Career Mobility Analysis With Uncertainty-Aware Graph Autoencoders: A Job Title Transition Perspective

Rui Zha, Chuan Qin, *Member, IEEE*, Le Zhang, Dazhong Shen, Tong Xu, *Member, IEEE*, Hengshu Zhu, *Senior Member, IEEE*, and Enhong Chen, *Senior Member, IEEE*

***Abstract*— Career mobility analysis aims at discovering the movement patterns of employees across different job positions or grades, which can benefit various human resource-related applications. Indeed, recent studies in this direction mainly focus on modeling individual career trajectories, while the macroguidance for labor market assessment has been largely ignored. To this end, in this article, we propose to study career mobility from a market-driven perspective based on large-scale online professional networks (OPNs). Specifically, we propose an uncertainty-aware graph autoencoders (UnGAEs) framework, which can simultaneously discover potential job title transition patterns and predict job durations. In this phase, we first construct a job title transition graph based on massive career trajectory data from OPNs. Then, considering the inherent uncertainty in career mobility, we introduce a novel uncertainty-aware graph encoder (UnGE) to represent job titles as Gaussian embeddings. Furthermore, we design two task-specific decoders that can preserve the asymmetric relationships between job titles, namely the gravity-inspired decoder (GID) and the energy-inspired decoder (EID), for predicting potential transition patterns and corresponding duration, respectively. In particular, both tasks are modeled through a specially designed multitask learning approach. Finally, extensive experiments on a real-world dataset clearly demonstrate the effectiveness of UnGAE compared with state-of-the-art baselines, as well as some potential applications such as job title benchmarking and career path planning.**

***Index Terms*— Career mobility analysis, graph autoencoders, job title transition.**

I. Introduction

IN THE past decades, career mobility has become increasingly prevalent and diverse in the fast-evolving business environment [1]. Analysis in this direction aims at discovering the movement patterns of employees across different job positions or grades, which benefits various human resource-related applications, such as salary analysis and job performance assessment [2], [3], [4]. Nevertheless, traditional approaches to career mobility analysis rely heavily on the experiences of domain experts and limited market survey data, making it difficult to handle dynamic scenarios with timely and fine-grained requirements.

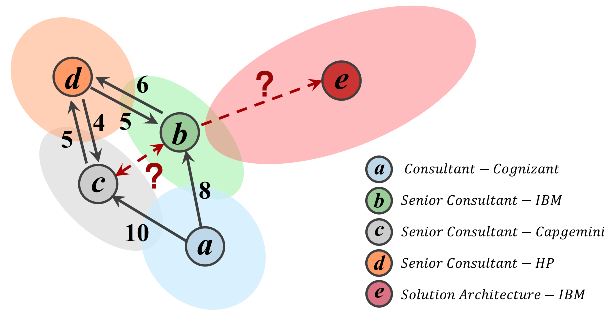


Fig. 1. Illustration of job title transition, where each node corresponds to a job title and the edge indicates the transition pattern. The average duration over the edge is characterized by quarters, while the shaded region reflects the range of the Gaussian distribution.

Recently, the rapid prevalence of online professional networks (OPNs) has enabled the accumulation of massive career trajectory data, providing an unparalleled opportunity for data-driven career mobility analysis. However, existing research mainly focuses on modeling the individual career mobility [5], [6], [7], which fails to provide macroguidance for labor market assessment. To tackle this limitation, we attempt to investigate career mobility from a market-driven perspective based on large-scale job title transition data. An illustrative example is shown in Fig. 1. In particular, we distance ourselves from previous studies by simultaneously predicting the potential job title-level transition patterns and their corresponding durations, contributing to a fine-grained investigation into the relationships of job titles.

Nevertheless, there still exist some critical challenges which lie in two major aspects. First, in terms of a specific job title, it may play different roles under various transition patterns. As a result, the representation of job titles would be uncertain according to distinct interacting neighbors. Specifically, we illustrate the example in Fig. 1 to explain this.

Manuscript received 12 September 2022; revised 12 December 2022; accepted 8 January 2023. Date of publication 17 February 2023; date of current version 31 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62222213 and Grant 62072423 and in part by the University of Science and Technology of China (USTC) Research Funds of the Double First-Class Initiative under Grant YD2150002009. *(Corresponding authors: Chuan Qin; Le Zhang.)*

Rui Zha, Dazhong Shen, Tong Xu, and Enhong Chen are with the School of Computer Science, University of Science and Technology of China, Hefei 230027, China (e-mail: zr990210@mail.ustc.edu.cn; sdz@mail.ustc.edu.cn; tongxu@ustc.edu.cn; cheneh@ustc.edu.cn).

Chuan Qin is with the Talent Intelligence Center, Baidu Inc., Beijing 100085, China (e-mail: chuanqin0426@gmail.com; zhuhengshu@gmail.com).

Le Zhang is with Baidu Research, Baidu Inc., Beijing 100085, China (e-mail: zhangle0202@gmail.com).

Hengshu Zhu is with the Career Science Laboratory, BOSS Zhipin, Beijing 100028, China (e-mail: zhuhengshu@gmail.com).

Digital Object Identifier 10.1109/TCSS.2023.3239038

Suppose that each job title can be characterized with a fixed point in the latent space, then *Senior Consultant—IBM* and *Senior Consultant—Capgemini* should be closer because they are in a similar interactive context determined by *Consultant—Cognizant* and *Senior Consultant—HP*. Accordingly, career mobility between these two job titles will be more likely to occur. However, as employees tend to choose jobs with higher levels in a career ladder [8], *Senior Consultant—IBM* is actually interacting with *Solution Architecture—IBM* instead. In this case, certain representations of the job titles fail to capture this employee's preference for career mobility. Second, the relationship between two job titles is complicated due to the directional diversity of job title transitions. As shown in Fig. 1, while a transition pattern exists from *Consultant—Cognizant* to *Senior Consultant—IBM*, the reverse one may not be allowed. Furthermore, even though there are strong two-way interactions between job titles *Senior Consultant—IBM* and *Senior Consultant—HP*, their required durations may also be completely different. Therefore, how to preserve such fine-grained asymmetric relationships in career mobility remains a nontrivial problem.

To address the above challenges, in this article, we propose an uncertainty-aware graph autoencoders (UnGAEs) framework for career mobility analysis, which can simultaneously discover potential job title transition patterns while also predicting the related job durations. Specifically, we first construct a job title transition graph based on the large-scale career trajectory data from OPNs, where each edge corresponds to the crowd transition pattern and the weight denotes the average duration. On such basis, our tasks can be formulated as link prediction and edge weight forecasting over the graph. Then, we introduce a novel uncertainty-aware graph encoder (UnGE) to model the job title interactions and yield comprehensive embeddings. In this phase, the Gaussian distribution is introduced to cope with the uncertain relationships under various neighbor perspectives. Along this line, we further develop two task-specific decoders, namely the gravity-inspired decoder (GID) and the energy-inspired decoder (EID), for predicting transition patterns and job durations, respectively. In particular, these two decoders are physics-inspired to preserve the asymmetric relationships between job titles. Finally, both tasks are modeled by a specially designed multitask learning paradigm.

In summary, the key contributions of our work can be summarized as follows.

1) We provide a fine-grained analysis of career mobility from the job title level. To this end, a novel framework model called UnGAE is proposed to simultaneously discover the potential transition patterns and forecast corresponding durations, leading to deeper insights into job title transitions.
2) We notably consider the uncertainty and asymmetry nature arising in the job title interactions. Specifically, we equip the message-passing scheme with Gaussian distributions to capture the inherent uncertainty. For the asymmetric relations, we further developed two physics-inspired decoders to make predictions.
3) We conducted extensive experiments on real-world datasets to demonstrate the effectiveness of our proposed model compared with state-of-the-art baselines. Moreover, several downstream applications also validate the significance of our study, such as job title benchmarking and career path planning.

