Pre-DyGAE: Pre-training Enhanced Dynamic Graph Autoencoder for Occupational Skill Demand Forecasting

Xi Chen[1,2*], **Chuan Qin**[2,3], **Zhigaoyuan Wang**[4], **Yihang Cheng**[2], **Chao Wang**[1,5], **Hengshu Zhu**[2†] and **Hui Xiong**[4†]
[1]School of Computer Science and Technology, University of Science and Technology of China
[2]Career Science Lab, BOSS Zhipin
[3]PBC School of Finance, Tsinghua University
[4]Artificial Intelligence Thrust, The Hong Kong University of Science and Technology (Guangzhou)
[5]Guangzhou HKUST Fok Ying Tung Research Institute Guangzhou, China
chenxi0401@mail.ustc.edu.cn,zwang901@connect.hkust-gz.edu.cn, xionghui@ust.hk,
{chuanqin0426, chengyihang544, chadwang2012, zhuhengshu}@gmail.com

Abstract

Occupational skill demand (OSD) forecasting seeks to predict dynamic skill demand specific to occupations, beneficial for employees and employers to grasp occupational nature and maintain a competitive edge in the rapidly evolving labor market. Although recent research has proposed data-driven techniques for forecasting skill demand, the focus has remained predominantly on overall trends rather than occupational granularity. In this paper, we propose a novel *Pre-training Enhanced Dynamic Graph Autoencoder (Pre-DyGAE)*, forecasting skill demand from an occupational perspective. Specifically, we aggregate job descriptions (JDs) by occupation and segment them into several timestamps. Subsequently, in the initial timestamps, we pretrain a graph autoencoder (GAE), consisting of a semantically-aware cross-attention enhanced uncertainty-aware encoder and decoders for link prediction and edge regression to achieve graph reconstruction. In particular, we utilize contrastive learning on skill cooccurrence clusters to solve data sparsity and a unified Tweedie and ranking loss to predict the imbalanced distribution. Afterward, we incorporate an adaptive temporal encoding unit and a temporal shift module into GAE to achieve a dynamic GAE (DyGAE), fine-tuned with a two-stage optimization strategy. Extensive experiments on four real-world datasets validate the effectiveness of Pre-DyGAE compared with state-of-the-art baselines.

1 Introduction

With the rapid development of technology, the nature of occupation is continually changing [Bisello *et al.*, 2022], e.g., the skill demand of the occupation. OSD forecasting plays an important role in research fields related to the labor market and economics, such as labor market evolution [Sus and Sylwestrzak, 2021], occupational mobility and development [Wang *et al.*, 2021b; Wang *et al.*, 2021a; Qin *et al.*, 2019], and enterprise recruitment [Bilan *et al.*, 2020; Qin *et al.*, 2022]. In particular, dynamic OSD forecasting is helpful for the development of individual career construction and career planning [Guo *et al.*, 2022a]. As for enterprises, forward-looking OSD is helpful for recruitment.

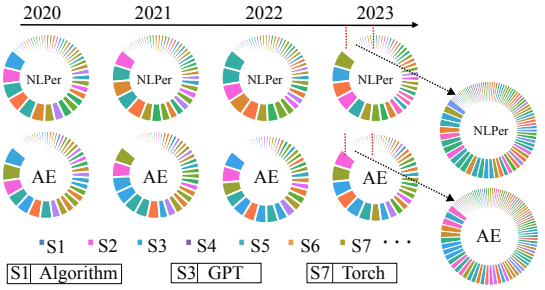

Figure 1: An illustration depicts the dynamic evolution of skill demands for NLP Algorithm Researchers (NLPer) and Algorithm Engineers (AE) from 2020 to 2023. The proportions of different colors on each donut chart represent the proportion of demand for different skills in the corresponding occupation.

Previous studies on the analysis between occupation and skill mainly leverage the survey or interview-based methods, e.g., Occupational Information Network (O*NET), which is very labor-intensive and involves human biases [Felten *et al.*, 2018]. Fortunately, recent years have witnessed the rapid development of online recruitment platforms, accumulating massive job postings across different occupations. This provides an unparalleled chance for implementing large-scale data-driven OSD forecasting. Along this line, although some research has been introduced for skill demand prediction [Wu *et al.*, 2019; Li *et al.*, 2020; Chao *et al.*, 2024] and skill trend analysis [Mahdavimoghaddam *et al.*, 2022], the skill demand forecasting at the occupational level is still largely

*This work was done when Xi Chen was an intern at Career Science Lab, BOSS Zhipin supervised by Chuan Qin.
†Corresponding Authors

under-explored. Predicting skill demand at an occupational granularity is essential as it acknowledges the varying trends in skill development across different occupations, highlighting that a singular prediction of overall skill trends falls short of meeting diverse demands.

Indeed, there are three major challenges for OSD forecasting. Firstly, the trends of identical skills across various occupations exhibit diversification, while the demands for distinct skills in occupations are intricately interconnected [Langley *et al.*, 2019]. This intricate interplay complicates the modeling in skill demand forecasting at the occupational level. Secondly, some pioneering skills often go unnoticed in enterprise recruitment job postings. This oversight makes it difficult to attain forward-looking forecasting on the skills that have not yet surfaced in occupations. Lastly, as shown in Figure 1 many skills occur in the long-tailed part, making accurate demand forecasting more challenging.

To address the above challenges, we introduce Pre-DyGAE, pre-training enhanced dynamic graph autoencoder for OSD forecasting. To be specific, we aggregate JDs by occupation and partition them into several timestamps. Firstly, leveraging JDs from the initial timestamps, we pre-train the graph autoencoder (GAE), consisting of an encoder, a link prediction decoder, and an edge regression decoder. The encoder is enhanced by semantically-aware cross-attention and Graph Convolution Neural Networks (GCNs) for robust representations of both occupations and skills. Besides, GAE is pre-trained with a link prediction task enhanced by contrastive learning on skill clusters and an edge regression task to predict the distribution of skill demand using a unified Tweedie and ranking loss. Subsequently, in the following timestamps, we add an adaptive temporal encoding unit and a temporal shift operator into the pre-trained GAE to achieve a DyGAE. Furthermore, we fix the GAE and fine-tune the DyGAE with a two-stage optimization strategy. Finally, we obtain the evolved representations and achieve the forecasting of future OSD. Extensive real-world dataset experiments confirm Pre-DyGAE significantly outperforms state-of-the-art baselines. The code is available online[1].

2 Related Work

2.1 Occupational Skill Analysis

Research on occupation and skill plays a crucial role in labor market analysis [Felten *et al.*, 2018; Sun *et al.*, 2021; Fang *et al.*, 2023; Qin *et al.*, 2024; Qin *et al.*, 2023; Chen *et al.*, 2021]. Recently, the labor market has undergone a rapid transformation driven by technological advancements. Several studies have examined the substitution of skills [Felten *et al.*, 2018] and the evolution of occupations, while others have explored occupational mobility and market changes post-economic crisis [Bisello *et al.*, 2022]. However, these studies often focus on localized phenomena of occupations and skills. The emergence of online recruitment platforms has enabled detailed statistical analysis of job postings. This data, collected from enterprise recruitment efforts, sheds light on the demand for talents and skills in the labor market [Zhu *et al.*, 2016; Zhang *et al.*, 2021; Guo *et al.*, 2022b; Chao *et al.*, 2024], helping to predict skill demand for specific jobs [Wu *et al.*, 2019]. Nonetheless, existing research tends to concentrate on skills or occupations separately, lacking a holistic analysis of both. Diverging from prior work, our study adopts an innovative perspective to explore the skill demand at the occupational level over time.

