Transactions on Computational and Scientific Methods | Vo. 4, No. 1, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Attrition-Aware Spatio-Temporal Graph Learning and Stochastic Programming for Multi-Horizon Workforce Demand Forecasting and Labor Cost Planning

You Lin
Cornell University, Ithaca, USA
lindoudou9117@gmail.com

Abstract: Labor is typically the largest controllable operating expense in modern enterprises, yet workforce planning remains siloed: demand forecasting, attrition prediction, and headcount budgeting are performed independently, causing systematic over-staffing or costly reactive hiring. This paper presents a unified framework that couples attrition-aware spatio-temporal graph learning with stochastic labor-cost optimization. The organization is modeled as a heterogeneous graph over roles, skills, and departments, on which a spatio-temporal graph transformer jointly forecasts workload demand, role-level staffing requirements, and cohort-level attrition hazards over horizons of 1 to 12 months; a cross-task attention module lets attrition signals recalibrate future capacity. Forecast distributions then feed a two-stage stochastic program that jointly optimizes planned and emergency hiring, overtime, contractor usage, and internal transfers under service-level and budget constraints, yielding auditable plans that are optimal for the sampled scenario set. On a 72-month benchmark for two enterprise-scale organizations (approximately 8,500 and 21,000 employees), synthesized from the public IBM HR Analytics attrition dataset with simulated operational demand, the framework reduces the weighted absolute percentage error (WAPE) of staffing-requirement forecasts by about 15% and 19% relative to independent long short-term memory (LSTM) and Prophet pipelines, and cuts simulated total labor cost by about 6–7% while maintaining a service level of approximately 98%, primarily by shifting spend from emergency hiring and overtime toward planned recruitment; the learned attention also yields faithful, interpretable indicators of team-level attrition risk.

Keywords: Demand forecasting, employee attrition, labor cost optimization, spatio-temporal graph neural networks, stochastic programming, workforce planning.

1. Introduction

Personnel cost is usually the largest controllable line item in an enterprise budget, so small planning improvements yield substantial savings. In practice, the decision chain is fragmented: business units forecast demand, human-resources (HR) analysts predict attrition, and finance sets headcount budgets, each using separate tools. This propagates errors: a forecast that ignores impending departures overestimates available capacity, and a budget that ignores seasonality either buffers with idle staff or forces costly last-minute hiring and overtime.

Existing research mirrors this fragmentation. Attrition is studied as a static tabular classification problem [1], producing per-employee risk scores that are rarely connected to capacity or cost. Demand is forecast with interpretable additive models such as Prophet [2] or with deep sequence models, but these treat each series

independently and ignore organizational structure and workforce dynamics. Staffing under demand uncertainty is addressed with two-stage stochastic programming [3], yet demand is typically exogenous and attrition is not modeled jointly. Reinforcement learning (RL) has been applied to operational planning in supply chains and related operations settings [4]; such policies are adaptive but are typically trained in simulators, motivating our emphasis on auditable, optimization-based plans.

We argue that forecasting and budgeting should be solved together, and that the organizational graph linking departments, roles, and skills carries signal that siloed models discard. Our framework has two coupled components. First, an attrition-aware spatio-temporal graph transformer represents the organization as a heterogeneous graph and jointly forecasts workload demand, staffing requirements, and cohort attrition hazards over 1 to 12 months, with cross-task attention through which predicted attrition recalibrates future capacity. Second, a two-stage stochastic program consumes the forecast distributions and jointly optimizes planned and emergency hiring, overtime, contractors, and internal transfers under service-level and budget constraints.

Our contributions are: (i) a heterogeneous spatio-temporal graph model that jointly forecasts demand, staffing, and attrition; (ii) a cross-task attention mechanism that turns attrition hazards into capacity corrections and exposes feature-level team risk attributions; (iii) a prediction–prescription coupling in which quantile forecasts and an explicit capacity identity drive a two-stage stochastic labor-cost program—the contribution is the coupling and its auditability, not stochastic programming itself; and (iv) an evaluation on a reproducible, publicly grounded semi-synthetic benchmark with released code.