A. Overview

The rest of this article is organized as follows. We first summarize the important related work in Section II. Then, in Section III, we describe the real-world data and suggest a strategy to aggregate the job titles. Along this line, we formally define our problem. Section IV shows the technical details and motivations of our proposed model. Moreover, we report the experimental setup and results in Section V. Besides, several downstream applications are also conducted in this phase. Finally, Section VI concludes the work.

II. Related Work

Here, we summarize the related work into two categories, namely career mobility analysis and network embedding.

A. Career Mobility Analysis

Career mobility analysis is one of the most important tasks in labor market assessment, which aims to discover the movement patterns across different job positions or grades. Traditional sociology-related research on career mobility analysis mainly focuses on observing and explaining the laws of mobility behavior [9]. For instance, Ng et al. [10] discussed the influence of different internal and external factors on individual career mobility. Lyons et al. [11] analyzed the changing nature of careers among recent generations. Lam et al. [2] explored the relationship between external career mobility and salary attainment. However, these approaches rely heavily on survey data and empirical analysis. Recently, the rapid growth of career trajectory data produced on OPNs has enabled a new paradigm for career mobility analysis with a data-driven perspective. On such basis, Li et al. [5] first proposed to integrate the personal career trajectory data and profile context to predict the employee's next career move. Meng et al. [6] later designed a hierarchical career-path-aware neural network to model personal career mobility. Moreover, Wang et al. [12] developed a career trajectory prediction framework by modeling the career path as a variable interval time series. Yet, all these methods are limited to the individual level and fail to provide macroguidance for labor market assessment.

To gain comprehensive insights into career mobility patterns, substantial effort has been posed to model mobility with a job title perspective, which can facilitate wide-range job-oriented applications. For instance, Xu et al. [13] utilized a Gaussian–Bayesian network to infer the posterior job title rank based on job trajectories. To match the job titles with similar responsibilities, Zhang et al. [14] further purposed a multiview representation learning method (Job2Vec) to obtain a comprehensive embedding for each job. However, the asymmetry and uncertainty in career mobility are largely neglected in existing research. To tackle these limitations, we propose a fine-grained approach for exploring the relationship across pairwise job titles. Specifically, for two given job titles, we attempt to identify the potential mobility pattern between them, as well as the corresponding duration.

B. Network Embedding

Network embedding is proposed to map the network into a low-dimensional space while preserving its inherent properties. A large effort has been made in this field, such as the matrix factorization-based models [15], [16] and random-walk-based models [17], [18]. Recently, representation learning with graph neural networks (GNNs) receive considerable attention. The domain paradigms can be divided into two categories, that is, convolutional GNNs (ConvGNNs) and graph autoencoders (GAEs) [19]. Typically, ConvGNNs represent each node following the idea of aggregating its neighbors' information. For instance, graph convolutional networks (GCNs) [20] and graph-SAGE [21] employ different aggregation strategies to obtain node representations. In addition, GAEs' purpose is to reconstruct the graph using the encoded information of each node. For example, GAEs leverage GCNs to encode both node structural and feature information, while variational GAE is a variational version of it [22]. Despite this, some research studies further pay attention to the long-range information propagation in graph models. For example, wave [23] is proposed with the equipment of recurrent techniques, and graph WaveNet [24] attempts to capture the long-range temporal sequences through an adaptive dependency matrix.

Nevertheless, most approaches in graph representation learning embed nodes as point vectors, failing to handle the uncertainties that may arise in real-world networks [25]. To alleviate this limitation, several approaches are developed from different angles, which can be substantially categorized into three main branches. The first one focuses on the uncertain knowledge graph, in which each relation fact is associated with a confidence score [26], [27]. They analyzed the uncertain relationships in knowledge formation to enhance the inference process. To provide explicit quantification, a number of research further explore the node-level uncertain labels in the graph classification problem [28], [29]. They investigate the uncertainties associated with class probabilities to minimize misclassification risk. In addition, there are also some studies that consider the implicit uncertainty with probabilistic representation techniques [30], [31], [32], which are close to our case. In this phase, considering the lack of discriminative information and preference diversity, Graph2Gauss [30] first represents each node as a Gaussian distribution with a personalized ranking formulation. Deep variational network embedding (DVNE) [31] captures the node uncertainty via Gaussian embeddings and projects them into Wasserstein space. Different from these studies, in this article, our UnGAE framework is tailor-designed for career mobility analysis. We take crowd diversity into consideration and equip the vanilla message-passing scheme with probabilistic representation.

III. Preliminaries

In this section, we first describe the dataset used in our article. Then, an aggregation strategy will be introduced to handle the noise in job titles. Finally, we formally define our prediction problem in the career mobility analysis.

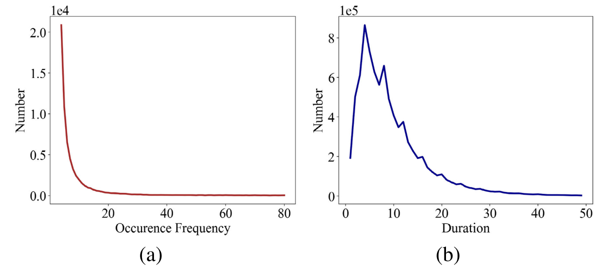


Fig. 2. Distribution of occurrence frequency and duration for each job title transition pattern. (a) Occurrence frequency. (b) Duration.

TABLE I
Statistical Details of Datasets

Dataset	IT	Finance	Merge
# Companies	565	560	1,125
# Pre-aggr. Titles	1,790,934	1,539,295	3,390,641
# Post-aggr. Titles	347,714	380,131	682,552
# Transitions	4,253,299	3,640,158	8,481,179

A. Data Description

The real-world data was collected from one of the largest online professional social networks, that is, LinkedIn. The platform allows both workers and employers to create profiles, including work and education information. Since we aim to investigate job title-level career mobility, the work experiences are extracted from the profiles of employees. Specifically, the data comprises the job titles as well as the beginning and ending dates of each employment history. To validate the robustness of our model, we conducted three datasets from different industries: IT, finance, and a merged one that covers the career mobility records across these two areas. Moreover, we analyzed the occurrence frequency and the duration distribution in terms of each pattern of job title transition over the merged dataset, as shown in Fig. 2. Here, we measured the job duration by quarter. Given that the career mobility records are sparse, studying the potential transition patterns is critical for gaining a comprehensive view of the labor market.

B. Job Title Aggregation

Typically, a job title contains two key components, that is, the company and the job position. In real-world job-related records, the job positions are usually hand-crafted and several distinct records may actually refer to the same job, contributing to some critical issues such as chaos and noise. To this end, we devised a *job position alignment strategy* for filtering out the redundant information. In general, the name of a job position can be split into three primary components as follows.

1) *Responsibility:* It serves as the core part of a job position and describes the responsibilities directly, such as "manager," "supervisor," denoted as *RES*.

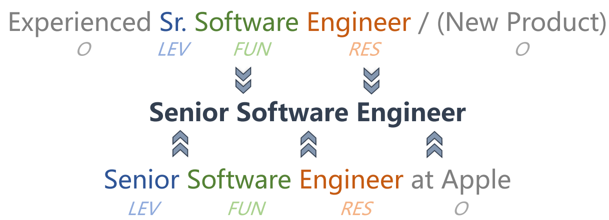


Fig. 3. Example of job title aggregation, in which *Experienced Sr. Software Engineer/(New Product)* and *Senior Software Engineer at Apple* are merged.

