Paradigm Shifts in Team Recommendation: From Historical Subgraph Optimization to Emerging Graph Neural Network (Half-Day Tutorial)

Mahdis Saeedi
University of Windsor, Canada
msaeedi@uwindsor.ca

Christine Wong
University of Windsor,Canada
wong93@uwindsor.ca

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

ABSTRACT

Collaborative team recommendation involves selecting experts with certain skills to form a team who will, more likely than not, accomplish a task successfully. To automate the traditionally tedious and error-prone manual process of team formation, researchers from several scientific spheres have proposed methods to tackle the problem. In this tutorial, while providing a taxonomy of team recommendation works based on their algorithmic approaches, we foremost perform a comprehensive study of the graph-based approaches that comprise the pioneering works in this field, then cover the graph neural network-based studies as the cutting-edge class of approaches. Further, we provide unifying definitions, formulations, and evaluation schema. Last, we introduce details of training strategies, benchmarking datasets, and useful open-source tools, as well as a performance comparison of the works. Finally, we identify directions for future works.

1 MOTIVATION

Team recommendation aims to automate forming teams of experts whose combined skills, applied in coordinated ways, can successfully solve difficult tasks. Successful teams have firsthand effects on creating organizational performance in academia, manufacturing, law, freelancing, and the healthcare sector, among others. Recommending a successful team whose members can effectively collaborate and deliver the outcomes within the specified constraints such as planned budget and timeline is challenging due to the immense number of candidates with various backgrounds, personality traits, and skills, as well as *unknown* synergistic balance among them; not all teams with *best* experts are necessarily successful [10].

Traditionally, teams were formed manually by relying on human experience and instinct, which is a tedious, error-prone, and suboptimal process due to *i*) hidden personal and societal biases, *ii*) a multitude of criteria to optimize, *iii*) an overwhelming number of candidates, among other reasons. The team formation can be heavily influenced by subjective opinions that already inherit hidden unfair societal biases, largely ignoring the *diversity* in recommended experts and resulting in discrimination and reduced visibility for an already disadvantaged *female* experts, disproportionate selection of *popular* experts, and racial/gender disparities. Additionally, since this process involves a multitude of criteria including project importance, budget, time constraints and team size limitations, the decision-making is all the more difficult. On top of these, expert candidates should be examined based on individual and relative factors such as technical abilities, availability, individual cost, personality traits, negotiation skills, and proactivity, among others, which makes manual team formation on a large scale almost impossible.

Together with business sectors like Linkedin[1], researchers in artificial intelligence and machine learning have long been endeavouring to develop computational models to analyze massive collections of experts and efficiently learn relationships between experts and their skills in the context of successful and *un*successful teams and excel at recommending *almost surely* successful teams. This has resulted in a rich body of various approaches grounded in computational and social science theoretical and conceptual frameworks wherein the problem definition of team recommendation remains the same essentially, while it has been referred to by such other names as team allocation, team selection, team composition, and team formation.

2 OBJECTIVES AND PRIOR TUTORIALS

Despite the substantial number of computational models for team recommendation, there is, however, yet to be a comprehensive tutorial with comparative analysis and critical reviews on approaches' applicability in real-world scenarios, especially when each comes with a domain-specific method with no standard implementation, incapable of accommodating different use cases. To start filling this gap, we provided a tutorial at `UMAP24`[2] centred on a narrowed scope of subgraph optimization objectives and experts being *online* skilled users. To foster future research in the field, we aim to present this *comprehensive* tutorial to review seminal solutions based on a novel taxonomy from a computational perspective with a special focus on the emerging graph neural network-based methods, as shown in Figure 1. In our tutorial, we bring forth a unifying and vetted methodology to the various definitions in this realm, criticize assumptions and comparative benchmarks, and point out shortfalls to smooth the path for future directions.

3 RELEVANCE TO IR COMMUNITY

Team recommendation problem falls under social information retrieval (**Social IR**), where the right group of experts are searched to solve the tasks at hand or only with the assistance of social resources. In this tutorial, *i*) we target *beginner* or *intermediate* researchers, industry technologists and practitioners with a broad interest in developing and applying AI-based methods for recommender systems, willing to have a whole picture of team recommendation techniques. *ii*) Furthermore, this tutorial targets audiences from the graph neural network (GNN) community for a comprehensive review of subgraph optimization objectives and calls them

