Diffusion-Based Generative Modeling for Expert Team Formation

Mohammad Hossein Saliminabi[a,*], Sajad Ebrahimi[b], Radin Hamidi Rad[c], Dimitrios Androutsos[a], Fattane Zarrinkalam[b] and Ebrahim Bagheri[c]

[a]*Toronto Metropolitan University, Toronto, ON, Canada*
[b]*University of Guelph, Guelph, ON, Canada*
[c]*University of Toronto, Toronto, ON, Canada*

ARTICLE INFO

Keywords:
Diffusion Models
Team Formation
Conditional Sampling

ABSTRACT

Forming effective expert teams is central to domains where solving complex problems requires diverse, complementary skills. However, automating this task is highly challenging due to sparse co-occurrence data, long-tailed expert participation, and the combinatorial complexity of unseen skill configurations. Existing graph-based, probabilistic, and neural approaches often struggle with generalization, fairness, and robustness, leading to biased selections that favor historically popular experts over more suitable candidates. To address these challenges, we propose a generative framework for expert team formation based on denoising diffusion probabilistic models. We cast team formation as skill-conditioned imputation (i.e., inpainting), where skills are treated as observed context and the expert component is generated via conditional diffusion sampling. This design enables our method to preserve semantic skill–expert alignment, mitigate data sparsity, and generate diverse yet contextually coherent teams. Extensive experiments on DBLP and DOTA2 datasets show that our model consistently outperforms state-of-the-art baselines, achieving over 3× higher recall (16.4% vs. 5.0%) and MAP (9.7% vs. 2.2%) on DBLP, while delivering more than 5× improvement in MRR (13.3% vs. 2.5%) on DOTA2. Fairness analysis further demonstrates that our method reduces average overlap with the top-100 most popular experts to 2.6, compared to 86.7 for the strongest baseline, and achieves near-optimal diversity with NDKL ≈ 0.1 under high non-popular expert ratios.
For **reproducibility purposes**, we made our code and model publicly available at `https://anonymous.4open.science/r/DiffTF-7A20`.

1. Introduction

Forming effective teams of experts is a foundational requirement across domains where complex, multi-dimensional problems demand coordinated contributions from individuals with diverse and complementary skills. Applications span assembling academic research groups, coordinating engineering project teams, curating contributors in open-source software ecosystems, and optimizing player compositions in competitive esports (Costa, Ramos, Perkusich, Dantas, Dilorenzo, Chagas, Meireles, Albuquerque, Silva, Almeida, and Perkusich, 2020). Automating this process, widely referred to as expert team formation, promises to reduce search overhead, enhance coordination, and reveal effective team configurations that may not emerge through manual processes (Balog, Fang, De Rijke, Serdyukov, and Si, 2012; Fu, Luo, Nan, and Li, 2025). As shown in Figure 1, for forming the network, each individual is a node, and past collaboration is modeled by connecting the associated nodes of individual members together to make an edge. Then, a graph search algorithm is employed to find a subgraph of the network, in which individual members collectively cover all required skills.

Beyond the specific task of expert team formation, this work addresses a broader class of information processing problems characterized by sparse, multi-source relational data and partial observability. Such problems arise widely in information retrieval, recommendation systems, social information processing, and multi-modal data integration, where a subset of variables is observed and the remaining structure must be inferred under uncertainty (Chen, Wang, McAuley, Jannach, and Yao, 2024; Khodabakhsh and Bagheri, 2025; Elías, Jiménez, Paganoni, and Sangalli, 2023). From this perspective, team formation serves as a concrete instantiation of a more general challenge, i.e., learning conditional distributions over structured outputs from heterogeneous and incomplete signals.

*Corresponding author
msaliminabi@torontomu.ca (M.H. Saliminabi); sebrah05@uoguelph.ca (S. Ebrahimi); radin@torontomu.ca (R. Hamidi Rad); dimitri@torontomu.ca (D. Androutsos); fzarrink@uoguelph.ca (F. Zarrinkalam); ebrahim.bagheri@utoronto.ca (E. Bagheri)
ORCID(s): 0000-0000-0000-0000 (M.H. Saliminabi); 0000-0002-0000-0000 (S. Ebrahimi); 0000-0002-9044-3723 (R. Hamidi Rad); 0000-0000-0000-0000 (D. Androutsos); 0000-0000-0000-0000 (F. Zarrinkalam); 0000-0002-5148-6237 (E. Bagheri)

The task of team formation is highly challenging and can be understood as a high-dimensional combinatorial optimization problem under sparse and biased data (Dorn, Skopik, Schall, and Dustdar, 2011; Sozio and Gionis, 2010). Three issues stand out. First, expert collaboration data is sparse and long-tailed, with most experts appearing only occasionally, while a small subset dominates historical records (Hamidi Rad, Fani, Bagheri, Kargar, Srivastava, and Szlichta, 2023; Boratto, Fenu, Marras, and Medda, 2023; Boratto, Faralli, Marras, and Stilo, 2022). This imbalance causes overfitting and entrenched popularity bias. Second, the space of skills is combinatorially large and features non-linear dependencies that are not captured well by linear factorization or simple similarity measures. Third, effective teams depend on latent properties such as complementarity and compatibility, which are rarely annotated and difficult to infer using conventional methods.

Prior research has pursued multiple directions in response to these challenges. Graph-based approaches search subgraphs covering skill requirements (Lappas, Liu, and Terzi, 2009; Kargar and An, 2011), offering structural expressiveness but facing scalability and generalization limits. Matrix factorization methods decompose expert–skill co-occurrence data into latent components (Koren, 2009; Wu, Jiang, Li, Li, and Liu, 2017), capturing global associations yet restricted by linear assumptions. Probabilistic models incorporate uncertainty in team composition (Rad, Seyedsalehi, Kargar, Zihayat, and Bagheri, 2022; Hamidi Rad et al., 2023), but rely on rigid priors that constrain flexibility. Neural architectures embed experts and skills in shared vector spaces (Dara, Rad, Zarrinkalam, and Bagheri, 2025), showing promise in representational capacity but tending to overfit to high-degree experts, thereby amplifying popularity bias. Importantly, these families of models are not generativeXu, Zhao, Yu, Zhang, Guo, and Yao (2024); Chen, Feng, Chen, Wu, Sun, Li, Gao, Zhang, and Xue (2025). They rank existing candidates rather than synthesizing novel and task-specific expert teams. More recent developments in temporal collaboration networks (Fani, Barzegar, Dashti, and Saeedi, 2024), curriculum learning (Barzegar, Kurepa, and Fani, 2025) and sequence-to-sequence modeling (Thang, Hosseini, and Fani, 2025) attempt to address entrenched issues of bias and interdependency, but they remain bound to historical collaboration patterns.

A common limitation across these approaches is that they implicitly treat team formation as a deterministic ranking or selection task over existing experts. Such formulations struggle to represent uncertainty, diversity, and unseen skill combinations inherent in sparse collaboration data. In this work, we instead adopt a *generative perspective* and model team formation as learning a skill-conditioned distribution over expert teams, enabling the synthesis of diverse and contextually valid teams rather than ranking fixed candidates. The methodological choices underlying our framework are not empirically motivated design heuristics, but principled responses to the structure of the team formation problem. Specifically, the generative formulation follows from the need to model uncertainty and diversity under sparsity; conditional inpainting reflects the asymmetry between known skills and unknown team composition; diffusion modeling enables robust estimation of complex conditional distributions; and embedding-space generation supports generalization beyond observed expert configurations.

At a high level, the central contribution of this work is conceptual rather than architectural where we argue that expert team formation is fundamentally a problem of conditional generation under uncertainty, not candidate ranking. The modeling choices that follow, including diffusion-based generation, embedding-space representations, and conditional inpainting, are introduced to operationalize this central idea rather than as independent contributions.

Adopting a conditional generative perspective naturally leads to diffusion probabilistic models, which are explicitly designed to learn complex, high-dimensional conditional distributions and support structured generation via inpainting. We argue that diffusion probabilistic models provide such a framework. We defer the technical details of the diffusion and inpainting formulation to Section 3. Adapting diffusion models to expert team formation allows us to formulate the task as skill-conditioned sampling, enabling the generation of diverse and unbiased teams guided directly by task requirements.

1.1. Objectives

From the perspective of conditional generative modeling, an effective team formation method must satisfy two key design requirements.

1. *Mitigating sparsity in team formation data.* Sparse and long-tailed distributions create high-dimensional but under-sampled feature spaces, which hinder traditional models' ability to learn robust mappings between skills and experts. Existing compression-based strategies risk discarding fine-grained associations essential for contextually valid team formation. We address this through dense representation learning embedded in diffusion models, which preserve semantic richness while enabling effective learning under sparsity. This requirement

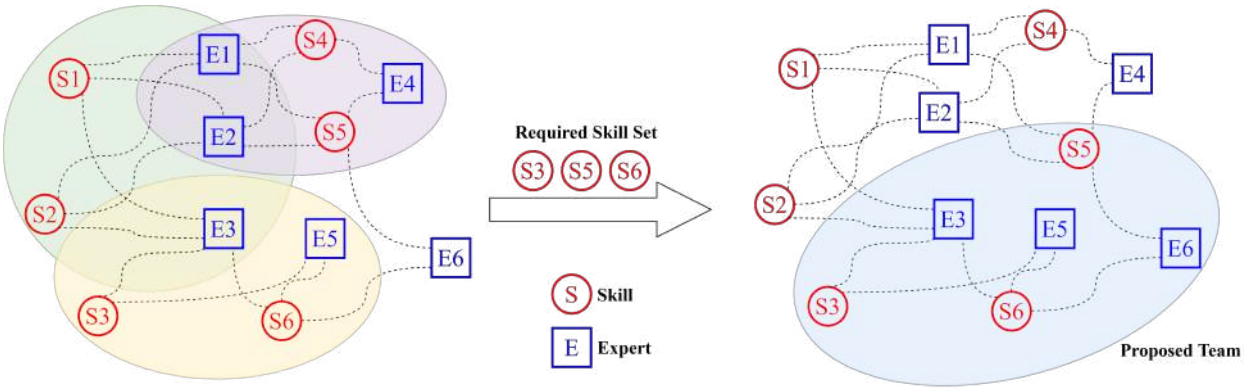


Figure 1: A slice of a collaboration network. The expert and skill nodes are shown along with papers they have collaborated on in the past. Given a set of required skills, the Team Formation task is proposing a group of authors that collectively cover the skills.

motivates operating in a continuous embedding space and adopting generative models that remain expressive under sparsity.

2. *Avoiding popularity bias in expert selection.* Existing methods disproportionately recommend frequently observed experts, perpetuating systemic bias. This undermines diversity, reduces fairness, and prevents the discovery of equally qualified but less visible experts. By framing team formation as a skill-conditioned inpainting problem, we anchor expert prediction directly on skill requirements, counteracting bias introduced by historical collaboration frequency. This motivates conditioning generation directly on skill requirements rather than relying on historical frequency or popularity.

To address these objectives, we frame team formation within a generative modeling paradigm that integrates semantic representation learning with diffusion processes. Embedding-based representations provide a structured space in which sparse and uneven co-occurrence data can be expressed in a form that captures semantic similarity and relational structure. Within this space, diffusion modeling offers a principled way to learn complex distributions by gradually refining noisy inputs into coherent outputs.

We propose to treat team formation as a conditional inpainting problem, where the known skill requirements act as constraints that guide the generative process. This formulation ensures that sampling is explicitly conditioned on task-specific needs rather than historical frequency, enabling more diverse and contextually valid expert teams. The theoretical strength of diffusion lies in its ability to balance fidelity to observed structures with exploration of the broader distribution, allowing it to overcome sparsity, mitigate popularity bias, and generalize to unseen skill combinations.

1.2. Contributions

We make a primary methodological contribution, supported by several secondary design and empirical contributions that operationalize and validate the core idea.

- **Core Contribution.** We formulate expert team formation as a *skill-conditioned generative modeling problem* and introduce a diffusion-based conditional inpainting framework that directly generates expert teams under data sparsity and long-tailed expert participation. This reframing shifts team formation from ranking-based selection to probabilistic synthesis of plausible team configurations.

- **Additional Contributions.**

  – An embedding-based representation of skills and experts that enables diffusion modeling in a continuous space while preserving semantic relationships.

- A conditional inpainting strategy that exploits the asymmetry between known task requirements and unknown team composition.
- A nearest-neighbor discretization mechanism that maps generated expert embeddings back to interpretable expert identities.
- A comprehensive experimental evaluation demonstrating improvements in effectiveness, fairness, and robustness across heterogeneous datasets.

2. Related Work

The task of team formation has been explored through multiple lines of research, most notably `graph-based` models, `collaborative filtering`, `neural architectures`, and `seq2seq` models. Each of these families emphasizes a different aspect of the problem, with graph-based methods focusing on structural relationships, collaborative filtering on co-occurrence patterns, neural architectures on latent representation learning, and seq2seq models on the sequential dependencies among expert selections.

`Graph-based approaches.` Among the earliest attempts to formalize the task of team formation are graph-based methods, which represent experts as nodes and their past collaborations as edges in a collaboration graph (Lappas et al., 2009). The objective of these methods is to extract connected subgraphs that jointly cover all required skills while minimizing communication or structural cost. Early contributions such as Lappas et al. (2009); Sozio and Gionis (2010) formalized the problem as variants of the *Steiner Tree*, relying on heuristics such as minimum spanning trees and shortest diameters. Later, Kargar, An, and Zihayat (2012) proposed minimizing the sum of pairwise distances, while Zihayat, An, Golab, Kargar, and Szlichta (2017) incorporated both edge weights (capturing past collaboration quality) and node weights (reflecting expert strength). More recent extensions have leveraged link prediction to form teams with no prior collaborations but strong potential for success (Keane, Ghaffar, and Malone, 2019). Despite their structural richness, these approaches face scalability challenges. Because they rely heavily on shortest-path computations over collaboration networks, their efficiency deteriorates when applied to large-scale graphs containing thousands of experts, diverse skill sets, and extensive collaboration histories (Kargar, Golab, Srivastava, Szlichta, and Zihayat, 2022). Even with indexing methods such as 2-Hop Cover Labeling (Akiba, Iwata, and Yoshida, 2013), the computational complexity remains prohibitive in practice.

`Collaborative filtering approaches.` To alleviate the scalability bottlenecks of graph-based methods, researchers have framed team formation as a recommendation problem. Collaborative filtering models avoid explicit graph traversal by inferring expert candidates from past co-occurrence patterns. For example, the Recurrent Recommender Network (RRN) (Wu et al., 2017) models sequences of collaborations using LSTMs, while the Group Expert Recommendation Framework (GERF) (Du, Meng, Zhang, and Lv, 2020) employs a Bayesian ranking model. These approaches are efficient and do not require the entire collaboration graph; however, they often struggle with infrequent or unseen skill combinations, leading to recommendations biased toward historically frequent experts.

`Neural approaches.` The shortcomings of collaborative filtering in handling sparsity and novelty motivated the adoption of neural models, which learn latent representations of experts and skills to capture hidden associations (Rad, Bagheri, Kargar, Srivastava, and Szlichta, 2021; Rad et al., 2022). Early work such as Sapienza, Goyal, and Ferrara (2019) employed autoencoders, however, these methods showed to overfit in sparse settings. For this purpose, variational neural models by Rad et al. (Hamidi Rad et al., 2023) introduced Bayesian uncertainty to improve robustness and avoid overfitting. In another direction, Sapienza et al. (2019) reconstructed collaboration graphs to generate expert embeddings. While these models showed promising performance improvements, they have been associated with bias towards more frequently seen experts in the network and potentially suffer from favoritism. More recently, Dashti et al. (Dashti, Saxena, Patel, and Fani, 2022) and Dara et al. (Dara et al., 2025) have emphasized generating diverse team compositions rather than ranking experts individually. Despite these efforts, these models still treat expert selection as independent predictions from a multi-hot vector, which reduces diversity and can reinforce popularity bias (Hemmatizadeh, Wong, Yu, and Fani, 2024). In addition to architectural and representational advances, there has also been work on training strategies that aim to mitigate popularity bias. For example, Barzegar et al. (Barzegar et al., 2025) introduced curriculum-based methods that guide neural models to learn progressively from easier cases involving popular experts to harder cases involving less visible experts.