2. Related Work

Multi-horizon forecasting. Additive models such as Prophet (Section I) are interpretable but treat each series independently, and transformer forecasters such as the Temporal Fusion Transformer (TFT) [5], built on the attention mechanism [6], improve multi-horizon accuracy yet remain largely univariate and model neither attrition nor organizational topology.

Spatio-temporal graph learning. Graph neural networks capture relational structure: attention over neighbors [7] improves expressiveness, and the heterogeneous graph transformer [8] adds type-dependent attention for multiple node and edge types. Prior spatio-temporal transformers target homogeneous sensor networks and single-task prediction; we adapt heterogeneous spatio-temporal learning to the workforce domain with multiple forecasting heads and explicit cross-task coupling.

Workforce optimization under uncertainty. Manpower planning has a long tradition of Markov and flow models of headcount [9], and staffing under demand uncertainty is commonly cast as a two-stage stochastic program (Section I) with a mature theory of recourse [10]; there, demand is exogenous and attrition is not jointly forecast. Our approach couples prediction and prescription in a single modular pipeline: the network is trained with forecasting losses, and the optimizer consumes its predictive distributions. This differs from decision-focused "predict-then-optimize" learning [11], which back-propagates a task loss through the optimization layer; we keep the stages separable for auditability and leave end-to-end training to future work. RL approaches (Section I) add adaptivity and are complementary to an auditable optimization stage.

3. Problem Formulation

We model an organization as a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node types $\mathcal{T} = \{department, role, skill\}$ and typed relations $\mathcal{R}$ (a role belongs to a department, a role requires a skill, and skills co-occur). Nodes carry type-specific features: department and role nodes include monthly workload, headcount, and HR-event series, whereas skill nodes carry utilization and scarcity indicators; a type-specific projection maps all types into a shared latent space. Given a lookback window of length L, the forecasting task is to predict, for horizons $h = 1, \ldots, H$ with $H = 12$, the workload demand y^d, the role-

level staffing requirement y^s, and the cohort attrition hazard h^a. The planning task chooses a labor plan that minimizes the expected cost while meeting a service-level target and a budget. Forecast quality is measured by the weighted absolute percentage error (WAPE),

$$\mathrm{WAPE} = \frac{\sum_i \left| y_i - \widehat{y}_i \right|}{\sum_i \left| y_i \right|},$$

summed over all evaluated (role, horizon) pairs i in the test folds; WAPE is scale-robust across roles of very different sizes.

4. Methodology

Figure 1 shows the framework. Vectors are set in bold (e.g., the node state $\mathbf{z}_v$), so the scalar planning variables introduced later, such as the transfer amount $z_{rr'}$, are distinct from them. Node features are first mapped by a type-specific projection so that departments, roles, and skills share a latent space:

$$\mathbf{g}_v^{(0)} = \mathbf{W}_{\tau(v)} \mathbf{x}_v + \mathbf{b}_{\tau(v)},$$

where $\tau(v)$ is the type of node v.