*https://fani-lab.github.io/

[1] www.linkedin.com
[2] https://www.um.org/umap2024/tutorials/

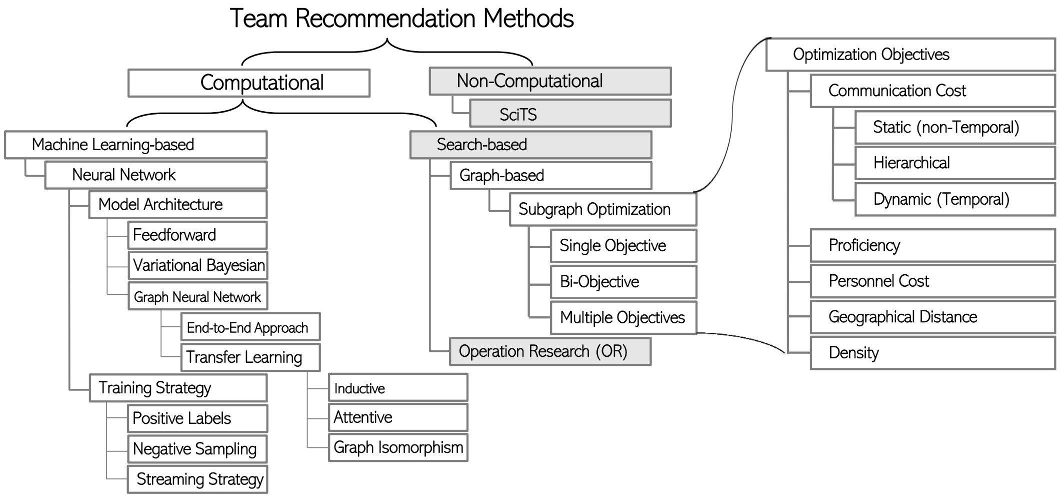

Figure 1: The taxonomy of team recommendation methods. The gray areas are excluded from our tutorial.

for further development of effective yet efficient graph neural networks with a special focus on team recommendation. Last, having regard to the unified comparative analysis, this tutorial enables *iii*) organizations and practitioners to compare different models and readily pick the most suitable one for their application to form collaborative teams of experts whose success is *almost surely* guaranteed. The target audience needs to be *familiar* with graph theory and machine learning. We will make no assumptions about the audience's knowledge of more advanced techniques. As such, sufficient details will be provided so that the content will be accessible and understandable to those with a fundamental understanding of such principles.

4 TUTORIAL OUTLINE

Foremost, we briefly introduce intuitive definitions of a *team* and some representatives, historical to modern and state-of-the-art methods for solving the team recommendation problem, motivating the importance of the problem, followed by a novel taxonomy of computational methods in this field. Next, we provide a list of unified notations and formalize *optimization objectives* based on which an optimum team is defined. From search-based methods, we focus on graph construction and subgraph optimization techniques, as they comprise the pioneering body of research. Then, we continue on learning-based methods, particularly graph neural network-based methods, which have been building up the following and yielding to state-of-the-art performances. Then, we discuss evaluation methodologies, including the datasets, metrics, and comparative baselines that exist in the literature. After the break, we continue the tutorial by presenting seemingly unrelated but highly valuable applications of team recommendation in education, research, and health care, followed by major lines of future research. Finally, we conclude the tutorial with hands-on experience on a benchmark library for neural team recommendation research.

```
[25 min] Pioneering techniques
    [5 min] Subgraph Optimization Objectives
    [20 min] Subgraph Optimization Techniques
        [05 min] Expert Network Construction
        [05 min] Communication Cost Minimization Algorithms
        [05 min] Multi-objective Optimization Algorithms
        [05 min] Community-based Optimization Algorithms
[65 min] Learning-based heuristics
    [35 min] Neural Model Architectures
        [05 min] Feedforward
        [10 min] Variational Bayesian
        [20 min] Graph Neural Network
    [15 min] Training Strategies
        [05 min] Positive Labels
        [05 min] Negative Sampling
        [05 min] Streaming Training
    [15 min] Evaluation Methodologies
        [05 min] Datasets
        [05 min] Effectiveness
        [05 min] Efficiency
Break
[20 min] Applications
    [10 min] Group Learning
    [05 min] Reviewr Assignment
    [05 min] Palliative Care
[20 min] Future Directions
    [10 min] Fair & Diverse Team Recommendation
    [05 min] Spatial Team Recommendation
    [5 min] End-to-End Architecture
[50 min] Hands-on OpeNTF & Adila
```

Figure 2: Our tutorial's outline.