Recent work has also explored learning under long-tailed distributions and generalization beyond observed data. In particular, the Graph Convolutional Mixture-of-Experts Learner Network (GCML) (Wang, Su, Wang, Wang, Yin, Shen, Lan, Yang, and Cao, 2025) addresses long-tailed domain generalization in classification settings by combining

multiple specialized experts through a graph convolutional architecture. GCML explicitly models different class-distribution regimes using a mixture-of-experts framework and leverages graph-based feature propagation to improve robustness under domain shift. While both GCML and our approach address data imbalance, they differ fundamentally in problem setting and methodology. GCML focuses on long-tailed classification and domain-invariant representation learning, whereas our work formulates team formation as a conditional generative problem and addresses long-tailed distributions in terms of expert participation and exposure through diffusion-based generation.

`Sequence-to-sequence models.` Beyond popularity bias, another limitation of existing neural approaches is their reliance on high-dimensional multi-label classification, which suffers from output sparsity and ignores sequential dependencies among expert selections. Thang et al. (Thang et al., 2025) reformulated the task as a sequence prediction problem, mapping a required skill sequence to an expert sequence using recurrent and transformer-based architectures. Their experiments across diverse domains (e.g., DBLP, USPTO, IMDb, GitHub) demonstrated that seq-to-seq models consistently outperform feedforward classifiers, especially when expert distributions are long-tailed. By conditioning each expert recommendation on previously selected experts, seq-to-seq models better capture the interdependencies inherent in real-world team formation.

While diffusion models have recently been explored for recommender systems the application domains and technical objectives differ fundamentally from expert team formation. DiffRec Wang, Xu, Feng, Lin, He, and Chua (2023) and related approaches Lin, Cao, Yu, and Zhang (2025) address the user-item preference prediction problem, where the goal is to estimate individual user preferences and generate ranked lists of items for personalized consumption. These methods operate within a user-item interaction matrix, learning to denoise corrupted interaction histories to predict future individual preferences. In contrast, expert team formation is a combinatorial set-to-set mapping problem that must satisfy structural constraints including complete skill coverage, team complementarity, and fairness under long-tailed expert distributions. Our task requires generating coherent teams of experts conditioned on required skills, where the output must be feasible as a functioning unit rather than a personalized ranking. Methodologically, while recommender diffusion models denoise user-item signals to yield item rankings for individual users, our approach performs skill-conditioned inpainting in a joint expert-skill embedding space to synthesize teams that collectively satisfy combinatorial task requirements. These differences distinguish our work from the recommender systems literature, even when both employ diffusion-based generative modeling as the underlying technical framework.

In summary, prior research has progressed along three main directions. Graph-based models emphasize structural expressiveness but suffer from prohibitive computational costs. Collaborative filtering approaches improve scalability yet remain vulnerable to sparsity. Neural architectures offer stronger generalization but tend to reinforce popularity bias and exhibit limitations in handling output sparsity. More recent efforts on curriculum-based learning (Barzegar et al., 2025) and sequence-to-sequence modeling (Thang et al., 2025) demonstrate a move toward revisiting training dynamics and output representations as a way of alleviating these persistent issues. Our work builds on this trajectory by advancing a generative framework based on diffusion models, where team formation is modeled as conditional sampling. By synthesizing expert vectors from noise under skill conditioning, the proposed approach promotes diversity, reduces dependence on frequent experts, and maintains robustness in sparse and imbalanced settings.

3. Formulation and Rationale

3.1. Problem Statement

Let $\mathcal{S} = \{s_1, s_2, ..., s_i\}$ denote a set of skills in a given domain and $\mathcal{E} = \{e_1, e_2, ..., e_j\}$ be a set of available experts. The skills required for a specific task can be expressed as a subset of the total possible skills $\mathbf{s} \subset \mathcal{S}; \mathbf{s} \neq \emptyset$, where the optimal team of experts for the task is $\mathbf{e} \subset \mathcal{E}; \mathbf{e} \neq \emptyset$. We formally define *the task of team formation* as finding a mapping of the Skill powerset to the Expert powerset through a function of parameters $f(\theta)$ such that $f(\mathbf{s}; \theta) \rightarrow \mathbf{e}$. Throughout this paper, we intend to learn an effective mapping function f based on historical teams that include skill-expert sets. We propose to learn f through a skill-conditioned Diffusion model (Sohl-Dickstein, Weiss, Maheswaranathan, and Ganguli, 2015; Ho, Jain, and Abbeel, 2020) performing an inpainting task (Lugmayr, Danelljan, Romero, Yu, Timofte, and Van Gool, 2022). Importantly, the supervision signal for learning f is derived from historically formed teams, considering that past collaborations provide an organic signal of effective team composition (Hamidi Rad et al., 2023; Kargar and An, 2011). The model is therefore trained to capture the *skill-conditioned* distribution of manually assembled team configurations observed in historical data. Consequently, the generated teams are intended to reflect historically validated formation patterns, rather than to optimize an external performance objective or to synthesize entirely unconstrained team structures. The model will be trained on $\mathcal{T} = (\mathbf{s}, \mathbf{e})$ pairs, and upon inference, only $\mathbf{s}$ will

be provided with random noise masking the corresponding **e**, our model will then *'inpaint'* the predicted experts vector that best matches the queried skills.

3.2. Rationale for Our Approach

The methodological design of our framework follows directly from the structural properties of the team formation problem. First, team formation datasets are characteristically sparse where most experts appear in only a handful of team configurations, and skill requirements vary widely across tasks. This results in *long-tailed distributions* over both the skill and expert spaces, making it difficult to learn robust representations of expert suitability or team dynamics. Second, this sparsity leads to over-representation of a small subset of frequently occurring experts, introducing strong *inductive biases* into learned models and increasing the risk of overfitting (Camuto, Willetts, Roberts, Holmes, and Rainforth, 2021; Jospin, Laga, Boussaid, Buntine, and Bennamoun, 2022). Third, existing graph-based or probabilistic models often rely on manually engineered structural assumptions such as fixed latent priors or rigid similarity metrics that fail to capture the high-dimensional, nonlinear relationships between required skills and expert capabilities, especially in low-observation settings.

To address these challenges, we propose a skill-conditioned Diffusion-based generative model for team formation. Our proposed model offers several technical advantages. Unlike variational autoencoders (VAEs) (Kingma and Welling, 2013) that impose a parametric form on the latent space (typically multivariate Gaussian) and attempt to optimize the evidence lower bound (ELBO), our diffusion-based model operates directly in the data space and models the score function via denoising. This allows it to capture complex, multi-modal distributions without imposing restrictive priors, making them more suitable for representing structurally diverse and rare team configurations while avoiding systematic bias towards popular experts. Moreover, diffusion models are known for their strong mode coverage and stability under limited supervision, allowing them to generalize more effectively in sparse domains such as expert-team formation (Song, Sohl-Dickstein, Kingma, Kumar, Ermon, and Poole, 2021; Dhariwal and Nichol, 2021). From a modeling perspective, this makes diffusion particularly well suited to team formation, where the conditional distribution over expert configurations is inherently multi-modal and cannot be adequately captured by single-latent or unimodal generative assumptions.

From a broader representation learning perspective, our framework can be viewed as learning a joint embedding space that integrates heterogeneous information sources—skills, experts, and historical interactions—into a unified continuous representation. The diffusion process then operates as a conditional generative mechanism over this space, enabling robust inference when portions of the input are missing or unreliable. This formulation aligns with a growing body of work in information processing that emphasizes learning expressive representations capable of supporting inference, generation, and reasoning across incomplete or weakly supervised data sources (Zhou, Wang, Ye, Guan, and Yu, 2025; Huang, Han, Xu, and Gan, 2022; Ouyang, Xie, Li, and Cheng, 2023).

Our approach further leverages the concept of *conditional sampling* within the diffusion framework. Rather than generating an entire team from scratch, we treat team formation as a masked generative task, where the skills **s** associated with a task are known and the corresponding experts **e** are masked. During training, the model learns to reconstruct the skill-expert pair vector from noisy perturbations. This strategy is analogous to image or text inpainting in other generative domains and provides a powerful inductive bias (Choi, Kim, Jeong, Gwon, and Yoon, 2021; Meng, He, Song, Song, Wu, Zhu, and Ermon, 2022). It anchors the generation process on the task requirements while learning the complex skill-to-expert mapping via iterative denoising. By doing so, the model avoids the collapse into popular expert selections and maintains fidelity to the task-specific composition of historically effective teams. Additionally, working in the data space rather than a latent space ensures that expert interdependencies such as complementary or synergistic skill overlaps are preserved during generation.

4. Proposed Approach

To address the challenges of sparsity, popularity bias, and non-linear dependencies identified in the team formation problem, we propose a diffusion-based generative framework. Our model operates directly in the space of skill–expert configurations and formulates team formation as a conditional sampling task. By iteratively denoising noisy representations into coherent expert vectors anchored on task-specific skills, the framework captures complex and multi-modal dependencies without imposing restrictive priors on latent structures. The conditional inpainting formulation ensures that generation is explicitly guided by the required skills, thereby reducing reliance on historical collaboration frequency and supporting the synthesis of diverse and contextually valid expert teams. While this design

introduces additional structural components compared to ranking-based or factorization methods, this complexity is a direct consequence of modeling team formation as conditional generation rather than candidate scoring. Each component serves a distinct role in addressing sparsity, uncertainty, and long-tailed expert participation, rather than incrementally increasing expressive power without purpose.

4.1. Skill-Conditioned Diffusion

At the core of our formulation is a denoising process that gradually learns to map from isotropic Gaussian noise back to coherent team configurations (Ho et al., 2020; Nichol and Dhariwal, 2021). To achieve this, we define a forward diffusion process that incrementally perturbs training samples, each consisting of a concatenated skill-expert pair $\mathcal{T} = (\mathbf{s}, \mathbf{e})$, with Gaussian noise over a fixed-length Markov chain of $\mathbf{T}$ steps. At each step t, noise is injected via a predefined variance schedule $\beta_1, \beta_2, ..., \beta_{\mathbf{T}}$, such that the sample is progressively corrupted into a high-entropy distribution. Formally, the forward process is defined as:

$$q\left(\mathcal{T}_{1:\mathbf{T}}|\mathcal{T}_0\right) = \prod_{t=1}^{\mathbf{T}} q\left(\mathcal{T}_t|\mathcal{T}_{t-1}\right), \qquad q\left(\mathcal{T}_t|\mathcal{T}_{t-1}\right) = \mathcal{N}\left(\mathcal{T}_t; \sqrt{1-\beta_t}\mathcal{T}_{t-1}, \beta_t\mathbf{I}\right) \tag{1}$$

Rather than computing this full sequence recursively, which would be computationally prohibitive, we leverage the closed-form sampling of Gaussian transitions and apply the reparameterization trick to directly express the noisy version of $\mathcal{T}_t$ at any step t as:

$$q\left(\mathcal{T}_t|\mathcal{T}_0\right) = \mathcal{N}\left(\mathcal{T}_t; \sqrt{\bar{\alpha}_t}\mathcal{T}_0, \left(1-\bar{\alpha}_t\right)\mathbf{I}\right) \tag{2}$$

$$\mathcal{T}_t = \sqrt{\bar{\alpha}_t}\mathcal{T}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \tag{3}$$

where $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$, $\alpha_t = 1 - \beta_t$, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. Since $\beta_t \in (0, 1)$ is a hyperparameter, this allows us to calculate all values of α_t in advance, thus enabling the noise sampling and subsequent calculation of $\mathcal{T}_t$ for any timestep t directly from the clean data $\mathcal{T}_0$.

As $t \to \mathbf{T}$, the signal is fully corrupted and $\mathcal{T}_{\mathbf{T}}$ approaches pure Gaussian noise. In the reverse process, we can see that $\bar{\alpha}_t \to 0$ as $t \to \infty$, thus allowing us to approximate $\mathcal{T}_t$ as an isotropic Gaussian distribution i.e. $\mathcal{T}_t \approx \epsilon$. We thus postulate that if we know the reverse distribution $q\left(\mathcal{T}_{t-1}|\mathcal{T}_t\right)$, we can sample $\mathcal{T}_T \sim \mathcal{N}\left(0, \mathbf{I}\right)$, run the reverse process, and acquire a new sample $q\left(\mathcal{T}_0\right)$, generating a novel team from the original team distribution. Since the true reverse transitions $q(\mathcal{T}_{t-1}|\mathcal{T}_t)$ are intractable due to their dependence on the full data distribution, we approximate them using a neural network parameterized by θ:

$$p_\theta\left(\mathcal{T}_{t-1}|\mathcal{T}_t\right) = \mathcal{N}\left(\mathcal{T}_{t-1}; \mu_\theta\left(\mathcal{T}_t, t\right), \Sigma_\theta\left(\mathcal{T}_t, t\right)\right) \tag{4}$$

We instantiate p_θ using a U-Net-style architecture (Ronneberger, Fischer, and Brox, 2015) that takes as input the current noisy sample $\mathcal{T}_t$ and timestep t, and outputs the predicted mean and variance of the reverse step distribution. This network is trained to iteratively denoise the expert vector conditioned on the known skill vector, learning to reconstruct plausible expert teams consistent with the task requirements. The model is trained by optimizing the variational lower bound on the data log-likelihood, resulting in the following evidence lower bound:

$$\begin{aligned}\log p_\theta\left(\mathcal{T}_0\right) \geq \mathbb{E}_{q(\mathcal{T}_0)}\Big[& \underbrace{\log p_\theta\left(\mathcal{T}_0|\mathcal{T}_1\right)}_{L_0} - \underbrace{KL\left(q\left(\mathcal{T}_{\mathbf{T}}|\mathcal{T}_0\right) \parallel p\left(\mathcal{T}_{\mathbf{T}}\right)\right)}_{L_{\mathbf{T}}} \\ & - \sum_{t=2}^{\mathbf{T}} \underbrace{KL\left(q\left(\mathcal{T}_{t-1}|\mathcal{T}_t, \mathcal{T}_0\right) \parallel p_\theta\left(\mathcal{T}_{t-1}|\mathcal{T}_t\right)\right)}_{L_t} \Big]\end{aligned} \tag{5}$$

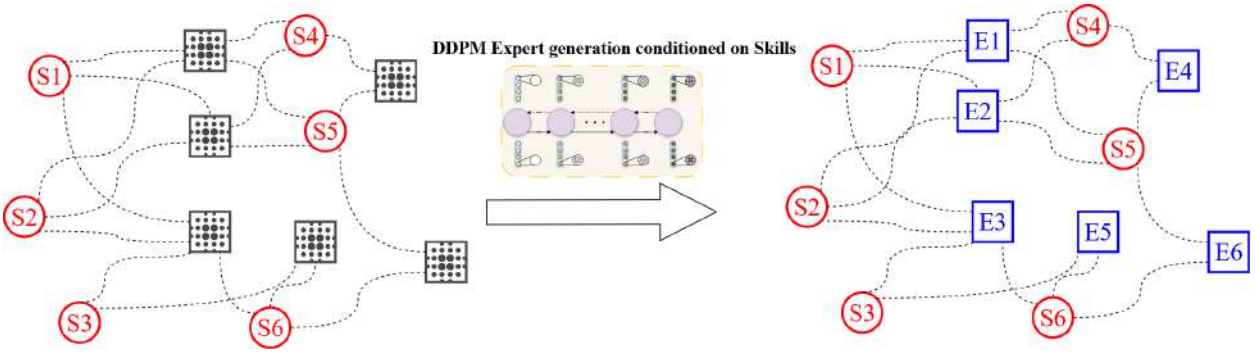


Figure 2: Visualization of the inpainting task. Experts (blue) are masked from the model; However, the Skills (red) remain known to our diffusion model during inference, at which it performs conditional sampling based on the known skills to generate the appropriate experts given those skills.

Here, L_0 quantifies the reconstruction error between the original data $\mathcal{T}_0$ and its immediate denoised form from $\mathcal{T}_1$. The term $L_{\mathbf{T}}$ ensures that the final state of the forward process aligns with the target noise distribution $\mathcal{N}(0, \mathbf{I})$, encouraging high entropy and sample diversity. The intermediate L_t terms constrain the learned reverse process to approximate the true (but unknown) posteriors at each denoising step. Finally, KL denotes the Kullback–Leibler divergence. To model conditional sampling, we adopt a skill-inpainting strategy: during inference, the skill vector $\mathbf{s}$ has a known noise profile, while the expert vector $\mathbf{e}$ is masked from the model and initialized with unknown noise that the model learns to recover.