[1]https://github.com/cx9941/Pre-DyGAE

2.2 Dynamic Graph Learning

Dynamic graph learning has garnered significant attention in recent research, which can model many real-world scenarios such as user-item interaction systems [Yu *et al.*, 2023] and traffic networks [Zhao *et al.*, 2020]. Many studies segment nodes based on temporal slices, analyzing static graphs within each slice and synthesizing temporal representations through methods like RNN or self-attention mechanisms, such as JODIE [Kumar *et al.*, 2019], Dyrep [Trivedi *et al.*, 2019], and CAWN [Wang *et al.*, 2021d]. Specially, focus on spatio-temporal sequences, GConv [Seo *et al.*, 2018], Dygrencoder [Taheri *et al.*, 2019] and TGCN [Zhao *et al.*, 2020] leveraged a combination of GCN and RNN to exploit spatial and temporal regularities. Besides, DCRNN [Li *et al.*, 2018] captures the spatial dependency using random walks on the graph and A3TGCN [Bai *et al.*, 2021] additionally adds the attention mechanism to assemble global temporal information. To learn the dynamic topology, EvolveGCN [Pareja *et al.*, 2019] uses a recurrent model to evolve the GCN parameters, and AGCRN [Bai *et al.*, 2020] captures fine-grained graph correlations in time series automatically. Some transformer-based methods, TCL [Wang *et al.*, 2021c] and DyGFormer [Yu *et al.*, 2023], learning from node historical first-hop interactions, efficiently benefit from long histories. Additionally, several studies have introduced temporal encoding mechanisms to effectively capture timing patterns in dynamic graphs, such as TGN [Rossi *et al.*, 2020], TCL [Wang *et al.*, 2021c] and GraphMixer [Cong *et al.*, 2023].

3 Preliminaries

In this section, we define the key concepts integral to our study, present the problem formulation of OSD forecasting, and give an overview of our proposed Pre-DyGAE.

Definition 1 (**Occupational Skill Demand**). *At timestamp t, each occupation o in the occupation set $\mathcal{O}$ contains collected JDs $\mathcal{P}_o^t$. Each JD p encompasses a set of skills from the skill set $\mathcal{S}$. Along this line, we define the OSD at timestamp t as: $\mathcal{R}_{o,s}^t = \sum_{p \in \mathcal{P}_o^t} \mathbf{1}(s \in p)/|\mathcal{P}_o^t|$, where $\mathcal{R}^t \in \mathbb{R}^{|\mathcal{O}| \times |\mathcal{S}|}$ and $|\cdot|$ denotes the size of a set.*

Definition 2 (**Occupational Skill Demand Graph**). *At timestamp t, OSD graph can be represented as an undirected bipartite graph $\mathcal{G}^t = (\mathcal{V}, \mathcal{E}^t, A^t)$, where $\mathcal{V} = \mathcal{O} \cup \mathcal{S}$, $\mathcal{E}^t$ is a set of edges between occupations and skills, and $A^t \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ is the weighted adjacency matrix. The edge $e_{o,s}^t$ between o and s is defined as $\{e_{o,s}^t \in \mathcal{E}^t | o \in O, s \in S,$ and $\mathcal{R}_{o,s}^t \neq 0\}$. The weighted adjacency matrix is defined as $A_{o,s}^t = A_{s,o}^t = \mathcal{R}_{o,s}^t$ if $e_{o,s}^t \in \mathcal{E}^t$, otherwise 0.*

Problem 1 (**Occupational Skill Demand Forecasting**). *Given an OSD graph sequence $(\mathcal{G}^1, \mathcal{G}^2, ..., \mathcal{G}^T)$, our objective becomes predicting the OSD graph $\mathcal{G}^{T+1}$ for the upcoming frame $T+1$. Specifically, we validate the effectiveness*

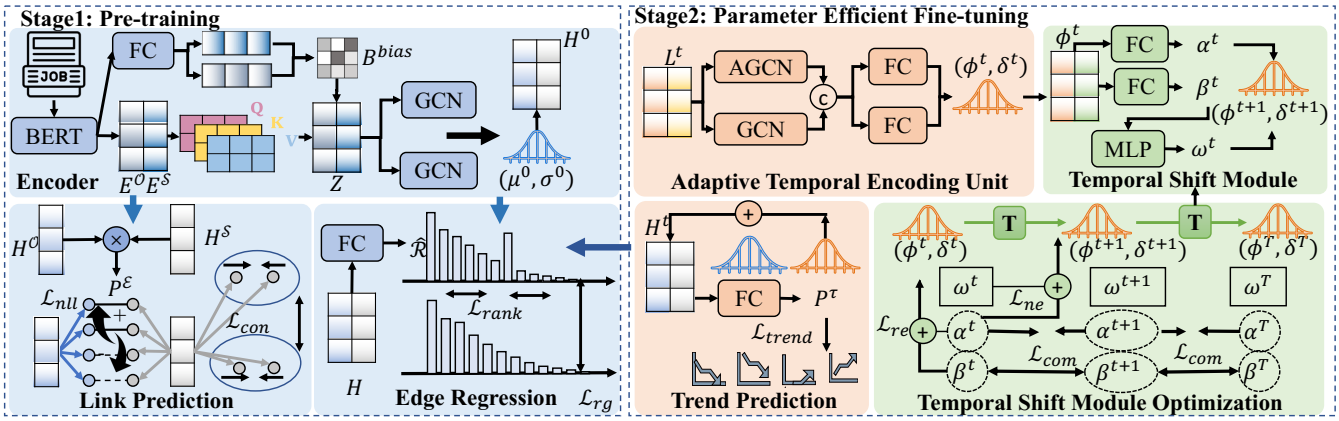

Figure 2: An illustration of our proposed Pre-DyGAE.

of dynamic graph forecasting via two common tasks in dynamic graph learning: (i) dynamic link prediction, which estimates the existence of $e_{o,s}^{T+1}$ between o and s; (ii) dynamic edge regression, which infers the weighted adjacency matrix $\{A_{o,s}^{T+1}|e_{o,s}^{T+1} \in \mathcal{E}^{T+1}\}$.

Solution overview In this paper, we introduce Pre-DyGAE, a two-stage framework comprising pre-training and fine-tuning phases. An overview of our Pre-DyGAE is graphically depicted in Fig. 2. Considering the tradeoff between utilizing data integrity knowledge and capturing evolutionary patterns, we aggregate JDs from initial timestamps $(\mathcal{P}^1, ..., \mathcal{P}^{\Delta t})$ into $\mathcal{P}^0$ to construct the initial graph $\mathcal{G}^0$ for pre-training and fine-tune on the following OSD graphs $(\mathcal{G}^{\Delta t+1}, ..., \mathcal{G}^T)$ for dynamic representations. In the pre-training stage, we employ a GAE to pre-train on $\mathcal{G}^0$, aiming to obtain robust representations. In particular, our encoder incorporates a semantically-aware cross-attention mechanism, utilizing textual information stored in JDs to capture intricate relationships between occupations and skills. Additionally, GCNs are applied to model uncertainty. To guide comprehensive graph reconstruction, we introduce a contrastive loss on skill cooccurrence clusters for sparse data and a unified Tweedie loss and ranking loss for imbalanced prediction. In the fine-tuning stage, we enhance GAE with an adaptive temporal encoding unit and a temporal shift module for DyGAE. This dynamic model is fine-tuned using a two-stage optimization strategy, allowing it to evolve representations from timestamp $\Delta t + 1$ to T. Finally, our framework enables the inference of future representations and the prediction of future OSD.