4.1 Pioneering Techniques

The foremost computational models for the team recommendation problem conceived in the operations research (OR) where objectives have to be optimized via an 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. A team is, however, relational and is a property of effective collaboration among the team members. In our tutorial, we have excluded these methods.

4.1.1 ***Subgraph Optimization Objectives.*** A step forward was to employ social network analysis (SNA) to incorporate experts' ties. In this stream, experts are modeled in an attributed weighted graph whose nodes are experts with individual skills, and weighted links are based on interpersonal attributes between experts. For the efficient representation of experts' ties as well as the synergistic interdisciplinary discoveries from social network analysis and graph theory, graphs have become a natural choice and pervasive in team recommendation literature. The graph-based approaches tackle the team recommendation problem by defining subgraph optimization of *objectives* where the different aspects of real-world teams are captured such as communication cost, budget, levels of proficiency, and geographical proximity. In our tutorial, we formalized more than 13 objectives in a unified framework with integrated notations for better readability and fostering conventions in this realm.

4.1.2 ***Subgraph Optimization Techniques.*** Subgraph optimization problems are NP-hard [9]. Therefore, heuristics have been developed to solve optimization in polynomial time through greedy and/or approximation algorithms. In our tutorial, we describe the seminal heuristics that have been followed by the majority of researchers in three groups: *i*) those that target minimizing communication cost only; *ii*) those that consider additional objectives such as personnel cost, expertise level and geographical proximity jointly with communication cost; and, *iii*) those considering maximizing the teams' density only.

4.2 Learning-based Heuristics

Recently, a paradigm shift to learning-based methods has been observed for team recommendation due to the advances in machine learning, graph neural networks in particular [5, 11]. These methods are different in that they learn the inherent structure of the ties among experts and their skills. Learning-based methods bring efficiency while enhancing efficacy due to the inherently iterative and online learning procedure, and can address the limitations of subgraph optimization solutions with respect to scalability, as well as dynamic expert networks [13–15]. In our tutorial, we explain this line of research categorized based on *i*) model architectures and *ii*) training strategies.

4.2.1 ***Neural Architectures.*** Neural team recommendation has started with autoencoders and is being followed through other neural-based architectures like a simple feedforward network whose parameters are learned by either maximum likelihood (MLE) optimization or maximizing a posterior (MAP) using Bayesian neural models [14–16]. Naturally enough, graph neural networks have also been receiving growing attention for the team recommendation problem for their expressive performance on the vector representation of the experts and their ties, and we will lay out their details in our tutorial.

4.2.2 ***Training Strategies.*** We explain various strategies to train neural model parameters, including *i*) negative sampling[3] and *ii*) streaming training[6]. Neural models learn from instances of teams labelled with success or failure. However, benchmark datasets in team recommendation may lack *un*successful teams. In the absence of explicit labels for unsuccessful teams, researchers proposed different negative sampling heuristics to draw virtually unsuccessful teams and show their synergy to the model convergence and improved inference during training and test, respectively.

To address the temporality of experts' interests, skills, and levels of expertise due to society's demands, novel technologies, and working experience, *streaming* training strategy has been proposed. Given the stream of experts' collaborations in each time interval, a neural model learns the vector representations for experts and skills at the time interval t to kick-start learning the vectors of the next time interval $t + 1$, allowing experts to change their vector positions in latent space up until *current* time interval to accurately predict experts' vector positions in the *future* time interval. In our tutorial, we explain the streaming training strategy, which puts a chronological order on teams during training to incorporate the temporal dependency of teams vs. randomly shuffled that assumes the independent and identically distributed (i.i.d) instances of teams (bag of teams) [11].

4.2.3 ***Evaluation Methodology.*** In this part of our tutorial, we lay out the methodologies used to evaluate the performance of the approaches. We discuss the well-known benchmark datasets in team recommendation literature and explain what has been considered as a team, and how they have been assumed successful to function as the gold truth in each of them. The effectiveness of neural team recommendation approaches has been evaluated quantitatively by classification and ranking metrics with reference to optimum (golden) teams, or lack thereof, using qualitative metrics such as skill coverage.

4.3 Applications

In the second part of our tutorial, we further explain novel applications of team recommendation besides its common use cases.

4.3.1 ***Group Learning.*** Team recommendation find immediate application in group-based learning environments. In online classes, where physical presence and interaction are absent, team recommendation connects students to share ideas and build relationships. This not only enhances their social skills but also combats the isolation that can sometimes accompany remote learning. Via working in teams, students are exposed to varying viewpoints, backgrounds, and problem-solving approaches [8]. 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. Moreover, team recommendation in online and large classes promotes accountability and responsibility [2]. When students work together towards a common goal, they are accountable not only to themselves but also to their team members. This dynamic motivates students to stay on track, meet deadlines, and contribute actively to the group's success.

