Translative Neural Team Recommendation: From Multilabel Classification to Sequence Prediction

Kap Thang
University of Windsor
Windsor, ON, Canada
thangk@uwindsor.ca

Hawre Hosseini
Five9
San Francisco, CA, USA
hawre.hosseini@five9.com

Hossein Fani
University of Windsor
Windsor, ON, Canada
hfani@uwindsor.ca

Abstract

Neural team recommendation has achieved state-of-the-art performance in forming teams of experts whose success in completing complex tasks is *almost surely* guaranteed. The proposed models frame the problem as a Boolean multilabel classification, mapping a subset of required skills, whose dense vector representations are transferred from expert collaboration graphs by graph neural networks, to the sparse occurrence (multi-hot) vector representation of an optimum subset of experts using multilayer feedforward neural networks. Such approaches, however, suffer from the curse of sparsity in the high-dimensional vector of optimum experts in the output layer. In this paper, we propose to reformulate the team recommendation problem into a sequence prediction task and leverage sequence-to-sequence models, including transformers, to map an input sequence of the required subset of skills onto an output sequence of the optimum subset of experts. Our experiments on four large-scale datasets from various domains, with distinct distributions of skills in teams, show that the sequence-to-sequence approach is consistently superior overall in a host of classification and information retrieval metrics. Our codebase is available at `https://github.com/fani-lab/OpeNTF/tree/nmt`.

1 Introduction

As modern projects have been surpassing the capacity of individuals, collaborative teams of experts have become vital in today's diverse landscape across academia [32, 60, 75], industry [1, 19, 35], law [6, 31], freelancing [24], and healthcare [61, 63], and the success of projects hinges on the effectiveness of teams. Assembling an effective team can be seen as social information retrieval (Social IR), where the right group of experts, rather than relevant information, is desired to accomplish a task at hand [29, 30]; a tedious, error-prone, and suboptimal process should it be manual, as it is predisposed to hidden personal and societal biases [50], falls short for an overwhelming number of experts, and fails to consider a multitude of criteria to optimize simultaneously [3]. Therefore, a rich body of computational methods, from operations research [4, 18, 21, 36, 66, 72, 73, 79] social network analysis [25, 39, 64] and more recently, machine learning [10, 22, 54, 56], have been proposed for team recommendation, also known as team allocation, team selection, team composition, and team configuration. Among such methods, neural models have brought state-of-the-art efficacy and efficiency due to the iterative and online learning procedure, and availability of training datasets.

By and large, proposed neural models frame the team recommendation problem as a *multilabel* Boolean classification task, learning the distributions of experts and their skill sets in the context of teams in the past to draw a subset of experts whose history of

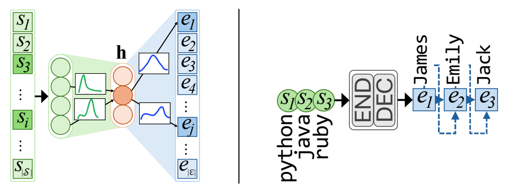

Figure 1: Multilabel [55] vs. sequence-to-sequence neural team recommendation.

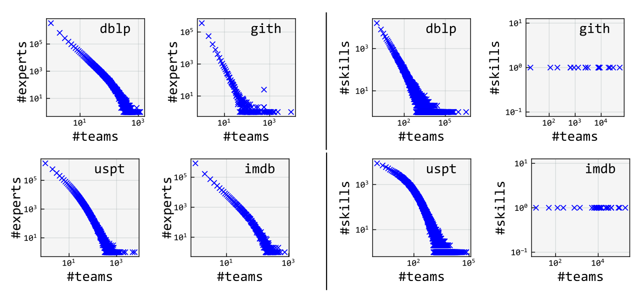

Figure 2: Distribution of teams over experts (left) and skills (right) for `dblp`, `uspt`, `imdb`, and `gith` datasets.

collaborations is statistically more likely successful. As seen in Figure 1 (left), they map a dense low-dimensional vector representation (embeddings) of a required subset of skills onto the output layer, which is an occurrence (multi-hot) vector representation of a successful (optimum) team. In the output layer, each expert is mapped to a label and would be recommended if their class's prediction probability is close to 1 [54, 55]. Such models, however, suffer from the curse of sparsity in the output layer; due to the large number of labels (experts), neural classifiers, which learn their parameters based on average loss values over all labels, may underfit as the average loss becomes near `0` over a very large number of experts. While researchers have tried weighted cross-entropy [56] and negative sampling heuristics [10] to fill the gap, such models still suffer from the lack of sufficient efficacy.

In this paper, we propose a sequence-to-sequence approach to the team recommendation problem. Unlike existing approaches, which first learn skill embeddings and then transfer them to a neural multilabel classifier, we reformulate the problem into an end-to-end sequence prediction task between pairs of ⟨sequence of required skills → optimum sequence of experts⟩ and employ recurrent and transformer-based encoder-decoder neural models, as seen in Figure 1 (right). Sequence-to-sequence approaches have gained significant traction for their efficacy not only in natural language tasks, but also in recommendation systems for sequence modelling like `transformer4rec` [13], `sas-rec` [38], and `bert4rec` [67].

Despite the similarity, applying the sequence-to-sequence approach for team recommendation withholds its own unique challenges: (1) training datasets have their own unique skill and expert sets, and therefore, the pretaining or finetuning approach, which is successful for tasks like product recommendation, has limited applications for team recommendations; (2) Moreover, a team recommender works with two distinct sets of skills and experts where their distributions over teams are highly domain-dependent. As seen in Figure 2 (left), the distributions of teams over experts in all datasets are long-tailed as many experts (researchers in `dblp`, developers in `gith`, inventors in `uspt`, and cast and crew in `imdb`) have participated in very few teams (papers in `dblp`, software repositories in `gith`, patents in `uspt`, and movies in `imdb`). However, with respect to the set of skills, while `dblp` and `uspt` suffer further from the long-tailed distribution of skills in teams, `gith` and `imdb` follow a more fair distribution, as shown in Figure 2 (right). Specifically, `gith` and `imdb` have a limited variety of skills (programming languages in `gith` and genres in `imdb`), which are employed by many teams. To our knowledge, no work has explored the sequence-to-sequence approach for the team recommendation task in the context of such challenges before our study.