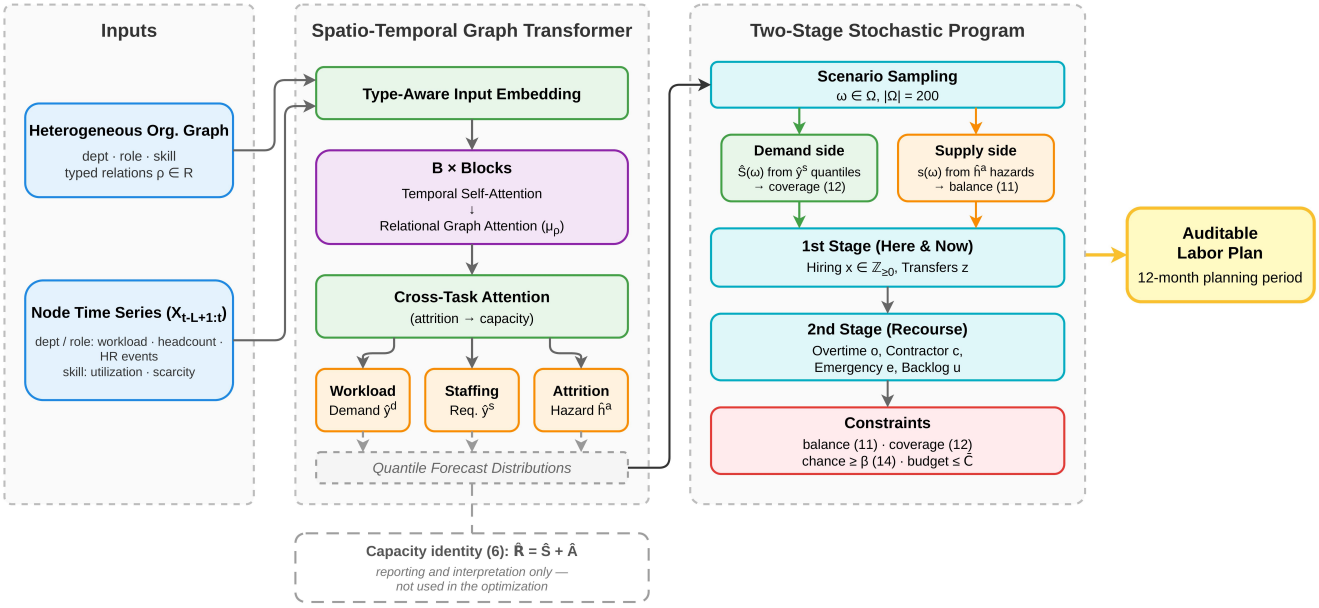

Figure 1. Overview of the framework. A heterogeneous organizational graph (departments, roles, and skills) and node time series are encoded by a spatio-temporal graph transformer whose blocks alternate temporal self-attention and relational graph attention; cross-task attention lets the attrition signal recalibrate capacity, and three heads emit workload-demand (), staffing (), and attrition-hazard () forecasts. The predictive distributions feed a two-stage stochastic program—first-stage hiring and transfers with second-stage overtime, contractor, emergency-hire, and backlog recourse—yielding an auditable labor plan for the 12-month planning period.

4.1 Spatio-Temporal Graph Transformer

We factorize spatio-temporal dependence by alternating a temporal encoder and a relational graph-attention layer. The temporal encoder applies multi-head self-attention over $\{\mathbf{g}_{v,t-L+1}, \ldots, \mathbf{g}_{v,t}\}$ to obtain a per-node

state $\mathbf{z}_v$. In the spirit of the heterogeneous graph transformer (Section II), relational attention aggregates typed neighbors, with a learnable relation prior μ_ρ per relation type $\rho \in \mathcal{R}$:

$$\alpha_{uv}^{(\rho)} = \underset{u \in \mathcal{N}_\rho(v)}{\mathrm{softmax}}\left(\frac{\left(\mathbf{W}_Q^{\tau(v)}\mathbf{z}_v\right)^\top \left(\mathbf{W}_K^{\tau(u)}\mathbf{z}_u\right)\mu_\rho}{\sqrt{d}} \right),$$

$$\mathbf{z}'_v = \mathbf{z}_v + \sigma\left(\sum_{\rho \in \mathcal{R}} \sum_{u \in \mathcal{N}_\rho(v)} \alpha_{uv}^{(\rho)}\, \mathbf{W}_V^{\tau(u)}\mathbf{z}_u \right),$$

where $\sigma(\cdot)$ is a nonlinear activation. Stacking B such blocks lets signals propagate along skill and reporting relations, so that a shortage of a scarce skill informs the staffing forecast of every role that requires it.

