Collaborative Team Recommendation for Skilled Users: Objectives, Techniques, and New Perspectives

MAHDIS SAEEDI, University of Windsor, Canada
CHRISTINE WONG, University of Windsor, Canada
HOSSEIN FANI, University of Windsor, Canada

Collaborative team recommendation involves selecting users with certain skills to form a team who will, more likely than not, accomplish a complex 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 to model skilled users in collaborative teams, we perform a comprehensive and hands-on study of the graph-based approaches that comprise the mainstream in this field, then cover the neural team recommenders 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 open-source tools, along with directions for future works.

1 MOTIVATION

Algorithmic search for collaborative teams, also known as team recommendation, aims to automate forming teams of skilled users whose combined skills, applied in coordinated ways, can successfully solve difficult tasks. Successful teams have firsthand effects on creating organizational performance in academia, manufacturing and the healthcare sector, to name a few. Business and employment-focused social media platforms like Linkedin[1], which are primarily used for professional networking and career development, have long been endeavouring to employ computational models to analyze massive collections of skilled users and efficiently learn relationships between users and their skills in the context of successful and unsuccessful teams and excel at recommending *almost surely* successful teams [4, 5]. 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 users with various backgrounds, personality traits, and skills, as well as unknown synergistic balance among them; not all teams with *best* skilled users are necessarily successful. 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 skilled users, among other reasons.

In an effort to automate team formation, researchers in different disciplines, such as psychology, management, engineering, and recently, machine learning, have proposed algorithmic solutions grounded in computational 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. In this tutorial, we provide a comprehensive and hands-on study of team recommendation works based on their algorithmic approaches to model skilled users in collaborative teams. 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.

[1]www.linkedin.com

Authors' addresses: Mahdis Saeedi, msaeedi@uwindsor.ca, University of Windsor, Canada; Christine Wong, University of Windsor, Canada, womg93@uwindsor.ca; Hossein Fani, hfani@uwindsor.ca, University of Windsor, Canada.

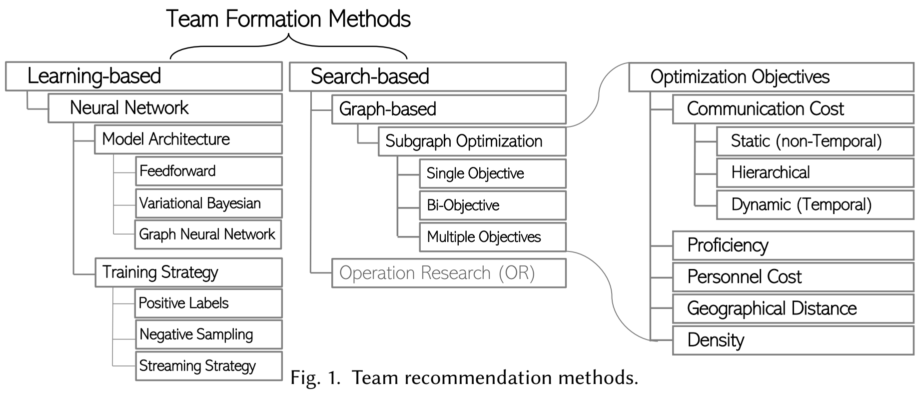

Fig. 1. Team recommendation methods.

2 PRIOR TUTORIALS

Despite the substantial number of algorithmic approaches to team recommendation, there is, however, yet to be a comprehensive tutorial with comparative analysis and critical reviews on approaches' applicability in real-world scenarios. The first of its kind and to foster future research in the field, we aim to present this tutorial based on a novel taxonomy from a computational perspective, as shown in Figure 1, discuss the evaluation metrics used, and identify any existing or ongoing comparative studies. We present a comprehensive overview of 34 seminal solutions to the team recommendation problem, including 13 proposed optimization objectives and 8 neural architectures after screening 63 algorithms from 126 papers.

3 TUTORIAL OUTLINE [180 MINUTES]

From Figure 1, we begin to introduce intuitive definitions of a *team* and some representative, 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, as explained hereafter.

3.1 Search-based Heuristics [75 minutes]

The foremost computational models of team recommendation were conceived in the Operations Research, where multiple objective functions have to be optimized via integer programming. Such work, however, overlooked the organizational and social ties among users. In our tutorial, we have excluded OR-based methods. A progressive step forward was to employ social network analysis to incorporate social ties, interpersonal attributes such as communication, collaboration attributes such as the number of projects and social attributes such as levels of friendship. In this stream, users are modelled in an attributed weighted graph whose nodes are users with individual attributes (e.g., skills), and weighted links are established either explicitly or inferred based on interpersonal attributes between users. For the efficient representation of users' social and collaborative ties as well as the synergistic interdisciplinary discoveries from social network analysis and graph theory, graphs have been firmly established not only in search-based methods but also in the newly emerged learning-based methods. Therefore, an overview of the graph-based literature is beneficial for the community and quite timely.

