October 11, 2024

RE: Revisions to Manuscript CSUR-2023-0992 for Consideration of ACM Computing Surveys

Dear Dr. Milano and Reviewers,

We would like to thank you for your careful evaluation of our manuscript. It is our pleasure to inform you that we have considered the received comments and carefully addressed them in the manuscript as indicated below. We believe the manuscript has improved because of the suggestions and hope that our revisions satisfy the requested changes. We would like to highlight some important points summarily:

- The journal's **35-page limit including references** required us to reorganize or adjust certain content to address the comments. For instance, we moved Figure 2 and Table 4 to Appendix B.3 and Appendix G, respectively.
- Our manuscript originally included **Appendices** containing Appendix A. Baselines to illustrate cross-comparison of graph-based Team Formation methods, Appendix B. Background on Team Formation Problem, Appendix C. Pitfalls in Evaluation, and Appendix D. Open-Source Implementation. We believe this content could have addressed some of the reviewers' comments. However, based on the reviewers' comments, it appears that the reviewers did not receive these appendices. We will make sure that the Appendices, along with our detailed responses to the reviewers' comments, are attached to the revised version of the manuscript.
- We clarify the scope of our survey by revising Section 1.2 Our Contributions to **1.2 Our Contributions and Scope Disclaimer** as well as adding a Venn diagram in Figure 1 (right), recognizing the significance of *learning-based* and *operations research-based* methods for Team Formation. However, due to the scope of the current survey and the page limit constraint, we suffice with a brief mention of these methods. Yet, we have targeted OR-based and neural-based techniques as a complement to this work in our immediate research direction.

Cordially,
The Authors.

Reviewer 1

R1.1

The work presents a recent survey of Subgraph Optimization for Collaborative Team Formation. The paper gives a good overview on the current state of the arts on the general Team Formation problems as well as its on-going work. Additionally, the justification of the work is very clear citing outdated surveys and lack of comparative studies between methods to name a few.

Response: Thank you for your encouraging feedback on our work. We appreciate your review.

R1.2

In Figure 1, the authors state the limitations of the survey. I feel that for completeness the authors should address these cited limitations as there would be the need for another survey.

Response: We are in total agreement. In the revised manuscript, we have revised the Subsection **1.2 Our Contributions and Scope Disclaimer**, explaining the pressing need for other surveys that cover the works outside the scope of this survey.

R1.3

From Figure 1, the authors need to give more explanation on Team Formation Method as grayed Operation Research. On positive note, in my opinion, the authors have covered works based on Operation research.

Response: Thank you for your comment. We acknowledge operations research category is vital and contains numerous significant contributions, especially those that leverage expert graphs and overlap with graph-based methods. However, we believe providing a detailed review of them, constituting 20+ papers to the best of our search efforts, requires a standalone survey. In the revised Subsection **1.2 Our Contributions and Scope Disclaimer**, we have expressed that such a review is urgently needed and is currently underway.

R1.4

I am aware of the state-space reduction approach. Here, the preprocessing graph search space is reduced by eliminating the experts with no required skills. More precisely, experts-skills matrix were utilized reduced both horizontally (selecting skills only required in the given task) and vertically (discarding experts not having any required skills for a given task) to obtain the sub-search space. Then, the full exhaustive search commences. Undoubtedly, such a work is computationally heavy and may not be practical for large datasets, it is interesting to see how the current survey classifies such a work (e.g. maybe search-based with preprocessing (see Figure 2)

Response: Thank you for your valuable suggestion. As you mentioned, majority of graph-based works utilize two approaches, mostly in combination, to reduce the computational complexity of large-scale expert graphs: (1) reducing the size of the expert graph by preprocessing the raw datasets based on thresholds for a minimum number of any or both of the following: a) project teams b) pairwise collaborations for experts, and limited subset of project teams based on their areas of specialty, frequent skills, timestamp associated with project teams, and/or graph samplings, followed by (2) heuristics during the optimization like exploring the neighbours of a specific node (e.g., the expert with the rarest skill) where the search is guided toward subparts of the entire graph. Since such approaches are employed by the majority of works in our survey, rather than creating a subcategory, we have added a new Subsection **3.5 Efficiency and Scalability Enhancement** in Section **3 Optimization Techniques**. Additionally, we have included the list of the graph-based Team Formation algorithms along with their filtering approaches for particular benchmark datasets in Appendix F.

R1.5

Table 3 gives a good overview of objectives employed in subgraph optimizations for Team Formation problems. The paper also rightfully highlights the issue of temporal dynamics which most work overlook the fact that experts' skills and social attributes constantly change over time. The authors rightly point out some consideration of hybrid algorithms apart from multi-objective for team formation, data sets, as well as evaluations (intrinsic, qualitative, quantitative (e.g.hypervolume etc.,).

Response: We sincerely appreciate your thoughtful and encouraging feedback on our work.

R1.6

I think the proposed future work can be expanded further to tackle new up and coming progress in the field of optimization. Apart from tackling multi-objective (MOO) issues, I think the future work should also consider the adoption of multi-factorial optimization (MFO). More precisely, in general, conventional optimization problems (like team formation) can be divided into two categories: single-objective optimization (SOO) problems and multi-objective optimization (MOO) problems. They are both committed to seeking the optimal solution of an optimization task. The difference is that SOO has only one objective function, while MOO needs to optimize multiple conflicting objective functions. MFO is devoted to implementing search on multiple optimization tasks simultaneously by seamlessly transferring knowledge between multiple optimization problems. Unlike SOO and MOO, MFO aims to seek out the optimal solutions for multiple tasks at once. In the case of the Team Formation problem, the search process may involve finding experts for multiple projects simultaneously from the common expert database. Typically using an evolutionary approach (but not limited to), MFO utilizes the shared population in a seamless manner to do parallel multitask search processes in an implicit manner.

Response: Your suggestion is accurate and insightful. After reviewing the graph-based works, also the OR-based and learning-based approaches, we found that no work has utilized multi-factorial optimization yet. We revised Section **3.3 Hybrid Optimization Algorithms** and have reminded the readers about such modern techniques [47, 51] and the fact that graph-based techniques can further be improved/adjusted utilizing it when forming teams of experts simultaneously. Also, in Section **5 Challenges and Future Directions**, we further add a new Subsection **5.3 Modern Hybrid Optimization Algorithms** and explain multi-factorial optimization and provide possible future paths.

Reviewer 2

R2.1

This survey paper presents a comprehensive overview of 18 seminal graph-based solutions to the Team Formation problem, including 13 proposed optimization objectives, after screening 63 algorithms from 126 papers. This paper has a detailed and solid description on the selected seminal subgraph optimization approaches to the team formation problem.

Although the authors have pointed out that the learning-based approach is in its early stages, it is suggested to include the learning-based approach. This will be more helpful for future research into the team formation problem.