2 Related Works

The work directly related to this paper can be broadly categorized into neural sequence modelling and neural team recommendation.

2.1 Neural Sequence Modelling

There has been extensive work on sequence-to-sequence models and transformers across nlp tasks [2, 14, 34, 48, 58, 70, 70, 76], speech recognition [7, 9, 15, 26], and computer vision [5, 17, 49, 53]. Herein, however, we focus on sequential recommendation systems [38, 67, 69, 78] for their direct relevance in team recommendation.

Hidasi et al. [27] were among the first to apply recurrent neural networks for sequence-based recommendations, stacking `gru` units with learning-to-rank losses to leverage user-item sequences where collaborative filtering fails. Similarly, Wu et al. [74] used `lstm` with `softmax` loss to capture temporal patterns. Tang et al. [69] applied convolution layers—horizontal for temporal patterns and vertical for item features—to produce effective user embeddings while addressing efficiency and scalability [38, 69]. A breakthrough came with Kang et al. [38] introducing `sas-rec`, a transformer-based model with left-to-right self-attention and point-wise ranking loss. Later, Sun et al. [67] proposed `bert4rec`, employing bidirectional self-attention with a point-wise ranking loss on item sequences. These transformers are now optimized and released by nvidia's `transformers4rec` [12, 13, 28, 38, 67] for product recommendations at scale. Despite extensive research and industrial successes of sequence modelling for user-item recommendation, its application for team recommendation is yet to be studied, and to our knowledge, we are the first to bridge this gap.

2.2 Neural Team Recommendation

Team recommendation has long attracted both social and computer science researchers. For instance, Paris et al. [52] and Stokols et al. [65] studied team interactions to cultivate effective teams from a *non*-computational perspective. Here, we focus on computational approaches, which fall into three categories: (1) search-based methods using operation research methods [4, 18, 21, 36, 66, 72, 73, 79] or expert network analysis [25, 39, 64], (2) reinforcement-based methods [20, 46, 80], and (3) learning-based methods [22, 41, 54–56, 62]. While search- and reinforcement-based techniques are theoretically sound, they often struggle with scalability, making learning-based approaches preferable. Within the learning-based category, simple feedforward networks [56] were used initially, which were later improved with variational Bayesian networks [11, 55, 56] to address popularity bias via Gaussian uncertainty. Dashti et al. [10] further enhanced performance by using negative sampling to mitigate the dominance of popular experts. However, these methods assume independent expert selection, overlooking team dynamics. To capture collaborative relationships, graph neural network-based methods emerged. Rad et al. [54] incorporated expert collaboration graphs with `metapath2vec` [16] for skill embeddings, and Kaw et al. [41] used deep graph infomax [71] with graph convolution networks with attention layers to improve upon skill embeddings. Despite these advances, current neural methods treat team recommendation as a multilabel classification problem, assuming experts can be selected independently, and hence fail to capture real-world team dynamics. Moreover, such models rely on high-dimensional multi-hot output representations for experts, leading to computational inefficiencies for a large pool of experts.

3 Problem Definition

Given a set of skills $\mathcal{S} = \{s_i\}$ and a set of experts $\mathcal{E} = \{e_j\}$, a team is a tuple $(\mathbf{s}, \mathbf{e})$ where an ordered list of experts $\mathbf{e} \subseteq \mathcal{E}$ collectively cover an ordered list of required skills $\mathbf{s} \subseteq \mathcal{S}$ to accomplish a task at hand. Further, $\mathcal{T} = \{(\mathbf{s}, \mathbf{e})_k\}$ indexes all instances of successful teams. For a given set of skills $\mathbf{s}$, the team recommendation problem aims at identifying an optimal subset of experts $\mathbf{e}$ such that their collaboration in the predicted team is successful. More concretely, the team recommendation problem is to learn a mapping function f of parameters $\boldsymbol{\theta}$ such that $\forall(\mathbf{s}, \mathbf{e}) \in \mathcal{T}; f_{\boldsymbol{\theta}}(\mathbf{s}) = \mathbf{e}$.

4 Proposed Approach

We propose to transform the team recommendation task into a sequence-to-sequence modelling task, mapping a dynamic-length input sequence of required skills onto a dynamic-length output sequence of predicted experts while leveraging the autoregression and global attention mechanisms, which capture dependencies *beyond* independent expert probabilities in multilabel classification.

We estimate $f_\theta(\mathbf{s})$ on a parallel dataset whose pairs of sequences are pairs of $(\mathbf{s}, \mathbf{e}) \in \mathcal{T}$, transforming the ordered list of required skills $\mathbf{s} = [s_i, \ldots, s_n]$ into an optimum ordered list of experts $\mathbf{e} = [e_j, \ldots, e_m]$. We then employ a sequence-to-sequence encoder-decoder neural architecture [8, 33, 37, 51, 68] to maximize the conditional probability $p(\mathbf{e}|\mathbf{s})$ to learn $f_{\boldsymbol{\theta}}$. The encoder maps the *sequence* of skills $[s_i, \ldots, s_n]$ onto $\mathbf{h}_n$ and the decoder generates the sequence of experts $[e_j, \ldots, e_m]$ from the $\mathbf{h}_n$, one expert at a time, decomposing the conditional probability $p(\mathbf{e}|\mathbf{s})$ as $\prod_{j=1} p(e_j|e_{<j}, \mathbf{s})$ and seeking the maximum probability among subsets of experts as an optimum team for $\mathbf{s}$, i.e., $f_\theta(\mathbf{s}) = \mathbf{e}$. The probability of generating