2) *Function:* It reflects diversified business functions, such as "sales," "r&d," denoted as *FUN*.
3) *Level:* It reveals the level of the position, such as "junior," "senior," denoted as *LEV*.

Except for these, the other words can be seen as immaterial information, such as "experienced," "at apple," denoted as O. Fig. 3 depicts how we aggregate these nonstandard job positions. Specifically, we first got unified expressions of job positions with the guidance of professional prior matching rules. For instance, "r&d" and "research and development" were aligned to reduce the sparsity of the data. Then, as we mentioned above, *RES* is the core part of job positions and we only kept the valid titles that contained *RES* words. Along this line, we further extracted the keywords comprised of *RES*, *FUN*, and *LEV* words in terms of each job position. Finally, different positions can be aligned if they hold the same keywords. In this case, job positions *Experienced Sr. Software Engineer/(New Product)* and *Senior Software Engineer at Apple* both refer to the *Senior Software Engineer* according to our strategy.

In practice, we also introduced manual annotation to validate the effectiveness of our *job position alignment strategy*. Specifically, 500 test cases are sampled from the IT and finance datasets, respectively. Then, we invited three human resource specialists to score on the aggregated job titles and calculated the average accuracy based on their evaluation results. As shown in Table I, the sparsity of job titles has decreased obviously after the aggregation operation. Besides, the accuracy of our strategy is consistently greater than 0.88 across all datasets, which further demonstrates the reliability of our strategy in filtering out the noise in job titles.

C. Problem Formulation

We first introduce how to obtain the job title transition graph. Specifically, we denote a career trajectory as $\mathcal{S}(u) = \{s_i | 1 \leq i \leq |\mathcal{S}(u)|\}$, where $|\mathcal{S}(u)|$ is the total number of experiences contained in $\mathcal{S}(u)$, and s_i is one work experience. Moreover, we denote s_i as $s_i = \{c_i, p_i, d_i\}$, where c_i, p_i, and d_i stand for the company, job position, and duration, respectively. Subsequently, each pair of adjacent work experiences can define a directed edge from $\{c_i, p_i\}$ to $\{c_{i+1}, p_{i+1}\}$, referring to a job title transition pattern. Since the duration needed may vary a lot in terms of a such transition pattern for different employees, we further utilize the average duration to represent the edge weight. For the confidence consideration, we only take the pair with an occurrence frequency greater than five into account in all trajectories. As a result, we can integrate the whole career trajectories and construct a novel job title transition graph, which is defined as follows.

Definition 1 (Job Title Transition Graph): The job title transition graph is defined as a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where each job title node $v_i \in \mathcal{V}$ refers to a standardized job position affiliated with the specific company, and each edge $e_{ij} \in \mathcal{E}$ indicates the job title transition pattern from v_i to v_j, while the weight w_{ij} represents the average job duration over this transition pattern.

Based on the job title transition graph, our problem can be defined as follows.

Definition 2 (Problem Definition): Given the observed job title transition graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with its corresponding weighted adjacency matrix A as training data, we aim to learn a mapping function f with a pair of nodes (v_i, v_j) as the input to estimate the existence of transition e_{ij} from v_i to v_j and the related job duration w_{ij} if existing.

IV. Technical Details

In this section, we introduce our approach called UnGAEs for career mobility analysis, which is shown in Fig. 4. Specifically, UnGAE captures the uncertainty and asymmetry in career mobility by utilizing a Gaussian distribution to characterize the role of the central node from the perspective of its neighbors. On such basis, two physics-inspired decoders are developed to discover potential job transition patterns and forecast corresponding durations.

A. Uncertainty-Aware Graph Encoder

In this phase, we introduce the UnGE module for learning the comprehensive representation of job titles in the transition graph $\mathcal{G}$. Following the existing methods where GCN has achieved remarkable performance on graph learning problems [19], we first adopt it to preserve the topological information of the data before and after mapping to a low-dimensional latent space. Specifically, given the initial feature matrix X of nodes, the aggregation process can be defined as follows:

$$H = \rho\big(D^{-1}(A + I)XW\big) \quad (1)$$

where I denotes the identity matrix, W denotes the trainable transformation matrix, and $\rho(\cdot)$ denotes the nonlinear activation function, such as rectified linear unit (ReLU). Considering the aforementioned asymmetry of career mobility, we also replace the usual symmetric normalization with the diagonal out-degree matrix $D = A + I$, where the element $D_{i,i}$ equals $\sum_j w_{ij} + 1$.

However, the uncertainty arising from the complex interactions between distinct job titles is largely ignored in this way. For a specific job title, the employee's preference may differ a lot according to distinct interacting neighbors. Consequently, the roles of job titles would vary greatly under different transition patterns, leading to the incapability of a single point in latent space for job title representation.

To tackle this limitation, we introduce a random variable $z_{i|j}$ for each directed edge e_{ij}, indicating the representation of

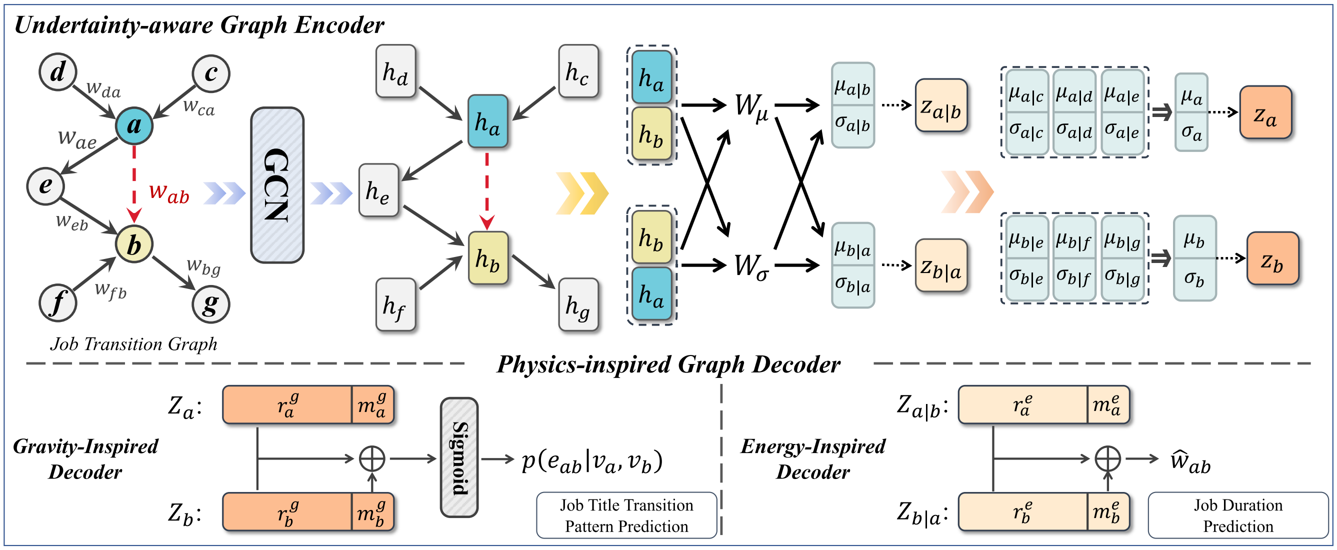


Fig. 4. Illustration of the proposed UnGAE framework, which consists of two components. The first one is UnGE, which is used to learn the comprehensive representation of job titles. The second one is a physics-inspired graph decoder, which is leveraged for discovering potential job title transition patterns and predicting job durations.

node v_i under the specific transition pattern e_{ij}. Moreover, $z_{i|j}$ is subject to the Gaussian distribution

$$z_{i|j} \sim \mathcal{N}\left(\mu_{i|j}, \sigma_{i|j}^2\right) \tag{2}$$

where $\mathcal{N}(x|\mu, \sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2. Specifically, we derive the mean vector $\mu_{i|j}$ and standard variance vector $\sigma_{i|j}$ based on the representation of both the source node and target node in edge e_{ij}

$$\mu_{i|j} = W_\mu\left[h_i; h_j\right], \quad \log \sigma_{i|j} = W_\sigma\left[h_i; h_j\right] \tag{3}$$

where W_μ, W_σ are two trainable transform matrices and $[\cdot;\cdot]$ denotes the concatenation operation. Likewise, we can obtain the conditional distribution of the node v_j from the perspective of the node v_i against the direction of edge by exchanging the symbols i and j in (2) and (3).