Response: Thank you for your suggestion. We are in total agreement on the pressing need to cover the emerging yet successful learning-based approaches. We have already identified 18+ learning-based works to the best of our searching efforts, and as mentioned in reply to Reviewer 1's comment in R1.2 about including operations research works, we believe that providing a detailed review on that category of works requires a standalone survey owing to the journal's page limit (35 pages *with* Reference section). Having said that, we revised Subsection **1.2 Our Contributions and**

Scope Disclaimer, where we briefly draw upon such works and have expressed that a review of them is urgently needed and is currently underway.

R2.2

In Sec. 2.2, the team formation problem is formalized as a subgraph optimization objective function with a set of constraints. While in general, most of the team formation problems are typically framed as constrained optimization problems, this paper focuses on objective functions but provides limited discussions on constraints. It is suggested to organize current research from the perspective of the constraints on team formation tasks as well.
Response: Thank you for your valuable suggestion. We have added the Subsection **2.4 Optimization Constraints** to clarify optimization constraints versus objectives. Additionally, we have included the list of the constraints employed in graph-based Team Formation algorithms along with references to the related papers in Appendix E.

R2.3

It is suggested to include a complete example to illustrate the team formation problem. It will be easy for novice readers to follow the team formation problem.
Response: Thanks for your suggestion. We have added an example along with the new figure (Figure 3 in this letter but Figure 2 in the revised manuscript) at the beginning of Section **2.2. Subgraph Optimization Objectives** as an explanation for subgraph optimization to find an optimum subgraph for a project team.

R2.4

In Figure 4, the cost of the minimum spanning tree should be 0.9, rather than 1.0 while that of the Steiner tree should be 0.8, rather than 0.9.
Response: Thank you for bringing this error to our attention. We have fixed the issue in both the figure and the text.

R2.5

In the evaluation methodology, this paper discusses the benchmark datasets, as well as quantitative and qualitative metrics, however, the task generation methods are missing. It is suggested to include the discussion of task generation methods.
Response: Thank you for your valuable suggestion. Following your recommendation, we have revised Section **4.2.1 Effectiveness (Accuracy)** according to how the instances of teams are selected as a gold team (ground truth) for evaluation. We explain the project generation methods where graph-based approaches generate *synthetic* subsets of skills, each of which is fed to an algorithm with the goal of finding an optimum subgraph of experts or a gold team via an exhaustive search over all the subgraphs graph [17, 25, 30, 40, 41, 72, 74, 81, 96, 111].

We make a contrast between project generation-based evaluation method and intrinsic evaluation where a portion of the dataset is selected as a gold set for evaluation. We also revise the **Quantitative** and **Qualitative** paragraphs as **Quantitative Metrics** and **Qualitative Metrics** to measure intrinsic and/or project generation-based evaluation methodologies.

R2.6

The wordings of qualitative metrics listed in Tables 5 and 6 are suggested to be consistent.
Response: Thanks for the suggestion. We revised qualitative metrics to be consistent throughout all the tables and the text body.

R2.7

Are the papers listed in Tables 2 and 5 the 18 selected seminal papers? If they are, there are only 17 papers listed in Table 2.

Response: Thank you for your insightful question. Your comment has also been raised by Reviewer 3 in R3.8. We have indeed selected 18 seminal solutions, but these are introduced across 17 seminal papers. By 'seminal solutions,' we are referring to optimization objectives, including Equations 2-11, 13-16, and 18, as well as hybrid optimization of objectives, as noted in Section **2.3 Hybrid (Multi-Objective) Optimization**. We acknowledge that our initial phrasing may have caused some confusion. To clarify, we revised the text to state:

"We present a comprehensive overview of 17 seminal graph-based research papers for the Team Formation problem, [...]"

R2.8

Some of the references do not have specific sources listed. For example, [113] Morteza Zihayat, Mehdi Kargar, and Aijun An. 2014. Two-Phase Pareto Set Discovery for Team Formation in Social Networks. In 2014 IEEE/WIC/ACM. The source should be IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) Some references have typos. For example, [96] Vijay V Vazirani. 2001. Approximation algorithms. Vol. 1. Springer. Vol 1. should be 1st ed.

Response: Thanks for pointing this out. All references have been reviewed and errors have been rectified.

Reviewer 3

R3.A

One of my greatest concerns is whether this paper is indeed a survey or a proposition of a new integrated model for solving the expert team formation problem. The authors attempt to make an integrated mathematical model, based on a number of relevant selected papers, to unify notions around sub-graph optimization applied to the expert team formation problem. This is a welcome contribution to the literature. However, across the paper, the proposed modeling does not clarify what percentage of the proposed papers adopt each proposed modeling element, which could help evaluate and/or justify the extent to which this model is indeed representative of the broader literature or whether it is a new mathematical model comprising selected modeling elements from different papers. The rest of the sections after the modeling, i.e. those referring to the optimization techniques and the benchmarks, are more in the spirit of a survey.

Response: Thanks for your comment. A main goal of ours was to introduce a unifying (inclusive) mathematical framework at the level of notions and definitions for the Team Formation problem as well as subgraph optimization objectives, rather than at the level of algorithms and algorithmic modeling. This appears feasible as notions should ideally not depend on the implementation aspects of an optimization algorithm. We can assure you that what you point out in your comment has in fact been a guideline for us in preparation of this survey. To clarify, let us provide examples for each of those two levels that we have considered.

At the definition level for the Team Formation problem (Section **2. Problem Formalization**), we foremost encapsulate all types of attributes that the literature has associated with experts as nodes' attributes in an *attributed* graph, where we make sure that the idea can be extended to new attributes as they may be introduced in the literature in the future. In the original papers, one or at most a subset of them are introduced in each work, where they are considered not as a part of the graph but as a separate information source for experts based on which an objective function is optimized. For example, Datta et al. [17] and Juarez et al. [70] have independently introduced an expert's capacity and an expert's expertise in their respective papers. While they were not cognisant of each other, we, in this survey, bring them together as expert's attributes. This facilitates for the definitions and equations in Section **2.2 Subgraph Optimization Objectives** to readily have access to any combinations of such attributes.

