

(iv) $\|\mathbf{U_P}^\top\mathbf{U_A} - \mathbf{V_P}^\top\mathbf{V_A}\|_{\mathrm{Frob}} = \mathrm{O}_\mathbb{P}\left(\frac{\log(n)\, r_\alpha(1,1/\beta)^2}{\rho n}\right)$.

Proof.

(i) Denoting by $\theta_1, \ldots, \theta_d$ the principal angles between the subspaces spanned by the columns of $\mathbf{U_P}$ and $\mathbf{U_A}$, a variant of the Davis–Kahan theorem (Yu et al. [2015], Theorem 4) states that

$$\max_{i\in[d]} |\sin(\theta_i)| \leq \frac{2d^{1/2}\left(2\sigma_1(\mathbf{P}_C) + \|\mathbf{A}_C - \mathbf{P}_C\|\right)\|\mathbf{A}_C - \mathbf{P}_C\|}{\sigma_d(\mathbf{P}_C)^2}.$$

By definition, $\theta_i = \cos^{-1}(\sigma_i)$, where the σ_i are the singular values of the matrix $\mathbf{U_P}^\top\mathbf{U_A}$. A standard result states that the non-zero eigenvalues of the matrix $\mathbf{U_A}\mathbf{U_A}^\top - \mathbf{U_P}\mathbf{U_P}^\top$ are precisely the $\sin(\theta_i)$ (each occurring twice) and thus the result follows after applying the bounds from Assumption 2 and Proposition 1. An identical argument gives the result for $\|\mathbf{V_A}\mathbf{V_A}^\top - \mathbf{V_P}\mathbf{V_P}^\top\|$.

(ii) Condition on a choice of latent positions. For any $i, j \in [d]$, and for any $t \in [T]$, let u and $v^{(t)}$ denote the ith and jth columns of $\mathbf{U_P}$ and $\mathbf{V_P}^{(t)}$ respectively, so that

$$\left(\mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\right)_{ij} = (1-\alpha)(\mathbf{T}_1 + \mathbf{T}_2) + \alpha\mathbf{T}_3,$$

where

$$\begin{aligned}
\mathbf{T}_1 &= \sum_{t=1}^{T}\sum_{k=2}^{n}\sum_{l=1}^{k-1}(u_k v_l^{(t)} + u_l v_k^{(t)})\left(\mathbf{A}_{kl}^{(t)} - \mathbf{P}_{kl}^{(t)}\right)\\
\mathbf{T}_2 &= \sum_{t=1}^{T}\sum_{k=1}^{n} u_k v_k^{(t)}\left(\mathbf{A}_{kk}^{(t)} - \mathbf{P}_{kk}^{(t)}\right)\\
\mathbf{T}_3 &= \sum_{t=1}^{T}\sum_{k=1}^{n}\sum_{l=1}^{p}\left(u_k v_{n+l}^{(t)} + u_{n+l} v_k^{(t)}\right)\left(\mathbf{C}_{kl}^{(t)} - \mathbf{D}_{kl}^{(t)}\right).
\end{aligned}$$

The term $\mathbf{T}_2$ can be seen to be $\mathrm{O}_\mathbb{P}(\rho T)$, and so we can disregard it for the purposes of our analysis. The terms in the remaining two sums are independent zero-mean random variables, where the terms in $\mathbf{T}_1$ are bounded in absolute value by $|u_k v_l^{(t)} + u_l v_k^{(t)}|$ and so we can apply Hoeffding's inequality to find that $|\mathbf{T}_1| = \mathrm{O}_\mathbb{P}(\log^{1/2}(n))$ (noting that since u and v have norm at most 1, the sum of the terms $|u_k v_l^{(t)} + u_l v_k^{(t)}|^2$ is at most 2). Similarly, one can use the fact that the sum of the terms $|u_k v_{n+l}^{(t)} + u_{n+l} v_k^{(t)}|^2$ is at most 2 and apply Assumption 5 to find that $|\mathbf{T}_3| = \mathrm{O}_\mathbb{P}(\frac{1}{\beta}\rho^{1/2}\log^{1/2}(n))$, and so deduce that

$$\left|(\mathbf{U_P}^\top(\mathbf{A}_C^{(t)} - \mathbf{P}_C^{(t)})\mathbf{V_P})_{ij}\right| = \mathrm{O}_\mathbb{P}\left(\log^{1/2}(n)\, r_\alpha(1, \rho^{1/2}/\beta)\right).$$

The result is then obtained by taking a union bound over all i, j and integrating over all possible choices of latent positions.

(iii) Observe that, since $\mathbf{V_A}^\top\mathbf{V_A} = \mathbf{I}_d$, we may write

$$\mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_A} = \mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)(\mathbf{V_A}\mathbf{V_A}^\top - \mathbf{V_P}\mathbf{V_P}^\top)\mathbf{V_A} + \mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\mathbf{V_P}^\top\mathbf{V_A}.$$

These terms satisfy

$$\|\mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)(\mathbf{V_A}\mathbf{V_A}^\top - \mathbf{V_P}\mathbf{V_P}^\top)\mathbf{V_A}\|_{\mathrm{Frob}} = \mathrm{O}_\mathbb{P}\left(T^{1/2}\log(n)\, r_\alpha(1,1/\beta)^2\right)$$

and

$$\|\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{V_P}\mathbf{V_P}^\top\mathbf{V_A}\|_{\mathrm{Frob}} = \mathrm{O}_\mathbb{P}\left(\log^{1/2}(n)\, r_\alpha(1, \rho^{1/2}/\beta)\right)$$

where we have applied Proposition 1, the results from parts (i) and (ii), and the fact that

$$\|\mathbf{MN}\|_{\mathrm{Frob}} \leq \max\big\{\|\mathbf{M}\|\|\mathbf{N}\|_{\mathrm{Frob}}, \|\mathbf{N}\|\|\mathbf{M}\|_{\mathrm{Frob}}\big\}$$

for any commensurate matrices $\mathbf{M}$ and $\mathbf{N}$. The first of these two terms dominates, from which the result follows, and an identical argument bounds the term $\|\mathbf{U}_{\mathbf{A}}^\top(\mathbf{A}-\mathbf{P})\mathbf{V}_{\mathbf{P}}\|_{\mathrm{Frob}}$.