4 Pre-training Enhanced Graph Autoencoder

In this section, our objective is to devise a pre-training model adept at capturing intricate relationships between occupations and skills to yield robust representations. The architecture consists of an encoder and two decoders. In the multi-task training framework, we jointly pre-train the model in a multi-task framework to guide comprehensive graph reconstruction.

4.1 Encoder

To fully leverage the semantic content of the data, we initial the primary embeddings of occupation o and skill s as E_o and E_s with their corresponding JDs:

$$E_o = \frac{\sum_{p \in \mathcal{P}_o^0} \text{Bert}(p)}{|\mathcal{P}_o^0|}, \; E_s = \frac{\sum_{p \in \mathcal{P}_s^0} \text{Bert}(p)}{|\mathcal{P}_s^0|}, \tag{1}$$

where $\mathcal{P}_s^0$ represents $\{p|s \in p, p \in \mathcal{P}^0\}$, $\text{Bert}(p)$ signifies the embedding of the p derived from a pretrained BERT model.

Semantically-Aware Cross-Attention Besides, a cross-attention mechanism is contributed to further harness the intricacies tied to skills and their representations in occupations. Let's represent the matrices of occupations and skills embeddings as $E^{\mathcal{O}}$ and $E^{\mathcal{S}}$. We obtain the semantically-aware cross-attention enhanced occupation and skill embeddings $Z^{\mathcal{O}}$ as:

$$\begin{aligned} &E^Q = E^{\mathcal{O}} W^Q, \; E^K = E^{\mathcal{S}} W^K, \; E^V = E^{\mathcal{S}} W^V, \\ &B = \frac{E^Q (E^K)^T}{\sqrt{d_K}}, Z^{\mathcal{O}} = \text{softmax}(B) E^V, Z^{\mathcal{S}} = E^{\mathcal{S}}, \end{aligned} \tag{2}$$

where W^Q, W^K, and W^V are trainable parameters and d_K is the dimension of E^K.

Furthermore, we introduce a semantically-aware cross-attention mechanism to further harness the semantic intricacies tied to skills and their representations in occupations. The mechanism can be reformulated as:

$$Z^{\mathcal{O}} = \text{softmax}(B + B^{bias}) E^V, \tag{3}$$

where the bias $B^{bias} \in \mathbb{R}^{|O| \times |S|}$ is learnable, pivotal for instilling deep semantic insights into the representations of occupations and skills. With a cosine similarity function $\text{Sim}(x, x') = \frac{x \cdot x'}{\|x\| \cdot \|x'\|}$, the bias B^{bias} is defined as:

$$B_{o,s}^{bias} = \text{Sim}(\text{FC}_1(E_o), \text{FC}_2(E_s)), \tag{4}$$

where $\text{FC}_i(\cdot)$ is a linear fully connected layer.

Uncertainy-Aware Modeling Furthermore, we utilize structural data to steer the integration of occupation and skill representations by GCN. The aggregation process is:

$$\text{GCN}(A^0, Z; W^0) = f(D^{-\frac{1}{2}}(A^0 + I)D^{\frac{1}{2}} Z W^0), \tag{5}$$

where Z is the concatenation of $Z^{\mathcal{O}}$ and $Z^{\mathcal{S}}$, f is a activation function, I is the identity matrix, W^0 is the trainable node transformation matrix, and D is the degree matrix of $A^0 + I$.

Considering the inherent uncertainty arising from multiple job postings associated with individual occupations, a deterministic representation may risk overfitting. To address this, we model the latent variable H^0 to adhere to a Gaussian distribution for deriving more robust representations:

$$\begin{aligned} H^0 \sim & \mathcal{N}(\mu^0, \sigma^0), \\ \mu^0 = \text{GCN}(A^0, Z; W^\mu), & \ \sigma^0 = \text{GCN}(A^0, Z; W^\sigma), \end{aligned} \tag{6}$$

where $\mathcal{N}(\mu, \sigma)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2, and W^μ and W^σ are two trainable node transformation matrices.

4.2 Decoders

Upon obtaining the representations, the decoders aim to reconstruct the input graph $\mathcal{G}^0$ in two aspects, link prediction to predict the relationships between occupations and skills and edge regression for predicting the specific values.

Link Prediction Decoder We model the likelihood $P^{\mathcal{E}}$ of associating skill s with occupation o as follows:

$$P^{\mathcal{E}}(e^0_{o,s} \in \mathcal{E}^0) = \text{sigmoid}(\text{MeanPooling}(H^0_o \odot H^0_s)), \tag{7}$$

where $\text{sigmoid}(\cdot)$ is an activation function for calculating the probability, $\odot$ denotes the element-wise multiplication operation, and $\text{MeanPooling}(\cdot)$ calculates the average value along the last dimension in the matrix.

Edge Regression Decoder Extending the capabilities of GAE, we utilize a linear layer with an activation function to estimate $\mathcal{R}^0_{o,s}$:

$$\hat{\mathcal{R}}^0_{o,s} = \text{softplus}(\text{FC}_3(H^0_o \odot H^0_s)), \tag{8}$$

where $\text{softplus}(\cdot)$ is an activation function as a smooth approximation of the rectifier linear unit activation function.

Through this approach, GAE decodes H^0 into distinct spaces for predicting both relationships between occupations and skills and OSD, facilitating mutual learning that contributes to the accomplishment of diverse tasks. The prediction within a static time slice also contributes to subsequent dynamic forecasting.

4.3 Pre-training Objectives

The pre-training objective for graph reconstruction involves training the primary tasks of link prediction and edge regression, incorporating two auxiliary tasks. The loss function is composed of a negative log-likelihood loss ($\mathcal{L}^0_{nll}$) and a collaborative-aware contrastive loss ($\mathcal{L}^0_{con}$) for link prediction with sparse data challenge, as well as a Tweedie loss ($\mathcal{L}^0_{rg}$) and a ranking loss ($\mathcal{L}^0_{rank}$) for imbalanced regression.

Negative Log-Likelihood Loss To achieve the accurate reconstruction of the graph edges $\mathcal{E}^0$, the negative log-likelihood loss $\mathcal{L}_{nll}$ is minimized:

$$\begin{aligned} \mathcal{L}^0_{nll} = & - \sum_{e^0_{o,s} \in \mathcal{E}^0} y^0_{o,s} \log(P^{\mathcal{E}}(e^0_{o,s})) \\ & - \sum_{e^0_{o,s} \notin \mathcal{E}^0} (1 - y^0_{o,s}) \log(1 - P^{\mathcal{E}}(e^0_{o,s})), \end{aligned} \tag{9}$$

where $y^0_{o,s} = 1$ if $e^0_{o,s} \in \mathcal{E}$ and $y^0_{o,s} = 0$ otherwise.