B. Physics-Inspired Graph Decoder

Next, we will leverage the representation learned from the UnGE module to serve downstream predicting applications. Specifically, we design two decoder schemes to handle different tasks inspired by physics theories.

1) Gravity-Inspired Decoder: In terms of job title transition pattern prediction, we focus on inferring the existence of directed edge for two given nodes v_i and v_j, that is, $p(e_{ij}|v_i, v_j)$. Based on the UnGE module, we can obtain the conditional representation $z_{i|j}$ of node v_i under the specific job title transition pattern e_{ij}. Then we aggregate all the conditional representation of node v_i to generate the overall representation, which is $z_i = (1/|ne(i)|)\sum_{j\in ne(i)} z_{i|j}$, where $ne(i)$ denotes neighbor set of node v_i. It is notable that despite the consideration of central node v_i while computing the parameters of $\mathcal{N}(\mu_{i|j}, \sigma_{i|j}^2)$, the variables $z_{i|j}$ are still independent of each other. As a result, since the sum of two independent normally distributed random variables is still subject to the Gaussian distribution, we can obtain the independent distribution of the latent variable z_i as follows:

$$\begin{aligned} z_i &\sim \mathcal{N}\left(\mu_i, \operatorname{diag}\left({\sigma_i}^2\right)\right) \\ \mu_i &= \sum_{j\in ne(i)} \frac{\mu_{i|j}}{\sqrt{|ne(i)||ne(j)|}} \\ {\sigma_i}^2 &= \sum_{j\in ne(i)} \frac{\sigma_{i|j}^2}{|ne(i)||ne(j)|}. \end{aligned} \tag{4}$$

To get the overall representation z_i, we opt for the "reparameterization trick" [33]. Specifically, we first sample the noise from the zero-mean Gaussian distribution, that is, $\mathcal{N}(0, I)$, then the representation z_i can be reparameterized by $z_i = \mu_i + \epsilon \odot \sigma_i$.

As we mentioned before, the job title transition is asymmetric, which means that the existence of edge e_{ij} and its reverse one e_{ji} are independent of each other. Commonly used methods fail to handle such situations due to the symmetry of the graph decoders, such as inner product operation [22], [34]. Inspired by [35], we instead adopt an asymmetric decoder alternatively to build a bridge between the job title transition pattern and Newton's theory of universal gravitation. Specifically, job title v_i is treated as a celestial object in space with mass m_i and location r_i. According to Newton's second law of motion, pairwise nodes (v_i, v_j) will be attracted to each other and the acceleration of node v_i in the direction of node v_j can be derived as follows:

$$a_{i\to j} = \frac{Gm_j}{R^2} \tag{5}$$

where G is the gravitational constant and R is the distance between v_i and v_j. We note that the less massive node is more likely to move to the greater one, while the larger the distance between nodes, the lower the acceleration. Intuitively, the acceleration equation can also describe the transition process between job titles. In the job title transition graph, the job title at the higher level usually shows more attractiveness, which can be analogized to the more massive object in the acceleration equation. In contrast, a large distance between two job titles demonstrates lower similarity and a smaller probability of the transition. Along this line, we utilize the acceleration of node v_i toward v_j caused by gravity to indicate the likelihood $p(e_{ij}|v_i, v_j)$. As long as the mass of v_i and v_j are different, the value of $a_{i\to j}$ and $a_{j\to i}$ will be quite different so that the asymmetry of job title transition can be captured.

In practice, we split the embedding z_i with dimension d into two components, where the last dimension represents the *mass* parameter, namely m_i^g. And the rest part indicates the *location* parameter, namely r_i^g. Then the distance R between v_i and v_j is defined as the Euclidean distance of the corresponding *location* parameters, that is, $\|r_i^g - r_j^g\|_2^2$. To ease the training process, we further choose $log(a_{i\to j})$ with an activation function to estimate $p(e_{ij}|v_i, v_j)$, thanks to the nice mathematical properties of taking the logarithm. Specifically, we estimate the possibility of potential transition from node v_i to v_j as follows:

$$p\big(e_{ij}|v_i, v_j\big) = \sigma\left(\underbrace{\log\Big(Gm_j^g\Big)}_{\tilde{m}_j^g} - \lambda_g \log\Big(\|r_i^g - r_j^g\|_2^2\Big)\right) \quad (6)$$

where $\sigma(\cdot)$ is the sigmoid activation function and we substitute $log(Gm_j^g)$ with $\tilde{m}_j^g$ for the ease of computations. Besides, we introduce hyperparameter $\lambda_g \in \mathbb{R}^+$ to balance the relative importance of mass parameter m_j and distance.

Afterward, we can compute the negative log-likelihood of all observed job transition edges given the representation set Z of all nodes $\mathcal{V}$ by the following reconstruction loss:

$$\mathcal{L}_g = -\sum_{e_{ij}\in\mathcal{E}} \log p\big(e_{ij}|v_i, v_j\big) - \sum_{e_{ij}\notin\mathcal{E}} \log\big(1 - p\big(e_{ij}|v_i, v_j\big)\big). \quad (7)$$

2) Energy-Inspired Decoder: In this phase, we aim to predict the job duration w_{ij} for a potential transition pattern e_{ij}. Indeed, the duration required from one job title to another reflects the difficulty of the specific movement. A longer duration usually indicates a greater gap between two jobs, which motivates us to draw an analogy in the job title transition with the energy transfer process of lifting an object from v_i up to v_j. Specifically, we estimate the job duration w_{ij} according to the energy required for moving v_i to v_j as follows:

$$w_{ij} = gm_i^e\|r_i^e - r_j^e\|_2^2. \quad (8)$$

Similar to the GID, here, g is the fixed gravity factor and $m_i^e \in \mathbb{R}^+$ denotes the mass parameter of the source node v_i. And $r_i^e, r_j^e \in \mathbb{R}^{d-1}$ denote the location parameters of the source node v_i and target node v_j, respectively. It is worth noting that m_i^e and r_i^e are split by the representation of node v_i condition on the job title transition pattern e_{ij}, that is, $z_{i|j} \in \mathbb{R}^d$, rather than the overall representation z_i. The rationale behind this is that the potential motivations of job title transition may vary a lot for distinct targets, even starting with the same job title node, leading to the impact on the duration needed. As a result, we use conditional representation for each edge independently.

To facilitate the calculation and optimization, we take the logarithm to limit the potential large values resulting from energy toward very marginal nodes with great massive. Besides, we introduce an additional hyperparameter $\lambda_e \in \mathbb{R}^+$ to balance the relative importance of the node distance in the embedding with respect to the mass parameter. Finally, the predicted job duration $\hat{w_{ij}}$ for the transition pattern e_{ij} is defined by

$$\hat{w}_{ij} = \underbrace{\log\big(m_i^e g\big)}_{\tilde{m}_i^e} + \lambda_e \log\big(\|r_i^e - r_j^e\|_2^2\big) \quad (9)$$

where $\tilde{m}_i^e = \log(m_i^e g)$ is used to simplify calculation.

