Acqui-hiring or Acqui-quitting: Data-Driven Post-M&A Turnover Prediction via a Dual-fit Model

Denghui Zhang
Rutgers Business School
denghui.zhang@rutgers.edu

Hao Zhong
ESCP Business School
hzhong@escp.eu

Jingyuan Yang
George Mason University
jyang53@gmu.edu

Abstract

Gaining highly skilled human capital is one of the primary reasons for corporate mergers and acquisitions (M&A), especially for knowledge intensive industry. However, the inevitable tensions brought by the divergent cultures and organizational misalignment during the M&A process result in high talent turnover rate and ultimately the integration failure. Hence, it is imperative to understand and prepare for the potential effects of M&A process on the employee turnover. To this end, we propose a novel dual-fit model induced heterogeneous Graph Neural Network (GNN) model to predict the talent turnover trend in the post-M&A process, by taking into account the complex relationship among the acquirer firm, the acquiree firm, and the acquired employees. Specifically, we creatively design a dual-fit model comprised of both the firm-level compatibility and employee-firm fit. Extensive evaluations on large-scale real-world data clearly demonstrate the effectiveness of our approach.

Introduction

Over the past few decades, mergers and acquisitions (M&A) have become one of the major strategies for business to grow and expand market shares. Driven by an essential need for growth, waves of M&A transactions have reached a historical high, particularly in the technology industry. For example, U.S. mergers and acquisitions sustained a vigorous pace in the first quarter of 2019, with 900 transactions and a total market valuation of $79.5 billion, a 35% increase year over year (YoY)[1]. "Acqui-hiring" is a trending M&A-based hiring strategy to effectively reinforce the talent pool and boost enterprise value[2] (Chen et al., 2021; Kim, 2020). To gain and sustain financial benefits of M&A transactions (e.g. revenue and market share gains), companies should also ensure that they have proper management practices and right workforce targets to maintain the acquired intangible assets, i.e., talents. Unfortunately, recent studies (Kim, 2020) revealed a notable "acqui-quitting" trend, i.e., 33% of acquired workers quit their jobs within the first year of their employers being acquired. From the moment M&A deals are

[1] https://www.aerotek.com/en/insights/power-through-m-and-a-disruption-with-a-strong-talent-strategy
[2] https://www.business-sale.com/insights/for-buyers/acquihiring-ma-strategy-to-boost-talent-pool-and-enterprise-value-221601

announced (or rumored), they can bring profound organizational and cultural changes and disruptions for employees through the integration process (Lee and Pennings, 1996). To sustain productivity and ensure long-term benefits, it is imperative to understand and foresee potential post-M&A talent flows.

Numerous studies have investigated M&A activities from various aspects, including financial performance outcomes, pre-merger firm-level compatibility, organizational culture and ex-ante M&A experiences, and M&A integration process (Das and Kapil, 2012; Trichterborn et al., 2016). While financial outcomes have received considerable attention in M&A studies (Thanos and Papadakis, 2012), post-M&A employee turnover is still under-explored. Only a handful of recent studies started focusing on M&A-related employee turnover by understanding employee attitudes through individual-level primary survey data (Kyei-Poku and Miller, 2013). There is a complex mechanism behind employee turnovers, whose driving factors may include M&A deal characteristics, managerial effects, organizational compatibility and cultural fit (Bauer and Matzler, 2014; Kim, 2020).

To tackle the post-M&A employee turnover prediction problem, there are several unique challenges: (1) While the majority of existing literature concentrates on firm-level characteristics, we argue that compatibility between employees and the acquirer firm cannot be overlooked. Indeed, the M&A turnover involve three primary entities, namely, the acquirer firm, the acquiree firm, and the acquired employees. We need a holistic understanding of employee attrition impacted by M&A events, considering not only the acquirer to acquiree compatibility, but also employees to acquirer firm fit. (2) To study this three-way relationship naturally requires various data describing these three entities comprehensively. However, such a collection of data is typically large-scale, heterogeneous and often unstructured, such as company profile descriptions. Thoughtful data preparation step is needed next to utilize abundant available unstructured data to extract meaningful information. (3) Traditional classification models cannot properly handle the complexity of the three-way relationship along with the heterogeneous and unstructured data. Advanced machine learning models are needed to effectively integrate and release the full potential of the comprehensive data.

To address the aforementioned challenges, our paper proposes a novel graph neural network-based method to examine the "fit" among these three parties and understand their impacts on employee turnover. In particular, we propose a *Dual-fit* model: an *Organization to Organization fit (O-O fit)* as the measure of firm-level compatibility and complementarity and a *Person to Organization fit (P-O fit)* as the "fit" measure between the acquired employees and the acquirer. Our focus here is whether we could effectively predict the impact of M&A on acquired employee turnover escalation, measured as the difference between pre-M&A vs. post-M&A turnover rates.

To this end, we obtain a large-scale heterogeneous dataset with more than 2,500 M&A transactions from *Crunchbase* and over 806K talents' employment profiles from *LinkedIn*, which enables us to perform fine-grained level analysis covering various factors. Then we perform data preprocessing and feature engineering to restructure the heterogenous data as well as to extract useful patterns based on the text data about company profiles and employee job records. Next, to extract the complex hidden relationship, we propose a novel dual-fit model induced heterogeneous Graph Neural Network (GNN) model to perform a fine-grained level analysis of turnover likelihood of various types of employees. The extracted informative node feature representations of the three-way relationships can reveal rich semantic, structural patterns that would have not been uncovered by the traditional classification models or homogeneous graph models. Finally, we conduct extensive experiments and ablation studies on real-world data to demonstrate that the proposed framework has superior prediction performance over state-of-the-art bench-marking methods, and provide insightful discussions to showcase the advantage of the dual-fit model design.

Literature Review

M&A and Employee Turnover. Mergers and acquisitions (M&A) are referred to the events of combining/merging two independent firms into one single entity. Extant literature aimed at understanding M&A events by studying motivations, organizational relatedness, and M&A effects on performance, especially on financial outcomes (e.g. stock price, profitability, and return on investment) or productivity (e.g. patents) (Narayanan et al., 2019; Schuler and Jackson, 2001). Although scholars have primarily focused on the impact of M&A on financial outcomes, there is a lack of attention on the employee-focused outcomes of M&A. It may be partially attributed to the difficulties in collecting the attitude and behavior data from employees throughout the M&A process. However, it is crucial to understand the employee side of M&A outcomes to evaluate the deal success. A few studies started to analyze the link between employee turnover and M&A performance by exploring human related factors such as culture, management, poor motivation (Krug et al., 2014; Kyei-Poku and Miller, 2013).