Collaborative-Aware Constrative Loss To solve the challenges of predicting relationships between occupations and less frequent skills, we further introduce a collaborative-aware contrastive loss. Specifically, by evaluating skill co-occurrence in $\mathcal{P}^0$, we apply a spectral clustering algorithm to divide the skills into C clusters and C_k is the k-th cluster contains skills with the closest co-occurrence degree. This approach tightens the representation of skills within a single cluster and thus addresses data sparsity, which is defined as:

$$\mathcal{L}^0_{con} = - \sum_{C_k \in C} \log \frac{\sum_{s,s' \in C_k} \exp(\text{sim}(H^0_s, H^0_{s'})/\gamma)}{\sum_{s,s' \in S} \exp(\text{sim}(H^0_s, H^0_{s'})/\gamma)}, \tag{10}$$

where γ is the temperature coefficient.

Tweedie Loss Considering the OSD follows long-tailed distribution, directly employing regression loss functions i.e., MAE, may hinder the model's ability to effectively capture the imbalances present in both the head and tail data in training. Here, we employ the Tweedie loss to address this issue.

$$\begin{aligned} \mathcal{L}^0_{rg} = \sum_{e^0_{o,s} \in \mathcal{E}^0 \cup \mathcal{E}^{neg}} \Bigg(& \frac{max(\mathcal{R}^0_{o,s}, 0)^{2-\eta}}{(1-\eta)(2-\eta)} - \\ & \frac{\mathcal{R}^0_{o,s}(\hat{\mathcal{R}}^0_{o,s})^{1-\eta}}{1-\eta} + \frac{(\hat{\mathcal{R}}^0_{o,s})^{2-\eta}}{2-\eta} \Bigg), \end{aligned} \tag{11}$$

where $\eta \in (1, 2)$ is a hyperparameter defining the power relation between Tweedie distribution mean and variance, and $\mathcal{E}^{neg}$ is the set of randomly sampled negative edges.

Ranking Loss Furthermore, the diverse skill demands of an occupation underscore the unique roles each skill plays. For instance, some foundational skills might be prevalent across different occupations, whereas the low occurrences could indicate some emerging demand. Therefore, we devise a ranking loss, as a pairwise loss [Wang *et al.*, 2023], to better capture the distinctions between skills.

$$\mathcal{L}^0_{rank} = \sum_{\rho^0_{o,s} > \rho^0_{o',s'}} \max(0, \hat{\mathcal{R}}^0_{o,s} - \hat{\mathcal{R}}^0_{o',s'} + \epsilon), \tag{12}$$

where ϵ is the margin value and $\rho^0_{o,s}$ is the rank. Specifically, if $\mathcal{R}^0_{o,s} \geq 0.1$, $\rho^0_{o,s} = 0$; if $0.1 > \mathcal{R}^0_{o,s} > 0.01$, $\rho^0_{o,s} = 1$; otherwise, $\rho^0_{o,s} = 2$.

Finally, the pre-training loss function $\mathcal{L}^0_{GAE}$ is defined:

$$\mathcal{L}^0_{GAE} = \lambda_{nll}\mathcal{L}^0_{nll} + \lambda_{con}\mathcal{L}^0_{con} + \lambda_{rg}\mathcal{L}^0_{rg} + \lambda_{rank}\mathcal{L}^0_{rank} + \mathcal{L}^0_{kl}, \tag{13}$$

where $\mathcal{L}^0_{kl} = \frac{1}{2}\sum_{i,j}({\mu^0_{ij}}^2 + (\sigma^0_{ij})^2 - 2\log(\sigma^0_{ij}) - 1)$ is a KL loss to minimize the discrepancy between the posterior and prior distributions of the latent variables, and λ_{nll}, λ_{con}, λ_{rg} and λ_{rank} are the hyperparameters to balance the effects of different modules.

5 Dynamic Graph Autoencoder

In this section, we incorporate an adaptive temporal encoding unit and a temporal shift module into GAE . Besides, we

fix GAE and apply parameter efficient fine-tuning on DyGAE with a two-step optimization strategy. Finally, we can evolve the dynamic representations underlying a dynamic graph sequence $(\mathcal{G}^{\Delta t+1}, ..., \mathcal{G}^T)$ and achieve OSD forecasting.

5.1 Temporal Feature Learning

Specifically, we design an adaptive temporal encoding unit to facilitate the learning of temporal representations. Additionally, we model the Gaussian distribution of temporal features, enhancing our ability to capture and represent dynamic uncertainties. Notably, we introduce a temporal shift module to effectively explore the temporal evolution of features, thereby achieving dynamic representations.

Adaptive Temporal Encoding Unit At timestamp t, the relationship between occupations and skills hinges on the prevailing graph structure $\mathcal{G}^t$ and the stable graph structure $\mathcal{G}^0$. To harness the wealth of structural information embedded within the relationship between occupations and skills, we employ the dual-view GCN propagation method, integrating it with the specified adjacency matrix A^t and a learnable graph A^a initialized with A^0. With an initial learnable embedding $L^t \in \mathbb{R}^{(|O|+|S|)\times d}$ where d is the feature dimension, the dual-view GCN propagation layer is devised as:

$$\begin{aligned} A^a &= \text{softmax}(\text{Relu}(M^t {M^t}^{\mathrm{T}})), \\ F^t &= \text{FC}_4([\text{GCN}(A^t, L^t; W^t); \text{GCN}(A^a, L^t; W^a)]), \end{aligned} \tag{14}$$

where $M^t \in \mathbb{R}^{(|O|+|S|)\times d}$ is a trainable parameter, F^t is the structure enhanced temporal hidden states, and $[\cdot;\cdot]$ denotes a concanate operation. Furthermore, We encode the temporal feature as a Gaussian distribution:

$$\phi^t = \text{FC}_5(F^t),\ \delta^t = \text{FC}_6(F^t), \tag{15}$$

where $\phi^t \in \mathbb{R}^{(|O|+|S|)\times d}$ and $\delta^t \in \mathbb{R}^{(|O|+|S|)\times d}$.

Dynamic Uncertainty-Aware Modeling While the stable representations capture the essential characteristics of occupations and skills, learned temporal characteristics enrich the representation with time-varying nuances. Leveraging the superposition property of normal distributions, we model the dynamic representations by incorporating the temporal features into the stable initial representation:

$$\mu^t = \mu^0 + \phi^t,\ \sigma^t = \sigma^0 + \delta^t,\ H^t \sim \mathcal{N}(\mu^t, \sigma^t). \tag{16}$$

This strategy allows us to capture intricate temporal patterns and facilitate robust representation learning.

Temporal Shift Module Following the assumption that variance does not depend on time [Gourru *et al.*, 2022], we infer $\hat{\delta}^{t+1} = \sum \delta^t / T$. To explore the evolution of ϕ^t, we first decompose ϕ^t into a stable term $\alpha^t = \text{FC}_8(\phi^t)$ and a trend term $\beta^t = \text{FC}_9(\phi^t)$. To predict ϕ^{t+1}, we add an MLP to evolve the trend term as $\omega^t = \text{MLP}(\beta^t)$. Then we forecast the ϕ^{t+1} as $\alpha^t + \omega^t$.

5.2 Parameter Efficient Fine-Tuning

Leveraging the pre-trained GAE, we initially focus on optimizing the temporal features in each timestamp. Following this, we delve into exploring the evolution pattern. This two-step optimization strategy allows for a comprehensive exploitation of the pre-trained model's advantages and facilitates the accurate capture of evolution characteristics.