(iv) Note that

$$\mathbf{\Sigma}_{\mathbf{P}}(\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}}) + (\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}})\mathbf{\Sigma}_{\mathbf{A}} = \mathbf{U}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)^\top\mathbf{U}_{\mathbf{A}}$$

and for any $i, j \in [d]$ the (i,j)th entry of the left-hand matrix is equal to

$$\big(\sigma_i(\mathbf{P}_C) + \sigma_j(\mathbf{A}_C)\big)(\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}})_{ij}.$$

Thus

$$\begin{aligned}\big|(\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}})_{ij}\big|^2 &= \frac{\big|(\mathbf{U}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)^\top\mathbf{U}_{\mathbf{A}})_{ij}\big|^2}{\big(\sigma_i(\mathbf{P}_C) + \sigma_j(\mathbf{A}_C)\big)^2} \\ &\leq \frac{2\left(\big|(\mathbf{U}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_{\mathbf{A}})_{ij}\big|^2 + \big|(\mathbf{V}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)^\top\mathbf{U}_{\mathbf{A}})_{ij}\big|^2\right)}{\big(\sigma_d(\mathbf{P}_C) + \sigma_d(\mathbf{A}_C)\big)^2}\end{aligned}$$

and so

$$\big\|\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}}\big\|_{\mathrm{Frob}} \leq \frac{2\max\left\{\big\|\mathbf{U}_{\mathbf{P}}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_{\mathbf{A}}\big\|_{\mathrm{Frob}}, \big\|\mathbf{U}_{\mathbf{A}}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_{\mathbf{P}}\big\|_{\mathrm{Frob}}\right\}}{\sigma_d(\mathbf{P}_C) + \sigma_d(\mathbf{A}_C)}$$

by summing over all $i, j \in [d]$ and noting that the Frobenius norm is invariant under matrix transposition. The result then follows by applying part (iii), Proposition 1 and Corollary 1.

□

Proposition 3. *Let* $\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} + \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}}$ *admit the singular value decomposition*

$$\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} + \mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top,$$

and let $\mathbf{W} = \mathbf{U}\mathbf{V}^\top$. *Then*

$$\max\left\{\|\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{W}\|_{\mathrm{Frob}}, \|\mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}} - \mathbf{W}\|_{\mathrm{Frob}}\right\} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log(n)\, r_\alpha(1, 1/\beta)^2}{\rho n}\right).$$

Proof. A standard argument shows that $\mathbf{W}$ satisfies

$$\mathbf{W} = \operatorname*{arg\,min}_{\mathbf{Q}\in\mathrm{O}(d)} \|\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{Q}\|_{\mathrm{Frob}}^2 + \|\mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}} - \mathbf{Q}\|_{\mathrm{Frob}}.$$

Let $\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} = \mathbf{U}_*\mathbf{\Sigma}_*\mathbf{V}_*^\top$ be the singular value decomposition of $\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}}$, and define $\mathbf{W}_* \in \mathrm{O}(d)$ by $\mathbf{W}_* = \mathbf{U}_*\mathbf{V}_*^\top$. Then, denoting by $\sigma_1, \ldots, \sigma_d$ the singular values of $\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}}$ and defining $\theta_i = \cos^{-1}(\sigma_i)$ as in Proposition 2 (i), we see that

$$\|\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{W}_*\|_{\mathrm{Frob}} = \|\mathbf{\Sigma}_* - \mathbf{I}_d\|_{\mathrm{Frob}} \leq d^{1/2}\sin^2(\theta_d) \leq d^{1/2}\|\mathbf{U}_{\mathbf{A}}\mathbf{U}_{\mathbf{A}}^\top - \mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^\top\|^2$$

and so

$$\|\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{W}_*\|_{\mathrm{Frob}} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log(n)\, r_\alpha(1, 1/\beta)^2}{\rho n}\right).$$

Also,

$$\|\mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}} - \mathbf{W}_*\|_{\mathrm{Frob}} \leq \|\mathbf{V}_{\mathbf{P}}^\top\mathbf{V}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}}\|_{\mathrm{Frob}} + \|\mathbf{U}_{\mathbf{P}}^\top\mathbf{U}_{\mathbf{A}} - \mathbf{W}_*\|_{\mathrm{Frob}}$$

and so

$$\|\mathbf{V}_\mathbf{P}^\top \mathbf{V}_\mathbf{A} - \mathbf{W}_*\|_{\text{Frob}} = \mathrm{O}_\mathbb{P}\left(\frac{\log(n)\, r_\alpha(1, 1/\beta)^2}{\rho n}\right).$$

by Proposition 2, part (iv).

Since by definition

$$\|\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A} - \mathbf{W}\|_{\text{Frob}}^2 + \|\mathbf{V}_\mathbf{P}^\top \mathbf{V}_\mathbf{A} - \mathbf{W}\|_{\text{Frob}}^2 \leq \|\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A} - \mathbf{W}_*\|_{\text{Frob}}^2 + \|\mathbf{V}_\mathbf{P}^\top \mathbf{V}_\mathbf{A} - \mathbf{W}_*\|_{\text{Frob}}^2$$

the result follows. □

Proposition 4. *The following bounds hold:*

(i) $\|\mathbf{W}\mathbf{\Sigma}_\mathbf{A} - \mathbf{\Sigma}_\mathbf{P}\mathbf{W}\|_{\text{Frob}} = \mathrm{O}_\mathbb{P}\left(T^{1/2}\log(n)\, r_\alpha(1, 1/\beta)^2\right)$.

(ii) $\|\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W}\|_{\text{Frob}} = \mathrm{O}_\mathbb{P}\left(\frac{T^{1/4}\log(n)\, r_\alpha(1,1/\beta)^2}{\rho^{1/2} n^{1/2}}\right)$.

(iii) $\|\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} - \mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W}\|_{\text{Frob}} = \mathrm{O}_\mathbb{P}\left(\frac{\log(n)\, r_\alpha(1,1/\beta)^2}{\rho n}\right)$.

Proof.

(i) Observe that

$$\mathbf{W}\mathbf{\Sigma}_\mathbf{A} - \mathbf{\Sigma}_\mathbf{P}\mathbf{W} = (\mathbf{W} - \mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A})\mathbf{\Sigma}_\mathbf{A} + \mathbf{U}_\mathbf{P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_\mathbf{A} + \mathbf{\Sigma}_\mathbf{P}(\mathbf{V}_\mathbf{P}^\top \mathbf{V}_\mathbf{A} - \mathbf{W})$$

and so

$$\|\mathbf{W}\mathbf{\Sigma}_\mathbf{A} - \mathbf{\Sigma}_\mathbf{P}\mathbf{W}\|_{\text{Frob}} = \mathrm{O}_\mathbb{P}\left(T^{1/2}\log(n)\, r_\alpha(1, 1/\beta)^2\right)$$

by Assumption 2, Corollary 1 and Propositions 2 and 3.

(ii) Note that

$$(\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W})_{ij} = \frac{(\mathbf{W}\mathbf{\Sigma}_\mathbf{A} - \mathbf{\Sigma}_\mathbf{P}\mathbf{W})_{ij}}{\sigma_j(\mathbf{A})^{1/2} + \sigma_i(\mathbf{P})^{1/2}},$$

and so the result follows by applying part (i), Assumption 2, Corollary 1 and summing over all $i, j \in [d]$.

(iii) Note that

$$(\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} - \mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W})_{ij} = \frac{\left(\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W}\right)_{ij}}{\sigma_i(\mathbf{P})^{1/2}\,\sigma_j(\mathbf{A})^{1/2}}$$

and so the result follows as in part (ii).