Turnover Theory. There are many important turnover theories in management literature to explain why employees leave their organizations (Hom et al., 2017), such as the *organizational equilibrium theory* (March and Simon, 1993), the *unfolding model* (Lee et al., 1999), and *job embeddedness theory* (Mitchell and Lee, 2001). The organizational equilibrium theory perhaps was the most foundational model, which emphasizes two major aspects: the ease of job movement and the personal intention of leaving (March and Simon, 1993). Turnover researchers are actively exploring other significant antecedents to explain employee turnover, for example, social relationship (Teng et al., 2019) and the organizational structure (Sun et al., 2019). Recently, Steigenberger and Mirc (2020) emphasized that organizational and especially under-studied occupational identification have strong influence on employee turnover decisions. Many turnover studies investigate the collective turnover, while there are much more voluntary turnover studies over involuntary ones (Hausknecht and Trevor, 2011). Also, turnover research has been conducted over various levels (individual, group/unit, and organization) and different industries and professions (e.g. IT professions (Joseph et al., 2007)). In our paper, we focus on the collective turnover at the firm level, with emphasis on the impacts of organizational level factors as well as the occupational employee group factors on the turnover.

Organizational Fit and Person-Organization Fit. There are many dimensions when evaluating the match between the target firm and the acquirer firm. A prominent strategic management research stream studies around the impact of the fit between the firm pairs on the M&A success (Bauer and Matzler, 2014; Homburg and Bucerius, 2006). In the school of strategic management literature, the core concept is a high *compatibility* (relatedness or similarity) in the management styles and organizational culture can effectively increases the value creation and boost synergy realization (Bauer and Matzler, 2014; Palich et al., 2000). In other words, the similarity increases the performance and also reduces the potential cultural conflicts during the integration process. A few studies focus on the measure of the similarity of the firm level characteristics between the acquirer and acquiree, such as the match or compatibility among the company pair as measured in several proximity metrics, which are discussed in Table 4 (King et al., 2004; Shi et al., 2016; Tuch and O'Sullivan, 2007), In addition to the firm level compatibility, we also need to consider the Person-Organization fit (P-O fit) when considering employee turnover. P-O fit has been defined as "the compatibility between the employee and organization" and has been described as "a multidimensional construct consisting of three determinants of fit: values, personality, and environment" (O'Reilly III et al., 1991; Westerman and Cyr, 2004). However, the changes brought by the M&A may cause employees from the acquired company feel misfit with the acquirer company based on the new visions, management styles, and culture shock (Buono and Bowditch, 2003). Although a vast amount of extant research studied P-O fit based on primary data such as interviews and surveys about the job satisfaction and employee turnover intention (Greenwood et al., 1994), in this work we will more rely on the objective data of P-O fit under the impact of M&A.

Graph Neural Networks. Our methodology relates to the broad literature of Graph Neural Networks (GNN) models. GNN models have recently caught wide and unprecedented attention in data mining and machine learning communities as there are a growing number of applications where data are represented in the forms of graphs/networks. According to the types of nodes and/or edges in a network, GNN models can be classified into *homogeneous network-based models* (only single type of node and edge) and *heterogeneous network-based models* (with multiple types of nodes and/or edges). As for homogeneous GNNs, convolutional GNNs appeared to be the mainstream and popular models: including DCNN (Atwood and Towsley, 2016), GCN (Kipf and Welling, 2016), and GAT (Veličković

Table 1. Descriptive Statistics

Variable	Count	Variable	Mean	Std	Min	Max
# M&A deals	2,566	# M&A deals per acquirer	1.38	1.31	1	24
# acquirers	1,861	# M&A deals per acquiree	1	0	1	1
# acquirees	2,566	# industry keywords per acquirer	3.73	1.89	1	13
# geo-locations	947	# industry keywords per acquiree	3.3	1.63	1	11
# industry keywords	48	# investors per acquirer	3.53	3.19	1	23
# investors	3,758	# investors per acquiree	3.52	2.61	1	18
# employees	806,536	# employees per acquiree	327.31	1,717.94	1	43,175
# employee groups	64	# employees per employee group	15,985.91	22,447.95	44	118,910
# job records	1,212,319	# employee groups per acquiree	11.32	11.52	1	60

et al., 2018). To cope with more complex heterogeneous networks, heterogeneous GNNs were later developed, which consists of proximity-preserving-based models, message-passing-based models, and relation-learning-based models (Yang et al., 2020). Our model fits into the class of message-passing methods, where representative models include *RGCN* (Schlichtkrull et al., 2018), *HAN* (Wang et al., 2019), and *HGT* (Hu et al., 2020).

In our study, we integrate both of O-O fit and P-O fit among the acquirer-acquiree-employee relationship into a heterogeneous graph neural network for a more holistic understanding of the M&A turnover. Based on the constructed comprehensive graph, a set of diverse driving factors in evaluating the dual fit has been incorporated together, such as the acquirer companies' past acquired experiences, the industry/business relatedness, top management team and so on. Comparing with the existing work, more nuanced relationship among these three parties can be captured in the GNN based network.

Data and Preliminary Analysis

M&A Data Collection

Our data are collected from two sources. The first is *Crunchbase*, a premier database of startup activities, investments and funding information. It is well-recognized and has rising potential for economic and managerial research (Butler et al., 2020; Dalle et al., 2017). We rely on this database to gather firm demographics, M&A deals, investments, and firms' key members. The second data source is *LinkedIn*, one of the major professional networking platforms. It has been used as a major data source in various studies to help understand career paths Lappas, 2020 and labor markets Liu et al., 2020b. Likewise, we acquired employee profiles (e.g., job titles and career history) from *LinkedIn*.

Our data collection process starts with sampling M&A deals. And we restrict our focus to the M&A deals completed after year 2000 (inclusive) and the acquirees founded after 1990 (inclusive) with headquarters in the United States. To avoid any discrepancies, the two data sources are then linked by ensuring exact matching of firm names. Individuals' profiles are largely retained for the sake of capturing more prominent career mobility patterns. Table 1 showcases our data sample's descriptive statistics.

Employee Group (EMG)

We note again that a main objective of our research is to understand the impact of M&A transactions on the turnover rate of various types of employees. We therefore carefully categorize employees according to their occupational genres, which include *functionality (FUN)* and *responsibility (RES)*. Given a job title, we extract *FUN* and *RES* using the method proposed by Liu et al., 2020a. The procedure of constructing employee groups is outlined as follows. We begin by classifying each job title according to its relationship to the two genres. Job term embeddings are then obtained by applying a pre-trained word embedding model, GloVe Pennington et al., 2014, which encodes semantic meaning into vectors. Next, mini-batch K-Means clustering is employed on the two genres of job term embeddings to generate genre-specific clusters for the job titles. To determine the proper number of clusters for two genres respectively, we employ the Elbow method Thorndike, 1953 and choose $K^{FUN} = 8$ and $K^{RES} = 8$. The common job terms in *FUN* and *RES* groups are documented in Section *Common Job Terms*. Our final employee groups are determined by the joint genre-specific cluster IDs. For example, suppose that firm i belongs to $C_i^{FUN} = 3$ and $C_i^{RES} = 5$, its employee group is defined as $C_i = \{3, 5\}$. With such clustering outcomes, we observe an average of 11.32 employee groups per acquiree and 16K employees per employee group, as indicated in Table 1.