Afterward, we can evaluate the performance of the job duration forecasting by the root mean square error (RMSE) loss

$$\mathcal{L}_e = \sqrt{\frac{\sum_{e_{ij}\in\mathcal{E}} \|\hat{w}_{ij} - w_{ij}\|^2}{|\mathcal{E}|}}. \quad (10)$$

C. Model Learning

The goal of UnGAE is to learn comprehensive representations for each job title, which can help to infer the potential transition pattern and, at the same time, estimate the corresponding duration. Intuitively, these two prediction tasks are correlated with each other, while the lower possibility of a job title transition pattern usually indicates a longer needed duration. Along this line, we learn two tasks jointly, and the overall optimization objective is defined as follows:

$$\mathcal{L} = \mathcal{L}_g + \mathcal{L}_{kl} + \lambda\mathcal{L}_e \quad (11)$$

where $\mathcal{L}_{kl}$ denotes the Kullback–Leibler divergence between the independent distribution of each node with the Gaussian prior $p(z) = \mathcal{N}(0, I)$, which can be computed analytically. Here, $\mathcal{L}_{KL}$ can be regarded as a regularization to close the distance between posterior and prior for the latent variables. Actually, the first two terms are the same as the loss function in [22], which aims to model the existence of edges in graph data based on a Gaussian latent variable model, but without consideration of the weight on each edge. In addition, the hyperparameter λ is used to balance the importance of the transition pattern and job duration prediction tasks. By minimizing the overall objective, the parameters of the UnGAE framework can be optimized.

V. Experiments

A. Experimental Setup

1) Dataset: Our datasets used in the experiments were collected from a well-known OPN.[1] As mentioned above, we first collected a total of 1125 companies from distinct industries. Then, we extracted the job titles from each company and leveraged the corresponding career trajectory data to construct three job title transition graphs as our datasets, that is, the IT dataset, the Finance dataset, and the merged one. The detailed

[1] https://github.com/zruiii/UnGAE

TABLE II
STATISTICS OF JOB TITLE TRANSITION GRAPHS

Datset	Number of nodes	Number of edges	Percentage of reciprocity
IT	7,077	26,744	16.77%
Finance	6,115	17,685	19.28%
Merge	13,232	45,768	17.90%

statistics of our datasets are shown in Table II. Note that the merged dataset contains more edges due to the interactions between job titles from different fields.

2) Baselines: We compared UnGAE with some state-of-the-art methods. Specifically, in this article, we aim to simultaneously infer potential job title transition patterns and estimate related durations, which can be formulated as link prediction and edge weight forecasting tasks on the job transition graph. Since there are few models for edge weight prediction, we picked six network embedding techniques and three recommendation methods alternatively as follows.

1) *DeepWalk* [17] learns node representations by treating random walks in a graph as the equivalent of sentences.
2) *Node2Vec* [18] extends DeepWalk with a biased random walk strategy to better explore the structure of graphs.
3) *Wave* [23] considers long-range message propagation in graph models with proposed wave procedure.
4) *Job2Vec* [14] proposes a multiview representation learning framework to represent job positions for the job title benchmarking task.
5) *Variational Graph Auto-Encoder (VGAE)* [22] learns interpretable latent representations for undirected graphs based on the variational autoencoder.
6) *Gravity* [35] is an extension of VGAE for directed network embedding, which is inspired by Newton's theory of universal gravitation.
7) *Graph2Gauss* [30] represents nodes as Gaussian distributions and leverages the natural ordering of nodes via a personalized ranking formulation.
8) *Digraph Inception Convolutional Network (DiGCN)* [36] extend spectral-based graph convolution to the directed graph and further simplify it with personalized PageRank.
9) *PMF* [37] is a traditional recommendation algorithm, in which users and items are represented as latent factors.
10) *Graph Convolutional Matrix Completion (GCMC)* [38] explores matrix completion in recommendation systems from the view of link prediction on graphs.
11) *Neural Graph Collaborative Filtering (NGCF)* [39] adopts GNN layers on the user–item interaction graph to refine user and item representations based on multihop neighbors' information.

Considering that some baseline methods cannot be directly applied to these two tasks simultaneously, we have made some adjustments to them. Specifically, we extended VGAE with a *source/target vectors* paradigm [35] to handle the asymmetric graph. Considering the d-dim latent vector returned from the encoder, we assume that the first $(d/2)$-dim denotes the source vector of node i, that is, $z_i^{(s)} = z_{i[1:(d/2)]}$. Likewise, the rest of the components correspond to the target vector, that is, $z_i^{(t)} = z_{i[(d/2)+1:d]}$. Afterward, we replace the symmetric decoder with $\hat{A}_{ij} = \sigma(z_i^{(s)T} z_i^{(t)})$. Moreover, we adopted DeepWalk, Node2Vec, Wave, VGAE, Graph2Gauss, DiGCN, and Gravity to handle the second task through the nonlinear transformation [i.e., two-layer multilayer perceptron (MLP)]. In addition, both DeepWalk and Node2Vec are trained from five random walks of length ten per node, with a window size of ten. We consider the information propagation within the tenth-order range. Furthermore, Graph2Gauss, VGAE, DiGCN, and Gravity were trained with similar settings with respect to our proposed UnGAE. In NGCF, we set the number of embedding propagation layers to two and the dropout to 0.4.

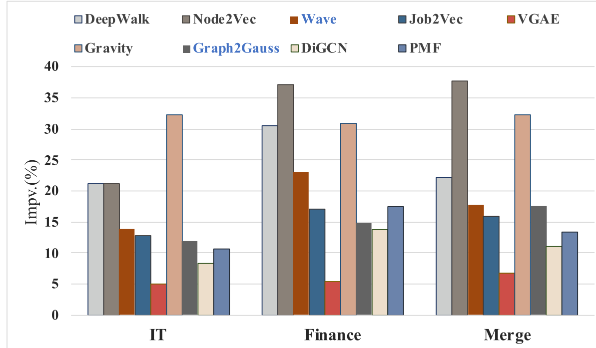


Fig. 5. Average improvement of UnGAE on four metrics for two tasks when compared with each baseline model. Since NGCF and GCMC can only apply to one task, their results are not shown here.

3) Evaluation Metrics: We chose area under the curve (AUC) and average precision (AP) to evaluate the performance of job title transition pattern prediction and used RMSE and mean absolute error (MAE) to evaluate the performance of job duration prediction.

4) Implementation Details: To guarantee the robustness of our experiments, we chose the most frequent 60% transition patterns as the training set, the lowest 20% as the test set to evaluate the performance, and the remaining 20% for tuning parameters. In our UnGAE, we set the dimension size d as 128 and tuned the hyperparameters by grid search. Our model was optimized by the Adam algorithm [40] with a learning rate of 0.01. And we set λ_g, λ_e and λ as 5, 0.1 and 0.01, respectively. Furthermore, to validate the significance of the improvement, we conducted a standard student t-test for the pair of UnGAE and each baseline. The code is available at: https://github.com/zruiii/UnGAE.