□

Proposition 5. *Let*

$$\begin{aligned}
\mathbf{R}_{1,1} &= \mathbf{U}_\mathbf{P}(\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W}) \\
\mathbf{R}_{1,2} &= (\mathbf{I} - \mathbf{U}_\mathbf{P}\mathbf{U}_\mathbf{P}^\top)(\mathbf{A}_C - \mathbf{P}_C)(\mathbf{V}_\mathbf{A} - \mathbf{V}_\mathbf{P}\mathbf{W})\mathbf{\Sigma}_\mathbf{A}^{-1/2} \\
\mathbf{R}_{1,3} &= -\mathbf{U}_\mathbf{P}\mathbf{U}_\mathbf{P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_\mathbf{P}\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} \\
\mathbf{R}_{1,4} &= (\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_\mathbf{P}(\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} - \mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W})
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{R}_{2,1} &= \mathbf{V}_\mathbf{P}(\mathbf{V}_\mathbf{P}^\top \mathbf{V}_\mathbf{A}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W}) \\
\mathbf{R}_{2,2} &= (\mathbf{I} - \mathbf{V}_\mathbf{P}\mathbf{V}_\mathbf{P}^\top)(\mathbf{A}_C - \mathbf{P}_C)(\mathbf{U}_\mathbf{A} - \mathbf{U}_\mathbf{P}\mathbf{W})\mathbf{\Sigma}_\mathbf{A}^{-1/2} \\
\mathbf{R}_{2,3} &= -\mathbf{V}_\mathbf{P}\mathbf{V}_\mathbf{P}^\top(\mathbf{A}_C - \mathbf{P}_C)^\top\mathbf{U}_\mathbf{P}\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} \\
\mathbf{R}_{2,4} &= (\mathbf{A}_C - \mathbf{P}_C)^\top\mathbf{U}_\mathbf{P}(\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} - \mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W})
\end{aligned}$$

and let $\hat{\mathbf{R}}_{1,i}$ and $\hat{\mathbf{R}}_{2,i}$ denote the restrictions to the first n and Tn rows of $\mathbf{R}_{1,i}$ and $\mathbf{R}_{2,i}$ respectively. Then the following bounds hold:

(i) $\|\hat{\mathbf{R}}_{1,1}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{T^{1/4}\log(n)\, r_\alpha(1,1/\beta)^2}{\rho^{1/2}n}\right)$ *and* $\|\hat{\mathbf{R}}_{2,1}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log(n)\, r_\alpha(1,1/\beta)^2}{T^{1/4}\rho^{1/2}n}\right)$.

(ii) $\|\hat{\mathbf{R}}_{1,2}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{T^{1/4}\log(n)\, r_\alpha(1,1/\beta)^2}{\rho^{1/2}n^{3/4}}\right)$ *and* $\|\hat{\mathbf{R}}_{2,2}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log(n)\, r_\alpha(1,1/\beta)^2}{T^{1/4}\rho^{1/2}n^{3/4}}\right)$.

(iii) $\|\hat{\mathbf{R}}_{1,3}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1,1/\beta)}{T^{1/4}\rho^{1/2}n}\right)$ *and* $\|\hat{\mathbf{R}}_{2,3}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1,1/\beta)}{T^{3/4}\rho^{1/2}n}\right)$.

(iv) $\|\hat{\mathbf{R}}_{1,4}\|_{2\to\infty}, \|\hat{\mathbf{R}}_{2,4}\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{3/2}(n)\, r_\alpha(1,1/\beta)^3}{\rho n}\right)$.

Proof. We give full proofs of the bounds only for the terms $\hat{\mathbf{R}}_{1,i}$, noting that any differences in the proofs for the terms $\hat{\mathbf{R}}_{2,i}$ can be derived. Observe that for all i, j we have $\|\hat{\mathbf{R}}_{i,j}\|_{2\to\infty} \leq \|\mathbf{R}_{i,j}\|_{2\to\infty}$.

(i) Note that $\mathbf{U_P} = \mathbf{P}_C\mathbf{V}_\mathbf{P}^\top\mathbf{\Sigma}_\mathbf{P}^{-1}$ and so using the relation $\|\mathbf{AB}\|_{2\to\infty} \leq \|\mathbf{A}\|_{2\to\infty}\|\mathbf{B}\|$ and the fact that the spectral norm is invariant under orthonormal transformations we find that $\|\mathbf{U_P}\|_{2\to\infty} \leq \|\mathbf{P}_C\|_{2\to\infty}\|\mathbf{\Sigma}_\mathbf{P}^{-1}\| = \mathrm{O}(n^{-1/2})$ by Assumption 2 and the fact that the Euclidean norm of any row of $\mathbf{P}_C$ is $\mathrm{O}(T^{1/2}\rho n^{1/2})$. Noting that

$$\begin{aligned}\|\hat{\mathbf{R}}_{1,1}\|_{2\to\infty} &\leq \|\mathbf{U_P}\|_{2\to\infty}\|\mathbf{U}_\mathbf{P}^\top\mathbf{U_A}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W}\| \\ &\leq \|\mathbf{U_P}\|_{2\to\infty}\left(\|(\mathbf{U}_\mathbf{P}^\top\mathbf{U_A} - \mathbf{W})\mathbf{\Sigma}_\mathbf{A}^{1/2}\|_{\mathrm{Frob}} + \|\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{1/2} - \mathbf{\Sigma}_\mathbf{P}^{1/2}\mathbf{W}\|_{\mathrm{Frob}}\right)\end{aligned}$$

the result follows by applying Proposition 3 and 4 and Corollary 1.
For $\hat{\mathbf{R}}_{2,1}$, we note that the Euclidean norm of any column of $\mathbf{P}_C$ is $\mathrm{O}(\rho n^{1/2})$, and so $\|\mathbf{V_P}\|_{2\to\infty} = \mathrm{O}((Tn)^{-1/2})$, and the rest of the proof follows in the same manner as before.

(ii) We begin by splitting the term $\mathbf{R}_{1,2} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3$, where

$$\begin{aligned}\mathbf{M}_1 &= \mathbf{U_P}\mathbf{U}_\mathbf{P}^\top(\mathbf{A}_C - \mathbf{P}_C)(\mathbf{V_A} - \mathbf{V_P}\mathbf{W})\mathbf{\Sigma}_\mathbf{A}^{-1/2} \\ \mathbf{M}_2 &= (\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}(\mathbf{V}_\mathbf{P}^\top\mathbf{V_A} - \mathbf{W})\mathbf{\Sigma}_\mathbf{A}^{-1/2} \\ \mathbf{M}_3 &= (\mathbf{A}_C - \mathbf{P}_C)(\mathbf{I} - \mathbf{V_P}\mathbf{V}_\mathbf{P}^\top)\mathbf{V_A}\mathbf{\Sigma}_\mathbf{A}^{-1/2}\end{aligned}$$