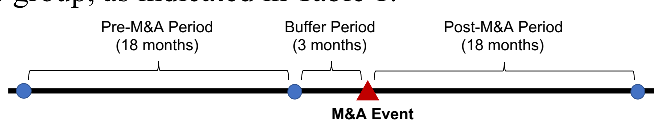

Figure 1. Illustration of three periods for M&A events.

Target Variable

Here, we elaborate how the target variable *Turnover Escalation* is defined in our study. We define *Turnover Escalation* as the significant rise of turnover rate in the post-M&A period compared with that in the pre-M&A period. Regarding any M&A event, we formally define three distinct periods as shown in Figure 1. To ensure sufficient data records, we set 18-month observation window for pre- and post-M&A periods. Meanwhile, we embed a 3-month buffer period prior to the M&A event date to eliminate possible turnover data contamination due to internal message leakage. Note that *Turnover Escalation* is computed at the level of *Acquirer - Acquiree - Employee_Group* (**ACR-ACE-EMG**). Given pre- and post-M&A periods, we first aggregate the number of turnovers in EMG group k for any M&A event between acquirer i and acquiree j, i.e., N_{ijk}^{pre} and N_{ijk}^{post}. With the total number of employees N_{jk}^{total} in EMG group k of acquiree j, we calculate the difference of turnover rates between pre-M&A and post-M&A periods as:

$$\Delta R_{ijk} = R_{ijk}^{post} - R_{ijk}^{pre} = \frac{N_{ijk}^{post}}{N_{jk}^{total}} - \frac{N_{ijk}^{pre}}{N_{jk}^{total}}. \quad (1)$$

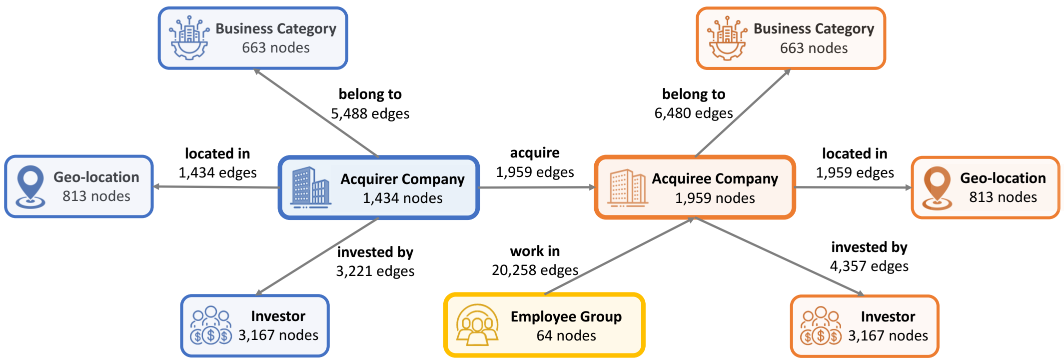

Figure 2. The meta graph of Heterogeneous Organization-Employee Graph

Lastly, we perform proportion tests of $\Delta R_{ijk} > 0$ and label *Turnover Escalation* $= 1$ if the difference is significant at the level of significance 0.01, otherwise *Turnover Escalation* $= 0$.

Methodology

We formulate the task of post-M&A turnover prediction as the following **binary classification problem**: given an acquirer company denoted by ACR_i, an acquiree company denoted by ACE_j, a group (type) of employees in acquiree company $\mathrm{EMG}_k \in \mathrm{ACE}_j$, assuming ACR_i will acquire or merge ACE_j, we aim to predict whether the turnover rate $\mathrm{R}_{\mathrm{EMG}_k}$ of the above employee group in the acquired company after the M&A event announced will increase significantly. More concisely, the input of our task is an Acquirer-Acquiree-EmployeeGroup triplet ($\mathbf{ACR}_i$, $\mathbf{ACE}_j$, $\mathbf{EMG}_k$), the output is a binary variable $\boldsymbol{y} \in \{0, 1\}$ where 1 means significant increase of turnover rate, 0 means no significant change. We apply attributes and features defined in Section *Raw Attributes and Feature Engineering* to represent all triplets.

Heterogeneous Organization-Employee Graph (HOEG)

Considering the flexibility and expressive power of graphs as well as the heterogeneity of the post-M&A data, we transform our data into heterogeneous graph data by building a Heterogeneous Organization-Employee Graph (HOEG).

Definition 1. Heterogeneous Graph. Heterogeneous graph is a type of graph consisting of different node types and link types. Let $G < V, E >$ denote a graph, where V denotes the node set, E denotes the edge set. Then $G < V, E >$ is heterogeneous when it contains a list of nodes types $V = \{V_1, V_2, ..., V_N\}$, where $N > 1$. Each type V_i contains n_i nodes: $\{t_{i,1}, t_{i,2}, ..., t_{i,n_i}\}$. Equally, it should also contain a list types of edges $E = \{E_1, E_2, ..., E_M\}$, where each type E_i contains m_i edges: $\{e_{i,1}, e_{i,2}, ..., e_{i,m_i}\}$.

For the post-M&A turnover trend prediction task, we define a special heterogeneous graph, namely, Heterogeneous Organization-Employee Graph (HOEG), to represent all heterogeneous objects in our

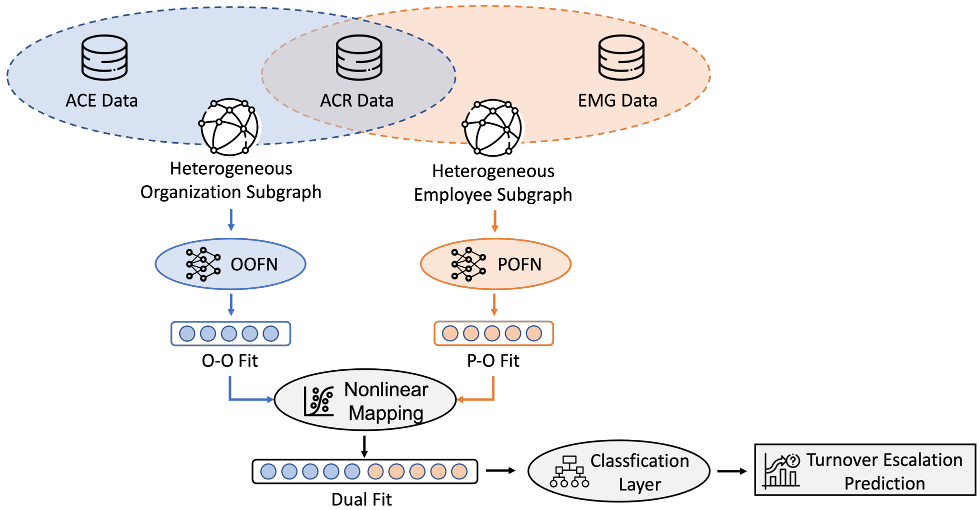

Figure 3. Overview of dual-fit heterogeneous GNN for triplet-based M&A turnover prediction

data. Figure 2 shows the meta graph of HOEG in which there are three types of *core node*, i.e., $\{V_1$=Acquirer Company, V_2=Acquiree Company, V_3=Employee Group$\}$, three types of *supplementary nodes*, i.e., $\{V_4$=Business Category, V_5=Geo-location, V_6=Investor$\}$. The core node types correspond to the input triplet (ACR$_i$, ACE$_j$, EMG$_k$) while the supplementary nodes correspond to objects in attribute columns of the input triplet. Figure 2 contains fewer nodes than the original data in Table 1 since we only built the graph on the training set.