3.1.1 Subgraph Optimization Objectives [25 minutes]. The graph-based approaches tackle the team recommendation problem by defining subgraph optimization of *objectives* on a graph 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.

3.1.2 *Subgraph Optimization Techniques [25 minutes].*

Subgraph optimization problems are proven to be NP-hard [7]. Therefore, different heuristics have been developed to solve this problem 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. As will be explained, optimization techniques can be studied in three groups: *i*) those that minimizes communication cost only [8]; *ii*) those that consider additional objectives such as personnel cost and geographical proximity jointly with communication cost [6]; and, *iii*) those considering maximizing the teams' density.

3.1.3 *Evaluation Methodology [25 minutes].*

Finally, we lay out the methodologies used to evaluate the performance of the graph-based approaches. We discuss the benchmark datasets, what has been considered as teams and how they have been assumed successful to function as gold truth, as well as quantitative and qualitative metrics that are utilized to measure the quality of the recommended teams by proposed approaches as compared to the gold truth.

3.2 Learning-based Heuristics [75 minutes]

Recently, a paradigm shift to learning-based methods has been observed for team recommendation due to the advances in machine learning, neural networks in particular. These methods are different from search-based solutions in that they learn the inherent structure of the ties among users and their skills. Wherein, all past (un)successful team compositions are considered as training samples to predict future teams and the team's performance. Learning-based methods bring efficiency while enhancing efficacy due to the inherently iterative and online learning procedure, and can address the limitations of search-based solutions with respect to scalability, as well as dynamic expert networks [10, 11]. In our tutorial, we explain this line of research, which has been mostly based on neural models.

3.2.1 *Neural Architectures [25 minutes].*

Neural team recommendation has started with Sapienza et al. [13], who employed an autoencoder, and is being followed by researchers through other neural-based architectures, whose parameters are learned by either maximum likelihood (MLE) optimization or maximizing a posterior (MAP) using Bayesian neural models. Graph neural networks have also been employed for the team recommendation problem. Graph neural network methods have provided an effective yet efficient way to solve the graph analytic problem by converting a graph into a low-dimensional vector space while preserving the graph structure. Graph neural network has shown expressive performance for a vast array of AI-hard problems such as knowledge graph and recommender systems. Not unexpectedly, its application in team recommendation has been receiving growing attention, and we will lay out their details in our tutorial.

3.2.2 *Training Strategies [25 minutes].*

Neural models learn from instances of teams that are 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, neural methods presume all instances of teams in the training dataset as successful (positive samples) and proceed with the training procedure. Meanwhile, other literatures have shown that leveraging not only positive samples (e.g., friendship in social networks) but also negative samples (e.g., distrust) convey complementary negative signals to the neural models and improve accuracy in various tasks in social network analysis and recommender systems. In this line, researchers proposed to follow the *closed-world* assumption and consider *no currently known successful team for the required subset of skills* as *'virtually'* unsuccessful [1]. In our tutorial, we will discuss the details of different negative sampling heuristics to draw virtually unsuccessful teams and will show their synergy to the model convergence and improved inference during training and test, respectively.

To address the temporality of users' interests, skills, and levels of expertise due to society's demands, novel technologies, and working experience, *streaming* training strategy [3] has been proposed. Given the stream of users' collaborations in each time interval, a neural model learns the vector representations for users and skills at time interval t to kick-start learning the vectors of the next time interval $t + 1$, allowing users to change their vector positions in latent space up until *current* time interval to accurately predict users' vector positions in the *future* time interval. In our tutorial, we explain streaming training strategy that put 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)[12].

3.2.3 Hands-On `OpeNTF` [25 minutes].

In our tutorial, we introduce publicly available libraries and tools for the task of team recommendation. Notably, we provide hands-on experience with `OpeNTF`[2][2], 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.

3.3 Challenges and New Perspectives [20 minutes]

3.3.1 `Adila`: Fair and Diverse Team Recommendation [10 minutes].

The primary focus of existing team recommendation methods is the maximization of the success rate for the recommended teams, largely ignoring diversity in the recommended users. There is little to no diversity-aware algorithmic method that mitigates *unfair* societal biases in team recommendation models. In our tutorial, we introduce notions of fairness, namely *i*) demographic parity and *ii*) equality of opportunity, along with the protected (sensitive) attributes concerning the users and study whether the state-of-the-art debiasing algorithms can mitigate the potential unfairness in the models' recommended teams. We introduce `Adila`[3][9], that enables further post-processing reranking refinements to the list of recommended users to reassure the desired fair outcome. `Adila` is also equipped with fairness evaluation metrics to measure the difference between the distribution of recommended teams over popularity or gender labels and a reference unbiased desired distribution. In tandem with *utility* metrics, which measure the efficacy of the recommended teams with respect to teams' success rate, `Adila` allows to explore the synergistic trade-offs between notions of fairness, on the one hand, and success rate on the other hand for the proposed solutions.