4.3.2 ***Reviewer Assignment.*** Another immediate application of team recommendation is in peer-review assignments [1] where a group of reviewers are paired with manuscripts within the reviewers' expertise for high-quality reviews while managing conflicts of interests. Like team recommendation, herein, research topics (skills) and reviewers (experts) are mapped into a latent space and, given a manuscript about a subset of research topics, team recommendation aims to recommend reviewers with top-k closest vectors to the vectors of the research topics [20].

4.3.3 ***Palliative Care.*** Another application of team recommendation is in healthcare, which assigns a team of caregivers to patients who seek help for their daily activities due to disease or disorders [18]. The challenge lies in optimally assigning care providers in teams to address patients' needs while considering factors such as communication, distance, and contract costs.

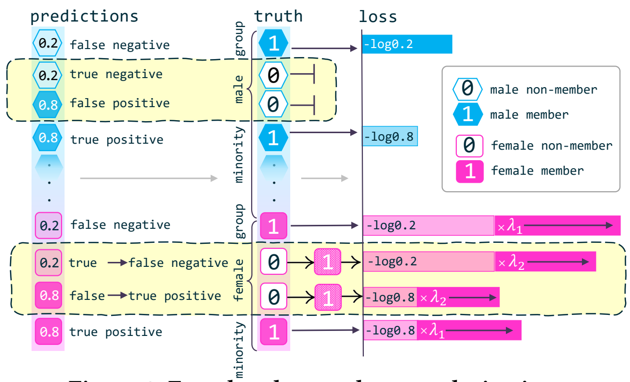

Figure 3: Female-advocate loss regularization.

4.4 Future Directions

Although remarkable progress has been made, several open issues and potential future directions are worth more research and will further unleash the great potential of recommending effective teams.

4.4.1 ***Fair and Diverse Team Recommendation****.* The primary focus of existing methods is the maximization of the success rate for the recommended teams, largely ignoring diversity in the recommended list of experts. 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 to no diversity-aware algorithmic method that mitigates unfair societal biases in team recommendation algorithms. In our tutorial, we introduce notions of fairness along with the protected attributes and study debiasing algorithms to mitigate the potential unfairness in the models' recommended teams.

4.4.2 ***Spatial Team Recommendation****.* In search of an optimal team, companies further look for experts in a region where the organization is geographically based, which leads to new challenges as it requires drilling down on the skills of experts while maintaining the condition of a given geolocation. In our tutorial, we bring forth the *spatial* team recommendation problem, where the goal is to determine whether the combination of skills and locations in forming teams has synergistic effects. Although remote work over online platforms has facilitated today's globalized work environment, geographical proximity remains important for face-to-face interactions, cultural understanding, time zone differences, and access to local resources such as availability of certain region-locked services [19], which impact team dynamics, coordination, and effectiveness.

4.4.3 ***End-to-End Graph Neural Network****.* Neural models that accept dense vector representation of skills in the input layer outperformed the sparse occurrence vectors. However, the dense vectors are learned separately in an unsupervised manner using a graph neural network, oblivious to supervised information about successful teams [12, 17]. However, the team recommendation problem can be reformulated into a link prediction in an expert graph to directly and jointly learn dense vectors of skills and experts and recommend an optimum subset of experts as a team through predicting links, eschewing the unnecessary complications by the two-phase graph representation learning and neural model fine-tuning.

4.5 Hands-On

We introduce publicly available libraries for team recommendation. Notably, we provide hands-on experience with `OpeNTF`[3] [4], an open-source benchmark library for neural models that: *i*) can efficiently preprocess large-scale datasets, *ii*) can be easily extended or customized to new neural methods, and *iii*) is extensible to experiments on new datasets from other domains. We also introduce `Adila`[4] [7], that enables further in-processing female-advocate loss regularization (Figure 3) and/or post-processing reranking to the list of recommended experts to reassure the desired fair outcome. `Adila` is equipped with fairness metrics, which, in tandem with *utility* metrics, allows exploring the synergistic trade-offs between notions of fairness and success rate.