4.2. Team Inpainting

Team formation is inherently asymmetric where task requirements (skills) are fully specified at query time, while the expert configuration is unknown and must be inferred. Conditional inpainting directly operationalizes this asymmetry by treating skills as observed context and experts as missing variables to be generated. To implement conditional sampling over expert teams given fixed skill requirements, we adopt an inpainting formulation within the diffusion framework (Lugmayr et al., 2022). A qualitative visualization of how the inpainting task would look in practice is depicted in Figure 2. The key idea is to know the noisy portion of the input, i.e., the skill vector $\mathbf{s}$, throughout the reverse process and let the model only denoise the unknown component, i.e., the expert vector $\mathbf{e}$. This formulation reflects the real-world asymmetry of the task: the skills $\mathbf{s}$ are provided as input, and the system must generate a team of experts that satisfies those constraints. During training, the model is exposed to full pairs $\mathcal{T}_0 = [\mathbf{s}; \mathbf{e}]$ and learns to denoise both jointly. During inference, however, only the expert portion is initialized with Gaussian noise and updated over time, while the skill vector's noise level remains known across all timesteps. This approach anchors generation on the task specification and enables the model to construct expert teams that are both contextually appropriate and structurally coherent.

At inference time, we perform conditional sampling by manually adding a portion of the initially sampled noise, according to the current timestep, to $\mathbf{s}$ at every timestep while progressively denoising the expert component. Specifically, at each timestep t, the state of the diffusion process is represented as:

$$\mathcal{T}_t = [\mathbf{s}_t; \mathbf{e}_t]$$

where $\mathbf{s}_t$ is noised according to the known noise scheduler and $\mathbf{e}_t$ is updated via the learned denoising model. The network predicts the noise component $\hat{\epsilon}_\theta(\mathcal{T}_t, t)$ and computes a denoised estimate for the expert vector $\hat{\mathbf{e}}_{t-1}$. This yields an estimate for the full input at the previous timestep, preserving the conditioning variable:

$$\mathcal{T}^*_{t-1} = [\mathbf{s}_{t-1}; \hat{\mathbf{e}}_{t-1}] \tag{6}$$

This procedure is repeated until $t = 0$, producing the final output:

$$\mathcal{T}_0 = [\mathbf{s}; \hat{\mathbf{e}}_0]$$

The result is a plausible expert vector $\hat{\mathbf{e}}_0$ conditioned on the known input $\mathbf{s}$, consistent with historical team patterns and reflective of skill-expert compatibility learned during training. This inpainting formulation allows the model

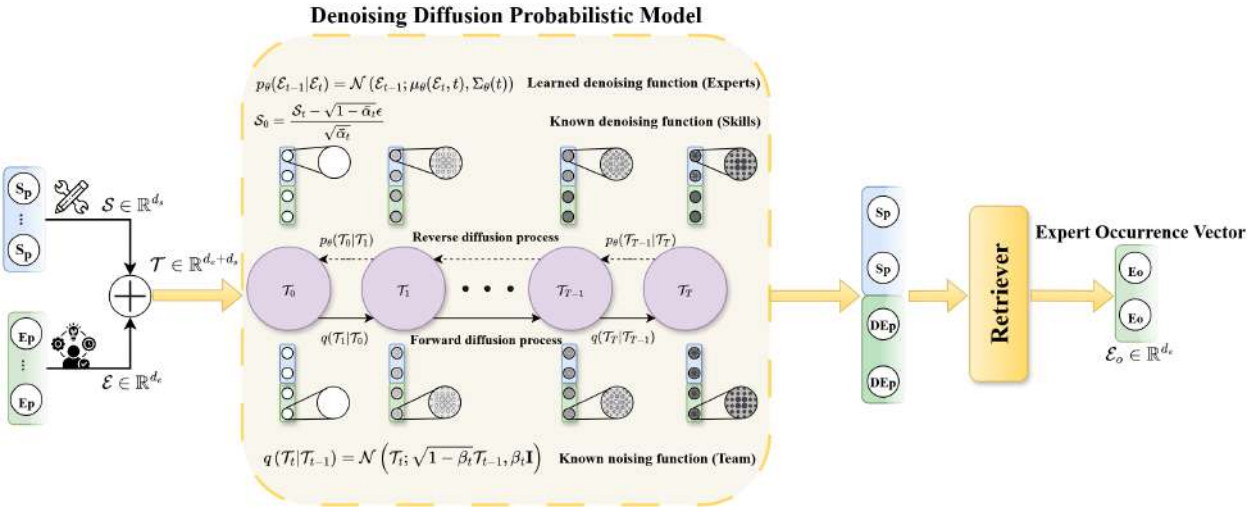


Figure 3: Overview of the architecture of our proposed model.

to faithfully preserve task specifications while generating expert teams that generalize beyond seen combinations, enabling high-quality team formation even in sparse and imbalanced settings.

4.3. Model Architecture

As described in the preceding sections, our model addresses the team formation task by learning a mapping $f(\mathbf{s};\theta) \rightarrow \mathbf{e}$, instantiated through a denoising diffusion process and operationalized via a conditional inpainting mechanism. Figure 3 provides an overview of this design. In essence, our core architecture is a denoising diffusion probabilistic model parameterized by a U-Net (Ronneberger et al., 2015) denoising function $\hat{\epsilon}_\theta$, which predicts the added noise at each timestep during the reverse diffusion process. The U-Net follows a symmetric encoder-decoder design with three downsampling layers, a bottleneck layer, and three corresponding upsampling layers. Each block consists of two ResNet submodules (He, Zhang, Ren, and Sun, 2016), a self-attention mechanism (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017), and a convolutional layer to enable effective feature aggregation across different resolutions. This design ensures that both local interactions and global patterns are captured during denoising, which is essential given the structural complexity of skill-expert relationships. This architectural complexity enables the model to capture both local skill–expert interactions and global team-level dependencies, which simpler architectures struggle to represent under sparse and combinatorial settings.

To efficiently represent high-dimensional and sparse co-occurrence vectors typically found in team formation datasets, we perform a preprocessing step that converts both skill and expert indicators into dense vector embeddings. Specifically, we train a Word2Vec model under a skip-gram setup to obtain continuous vector representations $\mathcal{V}_s$ and $\mathcal{V}_e$ for each skill and expert, respectively. These embeddings serve two critical purposes: (i) they reduce the input dimensionality, making the model more computationally efficient, and (ii) they capture semantic similarity and relational structure among skills and experts, which is essential for learning generalizable patterns.

The input to the diffusion model at each training step is the concatenated vector $\mathcal{T}_0 = [\mathcal{V}_s; \mathcal{V}_e] \in \mathbb{R}^{d_e+d_s}$, treated as a 1D flattened signal. As described in the Skill-Conditioned Diffusion subsection, the forward process gradually corrupts this input with Gaussian noise according to a precomputed schedule β_t. The reverse process, implemented by the U-Net, iteratively denoises the expert portion $\mathcal{V}_e$ while conditioning on the known skill embedding $\mathcal{V}_s$. This setup allows the model to treat team formation as an inpainting task, using known skills as context to reconstruct plausible expert teams.

During training, backpropagation is applied using the DDPM objective, which minimizes the expected squared error between the predicted and actual noise. At inference, the final denoised embedding $\hat{\mathcal{V}}_e$ is mapped back to the original discrete expert co-occurrence space using a nearest-neighbor search over the expert embedding dictionary. This post-processing step ensures that the output of the model is interpretable and compatible with downstream applications that require discrete expert identifiers.

Algorithm 1 Training Procedure

Require: Dataset $\mathcal{T} = \{(\mathbf{s}_i, \mathbf{e}_i)\}_{i=1}^{N}$, with $\mathbf{s}_i \subseteq \mathcal{S}$, $\mathbf{e}_i \subseteq \mathcal{E}$
Require: Embedding functions $\phi_s : \mathcal{S} \to \mathbb{R}^{d_s}$ and $\phi_e : \mathcal{E} \to \mathbb{R}^{d_e}$
Require: Diffusion steps $\mathbf{T}$, noise schedule $\{\beta_t\}_{t=1}^{\mathbf{T}}$, U-Net model $\hat{\boldsymbol{\epsilon}}_\theta$
1: **for** $i = 1$ to N **do** ▷ Iterate over each sample $(\mathbf{s}_i, \mathbf{e}_i) \in \mathcal{T}$
2: **Step 1: Embed skill and expert sets**
3: $\mathcal{V}_{si} \leftarrow \phi_s(\mathbf{s}_i) \in \mathbb{R}^{d_s}$
4: $\mathcal{V}_{ei} \leftarrow \phi_e(\mathbf{e}_i) \in \mathbb{R}^{d_e}$
5: $\mathcal{T}_0^{(i)} \leftarrow [\mathcal{V}_{si}; \mathcal{V}_{ei}] \in \mathbb{R}^{d_s+d_e}$
6: **Step 2: Sample timestep and noise**
7: $t \sim \text{Uniform}(1, \mathbf{T})$
8: $\boldsymbol{\epsilon}^{(i)} \sim \mathcal{N}(0, \mathbf{I})$
9: Compute: $\bar{\alpha}_t = \prod_{j=1}^{t}(1 - \beta_j)$
10: **Step 3: Forward process**
11: $\mathcal{T}_t^{(i)} \leftarrow \sqrt{\bar{\alpha}_t} \cdot \mathcal{T}_0^{(i)} + \sqrt{1 - \bar{\alpha}_t} \cdot \boldsymbol{\epsilon}^{(i)}$
12: **Step 4: Reverse process**
13: $\hat{\boldsymbol{\epsilon}_\theta}^{(i)} \leftarrow \hat{\boldsymbol{\epsilon}}_\theta(\mathcal{T}_t^{(i)}, t)$
14: $\hat{\mathcal{T}}_{t-1}^{(i)} \leftarrow \frac{1}{\sqrt{\alpha_t}} \left(\mathcal{T}_t^{(i)} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \hat{\boldsymbol{\epsilon}}_\theta^{(i)} \right)$
15: **Step 5: Compute loss**
16: $\mathcal{L}^{(i)} \leftarrow \|\hat{\boldsymbol{\epsilon}_\theta}^{(i)} - \boldsymbol{\epsilon}^{(i)}\|^2$
17: **end for**
18: **return** Trained model $\hat{\boldsymbol{\epsilon}}_\theta$

4.4. Training Procedure

The training process of our model is outlined in Algorithm 1. Given the task of learning the mapping function $f(\mathbf{s}; \theta) \to \mathbf{e}$, we structure the training as a noise prediction problem, in which the model learns to reverse a gradual corruption process applied to expert vectors while conditioning on a known set of skills. Each training sample $(\mathbf{s}_i, \mathbf{e}_i) \in \mathcal{T}$ consists of a subset of required skills and a corresponding expert team. In Lines 2–4, we embed these subsets to generate dense representations $\mathcal{V}_{si}$ and $\mathcal{V}_{ei}$ respectively. These vectors are then concatenated into a joint input $\mathcal{T}_0^{(i)}$. From Line 5 onward, we show the forward diffusion process by adding Gaussian noise to the expert portion of the input at a randomly selected timestep $t \in [1, \mathbf{T}]$. The noise level is determined using a pre-defined variance schedule β_t, and the corrupted sample $\mathcal{T}_t^{(i)}$ is computed using the closed-form DDPM approach (Line 11). The model then performs the reverse diffusion step by predicting the noise $\hat{\boldsymbol{\epsilon}}_\theta$ at that timestep (Line 12), and reconstructing the original latent using the predicted value in Line 13. The denoised vector $\hat{\mathcal{T}}_{t-1}^{(i)}$ is not directly used for generation during training. Instead, the model is optimized using a mean squared error loss between the predicted and true noise vectors, as described in Line 15. Unlike variational approaches that explicitly model a latent distribution and minimize KL divergence, our formulation does not treat the model parameters as distributions themselves. Instead, the U-Net learns to predict the noise added at each diffusion step, which implicitly defines the mean of the reverse denoising distribution. The model parameters θ therefore control the behavior of the reverse process through this learned noise estimator $\hat{\boldsymbol{\epsilon}}_\theta$.

4.5. Model Prediction

The inference process of our model is described in Algorithm 2. Given a new skill set $\mathbf{s} \subseteq \mathcal{S}$ at test time, our goal is to predict a plausible expert subset $\hat{\mathbf{e}}_q \subseteq \mathcal{E}$ that aligns with the required skills. To do so, we first generate a dense vector representation $\mathcal{V}_{sq} = \phi_s(\mathbf{s})$ using the same embedding function as used during training. Since the corresponding expert vector is unknown at inference time, we let $\mathcal{V}_e^{(\mathbf{T})}$ be a random sample from a standard Gaussian distribution (Line 2). From Lines 3–12, we show the reverse diffusion process where we concatenate the noisy skill and expert embeddings to form the input $\mathcal{T}_\mathbf{T} \in \mathbb{R}^{d_e+d_s}$ and progressively denoise the expert portion of the input while manually noising the skill portion (Lines 6-7). At each step t, the model predicts the noise to be removed from the expert vector using the trained denoising function $\hat{\boldsymbol{\epsilon}}_\theta$ (Line 8). This prediction is used to compute the denoised estimate $\mathcal{V}_e^{(t-1)}$ (Line 10). The updated

Algorithm 2 Inference Procedure

Require: Trained model $\hat{\epsilon}_\theta$
Require: Skill query $\mathbf{s} \subseteq \mathcal{S}$ with embedding $\mathcal{V}_s \leftarrow \phi_s(\mathbf{s}) \in \mathbb{R}^{d_s}$
Require: Diffusion steps $\mathbf{T}$, noise schedule $\{\beta_t\}_{t=1}^{\mathbf{T}}$
1: **Step 1: Initialize random expert vector**
2: $\mathcal{V}_e^{(\mathbf{T})} \sim \mathcal{N}(0, \mathbf{I}) \in \mathbb{R}^{d_e}$
3: **Step 2: Denoising process**
4: **for** $t = \mathbf{T}, \mathbf{T}-1, ..., 1$ **do**
5: Compute: $\bar{\alpha}_t = \prod_{j=1}^{t}(1-\beta_j)$
6: Manually noise skill vector $\mathcal{V}_s^{(t)} \leftarrow \sqrt{\bar{\alpha}_t} \cdot \mathcal{V}_s^{(0)} + \sqrt{1-\bar{\alpha}_t} \cdot \boldsymbol{\epsilon}^{(t)}$
7: Form noisy input $\mathcal{T}_t \leftarrow [\mathcal{V}_s^{(t)}; \mathcal{V}_e^{(t)}] \in \mathbb{R}^{d_s+d_e}$
8: Predict noise: $\hat{\boldsymbol{\epsilon}}_\theta \leftarrow \hat{\boldsymbol{\epsilon}}_\theta(\mathcal{T}_t, t)$
9: Manually denoise skill vector $\mathcal{V}_s^{(0)} \leftarrow \frac{\mathcal{V}_s^{(t)} - \sqrt{1-\bar{\alpha}_t} \cdot \boldsymbol{\epsilon}^{(t)}}{\sqrt{\bar{\alpha}_t}}$
10: Compute denoised expert vector: $\mathcal{V}_e^{(t-1)} \leftarrow \frac{1}{\sqrt{\alpha_t}} \left(\mathcal{V}_e^{(t)} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \hat{\boldsymbol{\epsilon}}_\theta \right)$
11: $\mathcal{T}_{t-1} \leftarrow [\mathcal{V}_s^{(t-1)}; \mathcal{V}_e^{(t-1)}]$
12: **end for**
13: **Step 3: Return predicted experts**
14: $\hat{\mathcal{V}}_e^0 \leftarrow \mathcal{T}_0[\mathcal{V}_e^0]$
15: $\hat{\mathbf{e}}_0 \leftarrow \text{Retriever}(\hat{\mathcal{V}}_e^0)$ ▷ via nearest neighbors in $\phi_e(\mathcal{E})$
16: **return** Predicted experts $\hat{\mathbf{e}}_0$

team $\mathcal{T}_{t-1}$ is then formed so that the query $\mathbf{s}$ is visible to the denoising model with a known noise profile throughout the process (Line 11). After completing all reverse steps, the final output vector $\mathcal{T}_0$ is produced, which consists of the known skill embedding and the predicted expert embedding $\hat{\mathcal{V}}_e$ (Line 14). This embedding is then passed to a retrieval module, which performs a nearest-neighbor search over the embedding dictionary $\phi_e(\mathcal{E})$ to identify the top candidate experts that best match the predicted expert vector (Line 15). The resulting subset $\hat{\mathbf{e}}$ forms the output of the model and represents the proposed team for the input skill query.