Adaptive Temporal Encoding Unit Optimization For dynamic OSD, trend changes usually include four situations: emergence, disappearance, growth, and decline. To explicitly allow dynamic representations to perceive changes in trends, for each o and s, we additionally add a linear layer with an activation function to map temporal features into four trend labels $\tau_{o,s}$ as above.

We estimate the likelihood of the trend label between o and s at t as $P^\tau(o, s, t) = \text{sigmoid}(\text{FC}_7(H_o^t \odot H_s^t))$. The trend-aware loss is defined as:

$$\mathcal{L}_{trend}^t = -\sum_{o\in O, s\in S}\sum_{k=1}^{4} \mathbf{1}(\tau_{o,s} = k)\cdot\log(P_k^\tau(o, s, t)). \tag{17}$$

In each timestamp $t \in [\Delta t + 1, T]$, we fix the GAE and fine-tune the adaptive temporal encoding unit via $\mathcal{L}_{tem}^t = \lambda_{trend}\mathcal{L}_{trend}^t + \mathcal{L}_{GAE}^t$ for the temporal feature distribution (ϕ^t, δ^t), where λ_{trend} is the hyperparameter, $\mathcal{L}_{GAE}^t$ is the pre-training loss in Eq.13 with H^t as the input of decoders and $\mathcal{G}^t$ as the target graph.

Temporal Shift Module Optimization Based on the temporal feature sequence $((\phi^{\Delta t+1}, \delta^{\Delta t+1}), ..., (\phi^T, \delta^T))$, we intend to predict the temporal feature in the next timestamp $(\phi^{T+1}, \delta^{T+1})$. Firstly we reconstruct the $\{\phi^t | t \in [\Delta t+1, T]\}$ with a reconstruction loss $\mathcal{L}_{re}$:

$$\mathcal{L}_{re} = \sum_{t\in[\Delta t+1,T]} \text{MSE}(\phi^t, \alpha^t + \beta^t), \tag{18}$$

where $\text{MSE}(x, x')$ is a function to compute the mean squared error between x and x'. Besides, to achieve next-step prediction, we introduce $\mathcal{L}_{ne}$ as:

$$\mathcal{L}_{ne} = \sum_{t\in[\Delta t+1,T-1]} \text{MSE}(\phi^{t+1}, \alpha^t + \omega^t). \tag{19}$$

Additionally, we formulate the loss λ_{com} to keep the stable terms similar and distinguish trend terms.

$$\mathcal{L}_{com} = \sum_{t,t'\in[\Delta t+1,T], t\neq t'} (\text{Cos}(\beta^t, \beta^{t'}) - \text{Cos}(\alpha^t, \alpha^{t'})), \tag{20}$$

where $\text{Cos}(x, x')$ is a function, defined as $\text{Cos}(x, x') = \frac{1}{m}\sum_{i=1}^{m} \frac{x_i \cdot x_i'}{\|x_i\|\cdot\|x_i'\|}$, to compute the average cosine similarity between x and x' and m is the number of row in x. During the training process, the three loss functions are combined as $\mathcal{L}_{shift} = \lambda_{re}\mathcal{L}_{re} + \lambda_{ne}\mathcal{L}_{ne} + \lambda_{com}\mathcal{L}_{com}$ where λ_{re}, λ_{ne}, and λ_{com} are the hyperparameters.

5.3 Inference Procedure

After the fine-tuning, we can infer the next step temporal features as $\hat{\phi}^{T+1} = \alpha^T + \omega^T$. Combining with the pre-trained representation distributions, we forecast the representation distributions of occupations and skills $\hat{H}^{T+1} \sim \mathcal{N}(\mu^0 + \hat{\phi}^{T+1}, \sigma^0 + \hat{\delta}^{T+1})$. Furthermore, the OSD graph $\hat{\mathcal{G}}^{T+1}$ can be inferred with the inferred likelihood $\hat{P}^{\mathcal{E}}(e_{o,s}^{T+1} \in \mathcal{E}^{T+1}) = \text{sigmoid}(\text{MeanPooling}(\hat{H}_o^{T+1} \odot \hat{H}_s^{T+1}))$ and the inferred edge weight $\hat{A}_{o,s}^{T+1} = \text{softplus}(\text{FC}_3(\hat{H}_o^{T+1} \odot \hat{H}_s^{T+1}))$.

Data	Metric	DCRNN	DyGrEncoder	AGCRN	A3TGCN	TGAT	CAWN	DyGFormer	CHGH	Pre-DyGAE
Dai	AUC%↑	66.42±0.25	72.84±2.89	64.05±0.91	61.91±11.11	90.47±0.09	91.17±0.43	90.83±0.38	70.76±1.18	**96.96±0.04****
	Hits@1%↑	11.99±0.44	9.52±0.88	12.59±0.98	1.93±1.91	38.62±7.50	39.30±1.90	29.05±0.85	21.90±1.56	**42.97±0.33****
	MRR%↑	16.62±1.01	13.60±2.44	17.24±2.11	4.00±6.52	41.29±3.29	43.06±0.05	41.16±0.21	25.00±1.74	**51.92±0.19****
	EGM%↓	0.09±0.00	0.23±0.05	0.09±0.00	0.15±0.08	0.07±0.00	1.12±0.55	1.17±0.97	0.11±0.06	**0.06±0.00****
	MAE%↓	0.74±0.02	0.93±0.05	0.89±0.02	1.10±0.10	1.16±0.00	1.83±0.47	2.02±0.80	0.35±0.01	**0.30±0.04***
	RMSE%↓	3.58±0.06	2.97±0.10	3.31±0.24	4.17±0.13	5.17±0.00	5.05±0.01	5.16±0.02	1.14±0.02	**0.66±0.27***
Fin	AUC%↑	67.58±0.49	70.15±2.02	65.87±0.95	68.39±2.06	92.17±0.20	92.26±0.03	92.26±0.17	71.51±2.54	**96.16±0.09****
	Hits@1%↑	15.08±0.45	15.07±1.85	14.99±0.85	5.11±2.86	46.14±2.66	45.20±5.16	28.25±20.81	23.37±3.84	**48.59±0.34***
	MRR%↑	20.52±1.26	19.05±5.44	20.47±1.38	8.84±6.62	48.91±0.64	48.65±0.72	45.09±3.85	26.73±4.17	**57.89±0.37****
	EGM%↓	0.12±0.00	0.17±0.01	0.14±0.01	0.39±0.08	0.13±0.00	16.18±10.27	1.29±1.06	0.18±0.04	**0.09±0.02****
	MAE%↓	0.67±0.03	0.88±0.01	1.02±0.03	1.58±0.11	1.54±0.00	16.70±10.24	2.71±1.17	0.49±0.05	**0.34±0.06****
	RMSE%↓	3.00±0.14	2.74±0.03	3.45±0.16	4.28±0.23	6.12±0.00	17.19±9.87	6.38±0.26	2.31±0.01	**0.47±0.24****
IT	AUC%↑	68.31±0.64	70.32±1.77	63.86±0.79	64.57±1.25	89.13±0.25	89.79±0.36	89.19±0.14	72.49±2.55	**96.04±0.06****
	Hits@1%↑	14.03±0.77	14.34±0.83	12.45±0.44	2.88±1.27	40.78±2.46	36.50±2.04	38.24±9.30	26.27±4.34	**44.34±0.67****
	MRR%↑	19.63±1.53	20.48±2.21	17.51±1.11	7.52±3.23	46.99±1.05	46.18±1.44	45.58±3.02	29.86±4.58	**52.59±0.45****
	EGM%↓	0.17±0.01	0.27±0.02	0.21±0.01	0.46±0.10	33.42±45.23	7.14±0.93	0.53±0.08	0.21±0.04	**0.11±0.02****
	MAE%↓	0.93±0.04	1.20±0.05	1.34±0.03	1.80±0.06	34.24±44.96	7.79±0.91	1.95±0.05	0.52±0.04	**0.35±0.07****
	RMSE%↓	3.36±0.14	3.92±0.64	4.11±0.10	4.87±0.17	36.33±42.38	8.95±0.71	6.53±0.01	1.39±0.02	**0.73±0.34***
Man	AUC%↑	68.20±0.89	72.73±3.24	65.45±1.39	69.75±1.73	89.18±0.24	90.12±0.11	90.47±0.28	71.14±1.76	**96.92±0.07****
	Hits@1%↑	12.87±0.99	13.74±1.62	14.04±1.71	3.13±1.26	35.07±2.21	41.80±1.51	28.54±1.90	22.47±2.39	**45.62±0.63****
	MRR%↑	17.12±1.73	18.25±3.90	18.40±4.06	6.86±2.44	45.86±0.85	49.03±1.86	44.86±0.80	25.69±2.66	**53.85±0.46***
	EGM%↓	0.12±0.00	0.23±0.06	0.14±0.01	0.32±0.05	0.12±0.00	2.58±1.22	7.69±9.58	0.16±0.03	**0.07±0.02****
	MAE%↓	0.66±0.01	0.89±0.07	0.94±0.04	1.29±0.03	1.31±0.00	3.20±1.14	8.51±9.22	0.40±0.03	**0.30±0.06****
	RMSE%↓	2.42±0.05	2.64±0.14	3.11±0.17	3.96±0.08	5.18±0.00	5.35±0.42	10.29±7.23	1.27±0.02	**0.68±0.18****

