Learning Collaborative Teams via Social Information Retrieval (Lecture-style Tutorial)

Mahdis Saeedi
School of Computer Science
University of Windsor, ON, Canada
msaeedi@uwindsor.ca

Ziad Kobti
School of Computer Science
University of Windsor, ON, Canada
kobti@uwindsor.ca

Hossein Fani
Faculty of Information
University of Toronto, ON, Canada
hossein.fani@utoronto.ca

Abstract

Team recommendation involves selecting skilled experts to form an *almost surely* successful collaborative team, or refining the team composition to maintain and/or excel at performance. To address the tedious and error-prone manual process, various computational approaches have been proposed, especially for the web-scale social networks and widespread online collaboration and diversity of interactions. In this tutorial, with a brief overview of pioneering subgraph optimization approaches and their shortfalls, we deliberately focus on the recent learning-based approaches, with a particular in-depth exploration of graph neural network-based methods. More importantly, and for the first of its kind, we then discuss team *refinement*, which involves structural adjustments or expert replacements to enhance team performance in dynamic environments. Finally, we discuss training strategies, benchmarking datasets, and open-source libraries, along with future research directions and real-world applications. Further resources are at `https://fani-lab.github.io/OpeNTF/tutorial/www26`

1 Motivation

Team recommendation aims to automate forming collaborative teams of experts who can successfully achieve tasks, given the real-world web-scale pool of candidates across diverse fields. For instance,

- Medical emergencies require forming ad-hoc teams of diverse professionals such as nurses and physicians to respond effectively to emergencies, and the challenge lies in automating the retrieval of the right group of caregivers considering expertise, communication efficiency, availability, and logistical constraints.
- Scientific publication depends on peer reviewers' assessments, yet the process currently remains a bottleneck due to the *manual* process, automating of which based on expertise, past performance, and load balancing, helps journal editors and conference organizers reduce administrative overhead, assign more accurate and fair reviewer matches, and accelerate publication cycles.
- Educators need to split students into collaborative teams to enhance peer learning, engagement, and social skill development, which has become of growing interest due to the proliferation of large-scale *online* classes. By automating team formation among students based on skill diversity, availability, and learning styles, teachers improve group efficacy and equitable collaboration.
- On online freelance platforms, with a rich source of experts, their skill profiles, past performance, and collaborative history, team recommendations are particularly valuable for identifying potential synergies among freelance members and assembling optimal teams for the large number of available projects and the dynamic nature of projects' requirements.

```
[10 min] Pioneering Techniques
    [10 min] Subgraph Optimization Objectives and Techniques
[60 min] Learning Team Recommendation
[35 min] Team Refinement
    [15 min] Structural Adjustment
    [20 min] Membership Substitution
[20 min] Training Strategies
[10 min] Evaluation Methodologies
    [05 min] Datasets
    [05 min] Effectiveness/Efficiency
[15 min] Future Directions
    [05 min] Fair & Diverse Team Recommendation
    [05 min] Spatial Team Recommendation
    [05 min] Multi-Objective Neural Team Recommendation
[20 min] Real-World Applications
[10 min] OpeNTF & Adila
```

Figure 1: Our tutorial's timeline.

Historically, collaborative teams were formed manually based on human experience and instinct; a tedious, error-prone, and suboptimal process due to unfair biases, multiple criteria to optimize like budget, time and team size limitations, and a large number of candidates, among other reasons, which make it all the more difficult and impossible for a web-scale pool of experts. Hence, artificial intelligence and machine learning researchers have long been developing computational models learning through massive collections of experts and efficiently learning relationships between experts and their skills in the context of successful and *un*successful teams and excel at recommending *almost surely* successful teams, and at helping existing teams maintain or even improve their success.