4.6. Computational Efficiency and Scalability

Given the structural complexity of the proposed framework, it is important to explicitly assess how this complexity translates into computational cost and whether the resulting performance gains remain practical at scale. As such, we discuss how the proposed diffusion-based team formation model scales with respect to the number of candidate experts and the size of the task dataset, as well as its behavior in practical large-scale settings.

In terms of *scaling with the number of experts*, the core diffusion model operates entirely in a continuous, fixed-dimensional embedding space. As a result, the computational cost of the forward and reverse diffusion processes does not depend on the number of candidate experts. During inference, the reverse diffusion procedure performs T denoising steps, each requiring a single forward pass of the denoising network, resulting in a diffusion cost of $\mathcal{O}(T \cdot C_{\text{UNet}})$.

The term C_{UNet} denotes the computational cost of a single forward pass of the U-Net denoiser. In our formulation, the denoiser operates on the concatenated representation $\mathbf{T}_t \in \mathbb{R}^{d_s+d_e}$, where d_s and d_e denote the dimensionalities of the skill and expert embeddings, respectively. Since the U-Net architecture is fixed in our model, the cost of a single forward pass scales linearly with the input dimensionality and can be expressed as $C_{\text{UNet}} = \mathcal{O}(d_s + d_e)$. Consequently, the total diffusion cost during inference is $\mathcal{O}\big(T \cdot (d_s + d_e)\big)$, which remains constant with respect to the number of experts and the dataset size.

Dependence on the expert pool size arises only in the final retrieval stage, where the predicted expert embedding is mapped to a discrete expert identity. Using an exact nearest-neighbor search, this step has a time complexity of $\mathcal{O}(|\mathcal{E}| \cdot d_e)$, where $|\mathcal{E}|$ is the number of candidate experts. In practical applications, this retrieval step can be implemented using approximate nearest-neighbor indexing techniques, yielding sublinear expected query time and enabling efficient scaling to large expert collections.

Table 1
Dataset Statistics.

Dataset	Teams	Experts	Exp./ Team	Skills	Skill/ Team	Skill/ Exp.
DBLP	10,674	1,887	2.47	2,000	6.63	2.68
DOTA2	6,390	2,727	5.00	3,057	27.76	5.55

When *scaling with task size and dataset growth*, task complexity influences the model only through the task embedding representation. Once embedded, each task is represented by a fixed-length vector, ensuring that both training and inference costs are independent of the raw task size (e.g., number of required skills or textual length of task descriptions). Let N denote the number of training instances. During training, each instance contributes a single diffusion loss term, resulting in a per-epoch training complexity of $\mathcal{O}(N \cdot C_{\text{UNet}})$, which scales linearly with dataset size while remaining constant with respect to individual task dimensionality.

Finally, combining the diffusion and retrieval stages, the overall inference complexity of the proposed method is $\mathcal{O}\big(T \cdot C_{\text{UNet}} + \text{Retriever}(|\mathcal{E}|)\big)$, where Retriever$(|\mathcal{E}|)$ denotes either linear-time exact retrieval or sublinear-time approximate retrieval, depending on the indexing strategy. From a practical perspective, this decoupling of continuous generation and discrete retrieval allows the framework to handle larger and more complex datasets without a combinatorial increase in computational cost. Both training and inference stages are amenable to parallelization on modern hardware, supporting efficient deployment in real-world expert recommendation systems.

Importantly, this analysis shows that the added modeling complexity does not translate into combinatorial inference cost. The diffusion process operates in a fixed-dimensional embedding space, and the only component that scales with the expert pool size is the final retrieval step, which can be efficiently indexed. As a result, the framework achieves substantial performance gains without sacrificing scalability or practical deployability.

5. Experimental Setup

5.1. Datasets

To evaluate the proposed team formation approach, we employed two datasets that are widely used in the literature (Hamidi Rad et al., 2023; Zihayat et al., 2017; Lappas et al., 2009; Khan, Golab, Kargar, Szlichta, and Zihayat, 2020; Dara et al., 2025), namely DBLP[1] and DOTA2[2]. The DBLP dataset contains bibliographic information on major Computer Science publications. We consider the authors of each publication as a team, with each author representing an expert. Following previous methods (Lappas et al., 2009; Kargar and An, 2011; Rad, Bagheri, Kargar, Srivastava, and Szlichta, 2022; Rad et al., 2022, 2021), the skill set of each expert is extracted from the titles of their publications. Pre-processing includes stop-word removal and stemming. The remaining $\{1, 2, 3\}$-grams are ranked based on their TF-IDF scores, and the top 2,000 keywords, manually reviewed for quality and redundancy, are retained as skills. We employ Word2Vec (Mikolov, Chen, Corrado, and Dean, 2013)(Skip-gram) embeddings to represent skills and experts, consistent with prior neural team formation studies and the baselines used in our evaluation (Khalil, Dai, Zhang, Dilkina, and Song, 2017; Kargar et al., 2022; Du et al., 2020; Wu et al., 2017). While more recent representation methods, including contextualized language models, can produce richer semantic representations, their use would introduce confounding factors and limit comparability with existing approaches. Recent work (Zhang, Hamidi Rad, Zihayat, and Bagheri, 2025) has explored LLM-based formulations of team formation, which represent a complementary but distinct modeling direction. In this study, we therefore adopt Word2Vec embeddings to ensure fair comparison and to isolate the impact of the proposed diffusion-based generative framework. The term distribution per article is provided in Table 1 to verify the representativeness of the keywords. The final dataset includes 10,674 teams, 1,887 experts, and 2,000 unique skills.

The DOTA2 dataset is comprised of detailed records of competitive matches in the DOTA2 video game. Each match involves two teams of five players. We consider each winning team as a valid team and its players as experts. The skill set for each match is derived from a range of features, including: (1) in-game configurations (selected heroes, tower status, and barracks status); and (2) game metadata (game version, server region, and server cluster). Matches with anonymous or infrequent players (appearing only once) were excluded. The final dataset included 6,390 teams,

[1] https://www.aminer.org/citation
[2] https://www.kaggle.com/datasets/devinanzelmo/dota-2-matches

Table 2
Comparison of proposed model to baselines for DBLP and DOTA2 datasets. The tabulated results represent the @10 value of the metrics.

Dataset	Method	Recall	MRR	MAP	NDCG
DBLP	Khalil et al.	1.11%	0.85%	0.42%	0.67%
	Kargar et al.	0.52%	0.11%	0.17%	0.29%
	Wu et al.	1.27%	0.97%	0.49%	0.80%
	Sapienza et al.	2.14%	2.38%	0.93%	1.59%
	Du et al.	1.69%	1.86%	0.73%	1.24%
	Fani et al.	0.64%	0.47%	0.19%	0.38%
	Barzegar et al.	3.25%	3.34%	1.29%	2.31%
	Rad et al.	5.03%	5.76%	2.19%	3.82%
	Proposed	**16.41%**	**15.42%**	**9.73%**	**13.08%**
DOTA2	Khalil et al.	0.61%	1.09%	0.27%	0.40%
	Kargar et al.	0.34%	0.27%	0.05%	0.23%
	Wu et al.	0.82%	1.25%	0.37%	0.59%
	Sapienza et al.	2.45%	3.82%	0.75%	1.85%
	Du et al.	2.40%	1.90%	0.73%	1.73%
	Fani et al.	0.60%	0.62%	0.12%	0.39%
	Barzegar et al.	1.60%	1.11%	0.22%	0.88%
	Rad et al.	2.80%	4.99%	0.99%	2.40%
	Proposed	**4.00%**	**13.34%**	**2.67%**	**5.09%**

2,727 experts, and 3,005 skills. Table 1 shows the statistics of the number of matches per player to illustrate player participation distribution. All datasets and code used in our experiments are publicly available for reproducibility and further research[3]

5.2. Evaluation Strategy

Similar to prior works (Dara et al., 2025; Hamidi Rad et al., 2023; Lappas et al., 2009; Zihayat et al., 2017), we evaluate the effectiveness of our proposed method using widely adopted information retrieval metrics: Recall, Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (NDCG). We employ a 10-fold cross-validation scheme. In each fold, the model is trained on 90% of the data and tested on the remaining 10%. A prediction is deemed successful only if the returned team of experts exactly matches the ground-truth team associated with the given skill set. This strict evaluation criterion ensures that the performance scores reflect true alignment with expected expert compositions.

In addition to standard retrieval-based evaluation, we adopt the Help-Hurt diagram to provide a comparative analysis of our method against baseline models. This diagram offers a head-to-head visualization of performance changes across individual queries. It illustrates the number of skill queries for which our model outperforms (help), underperforms (hurt), or performs equivalently to the baselines. This representation highlights not only the overall gain but also the consistency of improvements across different cases. Beyond *efficacy analysis*, we also evaluate our method in terms of *computational efficiency* during inference. In real-world applications, the speed of generating expert teams is often as critical as retrieval quality. We therefore report inference time performance with respect to the retrieval scores.

To evaluate the effectiveness of our approach against the inherent popularity bias problem in long-tailed expert distributions, we employ the Normalized Divergence Kullback-Leibler (NDKL) metric, which quantifies fairness by measuring distributional divergence between predicted and desired ratios of popular versus non-popular experts in formed teams by the fairness parameter q. Specifically, experts are classified as popular or non-popular based on their collaboration frequency relative to the dataset average, and NDKL measures how closely the model's recommendations match a desired ratio where q represents the target proportion of non-popular experts in the team. Lower NDKL values indicate better fairness performance, with NDKL approaching zero when the generated team composition perfectly matches the target distribution.

[3] https://anonymous.4open.science/r/DiffTF-7A20

5.3. Baselines

To benchmark the performance of our proposed model, we compare it against a diverse set of baselines, categorized into three major groups: **(1)** *graph-based*, **(2)** *collaborative filtering-based*, and **(3)** *neural-based* methods. This selection ensures a comprehensive evaluation across both traditional and modern approaches to team formation.

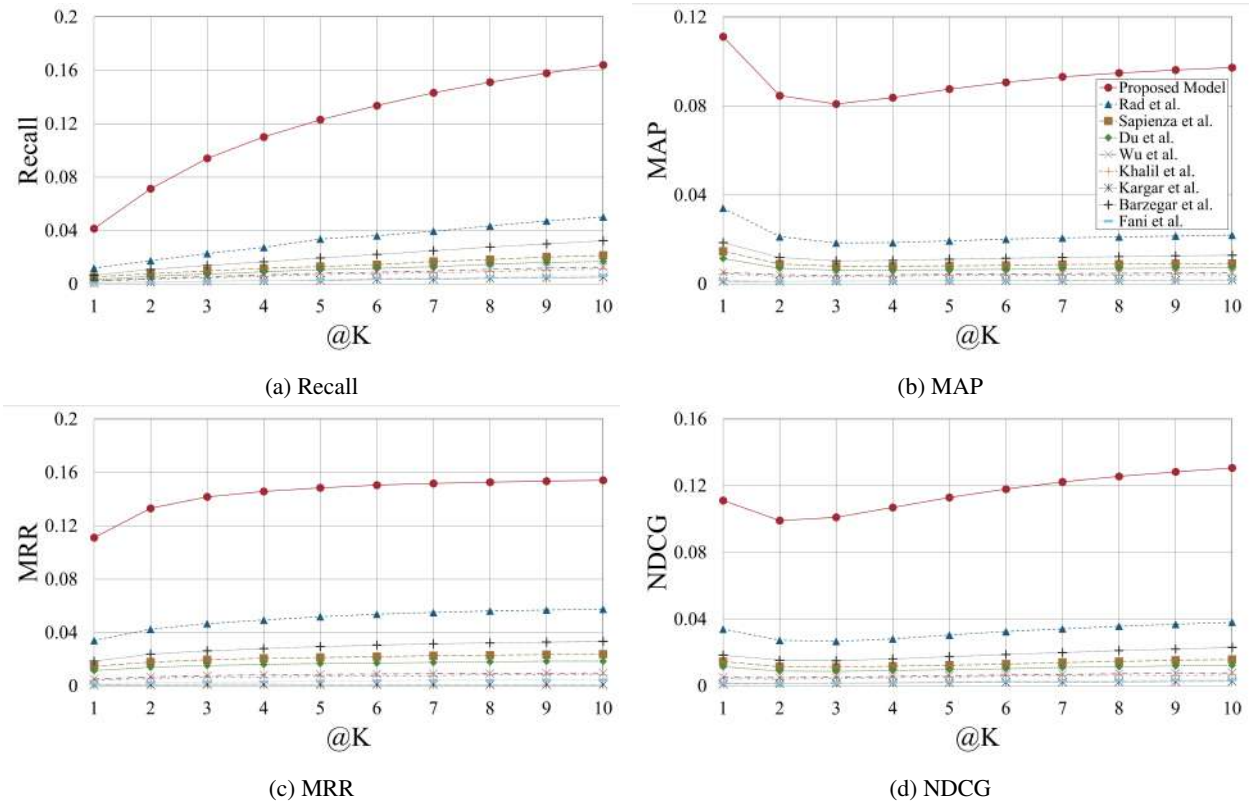
Graph-Based methods operate over a collaboration graph and aim to identify subgraphs (teams) that satisfy skill requirements through graph traversal or optimization heuristics.The state of the art graph based method includes `Kargar et al.` (Kargar et al., 2022), which formulates team formation as a keyword-aware subgraph identification problem. It searches for a connected subgraph that covers the required skill keywords while minimizing communication cost among team members. In contrast *collaborative filtering-based* methods treat team formation as a recommendation problem, where experts are suggested based on learned latent preferences for different skill combinations. We adopt two such baselines: `Recurrent Recommender Network (RNN)` by Wu et al. (Wu et al., 2017), which extends matrix factorization by integrating temporal dynamics using a recurrent LSTM-based framework to capture sequential patterns in collaborative behaviors, and `Group Expert Recommendation Framework (GERF)` by Du et al. (Du et al., 2020), which is a Bayesian learning-to-rank model that introduces Bayesian Group Ranking (BGR) to optimize feature weights. The model uses features representing expert-skill relationships and ranks candidates based on probabilistic relevance.

Finally, the *class of neural baselines* capture latent factors, such as collaboration patterns and domain expertise, by embedding entities (e.g., skills, experts, teams) into continuous vector spaces. Neural team formation models typically optimize an objective function that evaluates how well a generated team satisfies a skill requirement, balances expertise, and minimizes redundancy or communication cost. We include three state of the art baselines in this class: `Rad et al.` (Hamidi Rad et al., 2023), which leverages a variational Bayesian neural network to generate expert embeddings. It addresses data sparsity by encoding both skill and expert features into dense representations for effective team prediction. `Sapienza et al.` (Sapienza et al., 2019) that uses an autoencoder trained on the collaboration graph to learn expert representations. The autoencoder reconstructs the adjacency matrix and captures community structures relevant to team formation. `Khalil et al.` (Khalil et al., 2017), referred to as Structure2Vec (S2V), applies a graph neural network trained via reinforcement learning to generate expert embeddings. The model originally addresses the Group Steiner Tree problem, and we adapt the learned embeddings to solve the team formation task. In addition, we include the work of `Fani et al.` (Fani et al., 2024), which introduces a streaming training strategy that incrementally updates expert and skill embeddings over discrete time intervals to capture temporal evolution in expertise and collaboration patterns. This approach models temporal drift in a single ranking pipeline but does not explicitly learn to select or rerank among multiple candidate team configurations generated by heterogeneous base models. We also adopt `Barzegar et al.` (Barzegar et al., 2025), which proposes a loss-based curriculum learning strategy for neural team formation. The model dynamically reorders training samples based on their difficulty, allowing the neural model to gradually learn complex team formation patterns. While effective at improving convergence and robustness, the method assumes a static embedding space and does not explicitly model inter-list ranking interactions. For all baseline models, we adopted the default hyperparameter settings as specified in their original papers to ensure a fair and reproducible comparison.