In terms of notation, since graphs have been the core element of graph-based algorithms, we have endeavored to unify the varied notations relying on well-known notations from graph theory such that the notations within the Team Formation problem definition and subgraph optimization objectives can be connected with no extra additions. For example, Lappas et al. [81] and others [83] define the Expert Graph as $G(V, E)$, whereas a range of varied notations are introduced to denote an induced subgraph, e.g., T in [81], R in [83] or U in [17], as opposed to the standard notation for an induced subgraph $G[V]$. This unifying notation has the utility to also unify the Subgraph Optimization Objectives based on $\underset{V_p \in \mathcal{P}(G)}{min/max}\ \phi(G[V_p])$, ϕ being an optimization objective having regard to the attributes of expert nodes over all subgraph $G[V_p]$ of graph G. Such unification also reveals the interconnection or dependencies between optimization objectives. For instance, both Selvarajah et al. [34] and Juarez et al. [70] aim to maximize proficiency (referred to as mastery level in the former and expertise in the latter) as their objective function, denoted differently as $EL = \sum_{p=j_1,q=i_1}^{i_r,i_j} \xi\zeta_{V_p}^{S_q}$ and $Z(X') = \sum_{i\in X'} \sum_{a_j \in (X_i \cap A(T))} \frac{z_i(a_j)}{|X'|}$, respectively. Notably, Juarez et al.'s objective function is a normalized version of that of Selvarajah et al., as seen in Equations 9 and 10 of our unified notations. However, this relation is not immediately obvious from the main formulas.

Last, we unify the terminology in referring to the same elements of the Team Formation problem. For instance, Lappas et al. [81] and others [17, 25, 57, 83] refer to a subset of the required skills for a team as a task whereas Zihayat et al. [41] and others [30, 40] refer to the same entity as a *project*, although both are the same, semantically and mathematically.

Having provided the above explanation, we have revised instances where we refer to our survey's contribution about the unifying notations and have made sure we specify such contribution is at the levels of Team Formation and subgraph optimization objectives definitions. We have endeavored to lay a consistent conceptual foundation that will, among other things, make communication and conveyance of scientific ideas easier and, as a result, may yield easier integration and introduction of elements within the framework. At the same time, it can make fundamental additions distinguishable from incremental attribute additions within already introduced algorithmic approaches. We also provide further explanations on this topic by replying to related comments in R3.9, R3.11, R3.12, R3.13 below.

R3.B

At all cases, the paper needs a better definition of the scope to avoid over-generalization of its claims, for example regarding the type of problem it aims to address or the application area that its results apply to.

Response: Thanks for your comment, which has been also shared among other reviewers. In response, we have revised Section **1.2 Our Contributions and Scope Disclaimer** and Figure 1. Please refer to Review 1's comment R1.2 and our response above. Also, we have provided more explanation in related comments R3.5, R1.2, R1.3, R2.1, R3.6 below.

R3.C

Aside from elaborating on the papers and categorizing them, the paper could benefit from elaborating on its meta-review aspect. Specifically, section 5 (Challenges and Future Directions) could benefit from additional reflection from the authors' part on the literature covered to highlight research trends, gaps, and potential for future work.

Response: We are in total agreement. We studied the pitfalls in the evaluation methodology of the existing as well as the lack of open-source implementation but had to put them in Appendix C and D due to space limit. Based on Reviewer 1's comment in R1.6, we also explain modern optimization algorithms like multi-factorial optimization that can also be utilized for subgraph hybrid optimization. In your comments R3.16, R3.20 and R3.21, we provide more detailed responses.

DETAILED REVIEW REMARKS

R3.1

Paper title: The title includes the term "Collaborative team formation". The methods explored in the survey indeed concern sub-graph optimization for expert team formation (See comment 2 below). However, the use of the word collaborative seems incorrect or redundant. If the word "collaborative" is used to refer to the decision method of forming the teams, then it is incorrect since the way teams are formed in the papers covered by the survey is not collaborative, i.e. it is not performed through any collaboration but, rather, it is decided by the algorithm. If the word "collaborative" refers to the end result, which is a team, then it is redundant because a team collaborates by definition. I would recommend dropping the word "collaborative" from the title.

Response: Thanks for the suggestion and the detailed explanation. We meant to use the term *'collaborative'* to refer to the end result, that is, a team, as opposed to the process of forming a team. We agree that this word can cause ambiguity, as you have elaborated. In the original manuscript, we have included a brief background on the term *'team'* vs. *'group'* in Appendix B, where the vital part of a team that distinguishes it from a group is the accomplishment of a *shared* goal, which is hardly possible without collaboration among team members [112]. Therefore, to stress on *'collaboration'* as the necessary property of a team, we opt for *'collaborative team'*. We found that reviewers did not have access to the Appendices during the review process. We will attach the Appendices to the revised manuscript to make sure that they are accessible.

Nonetheless, the concern you raised is valid. We have removed *'collaborative'* from the title in the revised manuscript to prevent ambiguity.

R3.2

Overall across the paper: The team formation problem, as defined in the paper, is defined as gathering a team of experts to work on a project. However, the general Team Formation Problem as per relevant literature (i.e., not only graph-based one, but also from organizational sciences, for example) does not necessarily comprise experts or a single selected team. I would advise the authors to clarify that their definition of the team formation problem, which stems from the literature the survey focuses on, refers to a specific sub-problem, namely the *Expert* Team Formation Problem, where the goal is to identify the *single best team* (optimal subgraph) of *experts* to solve a given problem.

Response: We thank you for the suggestion. We have made adjustments in the revised manuscript to use *Expert* Team Formation instead of collaborative Team Formation. These adjustments are made in the title as well as sections of the manuscript. We revised the Subsection **1.2 Our Contributions and Scope Disclaimer** as the last part of Section **1 Introduction** and our contribution items to clarify this more explicitly. Also, throughout Section **2 Problem Formalization**, especially in Definition **2.3 Project Team**, we stress the fact that the main goal of the Expert Team Formation problem is to find a single optimum team of experts.

R3.3

In the Introduction, I recommend adding more references and examples to justify the claim that manual team formation, especially at large scale, is tedious and error-prone.

Response: Thank you for the suggestion. We have added more references and a real-world example to illustrate and motivate the algorithmic approach to Team Formation in Section **1 Introduction** in the second paragraph, as follows:

"It has long been well-established in the fields of psychology and cognitive science that human individuals have limited capacity to store and process information, making large-scale computation beyond their capacity [9, 91]. Also, research in industrial engineering [85] has shown that the size and variety of components within a task, like the task of forming a team, are crucial factors in determining its complexity, which are limited for human performance. The limitations of human intuition and cognitive biases have been further shown in complex decision-making tasks [14, 59, 62]. Such studies necessitate an algorithmic approach, at least as a supplement, when dealing with tasks of large-scale size and with diverse components like Team Formation [39]. For instance, major global corporations with thousands of employees like IBM have already invested in developing expert recommender systems such as SmallBlue [22] based on corporate emails correspondences to assist in locating experts among the company's broad geographically distributed presence in various countries for a given project team."

R3.4

Thank you for highlighting this point. When explaining the differences between the current survey and relevant surveys, the authors refer to the work of Wang et al. They mention that Wang's work lacks analyzing Team formation from the perspective of problem setting. Could you elaborate on this limitation? How does the present survey overcome this limitation?