Table 1: Statistics of the raw and preprocessed datasets as well as mapping of raw data properties to instances of team.

	dblp		uspt		imdb		gith	
teams $\mathcal{T}$	publications		patents		movies		software repos	
experts $\mathcal{E}$	authors		inventors		cast & crew		developers	
skills $\mathcal{S}$	keywords		subclasses		(sub) genres		prog. lang.	
success	published		issued		produced		released	
statistics	raw	filtered	raw	filtered	raw	filtered	raw	filtered
$\|\mathcal{T}\|$	4.9M	99K	7.1M	152K	507K	32K	133K	46K
$\|\mathcal{E}\|$	5.0M	14K	3.5M	13K	877K	2.0K	453K	1.2K
$\|\mathcal{S}\|$	90K	30K	242K	67K	28	23	20	20
#teams w/ one expert	769K	**0**	2.6M	**0**	323K	**0**	**0**	**0**
avg. #experts/team	3.06	3.29	2.51	3.79	1.88	3.98	5.52	7.53
avg. #skills/team	8.57	9.71	6.29	9.97	1.54	1.76	1.37	1.57

an expert at the decoder can be conditioned not only on $\mathbf{h}_n$ but also on all $\mathbf{h}_{<n}$ at the encoder, enabling the decoder to *attend* to all skills in the input sequence selectively [51]. To reduce the computational complexity at the encoder and the decoder, a sequence-to-sequence model may have *no* recurrent connections, like in transformers [70], enabling parallel calculation of $\mathbf{h}_{<n}$ at the encoder and $\mathbf{h}_{>=n}$ at the decoder, an architecture that yielded promising performance on machine translation and led to extensive research on sequence-to-sequence modelling [14, 59, 77].

5 Experiment

In this section, we present the details of our experiments toward addressing the following research questions:
RQ1. Does the sequence-to-sequence approach yield performance improvements over existing multilabel neural team recommenders?
RQ2. Which sequence-to-sequence model performs the best (worst) for team recommendation?
RQ3. How well does the sequence-to-sequence approach generalize across different domains in team recommendation?

5.1 Datasets

Our testbed includes four benchmark datasets in team recommendation literature: dblp [22, 41, 45, 47, 54–57] and uspt [22, 42], following similar long-tailed distributions for *both* experts and skills over teams, on the one hand, and imdb [10, 22, 43] and gith [22, 39, 40], following long-tailed distribution of experts but uniform distribution of skills over teams. Each dataset was preprocessed to ensure a team consisted of more than 3 experts, and each expert participated in at least 75 teams. Table 1 shows the mapping of raw data properties to the team $\mathcal{T}$, skill $\mathcal{S}$, and expert $\mathcal{E}$ sets, along with a summary of statistics.

5.2 Baselines

We compare two categories of baselines: (1) existing neural team recommenders including variational Bayesian feedforward neural network with multi-hot (bnn) [56] and dense (bnn_emb) [55] vector representation in the input layer; (2) sequence-to-sequence models including recurrent recommender network (rrn) [74], vanilla recurrent neural network with attention (rnn-att) [81], convolutional sequence-to-sequence (convs2s) [23], and the transformer [70].

Table 2: Hyperparameters and running settings for models.

	transformer	convs2s	rnn-att	rrn,bnn bnn_emb
batch size	128	8^+, 128	128	128
learning rate	- - - - - - -	Vaswani et al. [70]	- - - - - - - -	0.1
epochs	20	1^+, 20	20	20
optimizer	- - - - - - - - - - - - -	Adam	- - - - - - - - - - - - -	
hidden layer size	512	128	128, 512*	128
hidden activation	relu, softmax	glu	tanh, sigmoid	relu
output layer	128	128	128	$\|\mathcal{E}\|$
output activation	- - - - - - - - -	softmax	- - - - - - - - -	sigmoid

$^+$: convs2s model setting for uspt dataset.

The transformer recommends experts through its parallel self-attention mechanism, where each selection considers both the entire skill sequence and all previously predicted experts simultaneously, capturing dependencies regardless of their sequence position. The convs2s model, on the other hand, processes skill-expert relationships through stacked convolutional layers that create hierarchical representations. Lower layers capture local skill-expert matches, while deeper layers learn broader team composition patterns, capturing relationships between distant elements in the sequence efficiently. The rnn-att learns the expert selection sequentially, and in the hidden state, the expert choice is updated using attention to focus on which expert is relevant to which skill. The rrn model is an item-user recommender system that assumes dynamic (temporal) embeddings for users and items to capture behavioral trajectories using lstm for better prediction accuracy. We used opennmt-py [44] for implementation of sequence-to-sequence models except for rrn whose code has kindly been provided by its authors [74]. Table 2 summarizes the models' hyperparameters and running settings.

5.3 Evaluation

We randomly select 15% of teams for the test set and perform 3-fold cross-validation on the remaining teams for model training over 20 epochs for all the models in all datasets except for the convs2s in uspt dataset with 1 epoch due to intractable time complexity, which results in one trained model per fold. Given a team $(\mathbf{s}, \mathbf{e})$ from the test set, we compare the sequence of experts $\mathbf{e}'$, predicted by the model of each fold, with the observed subset of experts $\mathbf{e}$ and report the average performance of models on all folds in terms of classification metrics including precision, recall, as well as information retrieval metrics including normalized discounted cumulative gain (ndcg) and mean average precision (map) at $k \in \{2,5,10\}$ first generated sequence of experts. The final results are obtained by averaging the performance metrics across all folds for a robust evaluation of the model's predictive capabilities while minimizing fold-specific variation in the data distribution.

5.4 Results

Foremost, we acknowledge that all models achieve low values of evaluation metrics for practical applications of team recommendation, which is primarily due to the simplicity of the neural model architectures and the small number of training epochs given the intensive computational demands for such methods coupled with our limited computational resources; metric values are reported in % for ease of readability and comparison. Our main goal is to showcase the optimum solution setting, sequence-to-sequence vs. multilabel classification for team recommendation.