6. Findings

Rather than viewing the experimental results solely as aggregate performance comparisons, we interpret them as probes into how different modeling assumptions behave under key structural challenges of team formation. In particular, the experiments are designed to reveal (i) how models respond to extreme sparsity and long-tailed expert participation, (ii) how effectively they generalize to rare and unseen team configurations, and (iii) how architectural choices influence robustness and fairness beyond raw retrieval accuracy. The discussion below therefore focuses not only on numerical improvements, but on explaining why certain modeling choices lead to systematic advantages under these conditions.

The two datasets used in our evaluation exhibit markedly different structural characteristics, which influence both the difficulty of the team formation task and the behavior of competing models. DBLP represents an open-ended, evolving collaboration network with highly skewed authorship frequencies and variable team sizes, emphasizing long-tailed expert participation and sparse skill overlap. In contrast, DOTA2 involves fixed-size teams with rigid role constraints and dense interaction patterns, but limited historical diversity per expert. As a result, DBLP primarily stresses a model's ability to generalize under extreme sparsity, while DOTA2 tests its capacity to recover structured



(a) Recall (b) MAP

(c) MRR (d) NDCG

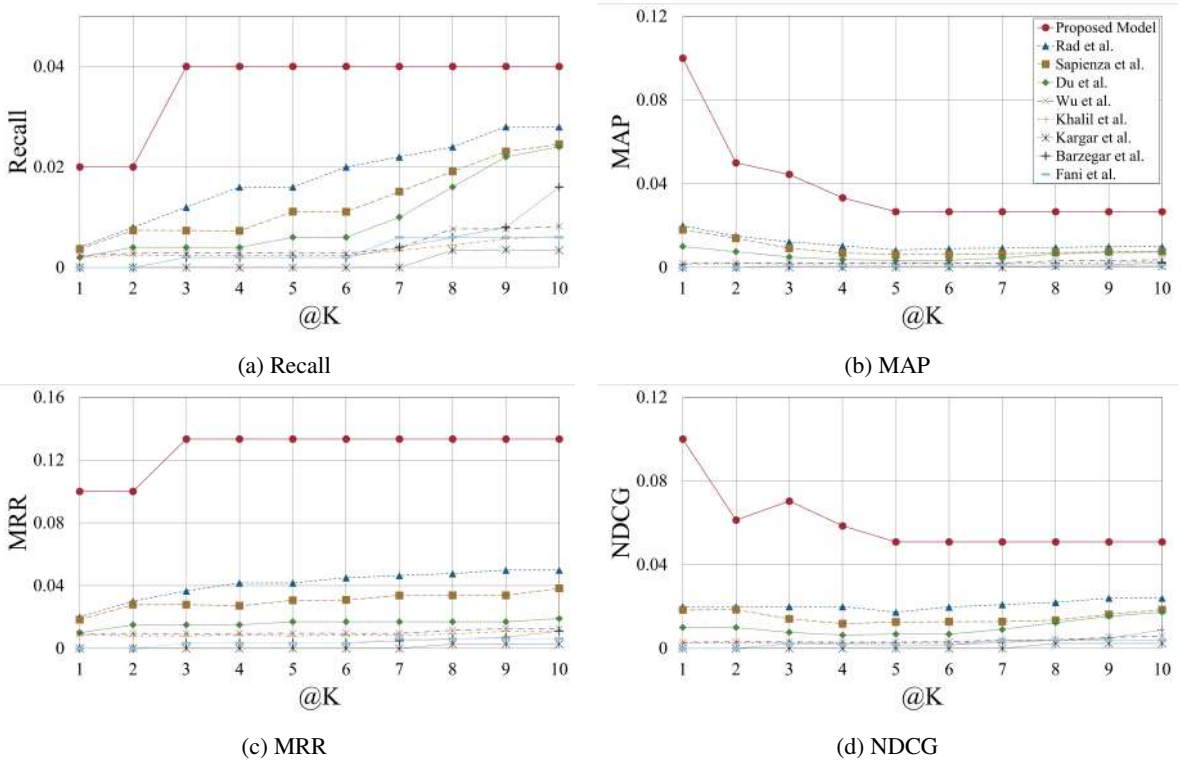
Figure 4: DBLP dataset performance results for Recall, MAP, MRR, and NDCG.

team compositions under strong combinatorial constraints. We therefore interpret results on these datasets through different analytical lenses.

6.1. Retrieval Performance

The evaluation outcomes for the DBLP and DOTA2 datasets are shown in Table 2. For better illustration, we also included trends for each of the methods' performances for the DBLP and DOTA2 datasets in Figures 4 and 5, respectively. Based on these results, we draw several key observations:

(1) Across both datasets, two of the encoder-decoder-based neural methods, Rad et al. (Hamidi Rad et al., 2023), Sapienza et al. (Sapienza et al., 2019), and our proposed approach consistently outperform other baselines. This advantage can be attributed to several factors. One major challenge in team formation datasets is data sparsity (Kargar et al., 2022; Hamidi Rad et al., 2023). Experts often appear in only a small number of collaborations relative to the overall dataset size, limiting the amount of training data available per expert. This issue is particularly evident in the DOTA2 dataset (Table 1), where the majority of experts participate in only a few matches. Neural models are better equipped to handle such sparsity due to their ability to learn latent representations and capture patterns in collaborative behavior. Their ability to both learn known collaborations and generalize to new skill combinations enables stronger performance in these settings (Rabin, Hussain, Alipour, and Hellendoorn, 2023). We note that the absolute values of Recall@10 on the DOTA2 dataset are relatively low across all methods. This behavior reflects the inherently combinatorial nature of expert team formation, particularly under strict exact-match evaluation criteria. Moreover, Recall@10 is computed under a strict exact-match criterion, where partial overlap with the ground-truth team is not considered a successful retrieval. As a result, even strong models yield low absolute scores under this evaluation setting. Importantly, all baselines considered are designed for the same team formation task and evaluated under identical conditions, making relative performance improvements the appropriate and meaningful indicator of model effectiveness.



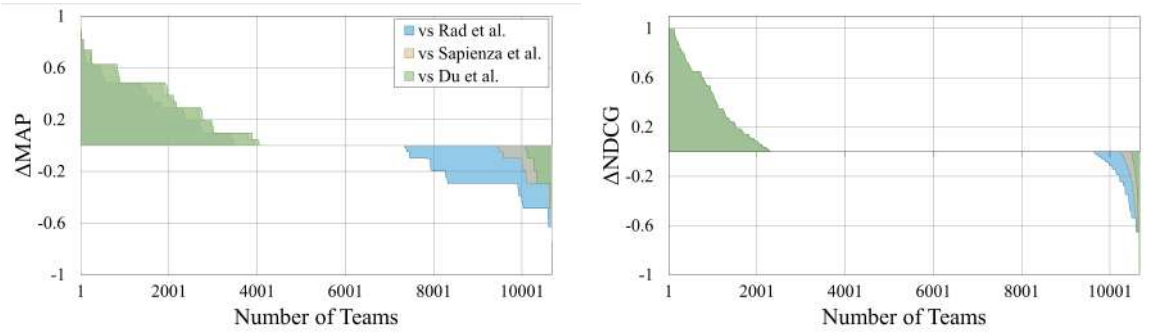
(a) Recall (b) MAP

(c) MRR (d) NDCG

Figure 5: DOTA2 dataset performance results for Recall, MAP, MRR, and NDCG.

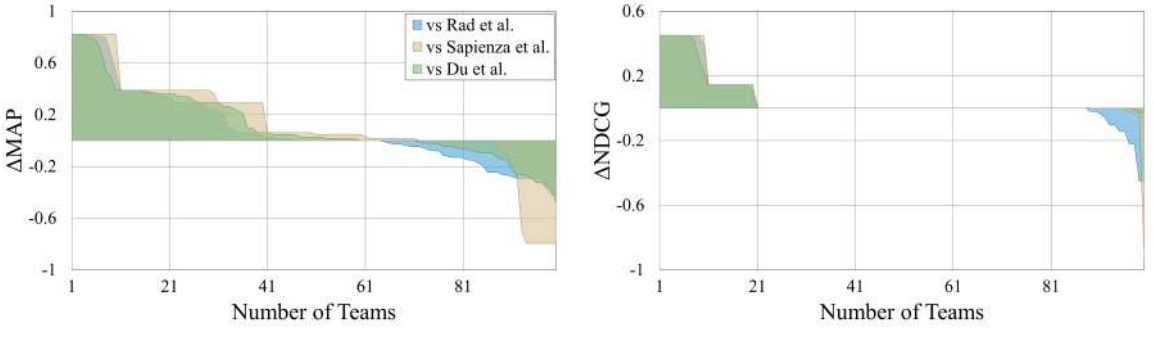
Importantly, the relative gains of the proposed model are most pronounced in rare and difficult team formation queries, where historical collaboration signals are weakest. In these cases, ranking-based and memorization-heavy models degrade sharply because they rely on repeatedly selecting a small subset of frequently observed experts. By contrast, the diffusion-based generative process explicitly models uncertainty and explores the conditional distribution of expert configurations, allowing it to recover plausible teams even when no close historical analog exists. This explains why improvements are largest precisely where sparsity and combinatorial uncertainty are most severe.

(2) Based on the results of the collaborative filtering-based approaches, namely Wu et al. (Wu et al., 2017) and Du et al. (Du et al., 2020), we find that these approaches exhibit different behavior in relation to the memorization–generalization tradeoff. Wu et al. employ recurrent neural networks that emphasize sequential modeling of historical collaborations, making them highly capable of memorization. In contrast, Du et al.'s GERF uses Bayesian inference to recommend experts based on probabilistic skill relevance, prioritizing generalization. While both strategies do not show competitive scores, GERF demonstrates stronger overall performance. This suggests that effective generalization is more critical than memorization in the team formation problem. In contrast, the primary strength of our proposed model lies in its hybrid architecture, which explicitly addresses the long-standing trade-off between memorization and generalization in expert team formation. We achieve this by combining two complementary strategies: skill and expert embeddings and a diffusion-based U-Net generator. The embedding mechanism (Le and Mikolov, 2014) encodes experts and skills based on contextual and co-occurrence information from sparse historical data. This helps preserve known collaboration patterns and previously observed expert-skill associations, thereby enhancing the model's ability to learn meaningful team configurations. Prior work has shown that such embedding-based models are highly effective in recovering and reinforcing implicit expert relations in sparse environments (Wu et al., 2017). On the other hand, the diffusion-based U-Net component offers strong generalization capabilities. Denoising diffusion models can learn structured latent representations and synthesize coherent outputs even from incomplete or noisy input data (Mi, Wang, Qian, Ye, Liu, Tulyakov, Aberman, and Xu, 2025; Batra and Sukhatme, 2025). In our context, this allows the model to propose expert teams that are not only effective but also novel. As a result, it expands beyond directly memorized cases while maintaining semantic validity. These properties are particularly valuable in real-world settings where input skill queries often involve rare or unseen combinations. By integrating



(a) MAP (b) NDCG

Figure 6: Help-Hurt diagram for DBLP dataset.



(a) MAP (b) NDCG

Figure 7: Help-Hurt diagram for DOTA2 dataset.

these two components, our model is able to both recall historically validated collaborations and generalize to unseen task requirements, resulting in robust and flexible team formation outcomes.

(3) Our method achieves superior scores across all top-k thresholds and evaluation metrics. This stable performance indicates that the model not only retrieves relevant experts for a given skill set but also scales well with the size of the predicted team. At lower values of k, where precision is critical, our model effectively prioritizes the most relevant experts who possess key skills and exhibit strong past collaborative records. As k increases, and the task demands a broader team composition, the model continues to maintain high recall by including additional candidates who are both contextually aligned with the required skill set and historically consistent with successful collaborations. This ability to preserve performance while expanding the solution space reflects the model's capacity to generalize beyond the most obvious choices. This indicates that the model maintains retrieval quality while expanding the candidate set, supporting flexible team formation across varying task sizes.

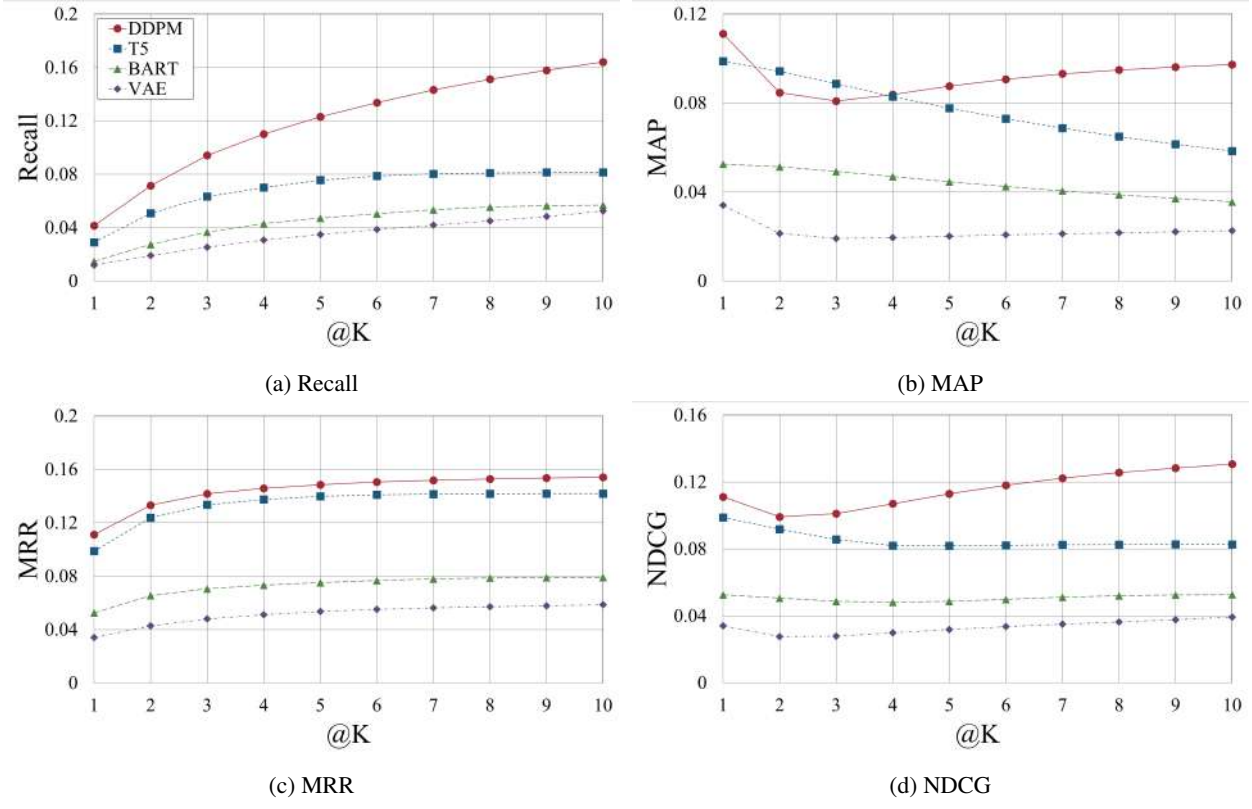
6.2. Performance Robustness

To assess the robustness of the performance improvements reported in Section 6.1, we conduct a detailed per-query analysis using Help-Hurt diagrams. This visualization helps us determine whether our model consistently outperforms competing methods across the entire set of team formation queries, rather than achieving improvements on only a subset of cases. For this analysis, we compare our proposed method against the three strongest baselines, namely Rad et al. (Hamidi Rad et al., 2023), Sapienza et al. (Sapienza et al., 2019), and Du et al. (Du et al., 2020). Since trends observed across all ranking metrics (Recall, MAP, MRR, NDCG) are consistent on both datasets, we focus on reporting results for NDCG and MAP for the sake of space.

Figures 6 and 7 present the Help-Hurt diagrams for the DBLP and DOTA2 datasets, respectively. In these figures, each point on the x-axis represents an individual team formation query, while the y-axis shows the percentage difference in performance between our method and the corresponding baseline for that query. The y-axis value is either positive

Table 3
Comparative results of our proposed DDPM backbone to other generative models. The tabulated results represent the @10 value of the metrics.