4.2 Cross-Task Attention and Multi-Task Objective

A shared trunk produces the task representations $\mathbf{z}^d$, $\mathbf{z}^s$, and $\mathbf{z}^a$. Rather than predicting these independently, we let the staffing representation attend to the attrition representation:

$$\tilde{\mathbf{z}}^s = \mathbf{z}^s + \mathrm{softmax}\left(\frac{\left(\mathbf{z}^s\mathbf{W}_Q^c\right)\left(\mathbf{z}^a\mathbf{W}_K^c\right)^\top}{\sqrt{d}} \right)\mathbf{z}^a\mathbf{W}_V^c.$$

Complementing this learned coupling, we expose an explicit, deterministic capacity identity that is used only for reporting and interpretation and does not enter the optimization: the total requirement equals the workload-driven staffing forecast $\widehat{S} \equiv \widehat{y}^s$ plus the headcount a role is expected to lose,

$$\widehat{R}_{r,t+h} = \widehat{S}_{r,t+h} + n_{r,t}\left(1 - \prod_{k=1}^{h}\left(1 - \hat{h}_{r,t+k}^a\right)\right),$$

where $n_{r,t}$ is the current headcount and the product is the survival probability implied by the discrete-time hazards. The demand and staffing heads emit quantile forecasts trained with the pinball loss ($\mathcal{L}^d$, $\mathcal{L}^s$); the attrition head emits discrete-time hazards [12] trained with the negative log-likelihood over at-risk periods,

$$\mathcal{L}^a = -\sum_{r,k} \delta_{r,k}\left[y_{rk}\log\hat{h}_{rk}^a + \left(1 - y_{rk}\right)\log\left(1 - \hat{h}_{rk}^a\right)\right],$$

with $\delta_{r,k} = 1$ while the cohort of role r is still at risk. The model is trained by minimizing

$$\mathcal{L} = \lambda_d\mathcal{L}^d + \lambda_s\mathcal{L}^s + \lambda_a\mathcal{L}^a + \gamma\|\theta\|_2^2.$$

4.3 Two-Stage Stochastic Labor-Cost Program

Point forecasts hide the tail risk that drives premium cost, so we propagate uncertainty into planning for a single aggregated period that represents the 12-month horizon. Scenarios $\omega \in \Omega$ are sampled from the predictive quantiles. First-stage decisions commit to integer planned hires $x_r \in \mathbb{Z}_{\geq 0}$ and internal transfers $z_{rr'} \geq 0$ (the cheaper, slower channel). Second-stage decisions react to each scenario with overtime $o_r(\omega)$, contractors $c_r(\omega)$, expedited emergency hires $e_r(\omega)$, and unmet backlog $u_r(\omega)$, all nonnegative, plus a coverage indicator $\xi_r(\omega) \in \{0, 1\}$; emergency hiring is the fast, premium counterpart of planned hiring,

and the two cost tiers serve as a proxy for the hiring lead-time gap rather than modeling explicit timing. With the here-and-now cost $C_1(x,z) = \sum_r \left[w_r (n_r + x_r) + \kappa_r^x x_r \right] + \sum_{r,r'} \kappa^z z_{rr'}$, the program is

$$\min_{x \in \mathbb{Z}^R_{\geq 0},\ z \geq 0} C_1(x,z) + \mathbb{E}_\omega[Q(x,z;\omega)],$$

$$Q(x,z;\omega) = \min_{o,c,e,u \geq 0,\ \xi \in \{0,1\}^R} \sum_r \left(\kappa_r^o o_r + \kappa_r^c c_r + \kappa_r^e e_r + \kappa_r^u u_r \right),$$

subject to, for every role r and scenario ω,

$$a_r(\omega) = n_r - s_r(\omega) + x_r + e_r(\omega) + \eta\, o_r(\omega) + c_r(\omega) + \sum_{r'} \left(z_{r'r} - z_{rr'} \right),$$