Table 1: OSD forecasting performance comparisons across four datasets. Each result was derived from four repetitive experiments. The best results are highlighted in **bold**, while the second-best results are underscored. Asterisks indicate statistical significance: * denotes performance improvement at the 0.05 level, and ** denotes a significant improvement at the 0.01 level, as determined by the paired t-test.

6 Experiments

6.1 Experimental Settings

Datasets The datasets used in our experiments were collected from the public information of one of the largest online recruitment platforms. We gathered JDs spanning a diverse range of occupations, covering the period from January 2020 to December 2023. Along this line, we constructed large-scale datasets from four industries, i.e., the Daily dataset (Dai), Finance dataset (Fin), IT dataset (IT), and Manufacturing dataset (Man). As mentioned above, we used half-year intervals as the basic time step, dividing each dataset into 8 timestamps. In our experiment setup, we selected $\Delta t = 4$, which entailed using JDs from the first 4 timestamps for pre-training and the subsequent 3 timestamps for dynamic temporal feature learning. Besides, more parameter experiments are in the Appendix, and we choose λ_{nll}, λ_{con}, λ_{rg}, λ_{rank}, λ_{trend}, λ_{re}, λ_{ne}, and λ_{com} as 1, 0.05, 100, 0.1, 1, 1, 1, and 1. Meanwhile, we partitioned the JDs from the first month of the final timestamp to create a validation set, while the remaining five months' JDs were used as a test set. Detailed statistics and divisions of our datasets are provided in the Appendix.

Evaluation Metrics We evaluated the effectiveness of our approach via dynamic link prediction and edge regression tasks. Therefore, we employed several link prediction metrics, including AUC [Zha *et al.*, 2023], MRR, and Hits@1 [Wang *et al.*, 2022], and regression metrics, including EGM [Yang *et al.*, 2021], MAE [Zhu *et al.*, 2016], and RMSE [Chang *et al.*, 2014]. In particular, EGM was used to measure the heavy-tailed nature of our data distributions.

Baselines We compared Pre-DyGAE with a state-of-the-art method, i.e., CHGH [Chao *et al.*, 2024]. In addition, we included dynamic graph learning baselines, including DCRNN [Li *et al.*, 2018], DyGrEncoder [Taheri *et al.*, 2019], AGCRN [Bai *et al.*, 2020], A3TGCN [Bai *et al.*, 2021], TGAT [Xu *et al.*, 2020], CAWN [Wang *et al.*, 2021d], and DyGFormer [Yu *et al.*, 2023]. To maintain fairness in comparison, each baseline was trained with data within the former 7 timestamps. Besides, we also compared some traditional methods, but due to not achieving competitive results and page limitations, we have put their experimental results in the Appendix.

6.2 Experimental Results

Performance Comparison Table 1 illustrates the comprehensive performance evaluation of Pre-DyGAE and baselines across four datasets. According to the results, there are several observations. Firstly, our proposed model has significant improvements over all baselines in all datasets, thus demonstrating the effectiveness of our framework in OSD forecasting. Specifically, the AUC metric showcases a remarkable performance of 96%. This confirms its capability to precisely predict future skills across a variety of occupations, a critical aspect for subsequent edge regression. Additionally, our approach demonstrates proficiency by improving the average RMSE in four datasets from 1.527% to 0.635%, an increase of 58% compared with the best-performing baseline, providing precise insights into the distribution of demand, and highlighting its capability to forecast future skill demand accurately. Moreover, the improvement in the EGM directly signifies the competence of our model in predicting unbalanced distribution. Thirdly, baseline models often show imbalanced results in both link prediction and regression metrics, indicating the complexity of integrating these two tasks. In contrast, our model excels in all metrics, demonstrating Pre-DyGAE's ability to improve performance on both tasks. Finally, it is noteworthy that RNN-based methods such as

Metrics	AUC%↑	MRR%↑	EGM%↓	MAE%↓
NGCF	70.15±0.09	14.23±0.10	0.03±0.00	0.51±0.01
LightGCN	76.45±1.10	15.74±1.00	0.34±0.05	0.86±0.08
MultVAE	77.22±0.00	12.08±0.00	0.02±0.00	0.51±0.00
w/o con	93.38±0.06	19.46±0.25	0.09±0.04	0.22±0.08
w/o rank	94.52±0.03	21.19±0.33	0.09±0.07	0.27±0.09
w/o bias	94.37±0.11	21.16±0.31	1.09±0.62	1.49±0.73
L1	94.44±0.15	21.54±0.21	0.06±0.03	0.12±0.02
GAE	**95.56±0.06**	**22.64±0.27**	**0.02±0.02**	**0.09±0.05**

Table 2: OSD graph completion performance on Fin dataset. We introduced four variants of our GAE: "w/o rank", "w/o con", "w/o bias", and "L1". The first three variants denote our model without the $\mathcal{L}_{rank}$ (Eq.12), $\mathcal{L}_{con}$ (Eq.10), and semantically-aware cross-attention, respectively. The last variant indicates the substitution of the Tweedie Loss with L1 loss in Eq.11.

DCRNN and CAWN outperform Transformer-based methods such as DyGFormer in these datasets. This is attributed to the Transformer's primary focus on addressing long-horizon autocorrelation while overlooking variable trends, aligning with the unique characteristics of our forecasting problem.