Dataset	Method	Recall	MRR	MAP	NDCG
DBLP	VAE	5.25%	5.87%	2.26%	3.93%
	BART	5.69%	7.91%	3.56%	5.27%
	T5	8.14%	14.18%	5.84%	8.26%
	DDPM	**16.41%**	**15.41%**	**9.73%**	**13.07%**
DOTA2	VAE	1.80%	1.24%	0.04%	1.08%
	BART	2.60%	1.97%	0.76%	1.75%
	T5	3.75%	4.86%	**3.15%**	3.87%
	DDPM	**4.00%**	**13.34%**	2.67%	**5.09%**



(a) Recall

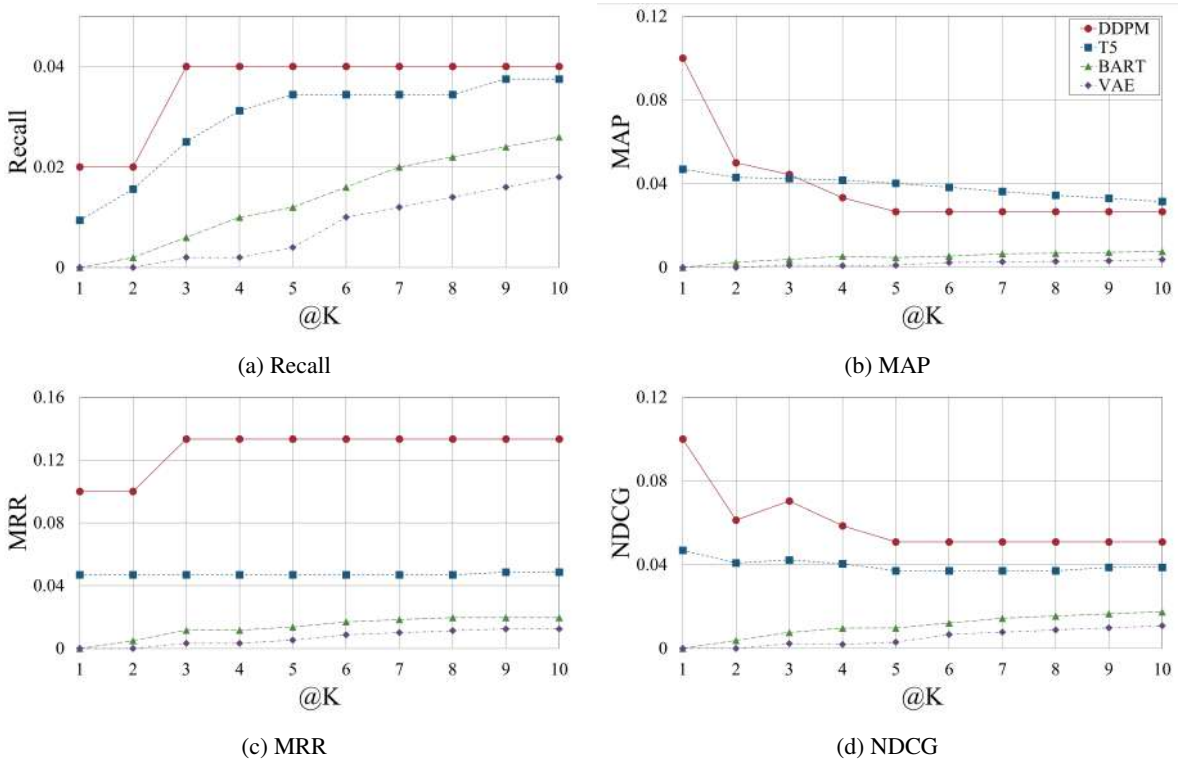
(b) MAP

(c) MRR

(d) NDCG

Figure 8: Comparison for the DBLP dataset performance results for Recall, MAP, MRR, and NDCG.

(help), negative (hurt), or zero, which indicates no change compared to the baseline. The diagrams clearly show that the improvements achieved by our model are not isolated to a few favorable cases but are observed consistently across a wide range of queries. For instance, on DBLP in particular, the advantage is significant. Our proposed method improves 3,942 queries and hurts only 1,218 on MAP versus Sapienza et al., and notably, achieves 4,050 helps versus just 596 hurts against Du et al. The NDCG results follow a similar trend, with over 2,274 helps and fewer than 424 hurts against both baselines. These results confirm that our improvements are not limited to a few favorable cases but generalize across a wide spectrum of team formation queries. The consistent dominance in the ”help” region of the diagram underscores the reliability of our approach in producing better-ranked teams.



(a) Recall (b) MAP

(c) MRR (d) NDCG

Figure 9: Comparison for DOTA2 dataset performance results for Recall, MAP, MRR, and NDCG.

6.3. Comparison to Other Generative Models

We further justify our choice of a DDPM-based approach by comparing it against alternative generative backbones. Specifically, we replace the diffusion backbone with a Variational Autoencoder (VAE), a Bidirectional and Auto-Regressive Transformer (BART) model, and a Text-to-Text Transfer Transformer (T5)-based generative model, while keeping the input representations, conditioning strategy, discretization via nearest-neighbor retrieval, and evaluation protocol identical.

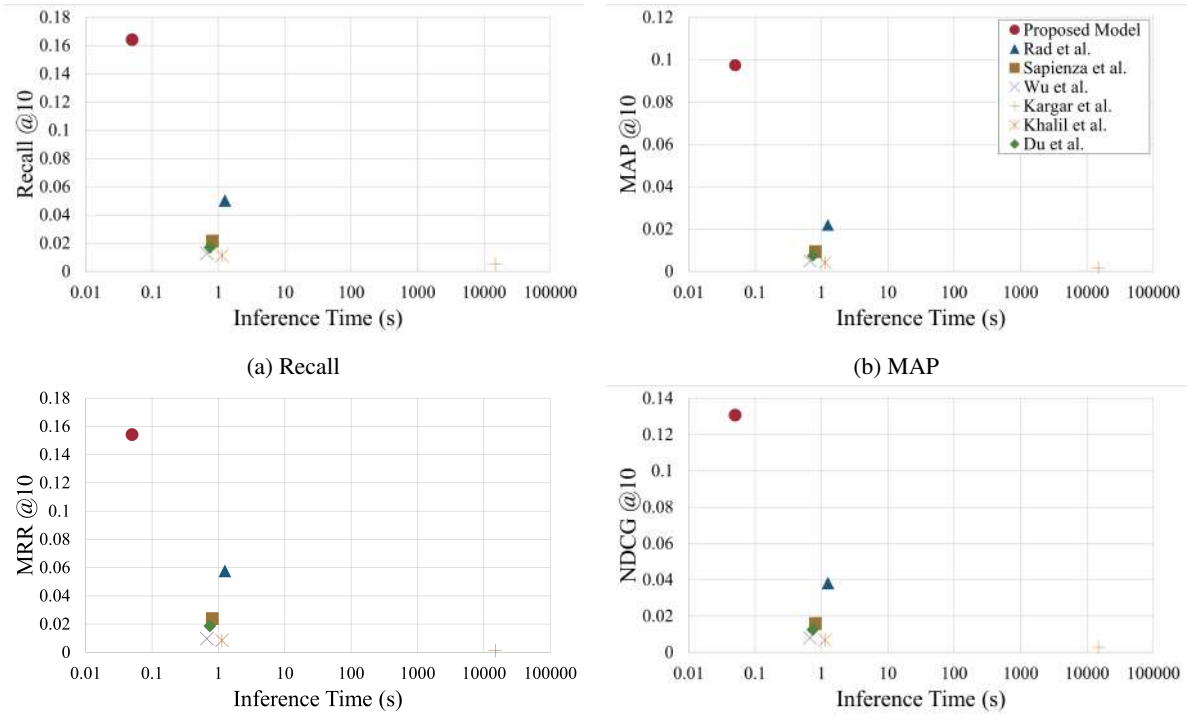
Table 3 reports results at @10 for our ranking metrics on both DBLP and DOTA2 datasets. Across all metrics and datasets, the DDPM backbone consistently outperforms the alternative generative models. On DBLP, DDPM achieves substantially higher Recall@10 than the strongest non-diffusion baseline (T5), with similarly pronounced gains observed for MAP and NDCG. A comparable trend is observed on DOTA2, where DDPM attains the highest results in three of the four evaluation metrics.

Figures 8 and 9 further illustrate performance trends across varying cutoff values k. On DBLP, the DDPM-based model dominates competing generative backbones across all values of k, exhibiting stable and monotonic improvements for the evaluation metrics. On DOTA2, a similar pattern emerges where DDPM consistently achieves higher Recall and NDCG across all cutoffs, while maintaining a clear advantage in MRR. Although MAP decreases as k increases for all methods, DDPM retains stronger overall ranking quality at lower cutoffs and remains competitive at higher values of k. In contrast, autoregressive and latent-variable models such as BART and VAE exhibit earlier saturation and weaker scaling behavior, particularly for NDCG and MRR.

These results are consistent with differences in how the compared models represent conditional distributions. VAEs rely on a single latent variable and can struggle to capture complex multi-modal structure, while autoregressive sequence-to-sequence models impose a fixed generation order. In contrast, DDPMs perform iterative refinement in the expert embedding space, allowing the model to progressively reconcile multiple skill constraints during generation.

6.4. Execution Efficiency

To assess the practical utility of our model in real-world settings, we evaluate its performance under computational time constraints. Specifically, we examine the trade-off between inference time and team formation quality. Figures 10



(a) Recall

(b) MAP

(c) MRR

(d) NDCG

Figure 10: Inference time versus performance tradeoff on the DBLP dataset.

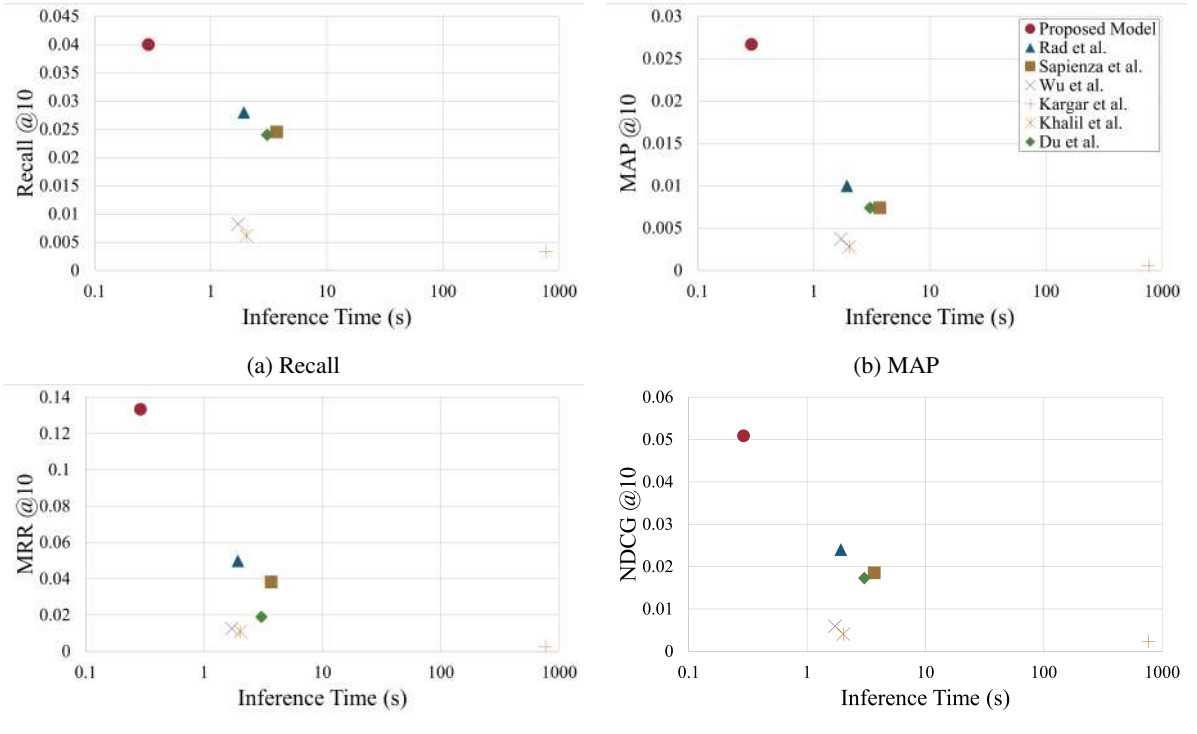
and 11 present this comparison using the DBLP and DOTA2 datasets, respectively. In both figures, the x-axis represents the average time (in seconds) required to generate a team, and the y-axis denotes the quality score. The optimal operating point lies in the upper-left quadrant, indicating both high effectiveness and low latency.

Across both datasets, our model consistently occupies the upper-left quadrant of the plots, reflecting a highly favorable trade-off between retrieval quality and inference efficiency. This performance stems from two key architectural choices: first, the use of dense embeddings for both skills and experts significantly reduces input dimensionality and accelerates forward passes through the network; second, the conditional sampling ensures that only the expert portion of the vector is actively sampled, reducing unnecessary computation. On the DBLP dataset, although the method by Rad et al. (Hamidi Rad et al., 2023) achieves reasonable quality teams due to its probabilistic modeling of team synergy, it incurs substantial computational overhead because it requires extensive sampling and optimization over candidate subsets. Our model, in contrast, generates expert vectors in a single, learned diffusion trajectory conditioned on the known skills, avoiding costly post-hoc search.

On the DOTA2 dataset, models such as those proposed by Wu et al. (Wu et al., 2017) and Khalil et al. (Khalil et al., 2017) exhibit faster inference, but this comes at the cost of output quality. These methods rely on heuristic rules or greedy selection mechanisms that do not model the interaction effects between team members, resulting in suboptimal team composition. Meanwhile, Rad et al. (Hamidi Rad et al., 2023) and Sapienza et al. (Sapienza et al., 2019) demonstrate stronger retrieval quality by leveraging graph-based structures or probabilistic team embeddings, yet their methods scale poorly with increasing expert space size or higher task complexity. Our approach avoids these issues by integrating the relational modeling strength of neural networks with the generative robustness of diffusion processes.

6.5. Fairness Assessment in Long-Tailed Distributions

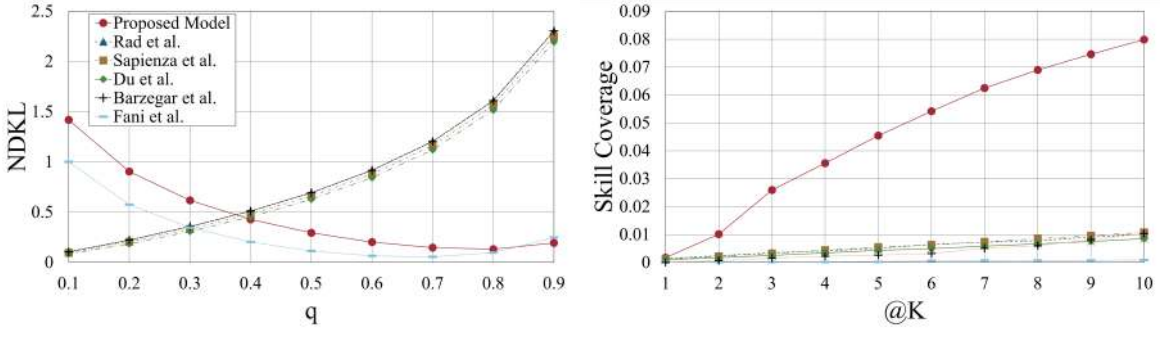
Our analysis of the DBLP dataset revelead that on average, each expert appears in 15 teams. We thus denote the popularity of an expert as whether they appear in more teams than the dataset average, a criteria that is met by 32% of the experts in the dataset, thus making it suitable for further fairness analysis. Figure 12a demonstrates the fairness performance across varying target ratios of non-popular experts (q). At each q, a low NDKL corresponds to better



(a) Recall (b) MAP

(c) MRR (d) NDCG

Figure 11: Inference time versus performance tradeoff on the DOTA2 dataset.