2 Relevance to Web Community

The team recommendation problem falls under social information retrieval (Social IR), where the *right* group of experts are searched to solve the tasks at hand. In this tutorial, (1) we target *beginner* or *intermediate* researchers, industrial practitioners with a broad interest in developing web-scale recommender systems, willing to have a whole picture of team recommendation techniques. (2) Furthermore, we target the graph neural network community for a comprehensive review of subgraph optimization objectives and call them for further development of effective yet efficient graph machine learning including gnn-based path embedding frameworks with a special focus on team recommendation. Last, this tutorial enables (3) organizations and practitioners to compare different approaches and readily pick the suitable one for their application to form or maintain successful teams. The target audience will be those *familiar* with graph theory and neural architectures. We provide sufficient details about advanced techniques, such as dynamic curriculum learning, multi-query graph matching methods or graph convolutional networks, so that the content becomes understandable to those with a fair knowledge of AI-ML fundamentals.

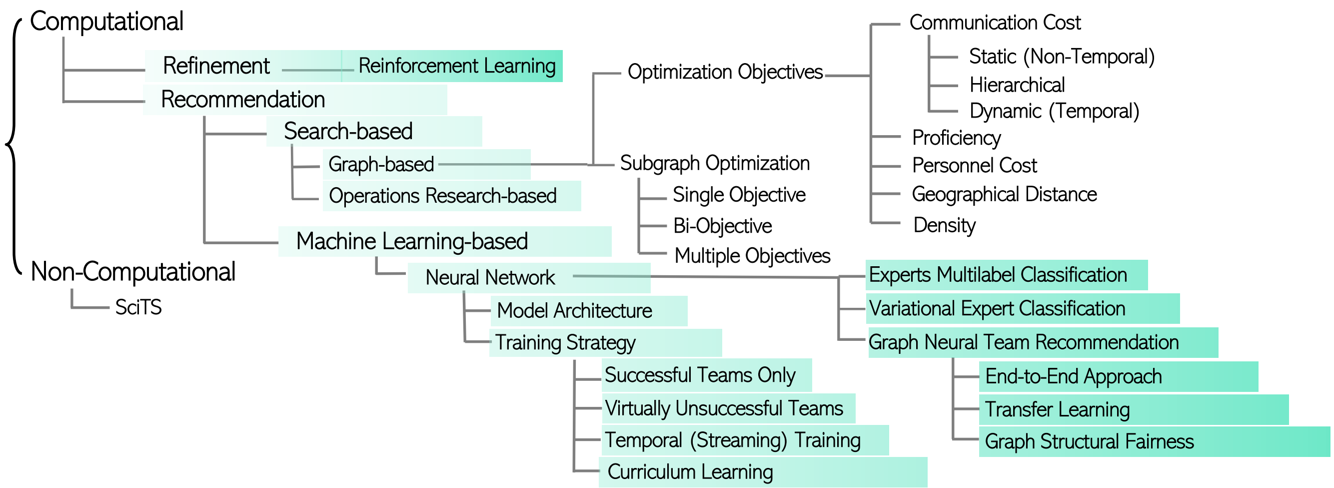

Figure 2: Taxonomy of team recommendation methods.

3 Prior Tutorials

Despite the abundance, there had been *no* comprehensive tutorial on team recommendation methods until recently when we began to fill the gap by tutorials at `UMAP24`[1], centered on *online* skilled users, `SIGIR-AP24`[2], based on a novel taxonomy from a computational perspective with a focus on real-world applications, and at `WSDM25`[3] and `CIKM25`[4], to review seminal solutions with a special focus on the emerging graph neural network-based methods. In this tutorial, from Figure 1 and 2, we expand on them based on our recently published [19] and ongoing surveys of neural approaches that leverage graph neural networks [10, 21], sequence-to-sequence models and transformers [20]. Notably, we further investigate team refinement methods, such as expert replacement through reinforcement learning in dynamic sport teams or online video games [22, 24]. We also present proposed methods to mitigate inherent popularity and gender biases via curriculum learning and loss regularization [5, 14, 17].