Dual-fit Heterogeneous Graph Neural Network

Overview. Figure 3 shows an overview of our Dual-fit Heterogeneous Graph Neural Network (DHGNN) for post-M&A turnover trend prediction. Our model mainly consists of two parts: Organization-Organization Fit Network (OOFN, O-O Fit Network) and Person-Organization Fit Network (POFN, P-O Fit Network). Given an input triplet (ACR$_i$, ACE$_j$, EMG$_k$), we first locate the corresponding core nodes as well as surrounding supplementary nodes in HOEG, these nodes automatically constitutes two heterogeneous subgraphs, i.e., heterogeneous organization subgraph and employee subgraph. The former subgraph is centered by acquirer, acquiree node ACR$_i$, ACE$_j$ while the latter is centered by acquirer node ACR$_i$, employee group EMG$_k$. O-O Fit Network (left part in Figure 3) will apply graph convolution operations on ACR$_i$, ACE$_j$ nodes to encode the attributes and heterogeneous neighborhood into the hidden representations. Later we further concatenate two hidden vectors of ACR$_i$, ACE$_j$ and apply non-linear layer to explicitly generate an O-O Fit score, which aims to model the compatibility, complementarity between the acquirer and acquiree. Similarly, P-O Fit Network (right part) will encode essential information of ACR$_i$ and EMG$_k$ into their node representations and generate a P-O Fit score, which models the compatibility between a certain employee group (in acquiree) and the acquirer company. In the last step, we combine P-O Fit and O-O Fit vectors into one unit by non-linear transformations and feed into the classification layer to make the final turnover trend prediction.

Heterogeneous Message Passing

The fundamental idea of most GNNs is Message Passing – aggregating feature information from a node's direct (first-order) neighbors, such as GCN (Kipf and Welling, 2016) or GAT (Veličković et al., 2018). The general message passing scheme given a node x_i in graph is defined as follows:

$$x_i' = \gamma\big(x_i, \rho_{j\in\mathcal{N}}\phi(x_i, x_j)\big) \tag{2}$$

where ϕ is message function, depends on node feature x_i, x_j. $\rho_{j\in\mathcal{N}}$ denotes the aggregation function (one can choose sum or average, etc,), γ is the update function, i.e., final transformation to obtain new attributes after aggregating message. First, each node in the graph computes a message for each of its neighbors. Then each node aggregates the messages it receives using a permutation-invariant function (i.e., the order of message does not matter). Upon receiving the messages, each node updates its attributes based on its current attributes and the aggregated messages.

Obviously, Message Passing assumes that the graph only contains one type of node and each node only contain one type of feature ($\mathcal{N}$ in Eq.(3) means homogeneous neighbor and x is homogeneous feature). The assumption is too strong for our setting. Actually, in HOEG, the Acquirer and Acquiree nodes contain numerical content, such as "Company Size" and "Company Age", whereas other node types only contain categorical content. As a result, we require different feature transformations to handle different types and dimensions of features.

To tackle the issues, we propose a novel Heterogeneous Message Passing method for HOEG. Specifically, we adopt a more flexible assumption that each node type may contain multiple features. Given a node type, let $\mathcal{X}_i = \{x_{i,1}, x_{i,2}, ..., x_{i,n}\}$ denotes the heterogeneous feature set for node v_i. We define a new message function that takes heterogeneous features:

$$\phi(v_i, v_j) = \frac{\sum_{x_{j,n}\in\mathcal{X}_j}\big[\overrightarrow{LSTM}\{\mathcal{F}(x_{j,n})\}\bigoplus\overleftarrow{LSTM}\{\mathcal{F}(x_{j,n})\}\big]}{|\mathcal{X}_j|} \tag{3}$$

where v_i is the center node, v_j is one of the neighbor node of v_i, $\mathcal{X}_j$ is feature set of v_j. As can be seen, we use bi-directional LSTM (bi-LSTM) (Hochreiter and Schmidhuber, 1997) to capture "deep" interactions among heterogeneous features and encode them into the same feature space. $\mathcal{F}(\cdot)$ denotes feature transformer that ensures equal length input features. $\bigoplus$ denotes vector concatenation.

Next, we define a new aggregation function that works for aggregating heterogeneous neighbor nodes. It consists of two steps: **Intra-type Aggregation** and **Inter-type Aggregation**. Intra-type aggregation first aggregates all neighbor nodes of the same type and then inter-type aggregation combines all different types. For a specific node v, we iterate all its neighbors and get a list of t_n types of nodes. For each type t, we first perform **intra-type aggregation**:

$$\rho_1^t(v) = \mathcal{G}^t_{v'\in\mathcal{N}_t(v)}\{\phi(v, v')\} \tag{4}$$

where $\mathcal{G}^t\{\cdot\}$ denotes the aggregator for node type t, which can be fully connected network or recurrent neural network. We use a fully connected network as the aggregation function. $\mathcal{N}_t(v)$ denotes the sampled type t neighbors for node v, here we adopt normalized sampling ratio to ensure balanced samples among different types, that is, we use a fixed sampling ratio r_0 times r_t the ratio of type t nodes in all graph nodes. $\phi(\cdot)$ denotes the content aggregation function defined previously.

Then we perform **inter-type aggregation** to aggregate the above intra-type aggregation results of all different types,

$$\rho_2(v) = \alpha^{v,v}\phi(v,v) + \sum_t \alpha^{v,t}\rho_1^t(v) \tag{5}$$

where $\phi(v,v)$ denotes the aggregated content embeddings of node v itself. $\alpha(v,t)$ is the learnable attention weights indicating the importance of the corresponding neighbor type t to node v, defined as follows:

$$\alpha^{v,t} = \frac{exp\{ReLU(W^\top[\phi(v,v) \bigoplus \rho_1^t])\}}{\sum_{t\in\mathcal{T}(v)\bigcup\{v\}} exp\{ReLU(W^\top[\phi(v,v) \bigoplus \rho_1^t])\}} \tag{6}$$

where $ReLU(\cdot)$ denotes the non-linear function of Rectified Linear Unit, $W^\top$ denotes the attention parameter, $\bigoplus$ denotes concatenation. Lastly, we use the $\rho_2(v)$ to update the embedding of node v, in other words, we use identity function as our updating function $\gamma(\cdot)$.

Basically, Eq.(4)-(7) constitutes the entire heterogeneous message passing method, which can aggregate heterogeneous neighbor nodes as well as their heterogeneous contents.

Organization-Organization Fit and Person-Organization Fit

Recall that the input to our model is a triplet of (ACR$_i$, ACE$_j$, EMG$_k$), we apply the O-O fit network and P-O fit network to obtain two fit vectors and fit scores, which model the M&A deal fitness from organization perspective and employee perspective, respectively.