Response: Within the problem setting of Wang et al., this fact is overlooked that a search for an optimum subgraph can be defined for different objectives, or even a combination of objectives (hybrid) such as bi-objectives or multi-objectives. They consider communication cost as the only objective that should be minimized for an optimum team. However, our problem setting is defined such as to generalize to any existing objectives or their combinations, like maximizing proficiency while minimizing both personnel cost and communication cost [111]. Furthermore, our problem setting presumes that experts hold *attributes* like authority level or workload capacity and defines the expert graph as an attributed graph, where all properties of the experts are modeled as node attributes that can be employed for definitions of a variety of subgraph optimization objectives. However, Wang et al., completely disregard the attributes of experts in their setting. We have added the following to the revised manuscript for the sake of clarification:

"[...] lacks analyzing Team Formation from the perspective of problem setting and the challenges in each setting like optimizing multiple objectives or the fact that experts can hold attributes that impact Team Formation like workload capacity; [...]"

R3.5

In defining the scope of the survey, the authors differentiate between graph-based team formation methods, Operations Research-based ones, and Learning-based ones. However, it is unclear how operations research-based methods are distinct from graph-based ones, given that OR is also used to solve graph problems (e.g. graph matching ones).

Response: Thank you for your comment. This point is also mentioned by Reviewer 1 in R1.2. We have revised Subsection **1.2 Our Contributions and Scope Disclaimer** and Figure 1, with a Venn diagram, for more clarity on the scope of our survey. Specifically, we acknowledge the overlap between OR-based and graph-based approaches. However, while some OR-based works like Camelo et al. [37] model the dataset as a graph structure, they opt for linear/non-linear integer/real programming methods as opposed to subgraph optimization. In this survey, we include works that not only model the data as graphs but also apply subgraph optimization methods.

Further diluting this distinction, among the learning-based methods, the authors mention GNN (graph neural network) methods; why are these excluded? Further clarification of the scope would be very useful for the survey in this context. Please also see comment 13 regarding better defining the scope of the survey.

Response: Thank you for the comment. Based on our conceptual categorization and taxonomy, graph neural networks-based works [45] belong to the learning-based category since, while they model the raw data as a graph, they use neural network estimators to learn and transfer low-dimensional vector representation of skills for supervised multilabel classifiers, instead of utilizing subgraph optimization techniques. Similar to our answer to comment R3.5, since such works use learning-based optimization, we have excluded them from this survey. In the revised Subsection **1.2 Our Contributions and Scope Disclaimer**, we further recognize the overlap between learning-based and graph-based approaches and clarify this point. We also acknowledge that graph neural networks-based methods are emerging techniques in this field. Therefore, in our immediate research direction, we have targeted them in a new standalone survey of learning-based methods, as a complement to this work.

R3.6

In Figure 1, it is not easily possible to distinguish the non-gray boxes (within scope) from the gray ones (out of scope). Further, the caption would benefit from additional explanations, e.g., concerning the second figure column (why are the specific variables chosen, out of the many variables that affect expert teamwork according to the literature cited by the authors)?

Response: Thank you for your feedback regarding the figure's readability. We apologize for the oversight. We have now revised it to ensure clarity. Also, we have added a Venn diagram to illustrate the overlapping works, and have revised the figure's caption. The variables assigned to each of the objectives were meant to blueprint the notations in the optimization technique section. The assignment of specific variables was done for better categorization. For instance, all solutions aiming to minimize communication cost are denoted by ϕ; as minimizing communication cost by sum of distance is ϕ_D, minimizing communication cost by sum of edge weights is ϕ_E, minimizing communication cost by diameter method is ϕ_R, and so on. Further, the notation for each category has been selected based on Greek letters and mathematical variables for notation unification on the surface, as opposed to acronyms that are chosen by different works differently. For instance, communication cost has been referred to as 'Cc' [81] or 'CC' [40], or '*sumDistance*' [72] for minimizing communication cost.

R3.7

Contributions section (page 3): "in our survey we introduce graph-based Team Formation algorithms" –> did you intend to say "we focus on graph-based Team Formation algorithms"? Because the survey does not introduce the graph-based team formation team notion.

Response: Thank you for your great comment. Yes, our *focus* has been on the Team Formation algorithms, and we have not introduced the graph-based team formation notion. We made the necessary changes to the text in the revised version of Section **1.2 Our Contributions and Scope Disclaimer**.

R3.8

Contributions section: (page 4) it is unclear why only 18 graph-based solutions were retained after screening 63 algorithms over 126 papers.

Response: We agree that our phrasing in the original manuscript raises ambiguity, as also mentioned by Reviewer 2 in R2.7 We found 126 papers addressing Team Formation problems using computational and non-computational models, based on our literature exploration at the time. Among those, 17 papers were identified within the scope of our survey, that is, modeling the data on a graph structure and proposing subgraph optimization algorithms to find an optimum team. Such papers include 63 optimization algorithms considering variations of their exact algorithms and the heuristics they used to address the efficiency of their algorithms. Such algorithms were proposed to optimize 18 optimization objectives, including Equations 2-11, 13-16, and 18, as well as hybrid optimization of objectives, as noted in Section **3.3 Hybrid (Multi-Objective) Optimization**. We have revised the sentence in the revised manuscript as follows:

"After screening 126 papers addressing the Team Formation problem using computational and non-computational models, we present a comprehensive overview of 17 seminal graph-based research papers to the Team Formation problem that fall within the scope of our survey, including 18 unique objectives that have been optimized via 63 subgraph optimization algorithms considering variations of their exact algorithms and the heuristics they used to address the efficiency of their algorithms."

R3.9

Contributions section (page 4, contribution 1): please justify further why standards and conventions are needed and why using different notations per application field, for instance, is problematic. It could very well be that by proposing a unified taxonomy, certain nuances of each application field are left out, resulting in oversimplification.

Response: Thank you for your comment. Our proposed unified framework and set of notations are designed to be a *minimal superset* of varied notations spread in papers, that is, it is a generalization to the existing notations/frameworks while respecting the specifications to each method. For example, we define Expert Graph (Definition 2.1) as an undirected *weighted* and *attributed* graph. This definition is inclusive of works that employ weighted graphs *with* and *without* experts' attributes like Lappas et al. [81] and Majmuder et al. [17], respectively. Meanwhile, such a definition is not loosely broad like including *directed* graphs, which has not, and probably will not, find application in graph-based Team Formation. Another example is a different use of uniform terminology for the same concept that has been defined or utilized with different names like authority that has been referred to by such other names like mastery level [34] and expertise level [70]. Moreover, standard notations help with finding overlaps or correlations among formulas, which can improve understanding, simplify analysis, validate models, and avoid redundancy, as explained in the comment R3.A when Selvarajah et al. [34] and Juarez et al. [70] aimed to maximize proficiency using two *seemingly* different objective functions but Juarez et al.'s objective function is a normalized version of that of Selvarajah et al. Last, standard notations allow seamless integration of new extensions to an existing problem settings by reusing the prior notations. We have made the following change to contribution 1:

"[...] To foster standards and conventions, which literature of this field lacks, we set forth a unified set of notations to formalize the problem and its different aspects and sub-tasks conveniently. Our notations are designed to be a minimal superset of varied notations in papers while respecting their specifications. Our unified notations help with finding overlaps or correlations among formulas, which can improve understanding, simplify analysis, validate models, and avoid redundancy while allowing seamless integration of novel ideas by reusing the prior notations."