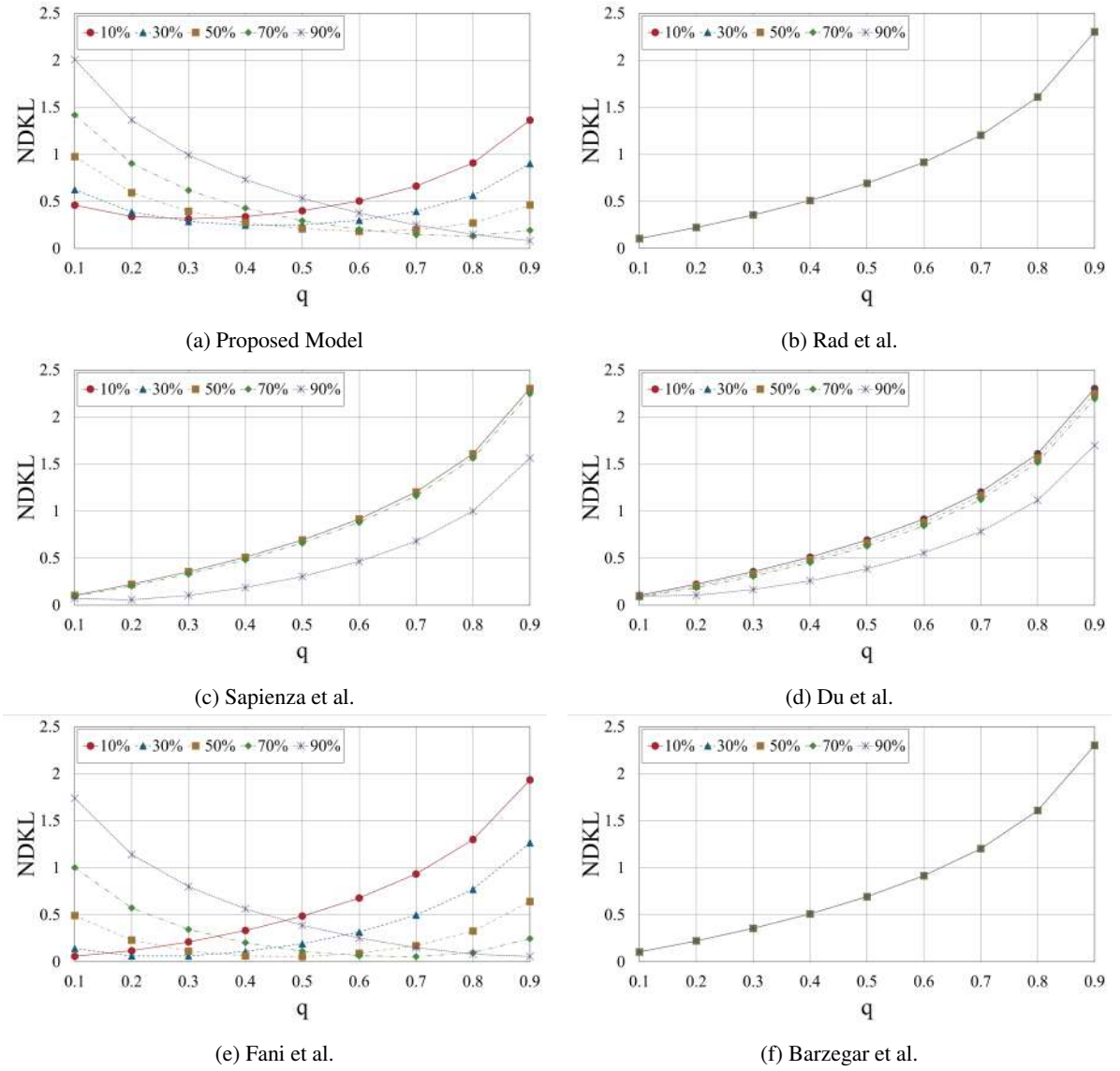


(a) NDKL (b) Soft Skill Matching

Figure 12: Fairness assessment through the NDKL metric and soft skill matching.

alignment between the model's generated team composition and the desired fairness distribution, while a high NDKL indicates divergence from the target ratio. In other words, the q value at which the NDKL minimizes for a model, corresponds to the average ratio of unpopular experts present in the teams formed by that model.

Our analysis shows that most baseline methods, with the exception of `Fani et al.`, exhibit monotonically increasing NDKL curves, indicating a persistent tendency to favor highly connected experts as fairness constraints become stricter. These methods remain influenced by the long-tailed structure of the data, continuing to rely on a small subset of frequently collaborating experts even as the desired exposure to underrepresented experts increases. By modeling collaboration as a time-varying network, `Fani et al.` partially alleviates this behavior and achieves a minimum NDKL at $q = 0.7$, reflecting reduced reliance on the most popular experts. However, this improvement in exposure comes at the cost of incomplete skill coverage, as reflected in the retrieval performance reported in Table 2 and Figures 4 and 5. In contrast, our proposed method achieves minimum NDKL score at a substantially higher $q = 0.8$,



(a) Proposed Model

(b) Rad et al.

(c) Sapienza et al.

(d) Du et al.

(e) Fani et al.

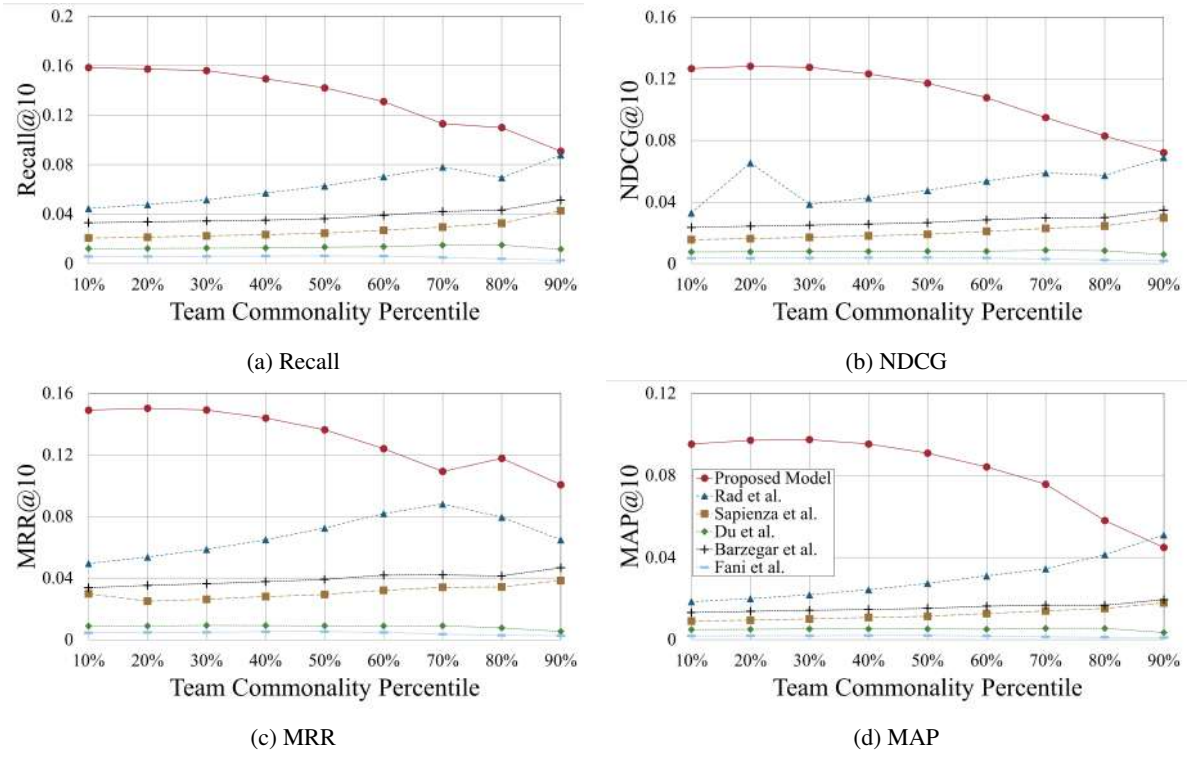
(f) Barzegar et al.

Figure 13: Detailed analysis of model fairness with variable expert popularity thresholds. At each threshold, the NDKL metric minimizes at the q that is equal to the ratio of unpopular experts present in the prediction.

which demonstrates that, on average, 80% of the experts generated by our model for each team are unpopular experts. Our proposed approach thus achieves a two-fold advantage, achieving superior performance in forming teams as evidenced by our evaluation metrics, while excelling at promoting non-popular experts and avoiding overfitting to the most common experts.

These fairness improvements are not incidental side effects of regularization, but a direct consequence of modeling team formation as conditional generation rather than candidate ranking. Because the diffusion process samples from a skill-conditioned distribution rather than scoring experts independently, it does not collapse onto historically dominant experts unless they are genuinely required by the skill configuration. This explains why the model simultaneously improves fairness and retrieval effectiveness: diversity emerges from distributional modeling, while relevance is preserved through skill-conditioned inpainting.

Figure 12b further validates that this fairness improvement does not compromise skill alignment quality through skill coverage analysis. The skill coverage metric measures the extent to which recommended teams completely cover the specified skill requirements, evaluating whether all required skills are represented in the team regardless of the



(a) Recall (b) NDCG

(c) MRR (d) MAP

Figure 14: Performance metrics@10 reported based on team rarity. Lower value on the horizontal axis represent rare and difficult teams where experts have the lowest collaboration history. Higher percentile indicates more common teams.

Table 4
Average overlap of the predicted experts by each method with the top-100 most popular experts across the DBLP dataset.

Method	Avg. Overlap
Barzegar et al.	94.21
Rad et al.	86.68
Sapienza et al.	11.45
Fani et al.	5.48
Du et al.	4.44
Proposed Model	**2.62**

specific experts chosen. Unlike exact-match recall metrics that require identical expert sets, skill coverage rewards models that form teams with different but equally qualified experts who collectively possess all necessary expertise.

We further extend our fairness analysis in Figures 13 and 14. In Figure 13, we vary the threshold for the percentage of total teams an expert must appear in to be considered a popular expert. At 10%, almost every expert is considered popular, while at 90%, almost every expert is considered unpopular. The goal of this experiment is to stress test the diversity of our model outputs in extreme situations regarding the distribution of experts. We see that similar to our results in Figure 12a, the previous approaches cannot escape long-tail overfitting even in the most lenient thresholds ($70-90\%$) and always fail to populate their teams with non-popular experts regardless of what collaboration threshold is used to define popularity. While `Fani et al.` improves this bias at higher, more lenient thresholds, their NDKL at the most strict thresholds 10% and 30% follows the same monotonic rise as other baselines, indicating that they also cannot mitigate the popularity bias at extremely sparse and long-tailed scenarios. In contrast, our model contains a higher ratio of unpopular experts at every threshold, including the 10% and 30% cases where the NDKL minimizes at $q = 0.3$ and $q = 0.5$, respectively. This result demonstrates that even in extremely strict evaluation where the candidate

pool of popular experts is very small, unpopular experts comprise $30-50\%$ of each team formed by our model, which is a significant improvement compared to the sub-10% ratio of other baselines.

The fairness behavior of the proposed model also differs across datasets in a manner consistent with their underlying structures. On DBLP, where popularity bias is extreme and historically entrenched, improvements in NDKL reflect the model's ability to break away from dominant authors and promote underrepresented experts without sacrificing skill coverage. On DOTA2, however, fairness improvements are constrained by fixed team size and role dependencies, which limit the extent to which unpopular players can be substituted without violating team viability. The model's consistent advantage across both datasets therefore demonstrates not uniform behavior, but adaptive responses to different structural constraints.

In Figure 14, we study the behavior of our model across team rarity, where we divide teams based on the collaboration history of experts in the team. Lower percentile teams are rare and difficult since their experts have the lowest outside collaboration history in other teams, and higher percentile teams are more common since the experts on those teams frequently collaborate in other teams. Our proposed model demonstrates consistently superior performance, particularly in rare and difficult teams where there is no collaboration pattern between the experts. As teams become more common, our model exhibits closer performance to our top baseline; This is expected since the correct experts at high percentiles often are the most popular experts, which dominate the prediction set of our top baselines.

The superior performance of our proposed model directly addresses the long-tail problem through our *predictive expert generation* paradigm. Rather than ranking the same subset of popular experts for every skill query, our conditional diffusion process learns to identify the most appropriate experts for specific skill configurations, effectively breaking free from the popularity constraints present in the training distribution.

6.6. Summary of Findings

The experimental results presented in Section 6 indicated that our proposed framework achieves the two central objectives outlined in the introductory section of this paper as follows:

(1) With respect to mitigating sparsity while preserving semantic richness, our embedding-space diffusion model consistently outperforms existing baselines across both DBLP and DOTA2 datasets. On DBLP, the model achieves notable improvements in recall and MAP over graph-based and factorization methods, even under extreme sparsity conditions where most experts appear in very few collaborations. On DOTA2, which exhibits highly skewed team distributions, the diffusion framework maintains high NDCG scores while neural and probabilistic methods degrade substantially. These results demonstrate that operating directly in the embedding space allows the model to capture nuanced skill–expert associations without excessive latent compression, ensuring robustness across datasets with differing sparsity and long-tail characteristics.

(2) In addressing popularity bias, the model achieves a marked improvement in fairness without sacrificing task performance. Popularity overlap analysis (Table 4) shows that the generated teams include on average only 2.62 of the top-100 most frequent experts, compared to 86.68 for the neural baseline of Rad et al. This indicates a notable reduction in over-reliance on frequently observed experts. Complementary fairness metrics provide further evidence, with normalized discounted KL divergence (NDKL) values approaching 0.1 in high-diversity evaluation settings, signifying near-optimal promotion of underrepresented experts while retaining full coverage of required skills. These findings confirm that the conditional inpainting formulation directs generation toward skill relevance rather than collaboration frequency, enabling the construction of diverse teams that remain effective under real-world conditions.

While the proposed framework introduces additional modeling components compared to traditional ranking-based approaches, our results indicate that this complexity is warranted by the resulting gains in robustness, fairness, and generalization. In particular, the diffusion-based generative backbone enables consistent improvements on rare and difficult team formation queries, where simpler models degrade substantially. At the same time, the controlled computational footprint of the model ensures that these gains do not come at the cost of prohibitive inference latency. This suggests that the added structural complexity is justified in settings characterized by sparse data, long-tailed expert participation, and high combinatorial uncertainty.

7. Concluding Remarks

The primary contribution of this work is a generative reframing of expert team formation, in which teams are synthesized via skill-conditioned diffusion rather than selected through ranking or optimization. By treating the

problem as a skill-conditioned sampling task, our approach captures complex dependencies between task requirements and expert configurations, overcomes limitations of sparse data and long-tailed distributions, and enables the synthesis of diverse and contextually appropriate teams. Through a combination of dense embeddings, skill-conditioned sampling, and skill-preserving denoising trajectories, our method offers a scalable and effective alternative to existing graph-based, probabilistic, and neural baselines. Empirical evaluations on multiple real-world datasets validate its advantages in both retrieval accuracy and inference efficiency. Our findings confirm that our predictive expert generation paradigm not only addresses fairness concerns but actually improves functional team composition by identifying the most skill-appropriate experts rather than defaulting to popular but potentially mismatched selections, directly mitigating the negative effects of long-tailed expert participation while maintaining superior performance.

7.1. Limitations

While the proposed diffusion-based framework demonstrates strong empirical performance and improved fairness in expert team formation, several limitations should be acknowledged:

- *Dependence on historical collaboration data.* The model is trained on historically observed teams and therefore inherits the assumptions and biases implicit in past collaboration patterns. Although the conditional generative formulation mitigates over-reliance on highly popular experts, the framework does not explicitly verify whether historically formed teams were optimal with respect to external performance criteria. As a result, the generated teams should be interpreted as plausible and historically consistent rather than guaranteed optimal solutions.

- *Static modeling of expertise.* In its current form, the framework assumes that expert skills and relationships are static over the training period. This limits applicability in environments where expertise evolves rapidly or where collaboration relevance decays over time. While the conditional diffusion formulation is amenable to temporal extensions, such dynamics are not explicitly modeled in the present study.

- *Fixed embedding representations.* The model relies on pre-trained embedding representations for skills and experts. Although this design improves computational efficiency and comparability with prior work, it constrains the expressiveness of the representation space and may limit performance in domains where richer contextual or semantic signals are available.