B. Experimental Results

1) Overall Performance: The comparison results on job title transition pattern and job duration prediction tasks are shown in Table III, which has clearly demonstrated the effectiveness of the proposed UnGAE framework. Overall, for the job title transition pattern prediction task, we observe that UnGAE improves the AUC and AP by an average improvement of 1.9% and 2.1%, respectively. As for the job duration prediction task, UnGAE shows an average of 4.4% and 6.8% achievement on the RMSE and MAE, correspondingly. Besides, Fig. 5

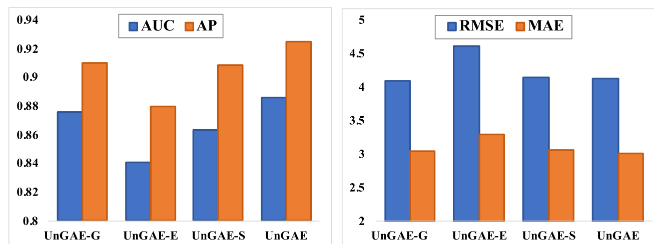
TABLE III

OVERALL PERFORMANCE OF JOB TITLE TRANSITION PATTERN PREDICTION (TASK 1) AND JOB DURATION PREDICTION (TASK 2) ON THREE DATASETS. NA DENOTES THAT THE METHOD CANNOT RUN ON OUR HARDWARE SETUP. † INDICATES THAT THE RESULT OF A PAIRED DIFFERENCE TEST IS SIGNIFICANT AT p-VALUE ≤ 0.05

Data	**IT Dataset**				**Finance Dataset**				**Merged Dataset**			
Task	Task 1		Task 2		Task 1		Task 2		Task 1		Task 2	
Method	AUC(%)	AP(%)	RMSE	MAE	AUC(%)	AP(%)	RMSE	MAE	AUC(%)	AP(%)	RMSE	MAE
DeepWalk	66.63	67.53	4.5557	3.3184	61.17	63.09	5.4364	3.6765	66.44	67.94	4.5952	3.3104
Node2Vec	66.67	67.38	4.5223	3.3290	60.36	60.45	6.0515	4.1691	61.78	61.95	5.8801	4.2417
Wave	71.67	77.40	4.3788	3.3673	63.69	69.66	4.8760	3.6111	69.60	75.71	4.5861	3.4078
Job2Vec	72.03	75.20	4.3588	3.1251	67.35	70.05	4.5884	3.2312	69.71	73.43	4.3718	3.1495
VGAE	85.70	87.61	4.4165	3.3050	85.65	87.75	4.6957	3.4071	85.92	88.37	4.5534	3.3337
Gravity	86.07	89.45	9.8178	8.8722	85.23	90.38	9.9790	8.7405	85.91	90.60	9.9374	8.8403
Graph2Gauss	83.07	82.37	4.7821	3.6922	82.35	80.78	5.3978	3.9870	78.64	76.79	5.054	3.695
DiGCN	78.41	80.21	4.2107	3.2401	70.35	73.84	4.4379	3.3184	75.67	78.49	4.1986	3.2278
PMF	79.87	83.93	5.1797	3.1646	72.8	77.68	5.5435	3.5027	78.45	82.73	5.2272	3.2594
NGCF	NA	NA	7.1980	5.6296	NA	NA	8.1275	6.4905	NA	NA	7.5401	5.9440
GCMC	NA	NA	6.0714	4.4843	NA	NA	6.7666	4.9197	NA	NA	6.2032	4.5416
UnGAE	† **88.56**	† **92.47**	**4.2101**	† **3.0789**	† **86.55**	† **91.28**	**4.4240**	† **3.0512**	† **87.21**	† **92.32**	**4.1115**	† **2.9412**

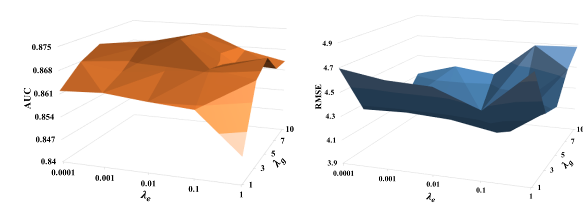
shows the average improvement of our proposed model on four metrics compared to other baselines, where the results further demonstrate the capability of UnGAE. It is worth noticing that our model has a better performance on both tasks simultaneously, whereas most baseline models can only handle one specific task. Accordingly, UnGAE can thus learn more comprehensive representation for each job title. In addition, the p-values in all experiments are small, demonstrating that the improvements are statistically significant for both tasks on different datasets.

Moreover, in terms of the transition pattern prediction task, random walk-based methods (DeepWalk and Node2Vec) get the worst performance. This may be caused by the imbalance between the in-degree and out-degree of nodes, which prevents them from capturing the asymmetric proximity sufficiently. Another long-range message propagation model, Wave, also performs worse than UnGAE. Since one may not change jobs frequently, long career paths only cover a small amount of data and local structure information is sufficient in our case. In addition, UnGAE and the variants of VGAE perform significantly better than PMF, indicating that leveraging the job title transition graph is effective for transition pattern learning. We also notice that Graph2Gauss is less effective, because it represents each node independently and cannot leverage the fine-grained diversity in node interactions. In terms of the job duration prediction, the performance of baselines is relatively similar, while gravity gets the worst performance, despite achieving comparable performance on the first task. We deem that the GID assumes a greater output for a more similar job title pair, which is contrary to the scenario of job duration prediction and demonstrates the necessity of designing the task-specific decoder. Furthermore, the performance on merged data is generally better than on a single industry. The result illustrates that a more complete structured job title transition graph enables a more powerful representation of network data, which in turn increases the predictive capability of methods.



(a) (b)

Fig. 6. Performance of UnGAE and its variants on job title transition pattern and job duration prediction tasks. (a) Job transition prediction. (b) Job duration prediction.



(a) (b)

Fig. 7. Performance of UnGAE with different parameter settings of λ_g and λ_e. (a) AUC. (b) RMSE.

2) Ablation Study: We further look into UnGAE to verify the effectiveness of each component. In particular, we conducted experiments on three variants as follows.

a) *UnGAE-G* replaces the GID with *source/target vectors* paradigm.

b) *UnGAE-E* replaces the EID with two-layer MLP for job duration prediction.

c) *UnGAE-S* learns two tasks separately instead of using the multitask learning strategy.

As illustrated in Fig. 6, replacing either GID or EID brings adverse effects on both prediction tasks. In this phase, compared with the VGAE that also utilizes the source/target vector paradigm, our UnGAE-G still gets better performance. Likewise, UnGAE-E is more effective in contrast with the baselines that use a two-layer MLP decoder. The results further

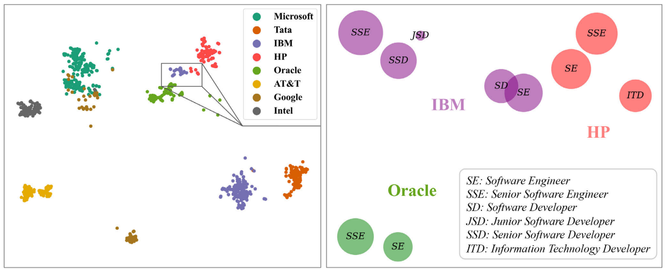


Fig. 8. Visualization of the job title representations. The left part is a global view of the embeddings for the IT dataset, and the right part is a fine-grained demonstration of some specific job titles.

demonstrate the expressiveness of our Gaussian representation of job titles. Besides, learning the duration prediction task separately performs significantly worse than using the multitask learning strategy, which implies the benefit of taking both tasks into consideration simultaneously.

3) Parameter Sensitivity: We investigate the parameter sensitivity of UnGAE on the IT dataset. Specifically, we evaluated the impact of parameters λ_g and λ_e together, which are introduced to tune the importance of node proximity with respect to mass attraction. We trained our UnGAE by varying λ_g from 1 to 10 and changing λ_e from 10^{-4} to 1. As shown in Fig. 7, we can observe that UnGAE achieves the best performances when λ_g and λ_e are 5 and 0.1, respectively. Moreover, it also demonstrates that when predicting the existence of a potential job transition pattern, λ_g tends to encourage the relative importance of symmetric node proximity in the GID. We deem that this may be related to the reciprocity of data, and the higher symmetry leads to the greater importance of the distance term in (6).