R3.10

Contributions section (page 4, contribution 2) and Problem Formalization (Section 2): Again, please clarify that the survey does not deal with the whole Team Formation Problem but with its problem variant that focuses on expert teams.

Response: Sure. We have revised Sections **1 Introduction** and **2 Problem Formalization**, as explained in comment R3.2.

R3.11

Please explain the meanings of E and w in the notation.

Response: Regarding E and w, we have explained them in the text body from Line #313-324:

"Next, E represents the connection between the experts as graph edges and $w(e_{v,v'}) : E \rightarrow \mathcal{R}$, is the weight function that maps an edge (connection) $e_{v,v'} \in E$ between two experts v and v' to its weight. Edges can represent experts' previous collaborations, followership in a social network, or spatial proximity and their weights $w(e_{v,v'})$ can be based on the number of joint projects,, [...]"

Also, in notation Table 1, E has been defined as the set of edges of the expert graph, and w is the weight function.

Additionally, considering one of the paper's purposes is to unify notations, why are multiple terms used interchangeably throughout the survey (e.g., network and graph, vertex, node, and expert)? This does not aid in achieving unification.

Response: Thanks for your comment. We primarily meant to unify the terms used in Team Formation literature based on well-established terms from graph theory, wherein some terms are commonly used interchangeably. In **Definition 2.1 Expert Graph**, we stated *"Accordingly, in this survey, the word 'vertex' or 'node' is used interchangeably with 'expert', and 'graph' with 'network'.*

Notwithstanding, we agree that it is more favorable to be consistent in the usage of such terms across our manuscript as well. We modified the text such that we always use *'node'* or *'expert node'* in the manuscript, as opposed to *'vertex'*, *'edge'* as opposed to *'link'*, and *'graph'* as opposed to *'network'* unless the term is outside our mathematical framework like referring to a social, academic, or open-source network or an approach like a neural network, graph neural network or social network analysis.

R3.12

It would be useful to accompany the choice of notation attributes and variables with the number of papers that support each, either in Table 1 or in the text. For example, what percentage of the papers covered uses the attribute "expert's level of authority" (a_v)? What percentage of papers utilize the attribute "rarest skill"? Providing this justification is important to ensure that the notation accurately represents the field (as expected in a survey paper) and is not a modeling proposition by the authors.

Response: We appreciate your comment. As explained in response to your comments in R3.9 and R3.A, attributes and optimization objectives in this survey are from existing graph-based works and are not part of our contributions. In Table 2, we have summarized attributes and objectives, referencing their respective papers. However, the choice of mathematical notations within a unified framework is our contribution. For instance, communication costs have been represented with various notations in different studies (e.g., *sumDistance* [1] or $CC-R$ [81]). We have standardized this objective by ϕ, with the subgraph optimization algorithm as a mathematical subscript like ϕ_D for minimizing communication cost via the sum of distance, or ϕ_E using the sum of edge weights. In the Team Formation literature, proposed works have used different notations for the same attributes or objectives, and hence, a notation from a paper falls short of representing that attribute or objective across the papers in the field. To bridge the gap, we put forward our unifying set of notations.

R3.13

Further on comment #12, the specific attributes chosen also outline the scope of applications that the proposed unified model can be applied to, i.e. it does not apply to all expert team formation problems since (for example) not all of them necessitate the expert's "level of authority" or "hiring budget". Concretely, it seems that the model and survey suit a specific set of application cases, with the most apparent one being the use case of authors collaborating on research papers. Please clarify the application cases in the paper scope and contributions section to avoid making overgeneralizing claims.

Response: Thank you for your comments. As we explained in R3.9, our unified notations are designed to be inclusive of all existing and, potentially, future works in Team Formation. In response to your comment, please note that the attributes are defined as the node's attributes in an expert graph. On the one extreme and in case of no attributes in a setting, the attributed graph reduces to a simple graph. From graph theory, while an attributed graph is defined to have as many attributes as available in a problem setting, an objective function can be defined based on a subset of the attributes, e.g., in the case of hybrid optimization (Section **2.3 Hybrid (Multi-Objective) Optimization**), a single attribute, e.g., Section **2.2.1 Maximizing Proficiency** (Equation 9), or no attribute, e.g., Section **2.2.1 Minimizing Communication Cost**, (Equation 2). As seen, in all such cases, the graph expert G remains the same.

R3.14

Please elaborate on how equation (20) addresses the problem of zero edge weights.

Response: Thank you for your comment. Mathematically, the range for Equation 20 is in $(0, 1] \cup \{\infty\}$, hence it never equal to zero. From the Team Formation problem setting, if two experts have had many collaborations in the past, Equation 20 becomes a very small number (minimum communication cost) but never yields 0. If two experts have a low number of collaborations, e.g., just one time, Equation 20 becomes 1 (a large communication cost). If there is no collaboration at all, Equation 20 becomes $1/0 = \infty$ in communication cost.

R3.15

In Figure 6, please explain in the caption what the reader should observe. In general, please add more explanatory text related to the survey text in the figure captions, and also, please include the reference of the work from which the figure has been taken or refers to (if applicable).

Response: We are in total agreement on a multi-line explanation of a figure in the caption. However, due to space limits, we had to shorten the figures' captions and leave the explanation in the text body of the manuscript. We brought back the figures' explanation in Figure 1 and Figure 5 (originally Figure 6) in the revised manuscript. For instance:

"*Figure 5.* Graph extension of the Steiner tree algorithm where (middle) the graph extends by adding additional skill nodes and connecting them to all of the experts associated with that skill, and then, (right) original nodes are replaced with a complete graph of the size $|S_v|$ in such a way that each node in the complete graph is assigned a specific skill."

Also, all figures presented are original and created by us. Notably, most research papers in our survey lack visualization of their proposed algorithms, making this a key strength of our manuscript.

R3.16

In the Optimization Techniques section (Section 3), in addition to detailing the papers individually, it would be beneficial for the authors to provide an overview of the area and identify its gaps. For example, it appears that the research community has ceased working on Pareto search solutions for multi-objective optimization methods since 2014 (see Table 3). Is this correct? If so, why, in the authors' opinion, has this occurred? Is there a research gap that future studies could address? It would be valuable for each section (and possibly subsection) to include such remarks; otherwise, the survey risks being merely a useful enumeration of papers without offering additional insight.