$$a_r(\omega) + u_r(\omega) \geq \widehat{S}_r(\omega),$$

$$a_r(\omega) - \widehat{S}_r(\omega) \geq -M \left(1 - \xi_r(\omega) \right),$$

$$\frac{1}{|\Omega|} \sum_\omega \xi_r(\omega) \geq \beta,$$

a budget $C_1(x,z) + \mathbb{E}_\omega[Q] \leq \bar{C}$ (set to the reactive-baseline cost and slack at every reported optimum), and transfers limited to skill-compatible pairs, $z_{rr'} = 0$ for all $(r,r') \notin \mathcal{C}$, and to a fraction $\zeta = 0.15$ of headcount, $\sum_{r'} z_{rr'} \leq \zeta n_r$. Here, $\widehat{S}_r(\omega)$ is the workload-driven staffing requirement realized in scenario ω, drawn from the predictive distribution of y^s; $s_r(\omega)$ is the number of employee separations realized in scenario ω, sampled from the predicted hazards $\hat{h}^a$; η converts overtime hours into full-time-equivalent (FTE) capacity; and M is a sufficiently large big-M constant. Crucially, the coverage constraint (12) uses the workload-driven requirement $\widehat{S}$, and attrition enters only through $s_r(\omega)$ on the supply side, so the capacity identity (6) is interpretive and never double-counts attrition. The backlog penalty κ^u enforces coverage, so at the optimum $u_r(\omega) \approx 0$. The result is a mixed-integer linear program (MILP) that is optimal with respect to the sample-average approximation (SAA) over Ω [13]; once transfers and the budget are relaxed, it is separable by role, which explains the fast solve times.

5. Experiments

5.1 Data and Setup

Because longitudinal, department-level HR and operations data are proprietary, we build a reproducible semi-synthetic benchmark. The public IBM HR Analytics dataset [14] is a fictitious cross-sectional record of 1,470 employees with 35 attributes and an overall attrition label (≈16%); it has no time dimension. We use it only to calibrate covariate–attrition relationships (overtime, tenure, compensation, and satisfaction) and synthesize a 72-month operational panel around them, so the benchmark is explicitly a simulation informed by public data. Monthly hazards (≈1.1–1.3%) follow from allocating each employee's attrition probability over time (exponential survival), and records are expanded to organizational scale by stratified resampling that preserves those relationships. Department-level workload is simulated with trend, annual seasonality, and demand shocks; staffing requirements follow via role productivity coefficients. Two

organizations are instantiated: Org-A and Org-B have about 8,500 and 21,000 employees across 8 and 14 departments, 62 and 118 roles, and 140 and 265 skills (210 and 397 graph nodes, respectively), on 72-month panels with 1.3% and 1.1% average monthly attrition. The benchmark generator and code are released for reproducibility (anonymized: https://anonymous.4open.science/r/awstgp-benchmark).

We use a lookback of $L = 12$ months and evaluate horizons 1–12 with a six-fold rolling-origin (expanding-window) protocol, reporting the mean and standard deviation across folds; the 72-month panel yields dozens of input–label windows per series. The stochastic program uses $|\Omega| = 200$ scenarios and a service target of $\beta = 0.98$. Cost parameters are normalized to a base wage $w = 1.0$ per FTE for the planning period: planned-hire premium $\kappa^x = 0.20$, contractor rate $\kappa^c = 1.45$, emergency-hire cost $\kappa^e = 2.30$, backlog penalty $\kappa^u = 4.00$, and transfer cost $\kappa^z = 0.30$; overtime $\kappa^o = 0.30$ with $\eta = 0.20$ is a marginal charge of $\kappa^o/\eta = 1.50$ per FTE on top of incumbent payroll. On a per-FTE basis, the channels are therefore ordered planned (1.20) < contractor (1.45) < overtime (1.50) < emergency (2.30). The MILP has on the order of 10^4–10^5 variables and is solved to a 0.5% optimality gap with Gurobi 11.0 [15] in under 90 s per instance.