4) Visualization: We visualize the learned representations for job titles in Fig. 8 to validate our motivation for modeling job title transition patterns. Specifically, we selected several leading IT companies and utilized t-distributed stochastic neighbor embedding (t-SNE) [41] to reduce the representation dimension of *location* parameter r_i^g for each job title. From the left part of Fig. 8, we observe that the job titles of each company can be well clustered, indicating that career mobility is more likely to occur within the company. Moreover, there is an overlap in the job title embeddings across several companies, such as textitIBM, *Oracle*, and *HP*. This may be caused by the close interaction in some positions. To validate our motivation of model design, we focus on these overlapping jobs in the right part of Fig. 8, where the marker scales are in proportion to *mass* parameter m_i^g. Compared with *Oracle*, *IBM* is closer to *HP* in terms of *Software Engineer*, which reveals a more general staff exchange between these two companies. In addition, the job title at a higher level shows a greater marker scale, which is in line with our hypothesis that greater m_i should be more attractive.

C. Case Study

1) Job Title Benchmarking: Here, we first show an example to verify that our method can be applied to the job title benchmarking task, which plays a vital role in company profile understanding. Following [14], we aim to match jobs with similar expertise levels across different organizations.

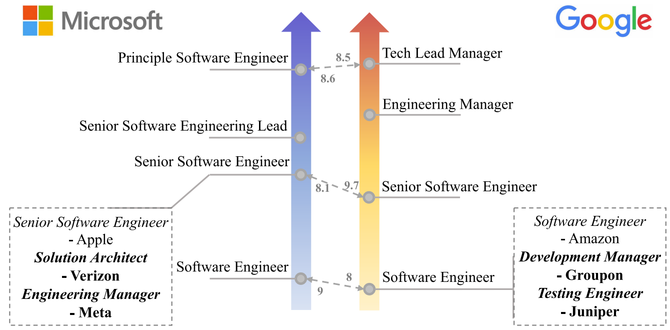


Fig. 9. Example of leveraging UnGAE on job title benchmarking, where the bold titles are unseen in the ground truth.

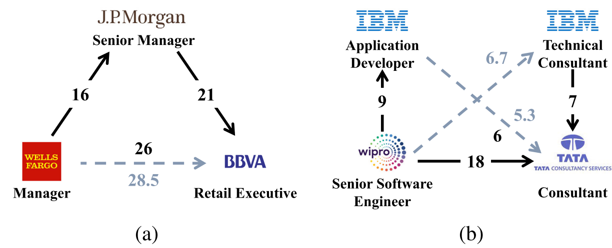


Fig. 10. Two examples of using UnGAE for career path planning, where the solid lines indicate the job title transition patterns in the training set, and the dotted lines represent the potential patterns found by UnGAE. In addition, the values indicate the ground-truth job duration (black) and predicted results (blue), respectively. (a) Case 1. (b) Case 2.

Here, we assumed that two job titles can be benchmarked if there is a close and short-duration interaction between them. Fig. 9 shows the job title hierarchies in *Microsoft* and *Google*, respectively. In this phase, we extracted the vertical rank structure within a company based on the average duration needed for a promotion. The results show that *Google* has a faster promotion system compared with *Microsoft*. In addition, we further selected a position from each of the companies and looked for other job titles that could be benchmarked across the whole labor market. The bold one represents the job titles that are unseen in the data, whereas our model matches them correctly. We also find that some literally different job titles are matched by UnGAE, while those titles may hold similar inherent responsibilities, such as *Software Engineer—Google* and *Development Manager—Groupon*.

2) Career Path Planning: In this phase, we aim to identify potential paths for achieving career goals. Fig. 10 shows two cases of career path planning generated by our UnGAE. As we can see in Fig. 10(a), it takes nearly 26 quarters to become *Retail Executive-BBVA* following the path {*Manager-WellsFargo* ⇒ *Senior Manager-J.P.Morgan* ⇒ *Retail Excutive-BBVA*}. Meanwhile, our prediction results also report the existence of the mobility from *Manager-WellsFargo* to *Retail Excutive-BBVA* directly with a shorter duration, which provides a "shortcut" for early career success. Moreover, Fig. 10(b) shows that two underlying paths with shorter time costs are predicted by our model in terms of the career mobility from *Senior Software Engineer—Wipro* to *Consultant—Tata*. In practice, such potential career mobility discoveries can provide an overall picture of the labor market, allowing valuable guidance for job seekers.

VI. Conclusion

In this article, we proposed an UnGAE framework for career mobility analysis, which can simultaneously predict the potential transition patterns and job durations. Specifically, we first constructed a job title transition graph based on the large-scale career trajectories. Then, we introduced an UnGE to represent the job titles as Gaussian embeddings with consideration of the inherent uncertainty in career mobility. We further designed two task-specific decoders, namely the GID and the EID, for predicting job transition patterns and job durations through a multitask learning approach. Finally, extensive experiments and case studies on real-world datasets demonstrate the effectiveness of UnGAE compared with state-of-the-art baselines as well as two labor market applications.

References

[1] Y. Guan, M. B. Arthur, S. N. Khapova, R. J. Hall, and R. G. Lord, "Career boundarylessness and career success: A review, integration and guide to future research," *J. Vocational Behav.*, vol. 110, pp. 390–402, Feb. 2019.

[2] S. S. K. Lam, T. W. H. Ng, and D. C. Feldman, "The relationship between external job mobility and salary attainment across career stages," *J. Vocational Behav.*, vol. 80, no. 1, pp. 129–136, Feb. 2012.

[3] T. W. H. Ng and D. C. Feldman, "Does longer job tenure help or hinder job performance?" *J. Vocational Behav.*, vol. 83, no. 3, pp. 305–314, Dec. 2013.

[4] A. G. Walder, "Career mobility and the communist political order," *Amer. Sociol. Rev.*, vol. 60, no. 3, pp. 309–328, 1995.

[5] L. Li, H. Jing, H. Tong, J. Yang, Q. He, and B.-C. Chen, "NEMO: Next career move prediction with contextual embedding," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 505–513.

[6] Q. Meng, H. Zhu, K. Xiao, L. Zhang, and H. Xiong, "A hierarchical career-path-aware neural network for job mobility prediction," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 14–24.

[7] L. Zhang et al., "Attentive heterogeneous graph embedding for job mobility prediction," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2021, pp. 2192–2201.

[8] C. Ryan, "The travel career ladder an appraisal," *Ann. Tourism Res.*, vol. 25, no. 4, pp. 936–957, Oct. 1998.

[9] N. Sicherman and O. Galor, "A theory of career mobility," *J. Political Economy*, vol. 98, no. 1, pp. 169–192, Feb. 1990.

[10] T. W. H. Ng, K. L. Sorensen, L. T. Eby, and D. C. Feldman, "Determinants of job mobility: A theoretical integration and extension," *J. Occupational Organizational Psychol.*, vol. 80, no. 3, pp. 363–386, Sep. 2007.

[11] S. T. Lyons, L. Schweitzer, E. S. W. Ng, and L. K. J. Kuron, "Comparing apples to apples: A qualitative investigation of career mobility patterns across four generations," *Career Develop. Int.*, vol. 17, no. 4, pp. 333–357, 2012.

[12] C. Wang, H. Zhu, Q. Hao, K. Xiao, and H. Xiong, "Variable interval time sequence modeling for career trajectory prediction: Deep collaborative perspective," in *Proc. Web Conf.*, Apr. 2021, pp. 612–623.

[13] H. Xu, Z. Yu, B. Guo, M. Teng, and H. Xiong, "Extracting job title hierarchy from career trajectories: A Bayesian perspective," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3599–3605.