Now,

$$\|\hat{\mathbf{M}}_1\|_{2\to\infty} \leq \|\mathbf{U_P}\|_{2\to\infty}\|\mathbf{A}_C - \mathbf{P}_C\|\|\mathbf{V_A} - \mathbf{V_P}\mathbf{W}\|\|\mathbf{\Sigma}_\mathbf{A}^{-1/2}\|$$

where

$$\begin{aligned}\|\mathbf{V_A} - \mathbf{V_P}\mathbf{W}\| &\leq \|\mathbf{V_A} - \mathbf{V_P}\mathbf{V}_\mathbf{P}^\top\mathbf{V_A}\| + \|\mathbf{V_P}(\mathbf{V}_\mathbf{P}^\top\mathbf{V_A} - \mathbf{W})\| \\ &= \|\mathbf{V_A}\mathbf{V}_\mathbf{A}^\top - \mathbf{V_P}\mathbf{V}_\mathbf{P}^\top\| + \|\mathbf{V}_\mathbf{P}^\top\mathbf{V_A} - \mathbf{W}\| \\ &= \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1,1/\beta)}{\rho^{1/2}n^{1/2}}\right)\end{aligned}$$

by Propositions 2 and 3 and our assumptions regarding the asymptotic growth of ρ, and so

$$\|\hat{\mathbf{M}}_1\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{T^{1/4}\log(n)\, r_\alpha(1,1/\beta)^2}{\rho^{1/2}n}\right)$$

by combining this with Proposition 1 and Corollary 1.
Next, note that

$$\|\hat{\mathbf{M}}_2\|_{2\to\infty} \leq \|(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\|_{2\to\infty}\|\mathbf{V}_\mathbf{P}^\top\mathbf{V_A} - \mathbf{W}\|\|\mathbf{\Sigma}_\mathbf{A}^{-1/2}\|.$$

To bound the term $\|(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\|_{2\to\infty}$, one can use an identical argument to that of Proposition 2(ii) to find that the absolute value of each term in the matrix is $\mathrm{O}_{\mathbb{P}}(\log^{1/2}(n)\, r_\alpha(1, \rho^{1/2}/\beta))$, and since each row has d elements, the bound also holds for the 2-to-infinity norm. By applying Proposition 3 and Corollary 1 we then find that

$$\|\hat{\mathbf{M}}_2\|_{2\to\infty} = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{3/2}(n)\, r_\alpha(1,1/\beta)^3}{T^{1/4}\rho^{3/2}n^{3/2}}\right).$$

To bound $\|\hat{\mathbf{M}}_3\|_{2\to\infty}$, let $\hat{\mathbf{M}} = (\hat{\mathbf{A}}_C - \hat{\mathbf{P}}_C)(\mathbf{I} - \mathbf{V_P}\mathbf{V_P}^\top)\mathbf{V_A}\mathbf{V_A}^\top$, so that $\hat{\mathbf{M}}_3 = \hat{\mathbf{M}}\mathbf{V_A}\mathbf{\Sigma_A}^{-1/2}$ and thus

$$\|\hat{\mathbf{M}}_3\|_{2\to\infty} \leq \|\hat{\mathbf{M}}\|_{2\to\infty}\|\mathbf{V_A}\mathbf{\Sigma_A}^{-1/2}\|.$$

The term $\|\mathbf{V_A}\mathbf{\Sigma_A}^{-1/2}\|$ is $\mathrm{O}_\mathbb{P}\left(\frac{1}{T^{1/4}\rho^{1/2}n^{1/2}}\right)$ by Corollary 1, so it remains to bound $\|\hat{\mathbf{M}}\|_{2\to\infty}$.

To do so, we claim that the Frobenius norm of the rows of $\hat{\mathbf{M}}$ are exchangeable and thus have the same expected value, which in turn implies that $\mathbb{E}[\|\hat{\mathbf{M}}\|_{\mathrm{Frob}}^2] = n\mathbb{E}[\|\hat{\mathbf{M}}_i\|^2]$ for any $i \in [n]$. Applying Markov's inequality, we therefore see that

$$\mathbb{P}(\|\hat{\mathbf{M}}_i\| > r) \leq \frac{\mathbb{E}[\|\hat{\mathbf{M}}_i\|^2]}{r^2} = \frac{\mathbb{E}[\|\hat{\mathbf{M}}\|_{\mathrm{Frob}}^2]}{nr^2}.$$

Now,

$$\|\hat{\mathbf{M}}\|_{\mathrm{Frob}} \leq \|\mathbf{M}\|_{\mathrm{Frob}} = \mathrm{O}_\mathbb{P}\left(T^{1/2}\log(n)\, r_\alpha(1, 1/\beta)^2\right)$$

by Propositions 1 and 2 (where we note that $(\mathbf{I} - \mathbf{V_P}\mathbf{V_P}^\top)\mathbf{V_A}\mathbf{V_A}^\top = (\mathbf{V_A}\mathbf{V_A}^\top - \mathbf{V_P}\mathbf{V_P}^\top)\mathbf{V_A}\mathbf{V_A}^\top$ to apply the latter) and so it follows that

$$\|\hat{\mathbf{M}}\|_{2\to\infty} = \mathrm{O}_\mathbb{P}\left(\frac{T^{1/2}\log(n)\, r_\alpha(1, 1/\beta)^2}{n^{1/4}}\right)$$

and

$$\|\hat{\mathbf{M}}_3\|_{2\to\infty} = \mathrm{O}_\mathbb{P}\left(\frac{T^{1/4}\log(n)\, r_\alpha(1, 1/\beta)^2}{\rho^{1/2}n^{3/4}}\right).$$

To see that our claim is true, let $\mathbf{Q} \in \mathrm{O}(n)$ be a permutation matrix, and let

$$\mathbf{Q}_* = \mathrm{diag}(\underbrace{\mathbf{Q}, \ldots, \mathbf{Q}}_{T}, \mathbf{I}_{Tp}) \in \mathrm{O}(T(n+p)).$$

Since the latent positions $\mathbf{Z}_i$ are assumed to be iid, the matrix entries of the pairs $(\mathbf{A}_C, \mathbf{P}_C)$ and $(\mathbf{Q}\mathbf{A}_C\mathbf{Q}_*^\top, \mathbf{Q}\mathbf{P}_C\mathbf{Q}_*^\top)$ have the same joint distribution, as these transformations simply correspond to a relabelling of the nodes, while the left- and right-singular vectors of $\mathbf{Q}\mathbf{A}_C\mathbf{Q}_*^\top$ (respectively $\mathbf{Q}\mathbf{P}_C\mathbf{Q}_*^\top$) are given by $\mathbf{Q}\mathbf{U_A}$ and $\mathbf{Q}_*\mathbf{V_A}$ (respectively $\mathbf{Q}\mathbf{U_P}$ and $\mathbf{Q}_*\mathbf{V_P}$).

Thus, since

$$\mathbf{Q}\hat{\mathbf{M}}\mathbf{Q}_*^\top = \mathbf{Q}(\hat{\mathbf{A}}_C - \hat{\mathbf{P}}_C)\mathbf{Q}_*^\top(\mathbf{I} - \mathbf{Q}_*\mathbf{V_P}\mathbf{V_P}^\top\mathbf{Q}_*^\top)\mathbf{Q}_*\mathbf{V_A}\mathbf{V_A}^\top\mathbf{Q}_*^\top$$

we deduce that the matrix entries of $\hat{\mathbf{M}}$ and $\mathbf{Q}\hat{\mathbf{M}}\mathbf{Q}_*^\top$ have the same joint distribution, proving our claim.

Combining these results, we see that

$$\|\hat{\mathbf{R}}_{1,2}\|_{2\to\infty} = \mathrm{O}_\mathbb{P}\left(\frac{T^{1/4}\log(n)\, r_\alpha(1, 1/\beta)^2}{\rho^{1/2}n^{3/4}}\right)$$

as required.