3.3.2 Spatial Team Recommendation [10 minutes].

In search of an optimal team, companies further look for skilled users in a region where the organization is *geographically* based, which leads to new challenges as it requires drilling down on the skills of users while maintaining the condition of a given geolocation. The majority of existing methods use skills as a primary factor while overlooking geographical location and the corresponding ties it leads to between users in a team. We conclude our tutorial by bringing forth the *spatial* team recommendation problem; that is, given a set of users, skills and geolocations, the goal is to determine whether the combination of skills and geolocations in forming teams has synergistic effects.

4 INTENDED AUDIENCE

Team recommendation problem falls under social information retrieval (Social IR) where we seek to find the right group of skillful users 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 user modeling and recommender systems who are 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 for further development of effective yet efficient graph neural

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

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 skilled users whose success is *almost surely* guaranteed.

Prerequisite Knowledge: The target audience needs to be *familiar* with graph theory and machine learning. Where appropriate, the tutorial will not make any assumptions about the audience's knowledge on more advanced techniques. As such, sufficient details will be provided as appropriate so that the content will be accessible and understandable to those with a fundamental understanding of such principles.

5 PRESENTERS

Mahdis Saeedi (She/Her) is an Assistant Professor at the Department of Engineering, Damavand University, Iran, and a Postdoctoral Fellow at the School of Computer Science, University of Windsor, where she works on subgraph optimization and graph neural networks. During her PhD studies, she focused on theoretical aspects of graph mining, including binomial edge ideals in bipartite and glued graphs. She has already published in the field's premier journals such as the Journal of Algebra, Number Theory and Applications and Transactions of Combinatorics.

Christine Wong (She/Her) is an undergraduate student in the School of Computer Science at the University of Windsor with a keen interest in AI-related research. Her active contributions to projects in Fani's Lab*!* have resulted in her involvement in a publication presented at the CIKM conference. Furthermore, her achievements include securing a silver medal in the UWill Discover 2023 competition at the University of Windsor.

Hossein Fani (He/Him) is 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, he addresses the ever-growing need for identifying, facilitating, and expanding effective interdisciplinary and collaborative teamwork, which is one of the pillars of growth in the scientific and industrial communities. Fani's research appears in top venues of the field such as Elsevier's IP&M, ACM's TOIS, and Wiley's JASIST journals, and CIKM and ECIR conferences. He also effectively translates his research and knowledge into tools and techniques for the industrial community. He leads industrial R&D work funded by NSERC Alliance and Mitacs Accelerate. His PhD work has resulted in a patent with USPTO US10885131.

REFERENCES

[1] Dashti, A., Samet, S., Fani, H.: Effective neural team formation via negative samples. In: CIKM. pp. 3908–3912. ACM (2022)

[2] Dashti, A., Saxena, K., Patel, D., Fani, H.: Opentf: A benchmark library for neural team formation. In: CIKM. pp. 3913–3917. ACM (2022)

[3] Fani, H., Barzegar, R., Dashti, A., Saeedi, M.: A training strategy for future collaborative team prediction. ECIR, Springer (2024)

[4] Geyik, S.C., Ambler, S., Kenthapadi, K.: Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In: KDD 2019. pp. 2221–2231. ACM (2019)

[5] Geyik, S.C., Guo, Q., Hu, B., Ozcaglar, C., Thakkar, K., Wu, X., Kenthapadi, K.: Talent search and recommendation systems at linkedin: Practical challenges and lessons learned. In: SIGIR 2018. pp. 1353–1354. ACM (2018)

[6] Kargar, M., An, A., Zihayat, M.: Efficient bi-objective team formation in social networks. In: ECML PKDD. vol. 7524, pp. 483–498. Springer (2012)

[7] Karp, R.M.: Reducibility among combinatorial problems. In: Complexity of Computer Computations: Proceedings of a symposium on the Complexity of Computer Computations. pp. 85–103. Springer (1972)

[8] Lappas, T., Liu, K., Terzi, E.: Finding a team of experts in social networks. In: SIGKDD. pp. 467–476. ACM (2009)

[9] Loghmani, H., Fani, H.: Bootless application of greedy re-ranking algorithms in fair neural team formation. In: BIAS. Communications in Computer and Information Science, vol. 1840, pp. 108–118. Springer (2023)

[10] Rad, R., Fani, H., Bagheri, E., Kargar, M., Srivastava, D., Szlichta, J.: A variational neural architecture for skill-based team formation. TOIS **42**(1), 1–28 (2023)

[11] Rad, R.H., Fani, H., Kargar, M., Szlichta, J., Bagheri, E.: Learning to form skill-based teams of experts. In: CIKM '20. pp. 2049–2052. ACM (2020)

[12] Rad, R.H., Nguyen, H., Al-Obeidat, F.N., Bagheri, E., Kargar, M., Srivastava, D., Szlichta, J., Zarrinkalam, F.: Learning heterogeneous subgraph representations for team discovery. Inf. Retr. J. **26**(1), 8 (2023)

[13] Sapienza, A., Goyal, P., Ferrara, E.: Deep neural networks for optimal team composition. Frontiers Big Data **2**, 14 (2019)