GAE Performance on OSD Graph Completion We conducted a validation experiment to assess the robustness of GAE's learned representations by predicting missing skill demand within the same time interval. JDs from the initial 4 timestamps were randomly split into training and test sets at a ratio of 50%, with corresponding occupational skill demand in triplet form $(o, s, \mathcal{R}^0_{o,s})$. For comparison, we selected NGCF, LightGCN, and MultVAE as the baselines [He *et al.*, 2020]. The results in Table 2 show that GAE improved AUC and MRR compared to all baselines, indicating effective representations of occupations and skills. Specifically, "w/o bias" significantly impacts EGM and RMSE, demonstrating the importance of the semantically-aware cross attention. The ranking loss distinguishes feature spaces, and contrastive learning on skill clusters enhances metrics, guiding recognition of sparse skills. Additionally, comparing L1 loss and the Tweedie loss suggests that the latter is more suitable for predicting imbalanced distributions. Comprehensive results across all datasets are detailed in the Appendix.

Ablation Studies We performed ablation studies to assess different modules in Pre-DyGAE, as depicted in Fig. 3. Results show Pre-DyGAE's performance compared to various backbones, highlighting GAE's robust representation and its ability to enhance capturing temporal features. Ablation experiments on the temporal module affirm our model's capability to capture demand distribution shifts caused by subsequent temporal changes based on the backbone representation.

Parameter Experiment We have carefully considered the balance between leveraging data integrity knowledge and capturing evolutionary patterns. As shown in Fig. 4, the trend of rising first and then falling suggests the advantages of stable representations for capturing subsequent changes and the necessity of preserving some future time slices for mining evolutionary patterns. The consistency in the number of time frames across different metrics in parameter experiments further supports the effectiveness of our pre-training and fine-tuning framework. More details can be seen in the Appendix.

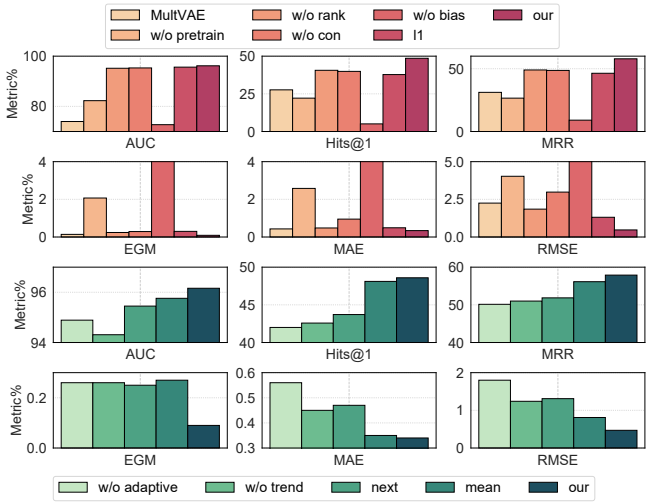

Figure 3: OSD Forecasting Ablation Study on Fin dataset. The red histogram examines the impact of backbone ablation on OSD forecasting, while the green histogram assesses the adaptive temporal encoding unit and temporal shift module's effectiveness. Besides, "w/o pre-train" denotes GAE without pretraining directly applied to learn the temporal features. And "w/o adaptive" and "w/o trendweight" denote the adaptive temporal encoding unit without the learnable graph $\mathcal{W}^a$ (Eq.14) and trend loss $\mathcal{L}_{trend}$, respectively.

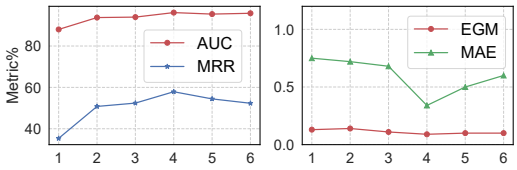

Figure 4: The influence of Δt on Fin dataset. We varied timestamps for aggregating JDs to construct $\mathcal{G}^0$ and pre-trained GAE with Δt ranging from 1 to 6. Pre-DyGAE performed best with $\Delta t = 4$.

7 Conclusion

In this paper, we addressed the critical task of forecasting skill demand in specific occupations, a key aspect in the dynamics of labor markets and economics. Specifically, we introduced Pre-DyGAE, a novel approach involving pre-training on initial timestamps, followed by parameter-efficient fine-tuning in subsequent timestamps. This approach leveraged the overall interaction between occupations and skills over an extended period, enabling the discernment of stable associations between them. The pre-training of GAE incorporated a semantically-aware cross-attention enhanced uncertainty-aware encoder and a multi-task pre-training framework. These components proved effective in establishing fundamental representations. Based on the pre-trained GAE, DyGAE excelled in capturing dynamic temporal features through a two-stage optimization process involving adaptive temporal encoding units and a temporal shift module. Extensive experiments on four real-world datasets have demonstrated the effectiveness of our Pre-DyGAE.

Acknowledgements

This work was partially supported by National Natural Science Foundation of China (Grant No.92370204), Guangzhou-HKUST(GZ) Joint Funding Program(Grant No.2023A03J0008), the National Key Research and Development Program of China (No. 2023YFF0725001), Education Bureau of Guangzhou Municipality, Guangdong Science and Technology Department, and Project funded by China Postdoctoral Science Foundation (Grant No.2023M730785).

References

[Bai *et al.*, 2020] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 2020.

[Bai *et al.*, 2021] Jiandong Bai, Jiawei Zhu, Yujiao Song, Ling Zhao, Zhixiang Hou, Ronghua Du, and Haifeng Li. A3t-gcn: Attention temporal graph convolutional network for traffic forecasting. *ISPRS International Journal of Geo-Information*, 2021.

[Bilan *et al.*, 2020] Yuriy Bilan, Halyna Mishchuk, Iryna Roshchyk, and Olena Joshi. Hiring and retaining skilled employees in smes: problems in human resource practices and links with organizational success. *Business: Theory and Practice*, 2020.

[Bisello *et al.*, 2022] Martina Bisello, Vincenzo Maccarrone, and Enrique Fernández-Macías. Occupational mobility, employment transitions and job quality in europe: The impact of the great recession. *Economic and Industrial Democracy*, 2022.

[Chang *et al.*, 2014] Biao Chang, Hengshu Zhu, Yong Ge, Enhong Chen, Hui Xiong, and Chang Tan. Predicting the popularity of online serials with autoregressive models. In *CIKM*, 2014.

[Chao *et al.*, 2024] Wenshuo Chao, Zhaopeng Qiu, Likang Wu, Zhuoning Guo, Zhi Zheng, Hengshu Zhu, and Hao Liu. A cross-view hierarchical graph learning hypernetwork for skill demand-supply joint prediction. In *AAAI*, 2024.

[Chen *et al.*, 2021] Liyi Chen, Chuan Qin, Ying Sun, Xin Song, Tong Xu, Hengshu Zhu, and Hui Xiong. Collaboration-aware hybrid learning for knowledge development prediction. In *WWW*, 2021.

[Cong *et al.*, 2023] Weilin Cong, Si Zhang, Jian Kang, Baichuan Yuan, Hao Wu, Xin Zhou, Hanghang Tong, and Mehrdad Mahdavi. Do we really need complicated model architectures for temporal networks? In *ICLR*, 2023.

[Fang *et al.*, 2023] Chuyu Fang, Chuan Qin, Qi Zhang, Kaichun Yao, Jingshuai Zhang, Hengshu Zhu, Fuzhen Zhuang, and Hui Xiong. Recruitpro: A pretrained language model with skill-aware prompt learning for intelligent recruitment. In *KDD*, 2023.