4 Proposed Tutorial (`180 minutes`)

We start our tutorial with a brief introduction to the pioneering graph-based team recommendation algorithms based on a taxonomy of computational methods, as shown in Figure 2, then continue to explore the learning-based team recommendation and team refinement methods, focusing on modern methods based on graph neural networks and reinforcement learning.

4.1 Pioneering Techniques (`10 minutes`)

The early computational models for team recommendation were developed in operations research (OR), optimizing objectives using integer linear and/or nonlinear programming (IP). Such work, however, was premised on the mutually independent selection of experts and overlooked the organizational and/or social ties. To bridge the gap, graph-based approaches have been proposed to recommend teams via subgraph optimization where the different aspects of real-world teams are captured like communication cost and geographical proximity [13].

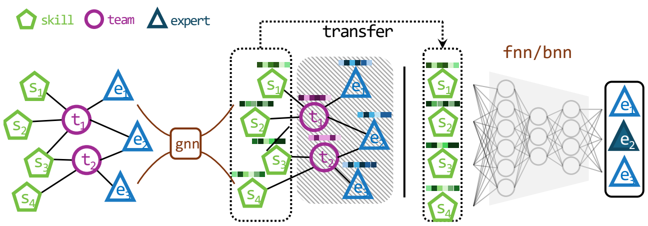

Figure 3: Dense vector representation learning for skills.

4.2 Learning Collaborative Teams (`60 minutes`)

Advances in machine learning, particularly graph neural networks, have led to learning-based methods for team recommendation [18]. These methods enhance efficiency while improving efficacy through their inherently iterative and online learning procedure, and address the limitations of subgraph optimization solutions with respect to scalability, as well as dynamic expert networks. In this tutorial, we detail them, particularly end-to-end graph neural network-based techniques, which have achieved state-of-the-art performances.

4.2.1 ***Variational/non-Variational Multilabel Classifier.*** Neural team recommendation has begun with neural-based multilabel classifiers, such as a simple feedforward network whose parameters are learned through either maximum likelihood optimization or Bayesian neural models [3, 10].

4.2.2 ***Graph Neural Team Recommendation.*** Next, graph neural networks have received growing attention in the team recommendation problem for their performance on learning the dense vector representation of the skills (Figure 3). The majority of approaches in this category have employed transfer learning techniques that involve pretraining dense vectors independently and feeding them into a neural classifier [3, 18].

4.2.3 ***End-to-End Team Recommendation.*** Following the success of end-to-end graph neural network approaches in tasks such as user-item recommendation and information retrieval, team recommendation studies also proposed end-to-end approaches via graph

[1] `www.um.org/umap2024/tutorials/`
[2] `www.sigir-ap.org/sigir-ap-2024/tutorial/`
[3] `www.wsdm-conference.org/2025/tutorials/`
[4] `www.cikm-conference.org/2025/tutorials/`

link predictions between team nodes and expert nodes in an expert collaboration graph. More recently, seq-to-seq and tranformer-based approaches have been proposed, wherein team recommendation has been reformulated as a sequence prediction task that directly maps a set of skills to a set of experts [20] (Figure 4), avoiding the unnecessary step of separately learning the skill node embeddings and transferring them to a neural classifier for team recommendation. Such end-to-end approaches tackle the training challenges of neural classifiers, like the curse of sparsity in the output layer and fragmented learning phases.

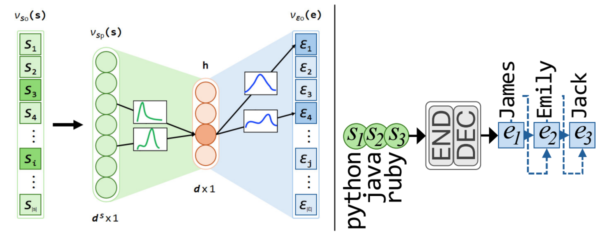

Figure 4: Multilabel vs. seq-to-seq team recommendation.