O-O Fit Network and P-O Fit Network. In O-O fit network, we first apply two layers of Heterogeneous Message Passing function (defined in Eq. (4)-(7)) on the acquirer node v_i^{acr} and acquiree node v_j^{ace} and their neighbor nodes to generate the aggregated feature embeddings x_i^{acr}, x_j^{ace},

$$x_i^{acr} = \rho_2(\rho_2(v_i^{acr})), \quad x_j^{ace} = \rho_2(\rho_2(v_j^{ace})), \tag{7}$$

The two embeddings above contains heterogeneous information such as “location”, “business category”, “investors”, etc. Then we concatenate two embeddings and further apply a non-linear transformation layer to obtain the O-O fit vector, i.e., $x_{i,j}^{acr,ace}$, which encodes the combinational information of acquire and acquiree. Based on $x_{i,j}^{acr,ace}$, we further calculate a scalar score (in $[0,1]$) as O-O fit measure. By later end-to-end training, the vector should learn what combinations of heterogeneous information of two companies will result in a good fit that keeps the post-M&A turnover rate stable. For P-O fit, we apply the similar procedures as O-O fit on the acquirer node v_i^{acr} and the employee group node v_k^{emg} to obtain the P-O fit vector and P-O fit score. The training and loss function details of our model can be found in Section *Loss Function and Model Training*.

Experimental Results

Based on the data described in Section *Data and Preliminary Analysis*, we built a unified experimental dataset in which each sample follows the format of $\{$triplet | raw attributes | handcrafted features | heterogeneous neighbor nodes | target label$\}$, which has the advantage of working for a wide range of baseline models as well as our graph-based model. We split the entire dataset into training/validation/test set using the ratio of 6:2:2. We then trained all models on the training set, tuned hyperparameters on the validation set (except mean-based models below), and computed evaluation metrics on the test set.

Baselines, Evaluation Metrics and Implementation Details

We perform experimental analysis and comparison on the following models:

- *Mean-based models*: We consider two mean-based models as our baselines. First, we compute each acquiree's post-M&A turnover escalation using the corresponding mean estimate of all acquirees in the same industry and with the same firm age. This model is denoted as **Industry+Age**. The second mean-based model is denoted as **EMG**, which relies solely on employ group (EMG), i.e., using mean estimates of all M&A deals with the same EMG as the predictions.
- *Conventional ML models*: We also built four conventional ML-based models on top of the handcrafted features as another set of baseline models, namely, Logistical Regression (**LR**), Support Vector Machine (**SVM**), Decision Tree (**DT**), and Random Forest (**RF**). We intend to demonstrate the best-achievable prediction performance using ML models without embedding-based features.
- *Existing GNN models*: To show the effectiveness of our proposed model, we compare it against two GNN-based models. First, we downgrade our heterogeneous network into a homogeneous one by ignoring type variation of nodes and edges to train a **GCN** model (Kipf and Welling, 2016). Besides, as an example of HGNN models, **RGCN** (Schlichtkrull et al., 2018) is incorporated given its popularity and fit to our problem.
- *Our models*: We include our Random Forest model built with embedding-based features, denoted as **Embedding+RF**. Then we conduct an ablation study and further examine two downgraded versions of our dual-fit model: **O-O Fit** and **P-O Fit**. Lastly, our proposed model **Dual-Fit** is included.

To evaluate the prediction performance of all models, we adopted a variety of classification metrics for a comprehensive evaluation: Precision, Recall, F1-score, AUC (Area under the ROC Curve). We omit Accuracy metric since it will be biased by our imblanced data (90% class 0). To obtain the best performance of our model, we empirically tuned the hyperparameters on the validation set and performed grid search over the following parameter values: learning rate = $\{0.00001, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.1\}$, batch size = $\{16, 32, 64, 128\}$, initial node feature dimension = $\{64, 128, 256, 512\}$,

Table 2. Overall Performances on Post-M&A Turnover Prediction

		Precision (%)		Recall (%)		F1-score (%)		AUC (%)
		Macro	Micro	Macro	Micro	Macro	Micro	
Mean-based models	Industry+Age	52.75	79.78	55.99	55.96	46.93	63.07	59.84
	EMG	51.93	79.11	54.23	52.97	44.88	60.46	56.62
Conventional ML models	LR	53.06	80.06	56.64	56.61	47.45	63.64	58.79
	SVM	52.57	79.64	55.61	55.65	46.64	62.82	58.05
	DT	57.95	83.76	66.44	66.41	55.81	71.78	70.88
	RF	58.37	84.03	**67.20**	67.18	56.49	72.40	72.88
Existing GNN models	GCN	68.56	83.11	58.97	85.32	59.43	82.7	72.11
	RGCN	68.98	83.24	59.12	85.88	59.81	83.3	72.98
Our models	Embedding+RF	69.5	84.58	57.8	87.1	59.04	82.76	72.85
	O-O Fit	69.85	83.44	59.36	86.89	60.58	83.98	73.39
	P-O Fit	66.18	82.13	56.71	85.16	58.92	82.33	71.54
	Dual-fit (DHGNN)	**70.51**	**84.96**	60.27	**87.71**	**62.31**	**84.98**	**74.53**

embedding dimension = $\{64, 128, 256\}$. To ensure robust results, we ran the fine-tuning 10 times and take the average. The optimal hyperparameters for our model are: learning rate $= 0.001$, batch size $= 512$, initial node feature dimension $= 128$, embedding dimension $= 128$. We also performed grid search for baseline models and reported their best performances.

Results Analysis

Overall Prediction Performance and Ablation Study. Table 2 shows the performance of all models using the default classification threshold (0.5) where bold numbers indicate the best results. Regarding the overall prediction performance on both classes (i.e., turnover escalation and non-escalation), we examined the AUC and F1-score and observed that: our complete model (Dual-fit) achieved the best results among all models. This confirms the superiority of our Heterogeneous GNN model over the mean-based models, conventional ML models as well as existing GNNs. Mean-based models are simple and fast, but does not have strong predictive power, whereas, ML models learn strong patterns that can generalize to test data. The results of our exploratory model (Embedding+RF) were secondary best, indicating that pre-trained embeddings can serve as feature augmentation, which aligns with our intuition of learning better M&A object embeddings using heterogeneous graph. We also conducted ablation studies to examine each component in our model. We compared our complete model with each of our sub-models (O-O fit and P-O fit) and observed that: O-O fit network alone yielded acceptable performance (better than RF+Embedding), whereas, P-O fit alone gave relatively poor results (worse than RF). This indicates O-O fit played a more important role than P-O fit in M&A fit modeling. As only combining them together resulted in better performances, it is evident that both O-O and P-O fit contributes to the superior performance of the entire model.