In response to **RQ1**, i.e., whether sequence-to-sequence models yield better performance vs. feedforward models, the results from Table 3 demonstrate that *all* sequence-to-sequence models outperform the feedforward ones across metrics and datasets. This can be attributed to conditioning of expert recommendation on the previously recommended experts in the output sequence, allowing for more robust prediction compared to the inherent sparse activations in the output layer of feedforward models. Notably, sequence-to-sequence models' relative improvements have been as high as `82x` in some datasets, suggesting the right track for the most suitable neural architecture for team recommendation. The only exception is `rrn`, which performs on par or even poorer than the feedforward baselines and will be discussed more in **RQ2**.

In response to **RQ2**, looking into the sequence-to-sequence models for the best (worst) performance, from Table 3, the `transformer` consistently outperforms other models across *most* datasets and metrics. This is because the `transformer` benefits from (1) its self-attention mechanism, which allows capturing the relevance of skills independently and simultaneously, (2) the encoder-decoder structure effectively captures the input skill domain and the output expert domain regardless of position in the sequence due to (1), and (3) the parallel processing. The runner-up, however, depends on the underlying distribution of skills over teams (the input sequence) in a dataset. In `dblp` and `uspt`, `convs2s` is the runner-up, whereas in `imdb` and `gith`, `rnn-att` is the second-best. The `convs2s` model's performance is particularly strong on large datasets like `uspt` with long-tailed distributions of skills and experts over teams. The `rnn-att` model's performance is affected by the dataset's distribution shape, struggling with long-tailed distributions but performing well on uniformly distributed skills like in `imdb`. Lastly, the worst sequence-to-sequence model is `rrn`, as mentioned in **RQ1**, which falls short of consistently outperforming compared to feedforward baselines across datasets and metrics. Specifically, `rrn` uses `lstm` without attention mechanism, which means all relevant skills must be in a fixed-size hidden state to be used as historical information.

To answer **RQ3**, i.e., whether the outperformance of sequence-to-sequence models can generalize to various domains, the results from Table 3 demonstrate superior performance compared to the baselines in all four datasets across metrics. However, the extent of improvement depends on the underlying dataset, esp., the distribution of skills over teams. In `imdb` and `gith`, where the skills are distributed uniformly in the input sequence, the improvement extends to `5x` on average across metrics. In contrast, when the skills are distributed in a long-tailed fashion, the improvement increases to about `82x`. For both types of distributions of skills, the improvement is at least `2x`. This showcases how well the sequence-to-sequence models generalize to various domains over the baselines.

6 Concluding Remarks and Future Work

In this paper, we studied the team recommendation problem as a sequence prediction task through sequence-to-sequence neural architectures. The results show that these architectures yield performance improvements over existing feedforward models, with the `transformer` model consistently outperforming other models across all datasets and metrics. The findings also highlight the importance of considering the distribution of skills over teams,

Table 3: Comparative results of multilabel vs. sequence-to-sequence neural team recommendation methods. Bold and underlined values indicate the best and second-best.

	k	transformer	convs2s	rnn-att	rrn	bnn	bnn_emb
				dblp			
%precision	2	**10.4119**	2.4998	3.6176	0.0570	0.0570	0.1124
	5	**7.0113**	1.6122	2.3581	0.0391	0.0663	0.1290
	10	**3.5392**	0.8242	1.1992	0.0472	0.0710	0.1251
%recall	2	**6.3457**	1.5071	2.1698	0.0380	0.0351	0.0668
	5	**10.5477**	2.4177	3.5115	0.0630	0.0993	0.1909
	10	**10.6397**	2.4760	3.5753	0.1552	0.2118	0.3699
%ndcg	2	**10.3611**	2.4770	3.5822	0.0478	0.0538	0.1083
	5	**10.4597**	2.4276	3.5184	0.0523	0.0806	0.1555
	10	**10.4824**	2.4487	3.5391	0.0959	0.1330	0.2397
%map	2	**5.9463**	1.3554	1.9412	0.0217	0.0242	0.0474
	5	**9.2909**	2.0008	2.8791	0.0281	0.0411	0.0792
	10	**9.3210**	2.0127	2.8930	0.0446	0.0558	0.1033
				uspt			
%precision	2	**41.7289**	28.5717	23.9729	0.0239	0.0657	0.3663
	5	**31.0677**	24.6530	17.7873	0.0383	0.0769	0.4123
	10	**16.5169**	15.2382	9.4717	0.0654	0.0910	0.3748
%recall	2	**23.1038**	13.9104	12.9871	0.0140	0.0353	0.1608
	5	**41.1643**	28.8167	23.0358	0.0500	0.0976	0.4509
	10	**42.6086**	33.7595	23.8896	0.1370	0.2212	0.8141
%ndcg	2	**41.6095**	28.3606	23.9146	0.0221	0.0655	0.3652
	5	**42.0309**	30.0325	23.8227	0.0408	0.0883	0.4531
	10	**42.1435**	31.4137	23.8270	0.0868	0.1481	0.6094
%map	2	**22.4053**	13.0305	12.4784	0.0096	0.0266	0.1212
	5	**38.6272**	24.4598	21.3567	0.0186	0.0433	0.2027
	10	**39.7591**	27.1994	22.0112	0.0340	0.0592	0.2583
				imdb			
%precision	2	1.5454	1.6097	**1.6985**	0.0000	0.2128	0.4255
	5	1.4574	1.4552	**1.4804**	0.8511	0.5106	0.5106
	10	**0.9035**	0.8998	0.9027	0.8511	0.4255	0.6383
%recall	2	0.7669	0.7952	**0.8193**	0.0000	0.1418	0.2837
	5	**1.8093**	1.8013	1.8043	1.4184	0.8511	0.8511
	10	**2.2085**	2.1926	2.1792	2.8369	1.3050	1.9574
%ndcg	2	1.5479	1.6173	**1.7003**	0.0000	0.1646	0.3292
	5	1.7364	1.7595	**1.7883**	0.8163	0.5699	0.5923
	10	1.9039	1.9222	**1.9333**	1.4606	0.7848	1.1358
%map	2	0.6172	0.6506	**0.6650**	0.0000	0.0709	0.1418
	5	1.0327	**1.0487**	1.0450	0.3191	0.2600	0.2813
	10	1.0914	**1.1041**	1.0975	0.6265	0.3148	0.4389
				gith			
%precision	2	**32.1596**	25.0590	29.7008	0.0000	3.0693	7.3267
	5	**21.6055**	16.9509	20.1806	0.1980	2.8515	4.7129
	10	**12.7104**	9.9503	12.0029	0.0990	2.6931	3.3861
%recall	2	**13.8543**	11.0787	12.8103	0.0000	1.2164	3.5441
	5	**22.2914**	18.0735	20.7963	0.0619	2.8846	5.1580
	10	**24.0868**	19.4837	22.6186	0.0619	5.1174	6.1885
%ndcg	2	**32.4291**	25.3569	29.7647	0.0000	3.1365	6.4753
	5	**28.2538**	22.3664	26.1975	0.1679	3.2893	5.8418
	10	**26.9900**	21.5849	25.1263	0.1090	4.2340	6.2665
%map	2	**12.9552**	10.5164	11.9759	0.0000	1.0104	2.3424
	5	**19.3215**	15.5615	17.9982	0.0206	1.5706	3.0822
	10	**20.7984**	16.6428	19.4844	0.0206	2.1633	3.3837