4.3 Team Refinement `(35 minutes)`

Reinforcement learning with neural policy estimators has been increasingly employed to learn the dynamics of real-world teams for structural modifications or team member replacements in order to maintain or even improve team effectiveness [22, 24]. Herein, the problem is formulated in a multi-agent setting where a group of agents synchronize their actions in a decentralized manner within a shared environment to achieve a common goal. The task cannot be completed by any individual agent alone; instead, it requires a team of agents.

4.3.1 ***Structural Adjustment.*** Herein, methods iteratively reconfigure team composition in response to evolving collaborative dynamics through a sequential decision-making or multi-armed bandit optimization task, where role allocation and/or hierarchical coupling are continuously updated to maximize collective performance [4, 23].

4.3.2 ***Membership Substitution.*** The second category of models learn to evaluate team member replacements by capturing both individual skill profiles and their influence on a team. During inference, agents iteratively refine substitution decisions to balance skill coverage, social cohesion, and organizational stability [24].

4.4 Training Strategies `(20 minutes)`

We explain strategies to train team recommenders, including (1) negative sampling [6] to learn from instances of *virtually* unsuccessful teams using different heuristics when benchmark datasets lack unsuccessful teams (negative samples), which showed synergy to the model convergence during training and improved performance during inference; (2) temporal training [8] that encode dynamic aspects of skills and experts in team recommendation, and (3) curriculum learning [5] that provide an order between experts from the *easy* popular experts to the *hard* non-popular ones to overcome models' popularity bias (Figure 5).

4.5 Evaluation Strategies `(10 minutes)`

We also discuss the benchmark datasets and what has been considered as successful teams as the ground truth. Also, we explore metrics utilized to measure the quality of the recommended teams.

4.6 Challenges and Future Work `(15 minutes)`

We discuss open issues and future directions as follows:

- **Fair and Diverse Teams:** Existing methods heavily influenced by subjective opinions which already inherit hidden *un*fair societal biases, primarily focus on maximizing the success rate of recommended teams, largely ignoring the *diversity* in recommended experts and resulting in discrimination and reduced visibility for already disadvantaged experts (e.g., females), disproportionate selection of *popular* experts, and racial disparities. Meanwhile, social science research provides compelling evidence about the synergistic effects of diversity on team performance; diversity breeds innovation and increases teams' success by enabling a stronger sense of community and support, reducing conflict, and stimulating more creative thinking. However, there is little work to mitigate biases in team recommender systems [14, 17]. In our tutorial, we introduce notions of fairness and protected attributes and study debiasing algorithms to mitigate the potential unfairness in the models' recommendations.
- **Spatial Team Recommendation:** In search of an optimal team, companies further look for experts in a region where the organization is *geographically* based. Existing methods use skills as a primary factor while overlooking geographical location. We bring forth the *spatial* team recommendation problem; that is, given a set of experts, skills and geolocations, the goal is to determine if the combination of skills and geolocations in forming teams has synergistic effects.
- **Multi-Objective Optimization:** In real-world team recommendation scenarios, balancing multiple, often conflicting objectives (e.g., team effectiveness and experts' workload distribution) requires a training process guided by a loss function that explicitly accounts for multiple objectives. However, existing neural team recommendation approaches, which commonly frame the problem as a classification or link prediction task, aim to maximize the coverage of the required skills and mainly rely on standard loss functions such as cross-entropy, and designing a task-aware loss function is overlooked.