Response: We cannot agree more with your comment. Indeed, the main purpose of our survey has been to review the existing works through a critical lens to identify the gaps and push further future research to bridge the gaps. As such, we brought forth the unifying notations, as well as visualizations to shed more light on the literature and identify the shortfalls. However, due to space limitations, we had to prioritize the gaps based on (1) their societal impact and immediate applications to real-world scenarios and (2) whether there is promise for potential solutions. We have explained the important ones in the manuscript, leaving the rest in the Appendices. We believe that Subsections **5.1 Fairness and Diversity** and **5.2 Temporality** are the research priorities in the Team Formation problem. The former has profound societal impacts in shaping how decisions are made in hiring for various domains like research, education and healthcare, among others. Also, there have been a growing number of studies on fairness-aware solutions that can be leveraged to fill this important gap in Team Formation algorithms. The latter is the immediate requirement of a Team Formation algorithm for real-world scenarios where skills and experts' skill sets are changing over time, with potential solutions. We also have explained other important gaps in the literature in Appendix **C Pitfalls in Evaluation** and Appendix **D Open-Source Implementation**.

Having said that, we have added a new Subsection **5.3 Modern Hybrid Optimization Algorithms** about modern optimization techniques that can be leveraged to improve the hybrid optimizations of objectives when forming multiple inter-related project teams are needed, as suggested by Reviewer 1 in R1.6. Specifically, we explained how the optimization techniques in existing graph-based Team Formation algorithms, like Pareto, fall short, and there have been better techniques like multi-factorial optimization.

PS. There was a typo in the last row of Table 3 where the position of the check mark was incorrect. Selvarajah et al. [34] have used Pareto search to solve the multi-objective problem.

R3.17

In Table 5, the "Number of non-dominated solutions" does not refer to any paper. Is this a typo?

Response: Thank you for raising this issue. Yes, it is a typo. 'Number of non-dominated solutions' belongs to Selvarajah et al. [34]; we have fixed Table 5 accordingly.

R3.18

Are all qualitative metrics of Table 5 present in Table 6 (and vice versa), and if not, why?

Response: Thank you for raising this issue. In the revised manuscript, we made the qualitative metrics consistent throughout the tables and typos are fixed.

R3.19

Sub-section 4.3 is particularly interesting and highlights a significant issue within the scientific community: the lack of universal benchmarks. It references Appendix A, but I was unable to find this appendix in the manuscript PDF. Please: (1) attach Appendix A to ensure proper evaluation of this subsection, and (2) elaborate further on this subsection, possibly including the authors' propositions.

Response: Thank you for your thoughtful feedback. We have included Appendix D, Open-Source Implementation, and included an outline for better benchmarking practices in the scientific community.

R3.20

Section 5 (Challenges and Future Directions) is one of the most interesting parts of the survey, but it requires significant reflection and elaboration, as it is among the most essential elements of the paper. For example, which modeling elements and algorithmic approaches have been consistently used over the years and can be considered most representative of the problem model? Which elements are case-specific? Which algorithmic approaches could be further explored in future research? Overall, what does a chronological examination of the literature reveal?

Response: Thanks for this enriching feedback. Based on this comment and within the constraints of the tight space limit, we have modified Section **5 Challenges and Future Directions**'s first paragraph, to include a general narrative of the evolution of the field, drawing on consistent fundamental approaches as well as incremental additions to improve performance.

R3.21

In Subsection 5.2, I advise the authors to expand their literature search for additional works and algorithms that incorporate an online property and/or consider the timeline for expert team formation. This could be achieved by snowballing from existing works and exploring additional application fields, such as expert crowdsourcing.

Response: Thank you for your timely suggestion, as online crowdsourcing platforms are indeed increasingly gaining traction. We have revised Subsection **5.2 Temporality** to include a paragraph on an immediate application for online, on-demand, and large-scale crowd work within crowdsourcing platforms, where experts hold portfolio and timeline, showcasing their successes and additional evidence of their expertise through previously established credentials over time.

[...] Temporal study of Team Formation problem finds immediate application for online, on-demand, and large-scale crowd work in crowdsourcing platforms [29, 48] where experts hold portfolio and timeline, showcasing their successes and additional evidence of their expertise through previously established credentials within time. In a crowdsourcing platform, candidate experts advertise their certified skills and bid prices for their participation while continuously adapting to changing task demands by upskilling or shifting their skills to remain relevant. While there has been Team Formation research on crowdsourcing platforms using social networks [29, 48], no work has considered the temporal aspect of skill set and experts' skills. [...]

R3.22

There are some minor typos (e.g., "the sought after a team"); I advise the authors to proofread the paper once more to enhance readability.

Response: We apologize for the oversight. We have proofread the manuscript and fixed the typos–here's a handful:

'Cunstruction' → *'Construction'*

"the sough after a team" → *"the search for a team"*

'cinsiders' → *'considers'*

"Barolli and Xhafa first collaboration" → *"Barolli and Xhafa's first collaboration"*

'Availabel' → *'Available'*

"has proxied" → *"has been estimated"*

"as is its required level of expertise" → *"x_{v_s} is its required level of expertise"*

A Survey of Subgraph Optimization for Expert Team Formation*†

MAHDIS SAEEDI, University of Windsor, Canada
HAWRE HOSSEINI, Thomson Reuters, Canada
CHRISTINE WONG, University of Windsor, Canada
HOSSEIN FANI ✉, University of Windsor, Canada

Team Formation is the search for gathering a team of experts who are expected to collaboratively work towards accomplishing a given project, a problem that has historically been solved in a variety of ways, including manually in a time-consuming and bias-filled manner, and algorithmically within such disciplines as social sciences, management, operations technology, and so forth. In the present effort, while providing a taxonomy to distinguish between search-based versus learning-based approaches, we survey graph-based studies from the search-based category, motivated as they comprise the mainstream. We present a unifying and vetted overview of the various definitions in this realm, scrutinize assumptions, and identify shortfalls. We start by reviewing initial approaches to the Team Formation problem to lay the conceptual foundations and set forth the necessary notions for a more grounded view of this realm. Next, we provide a detailed view of graph-based Team Formation approaches based on the objective functions they optimize. We lay out who builds on whom and how algorithms have evolved to solve previous works' drawbacks. Further, we categorize evaluation schemas and elaborate on metrics and insights that can be drawn from each. Referring to the evaluation schemas and metrics, we compare works and propose future directions.