[Felten *et al.*, 2018] Edward W. Felten, Manav Raj, and Robert Seamans. A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, 2018.

[Gourru *et al.*, 2022] Antoine Gourru, Julien Velcin, Christophe Gravier, and Julien Jacques. Dynamic gaussian embedding of authors. In *Proceedings of the ACM Web Conference 2022*, New York, NY, USA, 2022.

[Guo *et al.*, 2022a] Pengzhan Guo, Keli Xiao, Zeyang Ye, Hengshu Zhu, and Wei Zhu. Intelligent career planning via stochastic subsampling reinforcement learning. *Scientific Reports*, 2022.

[Guo *et al.*, 2022b] Zhuoning Guo, Hao Liu, Le Zhang, Qi Zhang, Hengshu Zhu, and Hui Xiong. Talent demand-supply joint prediction with dynamic heterogeneous graph enhanced meta-learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020.

[Kumar *et al.*, 2019] Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

[Langley *et al.*, 2019] Ann Langley, Kajsa Lindberg, Bjørn Erik Mørk, Davide Nicolini, Elena Raviola, and Lars Walter. Boundary work among groups, occupations, and organizations: From cartography to process. *Academy of Management Annals*, 2019.

[Li *et al.*, 2018] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR '18)*, 2018.

[Li *et al.*, 2020] Shan Li, Baoxu Shi, Jaewon Yang, Ji Yan, Shuai Wang, Fei Chen, and Qi He. Deep job understanding at linkedin. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020.

[Mahdavimoghaddam *et al.*, 2022] Jalehsadat Mahdavimoghaddam, Ayush Bahuguna, and Ebrahim Bagheri. Exploring the utility of social content for understanding future in-demand skills. *Proc. ACM Hum.-Comput. Interact.*, 2022.

[Pareja *et al.*, 2019] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B. Schardl, and Charles E. Leiserson. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs, 2019.

[Qin *et al.*, 2019] Chuan Qin, Hengshu Zhu, Chen Zhu, Tong Xu, Fuzhen Zhuang, Chao Ma, Jingshuai Zhang, and Hui Xiong. Duerquiz: A personalized question recommender

system for intelligent job interview. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2165–2173, 2019.

[Qin *et al.*, 2022] Chuan Qin, Kaichun Yao, Hengshu Zhu, Tong Xu, Dazhong Shen, Enhong Chen, and Hui Xiong. Towards automatic job description generation with capability-aware neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[Qin *et al.*, 2023] Chuan Qin, Hengshu Zhu, Dazhong Shen, Ying Sun, Kaichun Yao, Peng Wang, and Hui Xiong. Automatic skill-oriented question generation and recommendation for intelligent job interviews. *ACM Transactions on Information Systems*, 2023.

[Qin *et al.*, 2024] Chuan Qin, Le Zhang, Yihang Cheng, Rui Zha, Dazhong Shen, Qi Zhang, Xi Chen, Ying Sun, Chen Zhu, Hengshu Zhu, and Hui Xiong. A comprehensive survey of artificial intelligence techniques for talent analytics. *arXiv preprint arXiv:2307.03195v2*, 2024.

[Rossi *et al.*, 2020] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. In *ICML 2020 Workshop on Graph Representation Learning*, 2020.

[Seo *et al.*, 2018] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured Sequence Modeling with Graph Convolutional Recurrent Networks. In *Neural Information Processing*, 2018.

[Sun *et al.*, 2021] Ying Sun, Fuzhen Zhuang, Hengshu Zhu, Qi Zhang, Qing He, and Hui Xiong. Market-oriented job skill valuation with cooperative composition neural network. *Nature communications*, 2021.

[Sus and Sylwestrzak, 2021] Aleksandra Sus and Bartosz Sylwestrzak. Evolution of the labor market and competency requirements in industry 4.0 versus the covid-19 pandemic. *EUROPEAN RESEARCH STUDIES JOURNAL*, 02 2021.

[Taheri *et al.*, 2019] Aynaz Taheri, Kevin Gimpel, and Tanya Berger-Wolf. Learning to represent the evolution of dynamic graphs with recurrent models. In *Companion Proceedings of The 2019 World Wide Web Conference*, 2019.

[Trivedi *et al.*, 2019] Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. Dyrep: Learning representations over dynamic graphs. In *International Conference on Learning Representations*, 2019.

[Wang *et al.*, 2021a] Chao Wang, Hengshu Zhu, Qiming Hao, Keli Xiao, and Hui Xiong. Variable interval time sequence modeling for career trajectory prediction: Deep collaborative perspective. In *Proceedings of the Web Conference 2021*, 2021.

[Wang *et al.*, 2021b] Chao Wang, Hengshu Zhu, Peng Wang, Chen Zhu, Xi Zhang, Enhong Chen, and Hui Xiong. Personalized and explainable employee training course recommendations: A bayesian variational approach. *ACM Transactions on Information Systems (TOIS)*, 2021.

[Wang *et al.*, 2021c] Lu Wang, Xiaofu Chang, Shuang Li, Yunfei Chu, Hui Li, Wei Zhang, Xiaofeng He, Le Song, Jingren Zhou, and Hongxia Yang. TCL: transformer-based dynamic graph modelling via contrastive learning. *CoRR*, 2021.

[Wang *et al.*, 2021d] Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, and Pan Li. Inductive representation learning in temporal networks via causal anonymous walks. In *International Conference on Learning Representations*, 2021.

[Wang *et al.*, 2022] Xintao Wang, Qianyu He, Jiaqing Liang, and Yanghua Xiao. Language models as knowledge embeddings. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 2022.

[Wang *et al.*, 2023] Chao Wang, Hengshu Zhu, Chen Zhu, Chuan Qin, Enhong Chen, and Hui Xiong. Setrank: A setwise bayesian approach for collaborative ranking in recommender system. *ACM Transactions on Information Systems*, 2023.

[Wu *et al.*, 2019] Xunxian Wu, Tong Xu, Hengshu Zhu, Le Zhang, Enhong Chen, and Hui Xiong. Trend-aware tensor factorization for job skill demand analysis. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019.

[Xu *et al.*, 2020] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs. In *International Conference on Learning Representations (ICLR)*, 2020.

[Yang *et al.*, 2021] Yuzhe Yang, Kaiwen Zha, Ying-Cong Chen, Hao Wang, and Dina Katabi. Delving into deep imbalanced regression. In *International Conference on Machine Learning*, 2021.

[Yu *et al.*, 2023] Le Yu, Leilei Sun, Bowen Du, and Weifeng Lv. Towards better dynamic graph learning: New architecture and unified library. *Advances in Neural Information Processing Systems*, 2023.

[Zha *et al.*, 2023] Rui Zha, Chuan Qin, Le Zhang, Dazhong Shen, Tong Xu, Hengshu Zhu, and Enhong Chen. Career mobility analysis with uncertainty-aware graph autoencoders: A job title transition perspective. *IEEE Transactions on Computational Social Systems*, 2023.

[Zhang *et al.*, 2021] Qi Zhang, Hengshu Zhu, Ying Sun, Hao Liu, Fuzhen Zhuang, and Hui Xiong. Talent demand forecasting with attentive neural sequential model. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021.

[Zhao *et al.*, 2020] Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[Zhu *et al.*, 2016] Chen Zhu, Hengshu Zhu, Hui Xiong, Pengliang Ding, and Fang Xie. Recruitment market trend analysis with sequential latent variable models. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.