5.2 Baselines and Metrics

We compare against independent pipelines and ablations. Prophet and an LSTM sequence-to-sequence model forecast demand and staffing, each paired with a gradient-boosted-tree (GBT) attrition classifier; the TFT serves as a strong multi-task baseline. Ablations remove the skill graph (homogeneous variant) or the cross-task attention. Forecasts are scored with WAPE; attrition is scored with the area under the receiver operating characteristic curve (AUROC) for a six-month departure event at the role-cohort level, with baseline per-employee scores averaged to that level. Planning is scored by the normalized total labor cost, its composition, and service-level attainment (the fraction of role-months meeting the requirement).

5.3 Forecasting Results

Table 1 reports accuracy with per-fold standard deviations. The full model attains a staffing WAPE of 0.154, about 15% below the LSTM pipeline (0.181) and 19% below Prophet (0.191), and raises the attrition AUROC from 0.799 (GBT) to 0.863. Ablations isolate the two sources of gain: removing cross-task attention raises the staffing WAPE to 0.158, and further removing the skill graph raises it to 0.163. These absolute gaps are below the across-fold standard deviations, but because the folds are paired, the relevant quantity is the fold-wise difference: it is 0.004±0.003 (removing cross-task attention) and 0.005±0.004 (further removing the skill graph), where ± denotes the standard deviation of the difference across folds; both are significant at $p < 0.05$ under a paired t-test, and the full model ranks first in every fold. Error grows with the horizon for all methods (Figure 2), but the margin of the proposed model is largest at the long horizons that matter most for planning ahead of hiring lead times.

Table 1. Forecasting accuracy (mean ± standard deviation across six rolling-origin folds), averaged over both organizations and all horizons. Lower (↓) / higher (↑) is better.

Method	Workload WAPE ↓	Staffing WAPE ↓	Attrition AUROC ↑
Prophet + GBT	0.142±0.007	0.191±0.009	0.782±0.012
LSTM seq2seq + GBT	0.136±0.006	0.181±0.008	0.799±0.011
TFT (multi-task)	0.121±0.005	0.165±0.007	0.831±0.010
Ours, homogeneous graph	**0.119±0.005**	**0.163±0.006**	**0.839±0.009**
Ours, w/o cross-task attn.	**0.115±0.005**	**0.158±0.006**	**0.850±0.009**
Ours (full)	**0.113±0.004**	**0.154±0.006**	**0.863±0.008**

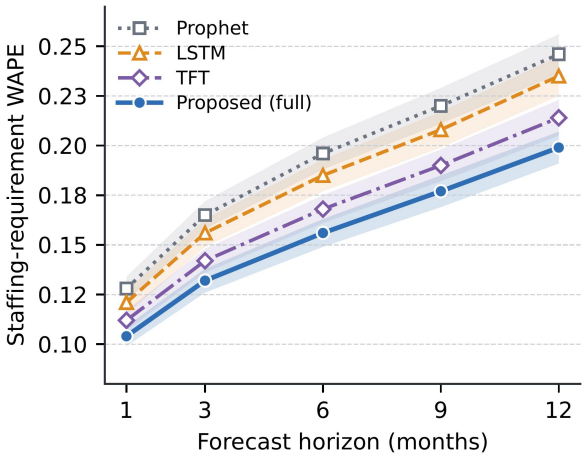

Figure 2. Staffing-requirement WAPE versus forecast horizon, averaged over both organizations; shaded bands are ±1 standard deviation across the six folds. The advantage of the proposed model over independent pipelines widens at long horizons.

5.4 Labor-Cost Planning Results

Figure 3 reports planning outcomes, normalized so that the reactive baseline equals 100. Coupling attrition-aware forecasts with the stochastic program lowers the total cost to 93.3 (a 6.7% pooled reduction) at a 98.1% service level. The savings come from reallocation rather than understaffing: planned recruitment rises from 6% to 16% of spend, while emergency hiring falls from 9% to 5%, overtime from 16% to 10%, and contractors from 13% to 9%. Deterministic planning on point forecasts improves over the reactive policy but under-hedges tail demand; optimizing over scenarios strikes a better cost–risk balance. Per organization, the reduction is 5.9% (Org-A) and 7.0% (Org-B), with realized service levels of 98.0% and 98.3%; the larger organization benefits more from pooling flexibility across roles. Service levels are out-of-sample realized attainment on held-out scenarios. Because the backlog penalty κ^u dominates the recourse cost, at the optimum $u_r(\omega) \approx 0$ and the in-sample coverage exceeds β (≈99.5% for the proposed plan); the chance constraint (14) is thus a non-binding floor, and coverage is driven in practice by κ^u. Out-of-sample attainment is lower, falling to 97.5% for the independent stochastic plan.