4.7 Applications `(20 minutes)`

We explain novel applications of team recommendation, including:

- **Learning Group Recommendation**: In online classes, where physical presence and interaction are absent, team recommendation connects students to improve their social skills and combat the isolation that can sometimes accompany remote learning [9, 16]. In large classes, where individual interactions with the instructor may be limited, group work ensures that students still have ample opportunities to engage with the material.
- **Reviewer Assignment**: Another application of team recommendation is in peer-review assignments [1, 2] where the reviewers are paired with manuscripts within their expertise for high-quality reviews while managing conflicts of interests [11]. Herein, research topics (skills) and reviewers (experts) are mapped into a latent space, and given a manuscript on a subset of research

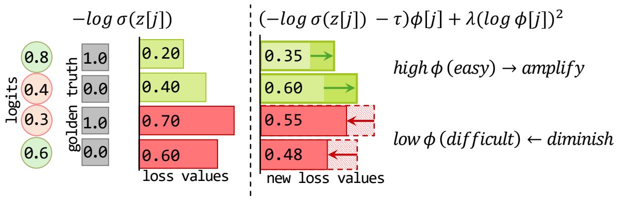

Figure 5: Standard (left) vs. dynamic curriculum (right) loss.

topics, a model recommends top-k optimum reviewers for the research topics.

- **Palliative Care:** Team recommendation is also applicable in healthcare to assign a team of caregivers to patients who seek help with their daily activities due to disease or disorders [12], or to form ad-hoc teams of clinical care experts for medical emergencies [15]. The challenge lies in optimally assigning care providers in teams to address patient needs while considering factors such as communication, distance, and contract costs.

4.8 OpeNTF and Adila (`10 minutes`)

We present publicly available libraries for team recommendation, including `OpeNTF`[5] [7], an open-source benchmark library for neural models that can efficiently preprocess large-scale datasets, can be easily extended or customized to new neural methods, and is extensible to new datasets from other domains. We also introduce `Adila`[6] [14], which enables further in-processing female-advocate loss regularization [17], and/or post-processing reranking [14] to the list of recommended experts to ensure a fair outcome.

5 Presenters

- **Mahdis Saeedi (She/Her)** is a Research Associate at the School of Computer Science and a Lecturer at the Department of Mathematics, University of Windsor. She has taught graph theory, linear algebra, graph representation learning and topics in AI at undergraduate and graduate levels. Her current research centers on graph machine learning. She has published in the field's premier venues, such as ACM Computing Surveys, WSDM, ECIR, SIGIR-AP, CIKM, Transactions of Combinatorics, and the Journal of Algebra.
- **Ziad Kobti (He/Him)** is a Professor at the School of Computer Science, University of Windsor, since 2015. He was the recipient of the Distinguished University Teaching Award and the Leadership Award. He has been the executive secretary, executive vice president, and the president of the Canadian AI Association, Canada's foremost AI community and the official arm of the AAAI in Canada. He is an associate editor of Wiley's Computational Intelligence and the guest editor of the Machine Learning and Knowledge Extraction Journal. He has expertise in machine learning for Social Network Analysis, funded by NSERC, NSF, CFI, and OCI.
- **Hossein Fani (He/Him)** is a Research Fellow at the Faculty of Information, University of Toronto, and has been an Assistant Professor at the School of Computer Science, University of Windsor. His research is at the intersection of Social Network Analysis, User Modeling, and Information Retrieval, with a diverse team of 15+ HQP, funded by NSERC-DG, NSERC-RTI, and CFI-JELF. His research appears in ACM Computing Surveys, Elsevier's IP&M, ACM's TOIS, Wiley's JASIST, and SIGIR, CIKM, WSDM and ECIR. He translates his research into techniques for the industry while leading R&D funded by NSERC Alliance and Mitacs Accelerate. He has been granted three patents by the USPTO, including `US10,885,131`, `US11,768,522`, and `US12,067,625`. Fani's teaching experience spans over 15+ years in countries with diverse cultures and educational systems. He has taught courses, including natural language communication and graph representation learning, in multi-section classes ranging from 30 (graduate) to 200+ (undergraduate) with in-person, hyflex, and online modalities. He has developed graduate courses, including big data analytics and the reciprocal role of AI, science, and society.

References

[1] Aksoy et al. 2023. Reviewer Assignment Problem: A Systematic Review of the Literature. *J. Artif. Intell. Res.* 76 (2023), 761–827.