1 Introduction

Algorithmic search for expert teams, also known as Team Formation, aims to automate forming teams of experts whose combined skills, applied in coordinated ways, can successfully solve difficult tasks [5, 112]. Some examples include a research team whose success can be measured by publications and citations in the scientific community, or the next blockbuster *'sci-fi'* movie with a touch of *'drama'* in the entertainment industry. Team Formation can also be seen as social information retrieval (Social IR) where the right group of experts are searched for and hired to solve the task at hand [67, 68]. Successful teams have firsthand effects on creating an organizational performance in academia [86], manufacturing [7], law [100], freelancing [4] and the healthcare sector [13], among others. Forming a successful team whose members can effectively collaborate and deliver the outcomes within constraints such as a 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 successful [105]. Further, the team members need to be *adaptive* to adjust to the environment and facilitate coordination [5].

Traditionally, teams were formed manually 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, and *iii*) an overwhelming number of candidates, among other reasons. On the one hand, the formation was heavily influenced by the team formers' 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 group [63], disproportionate selection of popular experts [49, 110], and over/under-representation and racial/gender disparities [78]. Additionally, since this process involves a multitude of criteria, including project importance, budget, time constraints

*We publicly released the artifacts in preparing this survey at `https://github.com/fani-lab/graph-based-team-formation-survey`.
†This survey includes appendices for further details.

and team size limitations, the decision-making is all the more difficult. On top of these, candidates for the formation of a team should ideally be examined based on several individual and relative factors such as technical abilities, availability, individual cost, productivity, behavior, personality traits, knowledge, negotiation skills, proactivity, and communicability, among others, which makes manual Team Formation on a large scale almost impossible. It has long been well-established in the fields of psychology and cognitive science that human individuals have limited capacity to store and process information, making large-scale computation beyond their capacity [9, 91]. Also, research in industrial engineering [85] has shown that the *size* and *variety* of components within a task, like the task of forming a team, are crucial factors in determining its complexity, which are limited for human performance. The limitations of human intuition and cognitive biases have been further shown in complex decision-making tasks [14, 59, 62]. Such studies necessitate an algorithmic approach, at least as a supplement, when dealing with tasks of large-scale size and with diverse components like team formation [39]. For instance, major global corporations with thousands of employees invest in developing expert recommender systems, such as IBM's SmallBlue [22], which exploits corporate email correspondences to assist in locating experts among the company's broad geographically distributed presence in various countries for a given project team.

In an effort to automate forming teams, researchers in different disciplines, such as operations research (OR) [11, 24, 35, 37, 104, 112], psychology [18, 55, 61, 77], management [3], engineering [10, 11], social network analysis [58, 72, 81, 102], and recently, artificial intelligence (AI) [15, 16, 43–45, 98] have long been endeavouring to find algorithmic solutions. 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 Formation remains the same essentially, while it has been referred to by such other names as team allocation, team selection, team composition, and team configuration. Despite the substantial number of algorithmic approaches to Team Formation, there is, however, yet to be a comprehensive survey with comparative analysis and critical reviews on approaches' applicability in real-world scenarios, especially when each comes with a domain-specific method and benchmark dataset with no standard implementation, incapable of accommodating different use-cases, not to mention the codebases and details are scarcely publicly available.

1.1 Existing Surveys

To the best of our knowledge, Wang et al. [108] and Juárez et al. [31] are the only surveys on *algorithmic* Team Formation. Wang et al. [108] is a pioneering survey that presents a comparative study of Team Formation algorithms on a reproducible, publicly available platform[1]. They re-implemented at-the-time (then) SOTA algorithms and evaluated them on four real-world datasets under the same experimental settings. However, Wang et al.'s work has the following limitations: *i*) lacks an analysis of Team Formation from the perspective of problem setting and the challenges in each setting like optimizing multiple objectives or the fact that experts can hold attributes that impact Team Formation like workload capacity; *ii*) only a limited number of methods using merely one optimization objective (communication cost), are covered, while other optimization objectives such as density or geographical proximity or methods that consider multiple optimization objectives in tandem (e.g., bi-objective optimizations) are overlooked; *iii*) it has been almost a decade since its publication, missing prominent state-of-the-art approaches to date.

More recently, Juárez et al. [31] has introduced a taxonomy with a special focus on the modeling aspects of the Team Formation problem at a conceptual level where teams are formed either through a match-making process between individual candidates and skills or via a search for a coherent community in the collaborative networks of candidates.

[1] `http://home.cse.ust.hk/faculty/wilfred/wangxinyu/TF.html`

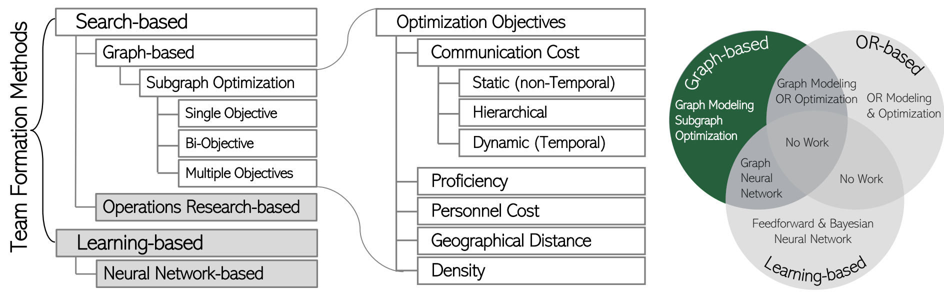


Fig. 1. (Left) A taxonomy of the computational Team Formation methods. (Right) A Venn diagram to highlight overlapping areas where operations research and learning-based methods leverage an expert graph for their optimizations. Gray areas are outside the scope of the current survey.

However, they forgo comparative technical study of the methods, especially *i*) the optimization objectives when forming an optimum team, *ii*) strategies to evaluate the performance of the methods, and *iii*) experimentation setups.

Team Formation has also been researched in social ecology, a field that studies the relationships between people and their environment and the interdependence of people and institutions [94, 103]~~. Paris et al. [94] have reviewed the theoretical perspectives on teamwork like the social-psychological approach pertaining to the social and psychological implications of team members' relationships/interactions with one another, or the~~ , and Science of Team Science (SciTS), which aims at understanding and enhancing the outcomes of large-scale collaborative *science* teams. Such studies are, however, *non*-computational where no algorithmic realization of such theoretical foundations has been explored.~~For instance, Paris et al. [94] propose a taxonomy of variables influencing team performance but forego the investigation of computational models for incorporating these variables when recommending experts for teams. Likewise, Stoklos et al. [103] explore different conceptual frameworks of team science in which several interpersonal, environmental, and organizational factors of collaboration are considered, such as the leadership styles of research directors, scientists' commitment to team research, and the availability of shared research lab; they propose strategies like brainstorming to create and integrate new ideas to deal with the cross-disciplinary biases and tensions that often arise in collaborative situations, and to negotiate and resolve conflicts.~~ In this survey, we consider such studies out of scope and only focus on algorithmic graph-based methods to form teams of experts applicable not only in science, as in SciTS, but also in any collaborative domain like the entertainment industry [2, 75] or crowdsourcing [29, 48].