as different sequence-to-sequence models exhibit strengths in different datasets and distribution types. Our future work includes investigating sequence-2-sequence models in the presence of additional contextual factors, such as geolocation (geo-aware team recommendation). Further, the sequence-to-sequence transformer-based approach allows us to study the order of skills or experts based on any prioritization criteria, such as their importance to a project, which is not accessible to existing neural approaches. We also plan to develop a framework that enables dynamic team formation and expert removal, allowing for real-time expert prediction, removal, or substitution—a capability that is uniquely enabled by the sequence-to-sequence approach and not feasible with existing feedforward neural models.

References

[1] Gholamreza Askari, Nader Asghri, Madjid Eshaghi Gordji, Heshmatolah Asgari, José Filipe, and Adel Azar. 2020. The Impact of Teamwork on an Organization's Performance: A Cooperative Game's Approach. *Mathematics* 8 (10 2020), 1–15. https://doi.org/10.3390/math8101804

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1409.0473

[3] Rodrigo Borges, Otto Sahlgren, Sami Koivunen, Kostas Stefanidis, Thomas Olsson, and Arto Laitinen. 2023. Multi-Objective Fairness in Team Assembly. In *New Trends in Database and Information Systems - ADBIS 2023 Short Papers, Doctoral Consortium and Workshops: AIDMA, DOING, K-Gals, MADEISD, PeRS, Barcelona, Spain, September 4-7, 2023, Proceedings (Communications in Computer and Information Science, Vol. 1850)*, Alberto Abelló, Panos Vassiliadis, Oscar Romero, Robert Wrembel, Francesca Bugiotti, Johann Gamper, Genoveva Vargas-Solar, and Ester Zumpano (Eds.). Springer, 106–116. https://doi.org/10.1007/978-3-031-42941-5_10

[4] Manoel B. Campêlo, Tatiane Fernandes Figueiredo, and Ana Silva. 2020. The sociotechnical teams formation problem: a mathematical optimization approach. *Ann. Oper. Res.* 286, 1 (2020), 201–216. https://doi.org/10.1007/S10479-018-2759-5

[5] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-End Object Detection with Transformers. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12346)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer, 213–229. https://doi.org/10.1007/978-3-030-58452-8_13

[6] Elizabeth Chambliss. 2010. Measuring law firm culture. *Social Science Research Network* 52 (2010). https://api.semanticscholar.org/CorpusID:151077727

[7] Chung-Cheng Chiu, Tara N. Sainath, Yonghui Wu, Rohit Prabhavalkar, Patrick Nguyen, Zhifeng Chen, Anjuli Kannan, Ron J. Weiss, Kanishka Rao, Ekaterina Gonina, Navdeep Jaitly, Bo Li, Jan Chorowski, and Michiel Bacchiani. 2018. State-of-the-Art Speech Recognition with Sequence-to-Sequence Models. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*. IEEE, 4774–4778. https://doi.org/10.1109/ICASSP.2018.8462105

[8] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). ACL, 1724–1734. https://doi.org/10.3115/V1/D14-1179

[9] Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. 2015. Attention-Based Models for Speech Recognition. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett (Eds.). 577–585. https://proceedings.neurips.cc/paper/2015/hash/1068c6e4c8051cfd4e9ea8072e3189e2-Abstract.html

[10] Arman Dashti, Saeed Samet, and Hossein Fani. 2022. Effective Neural Team Formation via Negative Samples. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, Mohammad Al Hasan and Li Xiong (Eds.). ACM, 3908–3912. https://doi.org/10.1145/3511808.3557590

[11] Arman Dashti, Karan Saxena, Dhwani Patel, and Hossein Fani. 2022. OpeNTF: A Benchmark Library for Neural Team Formation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, Mohammad Al Hasan and Li Xiong (Eds.). ACM, 3913–3917. https://doi.org/10.1145/3511808.3557526

[12] Gabriel de Souza Pereira Moreira, Sara Rabhi, Ronay Ak, Md Yasin Kabir, and Even Oldridge. 2021. Transformers with multi-modal features and post-fusion context for e-commerce session-based recommendation. *CoRR* abs/2107.05124 (2021). arXiv:2107.05124 https://arxiv.org/abs/2107.05124