The proof for $\hat{\mathbf{R}}_{2,2}$ follows similarly, but involves splitting the matrix $\hat{\mathbf{M}}' = (\hat{\mathbf{A}}_C - \hat{\mathbf{P}}_C)^\top(\mathbf{I} - \mathbf{U_P}\mathbf{U_P}^\top)\mathbf{U_A}\mathbf{U_A}^\top$ into T distinct blocks and applying the same argument to show that the rows of each individual block are exchangeable. Note that in this case we divide through by a factor of $T^{1/2}$, as the spectral norm of each individual block is $\mathrm{O}_\mathbb{P}(n^{1/2}\log^{1/2}(n)\, r_\alpha(1, 1/\beta))$, and consequently find that

$$\|\hat{\mathbf{R}}_{2,2}\|_{2\to\infty} = \mathrm{O}_\mathbb{P}\left(\frac{\log(n)\, r_\alpha(1, 1/\beta)^2}{T^{1/4}\rho^{1/2}n^{3/4}}\right).$$

(iii) We see that

$$\begin{aligned}\|\hat{\mathbf{R}}_{1,3}\|_{2\to\infty} &\leq \|\mathbf{U_P}\|_{2\to\infty}\|\mathbf{U_P}^\top(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V_P}\|_{\mathrm{Frob}}\|\mathbf{W}\mathbf{\Sigma_A}^{-1/2}\| \\ &= \mathrm{O}_\mathbb{P}\left(\frac{\log^{1/2}(n)\, r_\alpha(1, 1/\beta)}{T^{1/4}\rho^{1/2}n}\right)\end{aligned}$$

by Proposition 2 and Corollary 1. The proof for $\hat{\mathbf{R}}_{2,3}$ follows identically, noting the additional factor of $T^{1/2}$ in the denominator from $\|\mathbf{V_P}\|_{2\to\infty}$.

(iv) We see that

$$\begin{aligned}\|\mathbf{R}_{1,4}\|_{2\to\infty} &\leq \|(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_\mathbf{P}\|_{2\to\infty}\|\mathbf{W}\mathbf{\Sigma}_\mathbf{A}^{-1/2} - \mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W}\|_{\text{Frob}} \\ &= \mathrm{O}_\mathbb{P}\left(\frac{\log^{3/2}(n)\, r_\alpha(1,1/\beta)^3}{\rho n}\right)\end{aligned}$$

by applying Proposition 4 and bounding the term $\|(\mathbf{A}_C - \mathbf{P}_C)\mathbf{V}_\mathbf{P}\|_{2\to\infty}$ as in part (ii).

□

Theorem 1

Proof. We first consider the left embedding $\hat{\mathbf{X}}_\mathbf{A}$. Observe that

$$\hat{\mathbf{X}}_\mathbf{A} - \hat{\mathbf{X}}_\mathbf{P}\mathbf{W} = (\hat{\mathbf{A}}_C - \hat{\mathbf{P}}_C)\mathbf{V}_\mathbf{P}\mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W} + \sum_{i=1}^4 \hat{\mathbf{R}}_{1,i}$$

and so

$$\|\hat{\mathbf{X}}_\mathbf{A} - \hat{\mathbf{X}}_\mathbf{P}\mathbf{W}\|_{2\to\infty} \leq \sigma_d(\mathbf{P})^{-1/2}\|(\hat{\mathbf{A}}_C - \hat{\mathbf{P}}_C)\mathbf{V}_\mathbf{P}\|_{2\to\infty} + \sum_{i=1}^4 \|\hat{\mathbf{R}}_{1,i}\|_{2\to\infty}.$$

As shown in the proof of Proposition 5, $\|(\hat{\mathbf{A}}_C - \hat{\mathbf{P}}_C)\mathbf{V}_\mathbf{P}\|_{2\to\infty} = \mathrm{O}_\mathbb{P}(\log^{1/2}(n)\, r_\alpha(1,1/\beta))$, and consequently we see that

$$\begin{aligned}\|\hat{\mathbf{X}}_\mathbf{A} - \hat{\mathbf{X}}_\mathbf{P}\mathbf{W}\|_{2\to\infty} &= \mathrm{O}_\mathbb{P}\left(\frac{\log^{1/2}(n)\, r_\alpha(1,1/\beta)}{T^{1/4}\rho^{1/2}n^{1/2}}\right) + \sum_{i=1}^4 \|\hat{\mathbf{R}}_{1,i}\|_{2\to\infty} \\ &= \mathrm{O}_\mathbb{P}\left(\frac{\log^{1/2}(n)\, r_\alpha(1,1/\beta)}{T^{1/4}\rho^{1/2}n^{1/2}}\right)\end{aligned}$$

by using the bounds on $\|\hat{\mathbf{R}}_{1,i}\|_{2\to\infty}$ from Proposition 5.

Similarly, for the embeddings $\hat{\mathbf{Y}}_\mathbf{A}^{(t)}$ we find that

$$\hat{\mathbf{Y}}_\mathbf{A}^{(t)} - \hat{\mathbf{Y}}_\mathbf{P}^{(t)}\mathbf{W} = (\hat{\mathbf{A}}_C^{(t)} - \hat{\mathbf{P}}_C^{(t)})^\top\mathbf{U}_\mathbf{P}\mathbf{\Sigma}_\mathbf{P}^{-1/2}\mathbf{W} + \sum_{i=1}^4 \hat{\mathbf{R}}_{2,i}$$

and so

$$\|\hat{\mathbf{Y}}_\mathbf{A}^{(t)} - \hat{\mathbf{Y}}_\mathbf{P}^{(t)}\mathbf{W}\|_{2\to\infty} \leq \sigma_d(\mathbf{P})^{-1/2}\|(\hat{\mathbf{A}}_C^{(t)} - \hat{\mathbf{P}}_C^{(t)})^\top\mathbf{U}_\mathbf{P}\|_{2\to\infty} + \sum_{i=1}^4 \|\hat{\mathbf{R}}_{1,i}\|_{2\to\infty},$$

and an identical argument yields the same bound. □

B PROOF OF LEMMA 1

Proof. For node/time pairs (i,s) and (j,t) to be exchangeable in the attributed dynamic latent position model (Definition 3), for all $\mathbf{Z} \in \mathcal{Z} \cup \mathcal{I}$,

$$H(\mathbf{Z}_i^{(s)}, \mathbf{Z}) = H(\mathbf{Z}_j^{(t)}, \mathbf{Z}).$$

This implies that the corresponding rows of the mean attributed unfolded adjacency matrix are equal, $(\mathbf{P}_C^{(s)})_i = (\mathbf{P}_C^{(t)})_j$. Since $\mathbf{X}_\mathbf{P}(\mathbf{Y}_\mathbf{P}^{(t)})^\top = \mathbf{P}_C^{(t)}$ is a symmetric matrix for all $t \in [T]$,

$$\mathbf{Y}_\mathbf{P}^{(t)} = \mathbf{P}^{(t)}\mathbf{X}_\mathbf{P}(\mathbf{X}_\mathbf{P}^\top\mathbf{X}_\mathbf{P})^{-1},$$