Discussions on Fit Scores. We continue our discussions on the two fit scores by visualizing their distribution using a *Heat Map* in Figure 4. Each of the two fit scores are binned into 10 equal-width bins (thus 2D squares in the plot). In each square, the color shade indicates the proportion of *Turnover Escalation* cases, i.e., darker color indicates more turnover escalations. A 2D Gaussian filter is applied

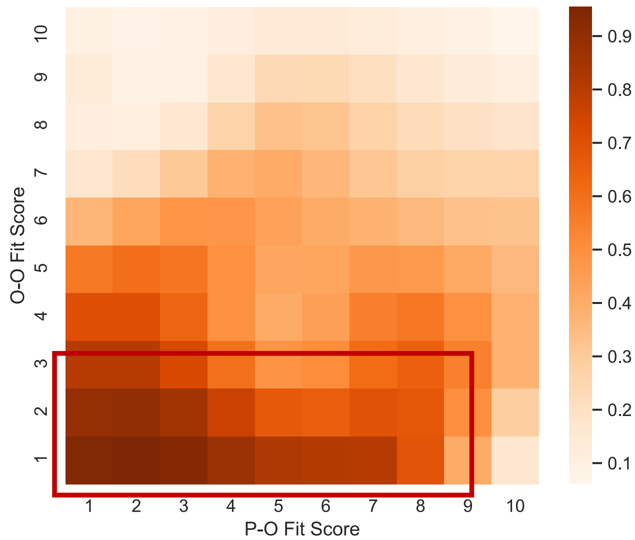

Figure 4. Overall distribution of fit scores

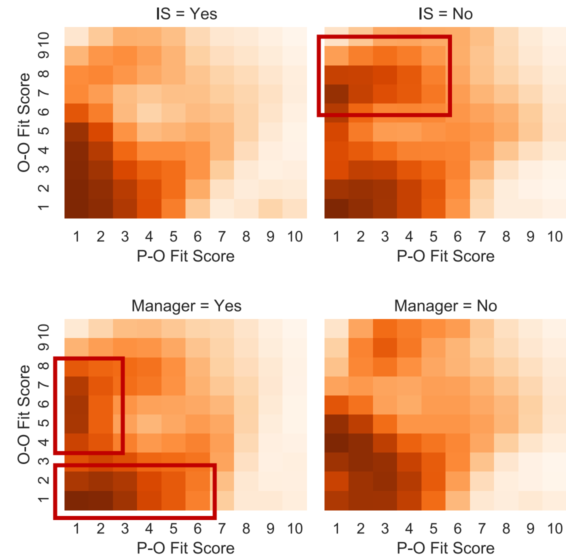

Figure 5. Distribution of fit scores by EMG

to increase the smoothness of the distribution. We have several interesting observations from Figure 4. First, there is a clear and smooth transition of color shades from bottom-left to top-right, which well reflects that our two fit scores largely coincide with true *Turnover Escalations*. Second, most of the darker squares reside in the near-diagonal regions, implying that we should not overlook either of the two fit scores when investigating post-M&A turnovers. Meanwhile, for the darker shades near the bottom (marked by a red box), we can observe that when O-O fit score is low and P-O fit varies from low to high, we may constantly get high turnover escalations, re-affirming our earlier argument that O-O fit potentially contributes more in identifying *Turnover Escalation* cases.

Owing to our model's unique capability of distinguishing different EMGs, we can investigate the distribution of the two fit scores in greater detail. We first segment all EMGs into IS-related (FUN #1, #2, #5 groups in Table 3) and non-IS-related groups. Then, we plot the distribution of the two fit scores for each group separately, as shown at the top of Figure 5. We observe a darker region in the top-left corner of the non-IS plot, which indicates that non-IS employees have a higher chance of quitting if they cannot fit in the new company even in the case of high O-O fit scores. Likewise, we re-segment all EMGs into Manager-related (RES #1, #2, #3, and #7 groups in Table 3) and non-Manager-related groups and show the fit score distribution at the bottom of Figure 5. We find that, in the plot of Manager-related groups, darker squares digress from the diagonal. It may imply that senior-level employees are more likely to quit if single fit scores is too low (i.e., the fit scores are notably unbalanced).

Conclusion

In this paper, we study how to predict post-M&A turnover escalation at a fine-grained level by considering both the merging firm-level fit and the person-level fit. To the best of our knowledge, this is the first work that tackles this problem using a heterogeneous graph neural network (GNN) approach to extract the dual-fit of the three-way relationship among acquirer, acquiree, and the employees. We propose a novel heterogeneous GNN approach with a dual-fit design (DHGNN), to extract informative node feature representations of the three-way relationships, which reveal rich semantic, structural patterns that would not be uncovered by the traditional classification models or homogeneous graph models. We conducted extensive experiments on real-world datasets to show that DHGNN significantly outperformed all the benchmarking methods based on classification metrics and the importance of our dual-fit design and effectiveness of the heterogeneous graph node embeddings. Our discussions on the two fit scores showcase the advantage of dual-fit model design and reveal interesting insights of post-M&A turnover escalation.

We discuss some limitations which help foster more future research. First, despite that our model demonstrates its superiority for the task at hand, it is still challenging to understand which factors/features are most effective. Given the heterogeneous network nature of our model, popular methods for model interpretability (e.g., SHAP Lundberg and Lee, 2017) are not directly adoptable. We are dedicated to enhance our model's interpretability as a critical next step. Moreover, our current method does not consider financial attributes mainly due to data unavailability for private companies. It would also be interesting to investigate how financial information could impact post-M&A turnover escalation.

References

Atwood, J. and D. Towsley (2016). "Diffusion-convolutional neural networks". In: *Advances in Neural Information Processing Systems* 29.

Bauer, F. and K. Matzler (2014). "Antecedents of M&A success: The role of strategic complementarity, cultural fit, and degree and speed of integration". In: *Strategic Management Journal* 35.2, pp. 269–291.

Buono, A. F. and J. L. Bowditch (2003). *The human side of mergers and acquisitions: Managing collisions between people, cultures, and organizations*. Beard Books.

Butler, J. S., R. Garg, and B. Stephens (2020). "Social networks, funding, and regional advantages in technology entrepreneurship: An empirical analysis". In: *Information Systems Research* 31.1, pp. 198–216.

Chen, D., H. Gao, and Y. Ma (2021). "Human capital-driven acquisition: evidence from the inevitable disclosure doctrine". In: *Management Science* 67.8, pp. 4643–4664.

Dalle, J.-M., M. Den Besten, and C. Menon (2017). "Using Crunchbase for economic and managerial research". In: *OECD Science, Technology and Industry Working Papers* 2017/08.

Das, A. and S. Kapil (2012). "Explaining M&A performance: a review of empirical research". en. In: *Journal of Strategy and Management* 5.3, pp. 284–330.

Greenwood, R., C. R. Hinings, and J. Brown (1994). "Merging professional service firms". In: *Organization Science* 5.2, pp. 239–257.

Hausknecht, J. P. and C. O. Trevor (2011). "Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications". In: *Journal of Management* 37.1, pp. 352–388.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". In: *Neural Comp.* 9.8, pp. 1735–1780.

Hom, P. W., T. W. Lee, J. D. Shaw, and J. P. Hausknecht (2017). "One hundred years of employee turnover theory and research." In: *Journal of Applied Psychology* 102.3, p. 530.