[13] Gabriel de Souza Pereira Moreira, Sara Rabhi, Jeongmin Lee, Ronay Ak, and Even Oldridge. 2021. Transformers4Rec: Bridging the Gap between NLP and Sequential / Session-Based Recommendation. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampín, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 143–153. https://doi.org/10.1145/3460231.3474255

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*

2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/V1/N19-1423

[15] Linhao Dong, Shuang Xu, and Bo Xu. 2018. Speech-Transformer: A No-Recurrence Sequence-to-Sequence Model for Speech Recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*. IEEE, 5884–5888. https://doi.org/10.1109/ICASSP.2018.8462506

[16] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 135–144. https://doi.org/10.1145/3097983.3098036

[17] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net. https://openreview.net/forum?id=YicbFdNTTy

[18] Edmund H. Durfee, James C. Boerkoel Jr., and Jason Sleight. 2014. Using hybrid scheduling for the semi-autonomous formation of expert teams. *Future Gener. Comput. Syst.* 31 (2014), 200–212. https://doi.org/10.1016/J.FUTURE.2013.04.008

[19] Amy Edmondson and Jean-François Harvey. 2017. Cross-boundary teaming for innovation: Integrating research on teams and knowledge in organizations. *Human Resource Management Review* 28 (03 2017). https://doi.org/10.1016/j.hrmr.2017.03.002

[20] Benjamin Ellis, Jonathan Cook, Skander Moalla, Mikayel Samvelyan, Mingfei Sun, Anuj Mahajan, Jakob N. Foerster, and Shimon Whiteson. 2023. SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (Eds.). http://papers.nips.cc/paper_files/paper/2023/hash/764c18ad230f9e7bf6a77ffc2312c55e-Abstract-Datasets_and_Benchmarks.html

[21] José G. M. Esgario, Iago E. da Silva, and Renato A. Krohling. 2019. Application of Genetic Algorithms to the Multiple Team Formation Problem. *CoRR* abs/1903.03523 (2019). arXiv:1903.03523 http://arxiv.org/abs/1903.03523

[22] Hossein Fani, Reza Barzegar, Arman Dashti, and Mahdis Saeedi. 2024. A Streaming Approach to 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 (Lecture Notes in Computer Science, Vol. 14608)*, Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis (Eds.). Springer, 325–340. https://doi.org/10.1007/978-3-031-56027-9_20

[23] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. 2017. Convolutional Sequence to Sequence Learning. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017 (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, 1243–1252. http://proceedings.mlr.press/v70/gehring17a.html

[24] Timothy Golden and Kristen Shockley. 2015. How Effective Is Telecommuting? Assessing the Status of Our Scientific Findings. *Psychological Science in the Public Interest* 16 (10 2015), 40–68. https://doi.org/10.1177/1529100615593273

[25] Kiarash Golzadeh, Lukasz Golab, and Jaroslaw Szlichta. 2024. Explaining Expert Search and Team Formation Systems with ExES. *CoRR* abs/2405.12881 (2024). https://doi.org/10.48550/ARXIV.2405.12881 arXiv:2405.12881

[26] Awni Y. Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, and Andrew Y. Ng. 2014. Deep Speech: Scaling up end-to-end speech recognition. *CoRR* abs/1412.5567 (2014). arXiv:1412.5567 http://arxiv.org/abs/1412.5567

[27] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent Neural Networks with Top-k Gains for Session-based Recommendations. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, Alfredo Cuzzocrea, James Allan, Norman W. Paton, Divesh Srivastava, Rakesh Agrawal, Andrei Z. Broder, Mohammed J. Zaki, K. Selçuk Candan, Alexandros Labrinidis, Assaf Schuster, and Haixun Wang (Eds.). ACM, 843–852. https://doi.org/10.1145/3269206.3271761

[28] Trong Dang Huu Ho and Sang Thi Thanh Nguyen. 2024. Self-Attentive Sequential Recommendation Models Enriched with More Features. In *Proceedings of the 2024 8th International Conference on Deep Learning Technologies (ICDLT '24)*. Association for Computing Machinery, New York, NY, USA, 49–55. https://doi.org/10.1145/3695719.3695727

[29] Damon Horowitz and Sepandar D. Kamvar. 2010. The anatomy of a large-scale social search engine. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti (Eds.). ACM, 431–440. https://doi.org/10.1145/1772690.1772735

[30] Damon Horowitz and Sepandar D. Kamvar. 2012. Searching the village: models and methods for social search. *Commun. ACM* 55, 4 (2012), 111–118. https://doi.org/10.1145/2133806.2133830

[31] Jia Hu and Robert Liden. 2014. Making a Difference in the Teamwork: Linking Team Prosocial Motivation to Team Processes and Effectiveness. *Academy of Management Journal* 58 (01 2014). https://doi.org/10.5465/amj.2012.1142

[32] Ying Huang, Xiaoting Liu, Ruinan Li, and Lin Zhang. 2023. The science of team science (SciTS): An emerging and evolving field of interdisciplinary collaboration. *El Profesional de la información* (2023). https://api.semanticscholar.org/CorpusID:257296123

[33] Sébastien Jean, KyungHyun Cho, Roland Memisevic, and Yoshua Bengio. 2015. On Using Very Large Target Vocabulary for Neural Machine Translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*. The Association for Computer Linguistics, 1–10. https://doi.org/10.3115/V1/P15-1001

[34] Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda B. Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *CoRR* abs/1611.04558 (2016). arXiv:1611.04558 http://arxiv.org/abs/1611.04558

[35] Almagul Kairgalievna and Nurul Mohammad Zayed. 2021. THE EFFECT OF TEAMWORK ON EMPLOYEE PRODUCTIVITY. https://api.semanticscholar.org/CorpusID:237007514

[36] Sunny Joseph Kalayathankal, John T Abraham, and Joseph Varghese Kureethara. 2019. A Fuzzy Approach To Project Team Selection. *International Journal of Scientific Technology Research* 8 (2019).