which implies that the corresponding rows of the noise-free dynamic embedding are equal, $(\mathbf{Y}_{\mathbf{P}}^{(s)})_i = (\mathbf{Y}_{\mathbf{P}}^{(t)})_j$. By Theorem 1,

$$\|\hat{\mathbf{Y}}_i^{(s)} - (\mathbf{Y}_{\mathbf{P}}^{(s)})_i \mathbf{W}\|,\ \|\hat{\mathbf{Y}}_j^{(t)} - (\mathbf{Y}_{\mathbf{P}}^{(t)})_j \mathbf{W}\| = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1, \log^\gamma(n))}{T^{1/4}\rho^{1/2}n^{1/2}}\right),$$

which, by the triangle inequality and the equality of the noise-free embeddings, shows that

$$\|\hat{\mathbf{Y}}_i^{(s)} - \hat{\mathbf{Y}}_j^{(t)}\| = \mathrm{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)\, r_\alpha(1, \log^\gamma(n))}{T^{1/4}\rho^{1/2}n^{1/2}}\right). \qquad \square$$

C DETAILS OF THE SIMULATED EXAMPLE OF SECTION 3.3

C.1 EXPERIMENTAL SETUP DETAILS

We define the attributed dynamic latent position model in Definition 3,

$$\mathbf{A}_{ij}^{(t)} \mid Z_i^{(t)}, Z_j^{(t)} \overset{\text{ind}}{\sim} \text{Bernoulli}(\mathbf{B}_{Z_i^{(t)}, Z_j^{(t)}}),$$
$$\mathbf{C}_{i\ell}^{(t)} \mid Z_i^{(t)} \overset{\text{ind}}{\sim} \text{Normal}(\mathbf{D}_{Z_i^{(t)}, \ell}, \sigma^2 \mathbf{I}),$$

with community edge probability matrix $\mathbf{B} \in \mathbb{R}^{3\times 3}$ and mean attribute matrix $\mathbf{D} \in \mathbb{R}^{3\times p}$,

$$\mathbf{B} = \begin{pmatrix} p_1 & p_0 & p_0 \\ p_0 & p_0 & p_0 \\ p_0 & p_0 & p_0 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix},$$

where $p_0 = 0.5$, $p_1 = 0.3$, $\boldsymbol{\mu}_1 = [\underbrace{0, \ldots 0}_{\text{0-19}}, \underbrace{1, \ldots 1}_{\text{20-74}}, \underbrace{0, \ldots 0}_{\text{75-149}}]$ and $\boldsymbol{\mu}_2 = [\underbrace{0, \ldots 0}_{\text{0-79}}, \underbrace{1, \ldots 1}_{\text{80-139}}, \underbrace{0, \ldots 0}_{\text{140-149}}]$.

C.2 HYPERPARAMETER SELECTION

We set the embedding dimension $d = 3$ for all the methods, the known number of communities. Experiments with other embedding dimension are given in `https://anonymous.4open.science/r/AUASE-80E4/README.md`. The selection of hyperparameters for each method shown in Figure 2 is outlined below:

- UASE - no other parameters.
- AUASE - visually chosen parameters (see Figure 5): $\alpha = 0.2$.
- DRLAN - visually chosen parameters (see `https://anonymous.4open.science/r/AUASE-80E4/README.md`): $\beta = 0.7$, $p = 2$, $q = 1$, $\alpha = [1, 0.1, 0.001, 0.0001]$, $\theta = [1, 1, 0.1, 0.001, 0.0001]$.
- DySAT - default parameters.

C.3 ADDITIONAL FIGURES

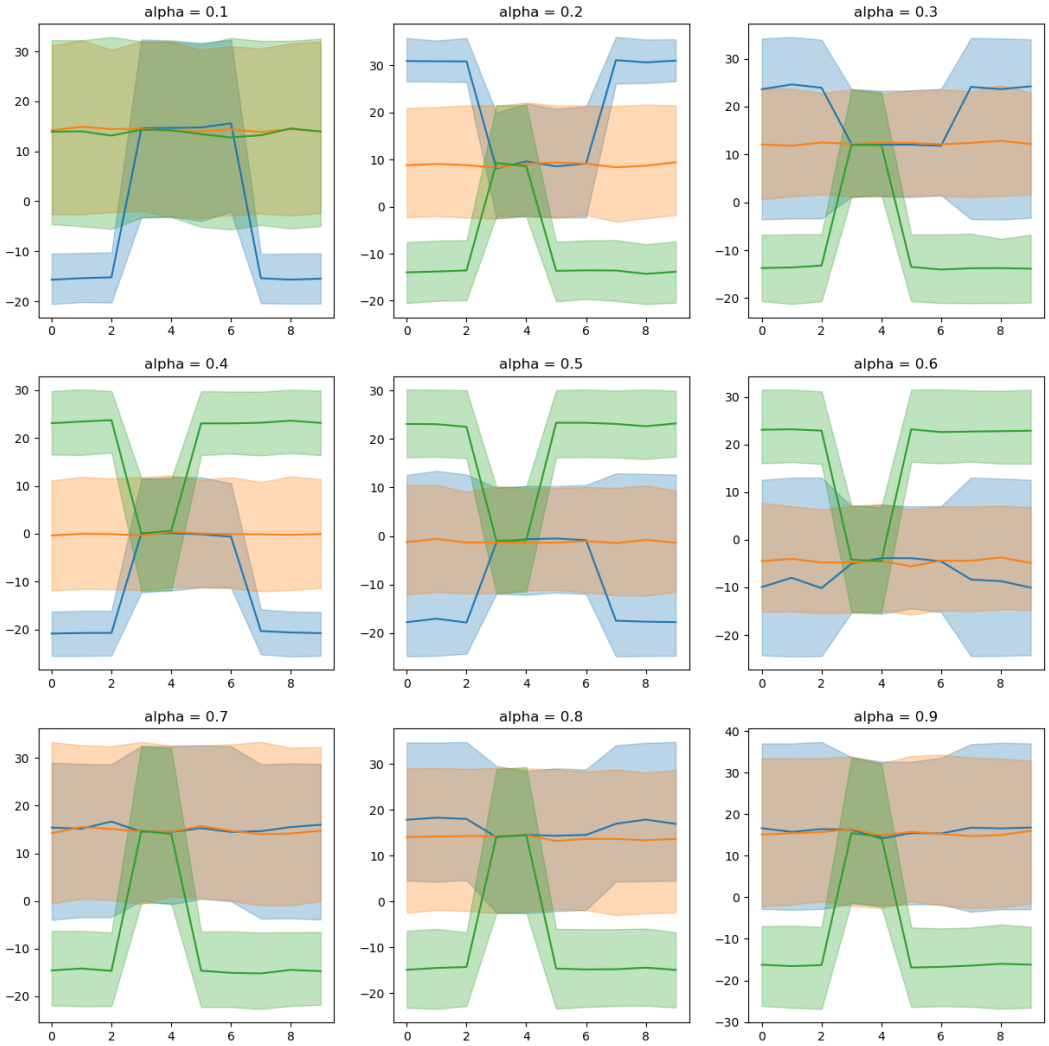

Figure 5: One-dimensional UMAP visualisation of the AUASE node embeddings for varying $\alpha \in [0.1, 0.9]$. The coloured lines show the mean embedding for each community with a 90% confidence interval.