- *Scalability of the discretization step.* While the diffusion process itself operates in a fixed-dimensional continuous space, the final nearest-neighbor retrieval step scales with the size of the expert pool. Although approximate indexing techniques can alleviate this cost, the current implementation does not explicitly evaluate large-scale retrieval optimizations.

7.2. Future Work

Several extensions of the proposed framework would further improve its applicability in dynamic and multi-domain environments where skill requirements and expert contributions evolve over time:

- *Temporal modeling.* In many real-world settings, expert availability, expertise, and collaboration patterns are inherently time-dependent. Incorporating temporal dynamics such as time-aware skill embeddings or sequential conditioning over past team formations would allow the model to adapt to evolving expertise and shifting collaboration structures. From a generative perspective, this could be achieved by conditioning the diffusion process on temporal context or by learning time-indexed denoising trajectories. Such extensions would enable the framework to support scenarios where historical relevance decays over time, improving robustness in rapidly evolving domains such as emerging research areas, fast-paced engineering teams, or competitive online environments.

- *Multi-task and multi-domain team formation.* Many practical applications require forming teams that satisfy multiple, potentially overlapping objectives, for example, balancing technical skills, domain diversity, and organizational constraints across different projects. Extending the current framework to a multi-task setting would allow the diffusion model to condition on multiple skill sets or task descriptors simultaneously. This would enable the generation of teams that are not only skill-complete for a single task but also reusable and adaptable across related domains, increasing the model's utility in large organizations or platform-based ecosystems.

- *Adaptive applicability through conditional generation.* Importantly, the conditional generative nature of the proposed framework makes these extensions natural rather than ad hoc. Temporal signals, task identifiers, or domain indicators can be incorporated as additional conditioning variables without altering the core generative mechanism. This flexibility suggests that the framework can evolve alongside the environments it is deployed in, supporting dynamic, multi-domain team formation scenarios without requiring fundamental redesign.

8. Statements and Declarations

This study was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2024-04733.

References

A. Costa, F. B. A. Ramos, M. Perkusich, E. Dantas, E. Dilorenzo, F. Chagas, A. Meireles, D. Albuquerque, L. Silva, H. O. Almeida, A. Perkusich, Team formation in software engineering: A systematic mapping study, IEEE Access 8 (2020) 145687–145712. doi:10.1109/ACCESS.2020.3015017.

K. Balog, Y. Fang, M. De Rijke, P. Serdyukov, L. Si, Expertise retrieval, Foundations and Trends in Information Retrieval 6 (2012) 127–256. doi:10.1561/1500000024.

Y. Fu, J. Luo, G. Nan, D. Li, Peer review expert group recommendation: A multi-subject coverage-based approach, Expert Systems with Applications 264 (2025) 125971. URL: https://www.sciencedirect.com/science/article/pii/S0957417424028380. doi:https://doi.org/10.1016/j.eswa.2024.125971.

X. Chen, S. Wang, J. McAuley, D. Jannach, L. Yao, On the opportunities and challenges of offline reinforcement learning for recommender systems, ACM Transactions on Information Systems 42 (2024) 1–26.

M. Khodabakhsh, E. Bagheri, Learning to rank under uncertainty: a robust neural approach, Knowledge and Information Systems 67 (2025) 12495–12545.

A. Elías, R. Jiménez, A. M. Paganoni, L. M. Sangalli, Integrated depths for partially observed functional data, Journal of computational and graphical statistics 32 (2023) 341–352.

C. Dorn, F. Skopik, D. Schall, S. Dustdar, Interaction mining and skill-dependent recommendations for multi-objective team composition, Data Knowl. Eng. 70 (2011) 866–891. doi:10.1016/j.datak.2011.06.004.

M. Sozio, A. Gionis, The community-search problem and how to plan a successful cocktail party, in: ACM SIGKDD 2010, Washington, DC, USA, July 25-28, 2010, 2010, pp. 939–948. doi:10.1145/1835804.1835923.

R. Hamidi Rad, H. Fani, E. Bagheri, M. Kargar, D. Srivastava, J. Szlichta, A variational neural architecture for skill-based team formation, ACM Trans. Inf. Syst. 42 (2023). doi:10.1145/3589762.

L. Boratto, G. Fenu, M. Marras, G. Medda, Practical perspectives of consumer fairness in recommendation, Information Processing Management 60 (2023) 103208. URL: https://www.sciencedirect.com/science/article/pii/S0306457322003090. doi:https://doi.org/10.1016/j.ipm.2022.103208.

L. Boratto, S. Faralli, M. Marras, G. Stilo, Guest editorial of the ipm special issue on algorithmic bias and fairness in search and recommendation, Information Processing Management 59 (2022) 102791. doi:10.1016/j.ipm.2021.102791.

T. Lappas, L. Liu, E. Terzi, Finding a Team of Experts in Social Networks, in: KDD, 2009, pp. 467–476. doi:10.1145/1557019.1557074.

M. Kargar, A. An, Discovering top-k teams of experts with/without a leader in social networks, in: ACM CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011, 2011, pp. 985–994. doi:10.1145/2063576.2063718.

Y. Koren, Collaborative filtering with temporal dynamics, in: ACM SIGKDD, Paris, France, June 28 - July 1, 2009, 2009, pp. 447–456. doi:10.1145/1721654.1721677.

C. Wu, M. Jiang, J. Li, P. Li, H. Liu, Recurrent recommender networks, in: WSDM, 2017. doi:10.1145/3018661.3018689.

R. H. Rad, S. Seyedsalehi, M. Kargar, M. Zihayat, E. Bagheri, A neural approach to forming coherent teams in collaboration networks, in: EDBT 2022, Edinburgh, UK, March 29 - April 1, 2022, 2022, pp. 2:440–2:444. doi:10.48786/edbt.2022.37.

M. Dara, R. H. Rad, F. Zarrinkalam, E. Bagheri, Retrieval-augmented neural team formation, in: ECIR 2025, Lucca, Italy, April 6-10, 2025, volume 15574 of *Lecture Notes in Computer Science*, 2025, pp. 362–371. doi:10.1007/978-3-031-88714-7_35.

E. Xu, K. Zhao, Z. Yu, Y. Zhang, B. Guo, L. Yao, Limits of predictability in top-n recommendation, Information Processing Management 61 (2024) 103731. URL: https://www.sciencedirect.com/science/article/pii/S0306457324000918. doi:https://doi.org/10.1016/j.ipm.2024.103731.

H. Chen, Z. Feng, S. Chen, H. Wu, Y. Sun, J. Li, Q. Gao, L. Zhang, X. Xue, Incorporating forgetting curve and memory replay for evolving socially-aware recommendation, Information Processing Management 62 (2025) 104070. URL: https://www.sciencedirect.com/science/article/pii/S0306457325000123. doi:https://doi.org/10.1016/j.ipm.2025.104070.

H. Fani, R. Barzegar, A. Dashti, M. Saeedi, A streaming approach tonbsp;neural team formation training, in: Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2024, p. 325–340. URL: https://doi.org/10.1007/978-3-031-56027-9_20. doi:10.1007/978-3-031-56027-9_20.

R. Barzegar, M. N. Kurepa, H. Fani, Adaptive loss-based curricula for neural team recommendation, in: W. Nejdl, S. Auer, M. Cha, M. Moens, M. Najork (Eds.), Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining, WSDM 2025, Hannover, Germany, March 10-14, 2025, ACM, 2025, pp. 914–923. URL: https://doi.org/10.1145/3701551.3703574. doi:10.1145/3701551.3703574.

K. Thang, H. Hosseini, H. Fani, Translative neural team recommendation: From multilabel classification to sequence prediction, in: N. Ferro, M. Maistro, G. Pasi, O. Alonso, A. Trotman, S. Verberne (Eds.), Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2025, Padua, Italy, July 13-18, 2025, ACM, 2025, pp. 2800–2806. URL: https://doi.org/10.1145/3726302.3730259. doi:10.1145/3726302.3730259.

M. Kargar, A. An, M. Zihayat, Efficient bi-objective team formation in social networks, in: P. A. Flach, T. D. Bie, N. Cristianini (Eds.), Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part II, volume 7524 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 483–498. URL: https://doi.org/10.1007/978-3-642-33486-3_31. doi:10.1007/978-3-642-33486-3_31.

M. Zihayat, A. An, L. Golab, M. Kargar, J. Szlichta, Authority-based Team Discovery in Social Networks, in: EDBT '17, 2017, pp. 498–501. doi:10.5441/002/edbt.2017.54.

P. Keane, F. Ghaffar, D. Malone, Using machine learning to predict links and improve steiner tree solutions to team formation problems, in: H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, L. M. Rocha (Eds.), Complex Networks and Their Applications VIII - Volume 2 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019, Lisbon, Portugal, December 10-12, 2019, volume 882 of *Studies in Computational Intelligence*, Springer, 2019, pp. 995–1006. URL: https://doi.org/10.1007/978-3-030-36683-4_79. doi:10.1007/978-3-030-36683-4_79.

M. Kargar, L. Golab, D. Srivastava, J. Szlichta, M. Zihayat, Effective keyword search over weighted graphs, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 601–616. doi:10.1109/TKDE.2020.2985376.

T. Akiba, Y. Iwata, Y. Yoshida, Fast exact shortest-path distance queries on large networks by pruned landmark labeling, in: K. A. Ross, D. Srivastava, D. Papadias (Eds.), Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2013, New York, NY, USA, June 22-27, 2013, ACM, 2013, pp. 349–360. URL: https://doi.org/10.1145/2463676.2465315. doi:10.1145/2463676.2465315.

Y. Du, X. Meng, Y. Zhang, P. Lv, Gerf: A group event recommendation framework based on learning-to-rank, IEEE Trans. Knowl. Data Eng. 32 (2020) 674–687. doi:10.1109/TKDE.2019.2893361.

R. H. Rad, E. Bagheri, M. Kargar, D. Srivastava, J. Szlichta, Retrieving skill-based teams from collaboration networks, in: ACM SIGIR 2021, Virtual Event, Canada, July 11-15, 2021, 2021, pp. 2015–2019. doi:10.1145/3404835.3463105.

A. Sapienza, P. Goyal, E. Ferrara, Deep Neural Networks for Optimal Team Composition, Frontiers in Big Data 2 (2019) 1–13. doi:10.3389/fdata.2019.00014.

A. Dashti, K. Saxena, D. Patel, H. Fani, Opentf: A benchmark library for neural team formation, in: ACM CIKM 2022, Atlanta, GA, USA, October 17-21, 2022, 2022, pp. 3913–3917. doi:10.1145/3511808.3557590.

F. Hemmatizadeh, C. Wong, A. Yu, H. Fani, Lady: A benchmark toolkit for latent aspect detection enriched with backtranslation augmentation, in: ACM SIGIR 2024, Washington DC, USA, July 14-18, 2024, 2024, pp. 1172–1178. doi:10.1145/3626772.3657894.

M. Wang, H. Su, S. Wang, S. Wang, N. Yin, L. Shen, L. Lan, L. Yang, X. Cao, Graph convolutional mixture-of-experts learner network for long-tailed domain generalization, IEEE Trans. Circuits Syst. Video Technol. 35 (2025) 6936–6947. URL: https://doi.org/10.1109/TCSVT.2025.3532309. doi:10.1109/TCSVT.2025.3532309.

W. Wang, Y. Xu, F. Feng, X. Lin, X. He, T.-S. Chua, Diffusion recommender model, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 832–841. URL: https://doi.org/10.1145/3539618.3591663. doi:10.1145/3539618.3591663.

J. Lin, Y. Cao, Y. Yu, W. Zhang, Diffusion models for recommender systems: From content distribution to content creation, in: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, KDD '25, Association for Computing Machinery, New York, NY, USA, 2025, p. 6074–6085. URL: https://doi.org/10.1145/3711896.3736554. doi:10.1145/3711896.3736554.

J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: ICML, 2015, pp. 2256–2265.

J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: NeurIPS 2020, volume 33, 2020, pp. 6840–6851.

A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, L. Van Gool, Repaint: Inpainting using denoising diffusion probabilistic models, in: CVPR, 2022, pp. 11461–11471. doi:10.1109/CVPR52688.2022.01117.

A. Camuto, M. Willetts, S. Roberts, C. Holmes, T. Rainforth, Towards a theoretical understanding of the robustness of variational autoencoders, in: A. Banerjee, K. Fukumizu (Eds.), Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, volume 130 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 3565–3573. URL: https://proceedings.mlr.press/v130/camuto21a.html.

L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, M. Bennamoun, Hands-on bayesian neural networks—a tutorial for deep learning users, IEEE Computational Intelligence Magazine 17 (2022) 29–48. doi:10.1109/MCI.2022.3155327.

D. P. Kingma, M. Welling, Auto-encoding variational bayes, CoRR abs/1312.6114 (2013).

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, in: ICLR, 2021.

P. Dhariwal, A. Nichol, Diffusion models beat gans on image synthesis, in: NeurIPS 2021, NIPS '21, Curran Associates Inc., Red Hook, NY, USA, 2021.

Z. Zhou, C. Wang, H. Ye, Y. Guan, T. Yu, Incomplete data, complete dynamics: A diffusion approach, arXiv preprint arXiv:2509.20098 (2025).

H. Huang, K. Han, B. Xu, T. Gan, Reconstructing diffusion networks from incomplete data., in: IJCAI, volume 2022, 2022, pp. 3085–3091.

Y. Ouyang, L. Xie, C. Li, G. Cheng, Missdiff: Training diffusion models on tabular data with missing values, arXiv preprint arXiv:2307.00467 (2023).

J. Choi, S. Kim, Y. Jeong, Y. Gwon, S. Yoon, Ilvr: Conditioning method for denoising diffusion probabilistic models, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 14367–14376. doi:10.1109/ICCV48922.2021.01410.

C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, S. Ermon, Sdedit: Guided image synthesis and editing with stochastic differential equations, in: ICLR, 2022.

A. Q. Nichol, P. Dhariwal, Improved denoising diffusion probabilistic models, in: International conference on machine learning, PMLR, 2021, pp. 8162–8171.

O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: MICCAI 2015, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, NeurIPS 30 (2017).

A. Khan, L. Golab, M. Kargar, J. Szlichta, M. Zihayat, Compact group discovery in attributed graphs and social networks, Inf. Process. Manag. 57 (2020) 102054. doi:10.1016/j.ipm.2019.102054.

M. Kargar, A. An, Keyword search in graphs: Finding r-cliques, Proc. VLDB Endow. 4 (2011) 681–692. doi:10.14778/2021017.2021025.

R. H. Rad, E. Bagheri, M. Kargar, D. Srivastava, J. Szlichta, Subgraph representation learning for team mining, in: ACM WebSci '22, Barcelona, Spain, June 26 - 29, 2022, 2022, pp. 148–153. doi:10.1145/3501247.3531578.

T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, in: Y. Bengio, Y. LeCun (Eds.), 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings, 2013. URL: http://arxiv.org/abs/1301.3781.

E. B. Khalil, H. Dai, Y. Zhang, B. Dilkina, L. Song, Learning combinatorial optimization algorithms over graphs, in: NeurIPS 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, pp. 6348–6358.

L. Zhang, R. Hamidi Rad, M. Zihayat, E. Bagheri, Say the task, build the team: Prompt-based team formation, in: Proceedings of the 17th International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2025), Lecture Notes in Computer Science, Niagara Falls, ON, Canada, 2025.

M. R. I. Rabin, A. Hussain, M. A. Alipour, V. J. Hellendoorn, Memorization and generalization in neural code intelligence models, Inf. Softw. Technol. 153 (2023) 107066. doi:10.1016/j.infsof.2022.107066.

Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on Machine Learning, 2014.

Z. Mi, K.-C. Wang, G. Qian, H. Ye, R. Liu, S. Tulyakov, K. Aberman, D. Xu, I think, therefore i diffuse: Enabling multimodal in-context reasoning in diffusion models, in: Forty-second International Conference on Machine Learning, 2025. URL: https://openreview.net/forum?id=2v91xhNdsz.

S. Batra, G. S. Sukhatme, Zero shot generalization of vision-based rl without data augmentation, in: Forty-second International Conference on Machine Learning, 2025. URL: https://openreview.net/forum?id=kQ42hGjrKn.