[37] Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent Continuous Translation Models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*. ACL, 1700–1709. https://aclanthology.org/D13-1176/

[38] Wang-Cheng Kang and Julian J. McAuley. 2018. Self-Attentive Sequential Recommendation. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*. IEEE Computer Society, 197–206. https://doi.org/10.1109/ICDM.2018.00035

[39] Mehdi Kargar and Aijun An. 2011. Discovering top-k teams of experts with/without a leader in social networks. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011*, Craig Macdonald, Iadh Ounis, and Ian Ruthven (Eds.). ACM, 985–994. https://doi.org/10.1145/2063576.2063718

[40] Mehdi Kargar, Lukasz Golab, Divesh Srivastava, Jaroslaw Szlichta, and Morteza Zihayat. 2022. Effective Keyword Search Over Weighted Graphs. *IEEE Trans. Knowl. Data Eng.* 34, 2 (2022), 601–616. https://doi.org/10.1109/TKDE.2020.2985376

[41] Sagar Kaw, Ziad Kobti, and Kalyani Selvarajah. 2023. Transfer Learning with Graph Attention Networks for Team Recommendation. In *International Joint Conference on Neural Networks, IJCNN 2023, Gold Coast, Australia, June 18-23, 2023*. IEEE, 1–8. https://doi.org/10.1109/IJCNN54540.2023.10191717

[42] Peter Keane, Faisal Ghaffar, and David Malone. 2020. Using machine learning to predict links and improve Steiner tree solutions to team formation problems - a cross company study. *Appl. Netw. Sci.* 5, 1 (2020), 57. https://doi.org/10.1007/S41109-020-00306-X

[43] Abeer Khan, Lukasz Golab, Mehdi Kargar, Jaroslaw Szlichta, and Morteza Zihayat. 2020. Compact group discovery in attributed graphs and social networks. *Inf. Process. Manag.* 57, 2 (2020), 102054. https://doi.org/10.1016/J.IPM.2019.102054

[44] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. OpenNMT: Open-Source Toolkit for Neural Machine Translation. In *Proceedings of ACL 2017, System Demonstrations*. Association for Computational Linguistics, Vancouver, Canada, 67–72. https://www.aclweb.org/anthology/P17-4012

[45] Yue Kou, Derong Shen, Quinn Snell, Dong Li, Tiezheng Nie, Ge Yu, and Shuai Ma. 2020. Efficient Team Formation in Social Networks based on Constrained Pattern Graph. In *36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, April 20-24, 2020*. IEEE, 889–900. https://doi.org/10.1109/ICDE48307.2020.00082

[46] Karol Kurach, Anton Raichuk, Piotr Stanczyk, Michal Zajac, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, and Sylvain Gelly. 2020. Google Research Football: A Novel Reinforcement Learning Environment. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 4501–4510. https://doi.org/10.1609/AAAI.V34I04.5878

[47] Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*, John

F. Elder IV, Françoise Fogelman-Soulié, Peter A. Flach, and Mohammed Javeed Zaki (Eds.). ACM, 467–476. https://doi.org/10.1145/1557019.1557074

[48] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR* abs/1907.11692 (2019). arXiv:1907.11692 http://arxiv.org/abs/1907.11692

[49] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*. IEEE, 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

[50] Hamed Loghmani and Hossein Fani. 2023. Bootless Application of Greedy Re-ranking Algorithms in Fair Neural Team Formation. In *Advances in Bias and Fairness in Information Retrieval - 4th International Workshop, BIAS 2023, Dublin, Ireland, April 2, 2023, Revised Selected Papers (Communications in Computer and Information Science, Vol. 1840)*, Ludovico Boratto, Stefano Faralli, Mirko Marras, and Giovanni Stilo (Eds.). Springer, 108–118. https://doi.org/10.1007/978-3-031-37249-0_9

[51] Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, Lluís Màrquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton (Eds.). The Association for Computational Linguistics, 1412–1421. https://doi.org/10.18653/V1/D15-1166

[52] Carol R. Paris, Eduardo Salas, and Janis A. Cannon-Bowers. 2000. Teamwork in multi-person systems: a review and analysis. *Ergonomics* 43 (2000), 1052 – 1075. https://api.semanticscholar.org/CorpusID:41152229

[53] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. 2018. Image Transformer. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018 (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer G. Dy and Andreas Krause (Eds.). PMLR, 4052–4061. http://proceedings.mlr.press/v80/parmar18a.html

[54] Radin Hamidi Rad, Ebrahim Bagheri, Mehdi Kargar, Divesh Srivastava, and Jaroslaw Szlichta. 2021. Retrieving Skill-Based Teams from Collaboration Networks. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 2015–2019. https://doi.org/10.1145/3404835.3463105

[55] Radin Hamidi Rad, Hossein Fani, Ebrahim Bagheri, Mehdi Kargar, Divesh Srivastava, and Jaroslaw Szlichta. 2024. A Variational Neural Architecture for Skill-based Team Formation. *ACM Trans. Inf. Syst.* 42, 1 (2024), 7:1–7:28. https://doi.org/10.1145/3589762

[56] Radin Hamidi Rad, Hossein Fani, Mehdi Kargar, Jaroslaw Szlichta, and Ebrahim Bagheri. 2020. Learning to Form Skill-based Teams of Experts. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux (Eds.). ACM, 2049–2052. https://doi.org/10.1145/3340531.3412140

[57] Radin Hamidi Rad, Shirin Seyedsalehi, Mehdi Kargar, Morteza Zihayat, and Ebrahim Bagheri. 2022. A Neural Approach to Forming Coherent Teams in Collaboration Networks. In *Proceedings of the 25th International Conference on Extending Database Technology, EDBT 2022, Edinburgh, UK, March 29 - April 1, 2022*, Julia Stoyanovich, Jens Teubner, Paolo Guagliardo, Milos Nikolic, Andreas Pieris, Jan Mühlig, Fatma Özcan, Sebastian Schelter, H. V. Jagadish, and Meihui Zhang (Eds.). OpenProceedings.org, 2:440–2:444. https://doi.org/10.48786/EDBT.2022.37