REFERENCES

[1] Meltem Aksoy, Seda Yanik, and Mehmet Fatih Amasyali. 2023. Reviewer Assignment Problem: A Systematic Review of the Literature. *J. Artif. Intell. Res.* 76 (2023), 761–827.
[2] Rob Burton, Ursula Kellett, Martha Mansah, and Deepa Sriram. 2024. A systematic review of online team based learning approaches in health professional education. *Nurse Education Today* (2024), 106156.
[3] Arman Dashti, Saeed Samet, and Hossein Fani. 2022. Effective Neural Team Formation via Negative Samples. In *CIKM*. ACM, 3908–3912.
[4] Arman Dashti, Karan Saxena, Dhwani Patel, and Hossein Fani. 2022. OpeNTF: A Benchmark Library for Neural Team Formation. In *CIKM*. ACM, 3913–3917.
[5] Sagar Kaw et al. 2023. Transfer Learning with Graph Attention Networks for Team Recommendation. In *IJCNN 2023*. IEEE, 1–8.
[6] Hossein Fani, Reza Barzegar, Arman Dashti, and Mahdis Saeedi. 2024. A Training Strategy for Future Collaborative Team Prediction. In *ECIR (Lecture Notes in Computer Science)*. Springer.
[7] Hamed Loghmani and Hossein Fani. 2023. Bootless Application of Greedy Re-ranking Algorithms in Fair Neural Team Formation. In *BIAS (Communications in Computer and Information Science, Vol. 1840)*. Springer, 108–118.
[8] Larry K Michaelsen, Arletta Bauman Knight, and L Dee Fink. 2023. *Team-based learning: A transformative use of small groups in college teaching*. Taylor & Francis.
[9] R Miller. 2013. *Complexity of computer computations: Proceedings of a symposium on the complexity of computer computations*. Springer Science & Business Media.
[10] Satyam Mukherjee, Yun Huang, Julia Neidhardt, Brian Uzzi, and Noshir Contractor. 2019. Prior shared success predicts victory in team competitions. *Nature human behaviour* 3, 1 (2019), 74–81.
[11] Hoang Nguyen, Radin Rad, Feras N. Al-Obeidat, Ebrahim Bagheri, Mehdi Kargar, Divesh Srivastava, Jaroslaw Szlichta, and Fattane Zarrinkalam. 2023. Learning heterogeneous subgraph representations for team discovery. *Inf. Retr. J.* 26, 1 (2023), 8.
[12] Radin Hamidi Rad, Ebrahim Bagheri, Mehdi Kargar, Divesh Srivastava, and Jaroslaw Szlichta. 2021. Retrieving Skill-Based Teams from Collaboration Networks. ACM, 2015–2019.
[13] Radin Hamidi Rad, Ebrahim Bagheri, Mehdi Kargar, Divesh Srivastava, and Jaroslaw Szlichta. 2022. Subgraph Representation Learning for Team Mining. In *WebSci*. ACM, 148–153.
[14] 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.
[15] Radin Hamidi Rad, Hossein Fani, Mehdi Kargar, Jaroslaw Szlichta, and Ebrahim Bagheri. 2020. Learning to Form Skill-based Teams of Experts. In *CIKM '20*. ACM, 2049–2052.
[16] Radin Hamidi Rad, Shirin Seyedsalehi, Mehdi Kargar, Morteza Zihayat, and Ebrahim Bagheri. 2022. A Neural Approach to Forming Coherent Teams in Collaboration Networks. In *EDBT*. OpenProceedings.org, 2:440–2:444.
[17] Anna Sapienza, Palash Goyal, and Emilio Ferrara. 2019. Deep Neural Networks for Optimal Team Composition. *Frontiers Big Data* 2 (2019), 14.
[18] Kalyani Selvarajah, Pooya Moradian Zadeh, Ziad Kobti, Mehdi Kargar, Mohd Tazim Ishraque, and Kathryn Pfaff. 2018. Team Formation in Community-Based Palliative Care. In *INISTA 2018, Thessaloniki, Greece, July 3-5, 2018*. IEEE.
[19] Kalyani Selvarajah, Pooya Moradian Zadeh, Ziad Kobti, Yazwand Palanichamy, and Mehdi Kargar. [n. d.]. A unified framework for effective team formation in social networks. *Expert Syst. Appl.* 177 ([n. d.]), 114886.
[20] Yixuan Xu, Steven Jecmen, Zimeng Song, and Fei Fang. 2023. A One-Size-Fits-All Approach to Improving Randomness in Paper Assignment. In *ANIPS*.

[3] `https://github.com/fani-lab/OpeNTF`
[4] `https://github.com/fani-lab/Adila`