Homburg, C. and M. Bucerius (2006). “Is speed of integration really a success factor of M&A? An analysis of the role of internal and external relatedness”. In: *Strategic Management Journal* 27.4, pp. 347–367.

Hu, Z., Y. Dong, K. Wang, and Y. Sun (2020). “Heterogeneous graph transformer”. In: *Proceedings of The Web Conference 2020*, pp. 2704–2710.

Joseph, D., K.-Y. Ng, C. Koh, and S. Ang (2007). “Turnover of information technology professionals: A narrative review, meta-analytic structural equation modeling, and model development”. In: *MIS Quarterly*, pp. 547–577.

Kim, J. D. (2020). *Startup acquisitions as a hiring strategy: worker choice and turnover*. en. SSRN Scholarly Paper ID 3252784. Rochester, NY: Social Science Research Network.

King, D. R., D. R. Dalton, C. M. Daily, and J. G. Covin (2004). “Meta-Analyses of Post-Acquisition Performance: Indications of Unidentified Moderators”. In: *Strategic Management Journal* 25.2, pp. 187–200.

Kingma, D. P. and J. Ba (2015). “Adam: A Method for Stochastic Optimization”. In: *Proceedings of the International Conference on Learning Representations*.

Kipf, T. N. and M. Welling (2016). “Semi-Supervised Classification with Graph Convolutional Networks”. In: *CoRR* abs/1609.02907. arXiv: 1609.02907.

Krug, J. A., P. Wright, and M. J. Kroll (2014). “Top management turnover following M&A: Solid research to date but still much to be learned”. In: *Academy of Management Perspectives* 28.2, pp. 147–163.

Kyei-Poku, I. A. and D. Miller (2013). “Impact of employee merger satisfaction on organizational commitment and turnover intentions: A study of a Canadian financial institution”. In: *International Journal of Management* 30.4, p. 205.

Lappas, T. (2020). “Mining career paths from large resume databases: evidence from IT professionals”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 14.3, pp. 1–38.

Lee, K. and J. M. Pennings (1996). *Mergers and acquisitions: strategic–organizational fit and outcomes*. Tech. rep. The Wharton School, University of Pennsylvania.

Lee, T. W., T. R. Mitchell, B. C. Holtom, L. S. McDaneil, and J. W. Hill (1999). “The unfolding model of voluntary turnover: A replication and extension”. In: *Academy of Management journal* 42.4, pp. 450–462.

Liu, J., Y. C. Ng, K. L. Wood, and K. H. Lim (2020a). “IPOD: a large-scale industrial and professional occupation dataset”. In: *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, pp. 323–328.

Liu, Y., G. Pant, and O. R. Sheng (2020b). “Predicting labor market competition: leveraging interfirm network and employee skills”. In: *Information Systems Research* 31.4, pp. 1443–1466.

Lundberg, S. M. and S.-I. Lee (2017). “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems*, pp. 4765–4774.

March, J. G. and H. A. Simon (1993). *Organizations*. John wiley & sons.

Mitchell, T. R. and T. W. Lee (2001). “The unfolding model of voluntary turnover and job embeddedness: Foundations for a comprehensive theory of attachment”. In: *Research in Organizational Behavior* 23, pp. 189–246.

Narayanan, S. et al. (2019). “Analysis of merger & acquisition frameworks from a deal rationale perspective in technology sector”. PhD thesis. Massachusetts Institute of Technology.

O’Reilly III, C. A., J. Chatman, and D. F. Caldwell (1991). “People and organizational culture: A profile comparison approach to assessing person-organization fit”. In: *Academy of Management Journal* 34.3, pp. 487–516.

Palich, L. E., L. B. Cardinal, and C. C. Miller (2000). “Curvilinearity in the diversification–performance linkage: an examination of over three decades of research”. In: *Strategic Management Journal* 21.2, pp. 155–174.

Pennington, J., R. Socher, and C. D. Manning (2014). “Glove: Global vectors for word representation”. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pp. 1532–1543.

Schlichtkrull, M., T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling (2018). “Modeling relational data with graph convolutional networks”. In: *European Semantic Web Conference*. Springer, pp. 593–607.

Schuler, R. and S. Jackson (2001). “HR issues and activities in mergers and acquisitions”. In: *European Management Journal* 19.3, pp. 239–253.

Shi, Z., G. M. Lee, and A. B. Whinston (2016). “Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence”. In: *MIS Quarterly* 40.4, pp. 1035–1056.

Steigenberger, N. and N. Mirc (2020). “Should I stay or should I go? Multi-focus identification and employee retention in post-acquisition integration”. In: *Human Relations* 73.7, pp. 981–1009.

Sun, Y., F. Zhuang, H. Zhu, X. Song, Q. He, and H. Xiong (2019). “The impact of person-organization fit on talent management: A structure-aware convolutional neural network approach”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1625–1633.

Teng, M., H. Zhu, C. Liu, C. Zhu, and H. Xiong (2019). “Exploiting the contagious effect for employee turnover prediction”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01, pp. 1166–1173.

Thanos, I. C. and V. M. Papadakis (2012). “The use of accounting-based measures in measuring M&A performance: a review of five decades of research”. In: *Adv. in Mergers and Acquisitions*. Vol. 10, pp. 103–120.

Thorndike, R. L. (1953). “Who belongs in the family”. In: *Psychometrika*. Citeseer.

Trichterborn, A., D. Zu Knyphausen-Aufseß, and L. Schweizer (2016). “How to improve acquisition performance: The role of a dedicated M&A function, M&A learning process, and M&A capability”. In: *Strategic Management Journal* 37.4, pp. 763–773.

Tuch, C. and N. O’Sullivan (2007). “The Impact of Acquisitions on Firm Performance: A Review of the Evidence”. In: *International Journal of Management Reviews* 9.2, pp. 141–170.

Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio (2018). “Graph Attention Networks”. In: *Proceedings of the International Conference on Learning Representations*.

Wang, X., H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu (2019). “Heterogeneous graph attention network”. In: *The World Wide Web Conference*, pp. 2022–2032.

Westerman, J. W. and L. A. Cyr (2004). “An integrative analysis of person–organization fit theories”. In: *International Journal of selection and assessment* 12.3, pp. 252–261.

Yang, C., Y. Xiao, Y. Zhang, Y. Sun, and J. Han (2020). “Heterogeneous Network Representation Learning: Survey, Benchmark, Evaluation, and Beyond”. In: *CoRR* abs/2004.00216. arXiv: 2004.00216.