[14] D. Zhang et al., "Job2vec: Job title benchmarking with collective multi-view representation learning," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 2763–2771.

[15] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. NIPS*, 2001, vol. 14, no. 14, pp. 585–591.

[16] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu, "Asymmetric transitivity preserving graph embedding," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1105–1114.

[17] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 701–710.

[18] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 855–864.

[19] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2020.

[20] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[21] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1024–1034.

[22] T. N. Kipf and M. Welling, "Variational graph auto-encoders," 2016, *arXiv:1611.07308*.

[23] M. K. Matlock, A. Datta, N. L. Dang, K. Jiang, and S. J. Swamidass, "Deep learning long-range information in undirected graphs with wave networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[24] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial–temporal graph modeling," 2019, *arXiv:1906.00121*.

[25] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 249–270, Jan. 2022.

[26] X. Chen, M. Chen, W. Shi, Y. Sun, and C. Zaniolo, "Embedding uncertain knowledge graphs," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 3363–3370.

[27] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 494–514, Feb. 2021.

[28] M. Stadler, B. Charpentier, S. Geisler, D. Zügner, and S. Günnemann, "Graph posterior network: Bayesian predictive uncertainty for node classification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 18033–18048.

[29] X. Zhao, F. Chen, S. Hu, and J.-H. Cho, "Uncertainty aware semi-supervised learning on graph data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12827–12836.

[30] A. Bojchevski and S. Günnemann, "Deep Gaussian embedding of graphs: Unsupervised inductive learning via ranking," 2017, *arXiv:1707.03815*.

[31] D. Zhu, P. Cui, D. Wang, and W. Zhu, "Deep variational network embedding in Wasserstein space," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 2827–2836.

[32] J. Jiang, D. Yang, Y. Xiao, and C. Shen, "Convolutional Gaussian embeddings for personalized recommendation with uncertainty," 2020, *arXiv:2006.10932*.

[33] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[34] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, "Adversarially regularized graph autoencoder for graph embedding," 2018, *arXiv:1802.04407*.

[35] G. Salha, S. Limnios, R. Hennequin, V.-A. Tran, and M. Vazirgiannis, "Gravity-inspired graph autoencoders for directed link prediction," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 589–598.

[36] Z. Tong, Y. Liang, C. Sun, X. Li, D. Rosenblum, and A. Lim, "Digraph inception convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17907–17918.

[37] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 20, 2007, pp. 1257–1264.

[38] R. van den Berg, T. N. Kipf, and M. Welling, "Graph convolutional matrix completion," 2017, *arXiv:1706.02263*.

[39] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2019, pp. 165–174.

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

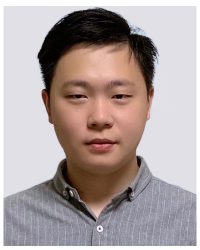
[41] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

Rui Zha received the B.S. degree in geophysics from the University of Science and Technology of China (USTC), Hefei, China, in 2020, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Technology.

His research interests include data mining, graph neural networks, and unsupervised learning.

Chuan Qin (Member, IEEE) received the Ph.D. degree in computer science and technology from the University of Science and Technology of China (USTC), Hefei, China, in 2021.

He is currently a Data Scientist at Baidu Inc., Beijing, China. He has authored more than 20 journals and conference papers in the fields of natural language processing and recommender system, including IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *ACM Transactions on Information Systems*, *Knowledge Discovery in Databases*, Special Interest Group on Information Retrieval (SIGIR), International World Wide Web Conferences (WWW), Neural Information Processing Systems (NeurIPS), Association for the Advancement of Artificial Intelligence (AAAI), International Joint Conference on Artificial Intelligence (IJCAI), and International Conference on Data Mining (ICDM).

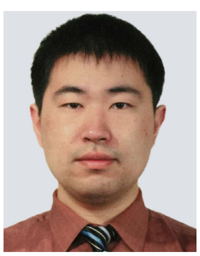


Le Zhang received the B.E. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2016, and the Ph.D. degree in computer science from the University of Science and Technology of China (USTC), Hefei, China, in 2022.

He is currently a Researcher at Baidu Research, Baidu Inc., Beijing, China. His general research interests include data mining and machine learning, with a focus on spatiotemporal modeling and its applications in business intelligence.

Dazhong Shen received the B.S. degree in mathematical sciences from the University of Science and Technology of China (USTC), Hefei, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Technology.

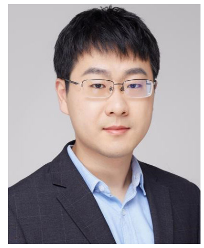
His research interests include generative models, variational inference, graph neural networks, and their data-intensive applications.



Tong Xu (Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China (USTC), Hefei, China, 2016.

He is currently working as a Professor at USTC. He has authored more than 100 top-tier journals and conference papers in the fields of social network and social media analysis, including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON MULTIMEDIA, *ACM Transactions on Multimedia Computing, Communications, and Applications*, *Knowledge Discovery in Databases* Special Interest Group on Information Retrieval (SIGIR), International World Wide Web Conferences (WWW), Association for the Advancement of Artificial Intelligence (AAAI), and International Joint Conference on Artificial Intelligence (IJCAI).

Dr. Xu was a recipient of the Best Paper Award of KSEM 2020 and the Best Student Paper Finalist of CICAI 2021.



Hengshu Zhu (Senior Member, IEEE) received the B.E. and Ph.D. degrees in computer science from the University of Science and Technology of China (USTC), Hefei, China, in 2009 and 2014, respectively.

Previously, he served as the Head of Baidu Talent Intelligence Center (TIC) and the Principal Scientist of Baidu Research, Baidu Inc., Beijing, China. He is currently the Head of BOSS Zhipin Career Science Lab (CSL). His research interests include data mining and machine learning, with a focus on developing advanced data analysis techniques for innovative business applications. He has published prolifically in refereed journals and conference proceedings, including PROCEEDINGS OF THE IEEE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON MOBILE COMPUTING, *ACM Transactions on Information Systems*, *ACM Transactions on Knowledge Discovery From Data*, ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), ACM Special Interest Group on Information Retrieval (SIGIR), International World Wide Web Conferences (WWW), International Joint Conference on Artificial Intelligence (IJCAI), and Association for the Advancement of Artificial Intelligence (AAAI). He is a reviewer of many leading academic journals and has served regularly on the organization and program committees of numerous conferences, including as a program Co-Chair of the KDD Cup-2019 Regular ML Track, and a founding Co-Chair of the International Workshop on Organizational Behavior and Talent Analytics (OBTA) and the International Workshop on Talent and Management Computing (TMC), in conjunction with ACM SIGKDD.

Dr. Zhu is the Senior Member of ACM and CCF. He was a recipient of the Distinguished Dissertation Award of CAS in 2016, the Distinguished Dissertation Award of CAAI in 2016, the Special Prize of President Scholarship for Postgraduate Students of CAS in 2014, the Best Student Paper Award of KSEM-2011, WAIM-2013, CCDM-2014, and the Best Paper Nomination of International Conference on Data Mining (ICDM)-2014.

Enhong Chen (Senior Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China (USTC), Hefei, China, in 1996.

He is currently the Director (Professor) of the Anhui Province Key Laboratory of Big Data Analysis and Application (BDAA), USTC. He is the Executive Dean of the School of Data Science, USTC, where he is also the Vice Dean of the School of Computer Science and Technology. He has published more than 200 refereed international conference papers and journal articles. His research interests include data mining, machine learning, and recommender systems.

Dr. Chen is a CCF Fellow. He was a recipient of the KDD'08 Best Application Paper Award and the International Conference on Data Mining (ICDM)'11 Best Research Paper Award.