D DETAILS OF THE REAL DATA ANALYSIS OF SECTION 3.4

D.1 EXPERIMENTAL SETUP DETAILS

DBLP. The raw data can be downloaded from `https://www.aminer.cn/citation` as *DBLP-Citation-network V3*. We consider papers published from 1996 to 2009. We group years 1996 - 2000 and 2001 - 2002, while we consider the remaining years as a single time point so that in each time point we have a similar order of magnitude of papers. We only consider authors who published at least three papers between 1996 and 2009.

To construct the adjacency matrices, we form an edge if two authors collaborated on at least one paper within that time point. For the attributes, we consider words from titles and abstracts of the papers published by an author within a certain time point. We remove stop words, words that have less than three characters, and any symbol that is not an English word. Then, we remove words that have an overall count of less than 100 and the ones that have an overall count of more than 5000.

We label each paper from the venue where it is published as follows:

- Computer Architecture: PPOPP, PPOPP/PPEALS, ASP-DAC, EURO-DAC, DAC, MICRO, PODC, DIALM-PODC.
- Computer Network: SIGCOMM, MOBICOM, MOBICOM-CoRoNet, INFOCOM, SenSys.
- Data Mining: SIGMOD Conference, SIGMOD Workshop, Vol. 1, SIGMOD Workshop, Vol. 2, SIGSMALL/SIGMOD Symposium, SIGMOD Record, ICDE, ICDE Workshops, SIGIR, SIGIR Forum.
- Computer Theory: STOC, SODA, CAV, FOCS.
- Multi-Media: SIGGRAPH, IEEE Visualization, ICASSP.
- Artificial Intelligence: IJCAI, IJCAI (1), IJCAI (2), ACL2, ACL, NIPS.
- Computer-Human Interaction: IUI, PerCom, HCI.

Then, each author is assigned a label at each time point based on the majority label of their papers published within that time point.

ACM. The raw data can be downloaded from `https://www.aminer.cn/citation` as *ACM-Citation-network V8*. We consider the years from 2000 to 2014; each year is a time point. The network and the attributes are constructed in the same manner as for DBLP. The labels are constructed as follows:

- Data Science: VLDB, SIGMOD, PODS, ICDE, EDBT, SIGKDD, ICDM, DASFAA, SSDBM, CIKM, PAKDD, PKDD, SDM, DEXA.
- Computer Vision: CVPR, ICCV, ICIP, ICPR, ECCV, ICME, ACM-MM.

Epinions. The raw data can be downloaded from `https://www.cse.msu.edu/~tangjili/datasetcode/truststudy.htm`. We consider the years from 2001 to 2011, and we consider users who establish at least two trust relationships in this time frame. Each year is a time point, and an edge is formed if the users trusted each other within that year.

The attributes are the words of the titles of the reviews published by each author within that year. We remove stop words, words with less than three characters, and words with an overall count of less than 100 or more than 10000.

To construct the labels, we consider the most frequent category of the reviews published by an author in a given year. We merge some sparse categories with similar, more popular ones; the resulting labels are *Books, Business & Technology, Cars & Motorsports, Computers & Internet, Education, Electronics, Games, Home and Garden, Hotels & Travel, Kids & Family, Magazines & Newspapers, Movies, Music, Musical Instruments, Online Stores & Services, Personal Finance, Pets, Restaurants & Gourmet, Software, Sports & Outdoors, Wellness & Beauty*.

ogbn-mag. The raw data can be downloaded from `https://ogb.stanford.edu/docs/nodeprop/`. The data spans years from 2010 to 2019, each year is a time point. We consider authors which published at least ten paper in the whole time frame.

To construct the adjacency matrices, we form an edge if two authors collaborated on at least one paper within that time point. Each paper is associated with a 128-dimensional word2vec feature vector, to construct the attributes matrices for each author we average the feature vectors of the papers they published in that time point. An author label is set to the label associated with the majority of the papers they published in that time point.

We use ogbn-mag data to construct a datasets which is suited to our dynamic tasks. The data format and the tasks are at author-level, different than the usual task performed on obgn-mag ('the task is to predict the venue of each paper, given its content, references, authors, and authors' affiliations'). That is because we are interested in dynamic tasks; authors are dynamic nodes, unlike papers, and their attributes and labels change over time. This creates a much harder task, hence our performance cannot be directly compared to other performances on obgn-mag to predict papers labels.

CONN, Glodyne, DySAT and DyRep are not computationally efficient enough to compute the embeddings of ogbn-mag. We report here the computational barriers for each methods.

- CONN: the authors' code is designed with dense matrices, which would require about 2T of memory for ogbn-mag matrices.
- Glodyne: we capped computational time after 48h.
- Dyrep: it requires an unreasonably large amount of memory. The largest dataset the authors' analyses has only ≈ 10000 nodes.

For all the data sets, the nodes that are not active within a certain time point are considered *Unlabelled*. These nodes are not used for training or testing.

D.2 METHOD COMPARISON

The following list shows unsupervised dynamic attributes methods we did not include in our method comparison and the reason why.

- [Li et al., 2017] - source code not available.
- [Xu et al., 2020b] - source code not available.
- [Luodi et al., 2024] - source code not available.
- [Liu et al., 2021] - the code to construct the motif matrices is unavailable, and there is no clear documentation on how to construct them for new datasets.
- [Tang et al., 2022] - the source code is undocumented and raises memory errors on the datasets considered in this paper.
- [Mo et al., 2024] - source code not available.
- [Wei et al., 2019] - source code not available.
- [Xu et al., 2020a] - First, the code does not allow for time-varying covariates. Second, it is extremely time-consuming for data sets with large p; for example, one epoch on one batch for DBLP takes more than 10 min, meaning the whole computation would be more than 1000 hours. The authors presented examples with a maximum $p = 200$, while for the data sets considered in the paper, p is at least 4000.
- [Ahmed et al., 2024] - source code not available.

D.3 HYPERPARAMETER SELECTION

We choose the embedding dimension using ScreeNOT giving $d = 12$, $d = 29$ and $d = 22$ for DBLP, ACM and Epinions, respectively (for DySAT, we set $d = 32$ for ACM). We use cross-validation to select α for AUASE, β for DRLAN and α for CONN.

- UASE - no other parameters.
- AUASE - no other parameters.
- DRLAN - we set all the parameters to the default set by the authors for node classification: $p = 2$, $q = 2$, $\alpha = [1, 1, 0.1, 0.1, 1]$, $\theta = [1, 10, 100]$.
- DySAT - we set all the parameters to the default set by the authors: epochs=200, valfreq=1, testfreq=1, batchsize=512, maxgradientnorm=1.0, useresidual='False',negsamplesize=10, walklen=20, negweight=1.0, learningrate=0.001, spatialdrop=0.1, temporaldrop=0.5, weightdecay=0.0005, positionffn='True', window=-1.