Appendix

Data Details

We have several major observations of our data sample:

- *M&A deals, acquirer and acquiree*: There are in total 2,566 M&A deals and 1,861 acquirers. Each acquiree associates with only one M&A transaction. The majority of M&A transactions occurred between 2010 and 2018 (Figure 6).
- *Geo-location*: The firms in our data sample are distributed in 947 different US cities, most of which are located at 'New York', 'San Francisco', and 'Austin' (Figure 7).
- *Industry keywords*: There are 48 distinct industry keywords attached to the firms in our dataset, which characterize firms' market sectors. These industry keywords are more granular categorizations compared with other standard industry codes, such as SIC code[3]. We provide a Word Cloud in Figure 8 showing common keywords, including 'Software', 'Internet Services', and 'Data and Analytics'. Note that, on average, each firm is associated with 3.5 industry keywords.
- *Employee and job records*: We have accumulated substantial employment-related data of acquiree firms, including 800K employees and over 1.2M job records. Each acquiree has an average of 327 employees and each employee has averagely 1.5 job records. Figure 9 outlines the dates on when these workers joined *(START year)* and departed *(END year)* their employers. We can observe that the bulk of career transitions occurred in 2010-2018, which chronologically align with the occurrences of most M&A transactions.
- *Employee groups*: To better understand turnover behaviors of different employee groups, we conduct a clustering analysis to generate 64 employee groups. Next subsection will discuss more analytical details.

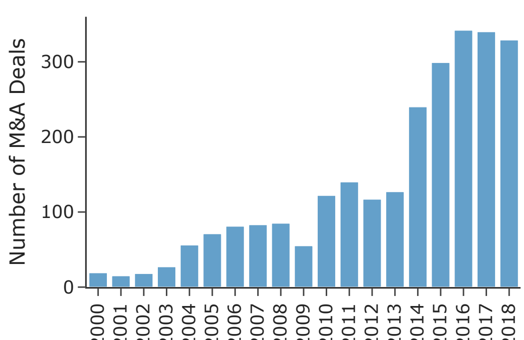

Figure 6. Distribution of M&A deals by year (2000-2018).

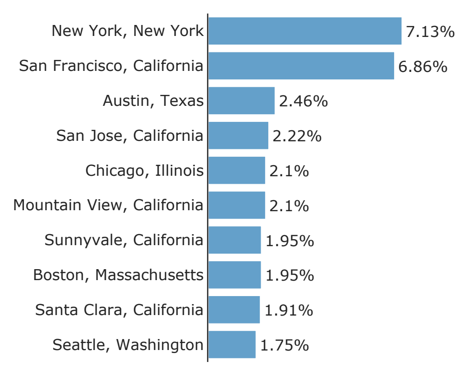

Figure 7. Percentage distribution of Top 10 most frequent geolocations.

[3]https://siccode.com/sic-code-lookup-directory

Figure 8. A word cloud visualization of industry keywords.

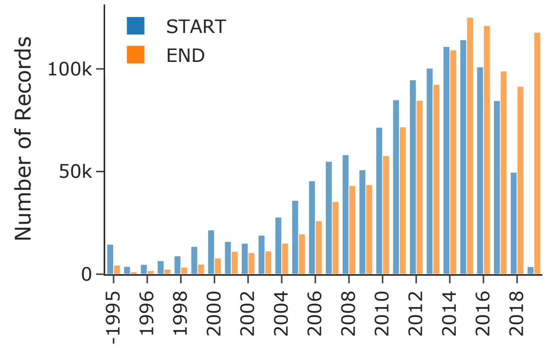

Figure 9. Distribution of job records by *START* and *END* years (1995-2019).

Common Job Terms

FUN and *RES* job terms are clustered into eight groups respectively. Table 3 showcases the common job terms in each group.

Table 3. Common Job Terms in *FUN* and *RES* Groups.

Group	Functionality	Responsibility
1	project, technical, engineering, development, support technical, technology, production	manager, manager senior, supervisor, manager regional, director management, associate manager
2	software, systems, solutions, enterprise, qa, security, solution, tech, development software, digital	lead, representative, director senior, leader, senior, associate senior, designer senior, lead senior
3	business, business development, strategic, corporate, business operations, strategy	vp, vice, rep, operator, controller, co, owner, clerk, vice president
4	customer service, service, customer, customer support, care customer, customer success	editor, writer, producer, artist, senior writer, editor senior, associate producer, editor writer
5	data, test, research, technology, field, information technology, healthcare	engineer, engineer senior, developer, technician, engineer lead, engineer principal, engineer staff
6	operations, product, support, team, program, quality, system, hr, network, channel	specialist, analyst, consultant, analyst senior, consultant senior, advisor, expert, consultant principal
7	supply chain, retail, commercial, electrical, store, products, storage, travel, purchasing	director, president vice, executive, head, executive senior, associate director, director regional
8	finance, financial, market, contract, compliance, accounting, accounts, credit, accounts strategic	intern, associate, assistant, trainer, recruiter, instructor, trainee, internship, student, generalist

Raw Attributes and Feature Engineering

Regarding all triplets (ACR_i, ACE_j, EMG_k) in our data, we have basic attributes and handcrafted features as shown in Table 4. The attributes reflect different aspects of the queried company pair and employee group, provide valuable information concerning the acquirer-acquiree fit and person-organization fit. For initial exploration, we developed a Random Forest (RF) model considering its good interpretability and robustness. More details can be found in Section *Experimental Results*.

Table 4. Raw Attributes and Hand-crafted Features for Post-M&A Turnover Prediction.

	Attributes & Features	Remarks
Firm-level	Business category	Industry keywords
	Geo-location	City, State, Country
	Firm size	Number of Employees
	Firm age	Number of Years since establishment
	Investors	Investors who have invested the firm
	Executives	Executive members in the firm
	Business proximity	Shared industry keywords
	Geographic proximity	Located in the same city, state, or country
	Investor proximity	Shared investors
	Relative size	Ratio of acquiree size over acquirer size
Employee-level	Job titles	Job function and responsibility terms
	Past employment records	Previous employers and job records

Loss Function and Model Training

Following the model design in Section *Organization-Organization Fit and Person-Organization Fit*, as the final step, for the input triplet (ACR_i, ACE_j, EMG_k), we use the O-O fit vector $x_{i,j}^{acr,ace}$ and P-O fit vector $x_{i,k}^{acr,emg}$ to generate O-O fit score s_{oo} and P-O fit score s_{po} respectively via fully-connected layers and non-linear activation function (sigmoid). Each score is in the range of $[0, 1]$. Lastly an overall fit score $s_{i,j,k}$ for the input triplet is obtained by averaging O-O fit and P-O fit scores, i.e., $s_{i,j,k} = (s_{oo} + s_{po})/2$. We train the entire model using the cross-entropy as the loss function and adopt Adam optimizer (Kingma and Ba, 2015) for mini-batch stochastic gradient descent,

$$\mathcal{L} = - \sum_{i,j,k \in \mathcal{B}} [y_{i,j,k} \cdot \log(1 - s_{i,j,k}) + (1 - y_{i,j,k} \cdot \log(s_{i,j,k}))] \tag{8}$$

where $\mathcal{B}$ denotes one random batch of the entire triplets data, $(i, j, k) \in \mathcal{B}$ stands for the indices for the triplet from the current batch. $y_{i,j,k}$ is the groundtruth label of turnover escalation. To be noted, the fit score $s_{i,j,k}$ has opposite optimization direction to the turnover escalation prediction variable. In other words, the larger the fit score is (i.e., $s_{i,j,k} \to 1$), the less likely a turnover escalation happen (i.e., $y_{i,j,k} = 0$).