[2] Arabzadeh et al. 2024. Reviewerly: Modeling the Reviewer Assignment Task as an Information Retrieval Problem. In *CIKM 2024*. 5554–5555.

[3] Ahmed et al. 2024. Vector Representation Learning of Skills for Collaborative Team Recommendation: A Comparative Study. In *WISE*.

[4] Almutairi et al. 2025. tAIfa: Enhancing Team Effectiveness and Cohesion with AI-Generated Automated Feedback. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work, CHIWORK 2025, Amsterdam, Netherlands, June 23-25, 2025*. 2:1–2:25.

[5] Barzegar et al. 2025. Adaptive Loss-based Curricula for Neural Team Recommendation. In *WSDM*.

[6] Dashti et al. 2022. Effective Neural Team Formation via Negative Samples. In *CIKM*. ACM, 3908–3912.

[7] Dashti et al. 2022. OpeNTF: A Benchmark Library for Neural Team Formation. In *CIKM*. ACM, 3913–3917.

[8] Fani et al. 2024. A Training Strategy for Future Collaborative Team Prediction. In *ECIR*. Springer.

[9] Gonzalo Candel et al. 2023. An Integer Linear Programming Model for Team Formation in the Classroom with Constraints. In *HAIS*.

[10] Hamidi Rad et al. 2024. A Variational Neural Architecture for Skill-based Team Formation. *ACM Trans. Inf. Syst.* 42, 1 (2024), 7:1–7:28.

[11] Katharina et al. 2023. RevASIDE: Evaluation of Assignments of Suitable Reviewer Sets for Publications from Fixed Candidate Pools. *J. Data Intell.* 4 (2023), 101–133.

[12] Kalyani Selvarajah et al. 2018. Team Formation in Community-Based Palliative Care. In *INISTA*.

[13] Lappas et al. 2009. Finding a team of experts in social networks. In *SIGKDD*. ACM, 467–476.

[14] Loghmani et al. 2023. Bootless Application of Greedy Re-ranking Algorithms in Fair Neural Team Formation. In *ECIR-BIAS 2023*, Vol. 1840. Springer, 108–118.

[15] McLeod et al. 2021. Hacking teamwork in health care: Addressing adverse effects of ad hoc team composition in critical care medicine. *Health care management review* 46, 4 (2021), 341–348.

[16] Michaelsen et al. 2023. *Team-based learning: A transformative use of small groups in college teaching.* Taylor & Francis.

[17] Moasses et al. 2024. vivaFemme: Mitigating Gender Bias in Neural Team Recommendation via Female-Advocate Loss Regularization. In *ECIR-BIAS 2024*.

[18] Nguyen et al. 2023. Learning heterogeneous subgraph representations for team discovery. *Inf. Retr. J.* 26, 1 (2023), 8.

[19] Saeedi et al. 2025. A Survey of Subgraph Optimization for Expert Team Formation. *ACM Comput. Surv.* (May 2025). Just Accepted.

[20] Thang et al. 2025. Translative Neural Team Recommendation: From Multilabel Classification to Sequence Prediction. In *SIGIR*.

[21] Vombatkere et al. 2025. A QUBO Framework for Team Formation. *CoRR* abs/2503.23209 (2025). https://doi.org/10.48550/ARXIV.2503.23209

[22] Yifan Zang et al. 2023. Automatic Grouping for Efficient Cooperative Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*.

[23] Zhou et al. 2018. In Search of the Dream Team: Temporally Constrained Multi-Armed Bandits for Identifying Effective Team Structures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*. 108.

[24] Bing Lv, Junji Jiang, Likang Wu, and Hongke Zhao. 2024. Team formation in large organizations: A deep reinforcement learning approach. *Decis. Support Syst.* 187 (2024), 114343. https://doi.org/10.1016/J.DSS.2024.114343

[5] `github.com/fani-lab/OpeNTF`

[6] `github.com/fani-lab/Adila`