- CONN - we set all the parameters to the default set by the authors: nlayer=2, epochs=200, activate="relu", batchsize=1024, lr=0.01, weightdecay=0.0, hid1=512, hid2=64, losstype=entropy, patience=50, dropout=0.6, drop=1. Note that we implemented CONN in the unsupervised version - *main_lp.py*.
- GloDyNE - we set all the parameters to the default set by the authors: limit=0.1, numwalks=10, walklength=80, window=10, negative=5, workers=32, scheme=4.
- DyRep - we set epochs to 20, batch size to 20, and we assume all of the links in Jodie as communication; we set all the other parameters to the default set by the authors. We implemented Dyrep without node features as it only supports non-time-varying attributes, and all three datasets we consider have time-varying attributes.

Note that for DySAT, we set $d = 32$ for ACM instead of $d = 29$. DySAT set the embedding dimension *structural_layer_config* = d as function of the hyperparameter *structural_head_config* = d_1, d_2, d_3 such that $d_1 \times d_2 = d$. For DBLP we set *structural_head_config = 4, 3, 2* giving $d = 12$, for Epinions we set *structural_head_config = 11, 2, 2* giving $d = 22$. However, the only option giving $d = 29$ is $d_1 = 29$, $d_2 = 1$. *structural_head_config* represent the dimensions of the layers, and two subsequent layers with such a wide gap would give DySAT an unfair disadvantage. Hence we set *structural_head_config = 8,4,4* and *structural_layer_config = 32* for ACM. Table 7 shows that our results are robust to the choice of embedding dimension.

D.4 ADDITIONAL TABLES

Table 3: AUC of link prediction on AUASE embeddings for varying α. For $\alpha = 1$ the method is using only the attributes and not the network, contrarily to AUASE for $\alpha \in (0, 1)$.

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
AUC	0.915	0.912	0.907	0.898	0.892	0.883	0.887	0.881	0.866	0.546

Method	Metrics	2010	2011	2012	2013	2014
CONN	F1	0.529	0.541	0.537	0.528	0.37
	F1 micro	0.535	0.569	0.553	0.542	0.431
	F1 macro	0.490	0.496	0.488	0.515	0.394
GloDyNE	F1	0.703	0.712	0.717	0.731	0.664
	F1 micro	0.584	0.589	0.622	0.559	0.486
	F1 macro	0.509	0.515	0.535	0.481	0.441
DRLAN	F1	0.483	0.495	0.527	0.479	0.39
	F1 micro	0.627	0.599	0.619	0.556	0.466
	F1 macro	0.385	0.417	0.450	0.424	0.403
DySAT	F1	0.662	0.557	0.583	0.397	NA
	F1 micro	0.680	0.551	0.578	0.556	NA
	F1 macro	0.626	0.548	0.558	0.357	NA
UASE	F1	0.755	0.728	0.753	0.804	0.68
	F1 micro	0.772	0.746	0.771	0.81	0.688
	F1 macro	0.727	0.702	0.722	0.798	0.685
AUASE	F1	0.809	0.797	0.790	0.859	0.878
	F1 micro	0.806	0.794	0.787	0.858	0.877
	F1 macro	0.799	0.790	0.779	0.858	0.875

Table 4: Micro, macro and weighted F1 of node classification on ACM.

Method	Metrics	2007	2008	2009	2010	2011
CONN	F1	0.062	0.081	0.070	0.067	0.067
	F1 micro	0.114	0.133	0.137	0.129	0.131
	F1 macro	0.030	0.038	0.034	0.032	0.032
GloDyNE	F1	0.194	0.176	0.141	0.140	0.154
	F1 micro	0.211	0.189	0.156	0.155	0.165
	F1 macro	0.113	0.103	0.087	0.086	0.096
DRLAN	F1	0.092	0.076	0.106	0.067	0.083
	F1 micro	0.119	0.086	0.135	0.098	0.111
	F1 macro	0.049	0.039	0.052	0.032	0.044
DyRep	F1	0.118	0.107	0.105	0.095	0.101
	F1 micro	0.15	0.146	0.142	0.132	0.147
	F1 macro	0.074	0.065	0.065	0.054	0.057
DySAT	F1	0.062	0.039	0.036	0.011	NA
	F1 micro	0.069	0.095	0.075	0.042	NA
	F1 macro	0.035	0.019	0.021	0.009	NA
UASE	F1	0.224	0.204	0.203	0.196	0.178
	F1 micro	0.224	0.204	0.203	0.196	0.178
	F1 macro	0.112	0.085	0.090	0.101	0.076
AUASE	F1	0.318	0.304	0.279	0.266	0.263
	F1 micro	0.332	0.319	0.294	0.279	0.277
	F1 macro	0.197	0.176	0.167	0.163	0.154

Table 5: Micro, macro and weighted F1 of node classification on Epinions.

Method	Metrics	2007	2008	2009
CONN	F1	0.242	0.189	0.141
	F1 micro	0.259	0.219	0.181
	F1 macro	0.137	0.127	0.110
GloDyNE	F1	0.138	0.252	0.077
	F1 micro	0.178	0.270	0.106
	F1 macro	0.117	0.125	0.064
DyRep	F1	0.193	0.319	0.110
	F1 micro	0.258	0.367	0.190
	F1 macro	0.192	0.194	0.091
DRLAN	F1	0.135	0.241	0.150
	F1 micro	0.193	0.288	0.214
	F1 macro	0.115	0.163	0.148
DyRep	F1	0.193	0.319	0.110
	F1 micro	0.258	0.367	0.190
	F1 macro	0.192	0.194	0.091
DySAT	F1	0.029	0.267	NA
	F1 micro	0.051	0.299	NA
	F1 macro	0.055	0.133	NA
UASE	F1	0.433	0.707	0.376
	F1 micro	0.488	0.717	0.446
	F1 macro	0.460	0.525	0.424
AUASE	F1	0.584	0.757	0.564
	F1 micro	0.590	0.768	0.584
	F1 macro	0.545	0.554	0.543

Table 6: Micro, macro and weighted F1 of node classification on DBLP.

Method	DBLP			ACM			Epinions		
	$d = 32$	$d = 64$	$d = 128$	$d = 32$	$d = 64$	$d = 128$	$d = 32$	$d = 64$	$d = 128$
CONN	0.532	0.589	0.597	0.766	0.779	0.782	0.165	0.176	0.183
GloDyNE	0.383	0.386	0.394	0.657	0.661	0.662	0.285	0.288	0.269
DRLAN	0.657	0.692	0.723	0.785	0.801	0.830	0.242	0.258	0.277
DyRep	0.274	-	-	-	-	-	-	-	-
DySAT	0.752	0.780	0.812	-	-	-	-	-	-
UASE	0.763	0.774	0.781	0.844	0.849	0.852	0.229	0.228	0.234
AUASE	0.853	0.864	0.872	0.927	0.936	0.940	0.306	0.335	0.349

Table 7: Accuracies of node classification on DBLP, ACM and Epinions for $d = 32$, $d = 64$, and $d = 128$. For DySAT and DyRep, we only report partial results due to the intensive computational times