[58] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[59] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 21 (2020), 140:1–140:67. https://jmlr.org/papers/v21/20-074.html

[60] Betsy Rolland, Sarah Hohl, and LaKaija Johnson. 2021. Enhancing translational team effectiveness: The Wisconsin Interventions in Team Science framework for translating empirically informed strategies into evidence-based interventions. *Journal of Clinical and Translational Science* 5 (07 2021). https://doi.org/10.1017/cts.2021.825

[61] Eduardo Salas, Marissa L. Shuffler, Amanda L. Thayer, Wendy L. Bedwell, and Elizabeth H. Lazzara. 2015. Understanding and improving teamwork in organizations: a scientifically based practical guide. *Human Resource Management* 54 (2015), 599–622. https://api.semanticscholar.org/CorpusID:146805565

[62] Anna Sapienza, Palash Goyal, and Emilio Ferrara. 2019. Deep Neural Networks for Optimal Team Composition. *Frontiers Big Data* 2 (2019), 14. https://doi.org/10.3389/FDATA.2019.00014

[63] Jan B. Schmutz, Laurenz L. Meier, and Tanja Manser. 2019. How effective is teamwork really? The relationship between teamwork and performance in healthcare teams: a systematic review and meta-analysis. *BMJ Open* 9 (2019). https://api.semanticscholar.org/CorpusID:202568898

[64] Mauro Sozio and Aristides Gionis. 2010. The community-search problem and how to plan a successful cocktail party. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, Bharat Rao, Balaji Krishnapuram, Andrew Tomkins, and Qiang Yang (Eds.). ACM, 939–948. https://doi.org/10.1145/1835804.1835923

[65] Daniel Stokols, Kara Hall, B.K. Taylor, and R.P. Moser. 2008. The science of team science. *Am. J. Preventive Med.* 35 (01 2008), S78–S89.

[66] Damjan Strnad and Nikola Guid. 2010. A fuzzy-genetic decision support system for project team formation. *Appl. Soft Comput.* 10, 4 (2010), 1178–1187. https://doi.org/10.1016/J.ASOC.2009.08.032

[67] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, Wenwu Zhu, Dacheng Tao, Xueqi Cheng, Peng Cui, Elke A. Rundensteiner, David Carmel, Qi He, and Jeffrey Xu Yu (Eds.). ACM, 1441–1450. https://doi.org/10.1145/3357384.3357895

[68] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger (Eds.). 3104–3112. https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html

[69] Jiaxi Tang and Ke Wang. 2018. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018*, Yi Chang, Chengxiang Zhai, Yan Liu, and Yoelle Maarek (Eds.). ACM, 565–573. https://doi.org/10.1145/3159652.3159656

[70] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 5998–6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[71] Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. 2019. Deep Graph Infomax. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=rklz9iAcKQ

[72] Lin Wang, Yifeng Zeng, Bilian Chen, Yinghui Pan, and Langcai Cao. 2020. Team Recommendation Using Order-Based Fuzzy Integral and NSGA-II in StarCraft. *IEEE Access* 8 (2020), 59559–59570. https://doi.org/10.1109/ACCESS.2020.2982647

[73] Hyeongon Wi, Seungjin Oh, Jungtae Mun, and Mooyoung Jung. 2009. A team formation model based on knowledge and collaboration. *Expert Syst. Appl.* 36, 5 (2009), 9121–9134. https://doi.org/10.1016/J.ESWA.2008.12.031

[74] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J. Smola, and How Jing. 2017. Recurrent Recommender Networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, February 6-10, 2017*, Maarten de Rijke, Milad Shokouhi, Andrew Tomkins, and Min Zhang (Eds.). ACM, 495–503. https://doi.org/10.1145/3018661.3018689

[75] Lingfei Wu, Dashun Wang, and James A. Evans. 2017. Large Teams Have Developed Science and Technology; Small Teams Have Disrupted It. *CoRR* abs/1709.02445 (2017). arXiv:1709.02445 http://arxiv.org/abs/1709.02445

[76] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *CoRR* abs/1609.08144 (2016). arXiv:1609.08144 http://arxiv.org/abs/1609.08144

[77] Andrew Yates, Rodrigo Frassetto Nogueira, and Jimmy Lin. 2021. Pretrained Transformers for Text Ranking: BERT and Beyond. In *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, Liane Lewin-Eytan, David Carmel, Elad Yom-Tov, Eugene Agichtein, and Evgeniy Gabrilovich (Eds.). ACM, 1154–1156. https://doi.org/10.1145/3437963.3441667

[78] Fajie Yuan, Xiangnan He, Haochuan Jiang, Guibing Guo, Jian Xiong, Zhezhao Xu, and Yilin Xiong. 2020. Future Data Helps Training: Modeling Future Contexts for Session-based Recommendation. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen (Eds.). ACM / IW3C2, 303–313. https://doi.org/10.1145/3366423.3380116

[79] Armen Zakarian and Andrew Kusiak. 1999. Forming teams: An analytical approach. *IIE Transactions* 31 (01 1999), 85–97. https://doi.org/10.1023/A:1007580823003

[80] Yifan Zang, Jinmin He, Kai Li, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. 2023. Automatic Grouping for Efficient Cooperative Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (Eds.). http://papers.nips.cc/paper_files/paper/2023/hash/906c860f1b7515a8ffec02dcdac74048-Abstract-Conference.html

[81] Jie Zhou, Ying Cao, Xuguang Wang, Peng Li, and Wei Xu. 2016. Deep Recurrent Models with Fast-Forward Connections for Neural Machine Translation. *Trans. Assoc. Comput. Linguistics* 4 (2016), 371–383. https://doi.org/10.1162/TACL_A_00105