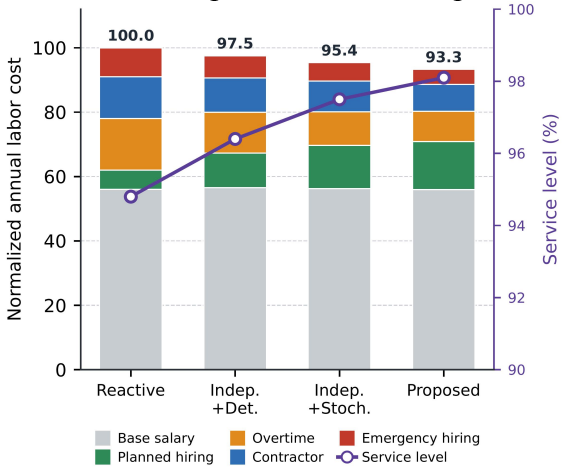

Figure 3. Composition of the normalized labor cost for the aggregated 12-month period under four strategies (total above each bar; service level, purple line, right axis). Bars show shares of each strategy's own total: "base salary" is incumbent payroll ; "planned hiring" is the wage plus premium of new hires, ; and the remaining segments are recourse costs (transfers, below 1%, are omitted). As the total cost falls

from 100.0 to 93.3, spend shifts from emergency hiring and overtime toward planned recruitment while the service level rises to 98.1%.

5.5 Interpretability

The attention that the attrition head places on candidate driver features is inspectable (Figure 4). It concentrates on interpretable, actionable factors: overtime load in customer support and a compensation gap in sales receive the strongest weight. Two caveats apply. First, in this semi-synthetic setting these factors are, by construction, among the generative drivers, so the result validates the faithfulness of the attention rather than discovering new causes; attention weights are indicative, not causal. Second, we make no prospective claim that these signals lead attrition by a fixed horizon—establishing genuine lead time requires longitudinal real data. Subject to these caveats, faithful attributions can help planners direct the retention budget toward the teams most at risk.

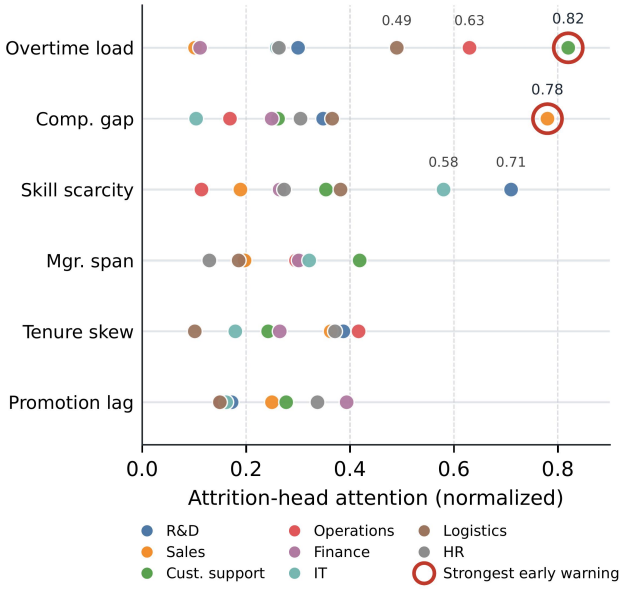

Figure 4. Attention that the attrition-hazard head places on candidate driver features (Org-A), shown as a dot plot with one row per driver and one point per department. These are input covariates, so this is a feature-attribution view, distinct from the inter-node relational attention of (3)–(4); each score is averaged over heads and the lookback window and min–max scaled to per driver. The two points circled in red—overtime in customer support and a compensation gap in sales—are, by construction, among the generative drivers.

6. Conclusion

We presented a framework that couples attrition-aware spatio-temporal graph learning with two-stage stochastic optimization for workforce forecasting and labor-cost planning. Modeling the organization as a heterogeneous graph and forecasting demand, staffing, and attrition jointly improves staffing accuracy over independent pipelines and yields interpretable, faithful risk indicators; feeding the forecast distributions into a stochastic program reduces the simulated labor cost while maintaining service levels, mainly by substituting planned recruitment for emergency hiring and overtime. The evaluation is semi-synthetic, so absolute figures depend on the simulated assumptions; the optimization is optimal only for the sampled scenario set, the quantile-marginal scenarios do not model cross-role or temporal correlation, and the service-level and budget constraints are not tight at the reported optima. Future work includes validation on proprietary longitudinal records, an explicit multi-period formulation with per-role hiring lead times, a sensitivity analysis over the service target and penalty, correlated and distributionally robust scenarios, and fairness-constrained planning.

References

[1] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. De Luca, “Predicting employee attrition using machine learning techniques,” Computers, vol. 9, no. 4, art. no. 86, 2020, doi: 10.3390/computers9040086.

[2] S. J. Taylor and B. Letham, “Forecasting at scale,” The American Statistician, vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.

[3] K. Kim and S. Mehrotra, “A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management,” Operations Research, vol. 63, no. 6, pp. 1431–1451, 2015, doi: 10.1287/opre.2015.1421.

[4] B. Rolf, I. Jackson, M. Müller, S. Lang, T. Reggelin, and D. Ivanov, “A review on reinforcement learning algorithms and applications in supply chain management,” International Journal of Production Research, vol. 61, no. 20, pp. 7151–7179, 2023, doi: 10.1080/00207543.2022.2140221.

[5] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, “Temporal fusion transformers for interpretable multi-horizon time series forecasting,” International Journal of Forecasting, vol. 37, no. 4, pp. 1748–1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017, pp. 5998–6008. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[7] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” in Proc. International Conference on Learning Representations (ICLR), 2018. [Online]. Available: https://openreview.net/forum?id=rJXMpikCZ

[8] Z. Hu, Y. Dong, K. Wang, and Y. Sun, “Heterogeneous graph transformer,” in Proc. The Web Conference 2020 (WWW), 2020, pp. 2704–2710, doi: 10.1145/3366423.3380027.

[9] D. J. Bartholomew, A. F. Forbes, and S. I. McClean, Statistical Techniques for Manpower Planning, 2nd ed. Chichester, U.K.: Wiley, 1991, ISBN: 978-0-471-92879-9. [Online]. Available: https://books.google.com/books?id=e_-qAAAAIAAJ

[10] J. R. Birge and F. Louveaux, Introduction to Stochastic Programming, 2nd ed. New York, NY, USA: Springer, 2011, doi: 10.1007/978-1-4614-0237-4.

[11] A. N. Elmachtoub and P. Grigas, “Smart ‘predict, then optimize’,” Management Science, vol. 68, no. 1, pp. 9–26, 2022, doi: 10.1287/mnsc.2020.3922.

[12] P. D. Allison, “Discrete-time methods for the analysis of event histories,” Sociological Methodology, vol. 13, pp. 61–98, 1982, doi: 10.2307/270718.

[13] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello, “The sample average approximation method for stochastic discrete optimization,” SIAM Journal on Optimization, vol. 12, no. 2, pp. 479–502, 2002, doi: 10.1137/S1052623499363220.

[14] IBM, “IBM HR Analytics employee attrition and performance dataset,” Kaggle, 2017. [Online]. Available: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

[15] Gurobi Optimization, LLC, “Gurobi Optimizer Reference Manual,” 2024. [Online]. Available: https://www.gurobi.com