1.2 Our Contributions and Scope Disclaimer

~~Effective collaboration among individuals is a crucial aspect of Team Formation. Graphs provide a useful way to represent the connections between people, making them valuable tools for studying Team Formation. In our survey, we introduce graph-based Team Formation algorithms. We explain how we construct the graph, discuss the evaluation metrics used, and identify any existing or ongoing comparative studies. These concepts have not been extensively explored in previous surveys, making our research original and innovative. A comprehensive *technical* review of proposed methods in terms of their formalization, evaluation methodology, and benchmark test beds is indeed needed to foster future research in the field.~~ In this survey, we present a novel taxonomy from a computational perspective, as shown in Figure 1. Team Formation approaches can be distinguished based on the way optimizations are performed: *i*) search-based, where the search for an *almost surely* successful team (optimum team) is carried out over the subgraphs of an expert graph using subgraph optimization methods, or it is performed on subsets of experts as variants of the set cover problem [19, 55, 55] using operations research (OR) techniques including integer linear/nonlinear programming, and *ii*) learning-based, where machine learning approaches are used to learn the distributions of experts and skills in the context of previous (un)successful teams in order to draw future successful teams.

Within the search-based category, operations research, by and large, optimizes the mutually independent selection of experts, overlooking the organizational and collaborative ties among individuals [24, 35, 71]. However, graph-based methods rely on the premise that a team is inherently relational and is a property of the interaction among the experts and how effectively they can collaborate. Following Cheatham et al. [8] and the seminal work by Lappas et al. [81], graph-based methods took the stage and became canonical in Team Formation literature. Recently, learning-based methods have been proposed to bring efficiency while enhancing efficacy due to their iterative and online learning procedure [15, 44]. Nonetheless, since effective collaboration among experts is a crucial aspect of Team Formation, graphs provide a useful way to represent the connections between experts, making them valuable tools for studying Team Formation. Therefore, we can observe the synergistic integration of expert graph in operations research [21, 37] and learning-based methods [43, 45, 98]. As seen in Figure 1 (right), there is an overlap between graph-based and operations research-based methods, where optimization functions have been defined based on linear or nonlinear equation of Boolean variables representing edges on the expert graph [37]. Additionally, learning-based methods utilize the expert graph to learn vector representations of skills using graph neural networks (GNNs), which helps reduce the complexity of neural models at the input layer [28]. Specifically, graph neural networks (GNN) [42, 53] provide an effective and efficient general framework for solving graph analytics problems by converting a graph into a low-dimensional vector space while preserving its graph-structured information. Having demonstrated strong performances across a wide range of problems, including natural language processing. [36], knowledge graphs [54], recommender systems [46], graph neural networks are gradually finding their application in Team Formation [28]. Successful as they are, learning-based literature is in its early stages with little work, as shown in Appendix B.3, Figure 9.

This survey pertains to the graph-based Team Formation algorithms, that is, those that employ graphs to model the experts' collaboration ties followed by subgraph optimization algorithms, as they comprise the mainstream body of research. We exclude works that are based on operations research and learning-based methods, as they differ fundamentally from subgraph optimization algorithms. We recognize the importance of these areas and the wealth of work they include, but a thorough analysis of them is beyond the scope of this work and merits separate surveys. We identify any existing or ongoing comparative studies within this scope, explain how expert graphs are formed, discuss the objectives for an optimum team, subgraph optimization algorithms and their complexities followed by workarounds for efficiency, and finally, elaborate on evaluation metrics used. These concepts have not been extensively explored in previous surveys, making our research original and innovative. A comprehensive *technical* review of the proposed methods in the literature in terms of their formalization, evaluation methodology, and benchmark test beds was indeed needed to foster future research in the field.

After screening 126 papers addressing Team Formation problems using computational and non-computational models, we present a comprehensive overview of 17 seminal graph-based research papers on the Team Formation problem within the scope of our survey, including 18 unique objectives, to be optimized via 63 subgraph optimization algorithms considering variations of their exact algorithms and the heuristics they used to address the efficiency of their algorithms. Our survey brings forth a unifying and vetted methodology to review the various definitions of the notions in this realm, criticizes assumptions and comparative benchmarks, and points out shortfalls to smooth the path for future research directions. It targets the information retrieval (IR) and recommender systems (RecSys) research communities to propose new Team Formation solutions and evaluate their effectiveness compared to existing methods and on datasets from various domains. Further, having regard to the unified comparative analysis, organizations and

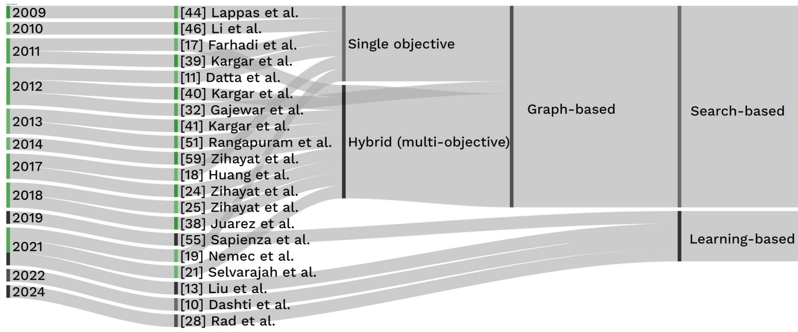


Fig. 2. ~~Team Formation methods in time; a substantial number of methods are graph-based.~~ Moved to Appendix B.3

practitioners can compare different models and readily pick the most suitable one for their application to form teams of experts whose success is almost surely guaranteed.

Below, we summarize our contributions in this survey:

(1) We study graph-based approaches to the problem of Team Formation utilizing a taxonomy of computational methods. To foster standards and conventions, which literature of this field lacks, we set forth a unified set of notations to formalize the problem and its different aspects conveniently. Our notations are designed to be a *minimal superset* of varied notations in papers while respecting their specifications. Our unified notations help with finding overlaps or correlations among formulas, which can improve understanding, simplify analysis, validate models, and avoid redundancy, while allowing seamless integration of novel ideas by reusing prior notations.
(2) We provide a detailed study of the graph-based optimization techniques used for tackling the Team Formation problem. In doing so, we adopt a chronological view that reflects the advancements over time, drawing upon the drawbacks of each method and how the subsequent works have sought to overcome them.
(3) While introducing the available datasets in this field and their drawbacks, we perform a comparative analysis of the performance of the various techniques in the field of graph-based Team Formation. Further, we show how the field suffers from the void of proper benchmarking, baselining and performance comparisons.
(4) Based on various open issues and challenges in the literature, we categorically set forth three major lines of research as future directions, namely *i*) fairness-aware, *ii*) time-sensitive, and *iii*) multi-factorial Team Formation. For each line of work, we thoroughly outline the shortcomings in the literature and value solutions to these problems. We believe that these future directions deserve more attention from the academic sphere.

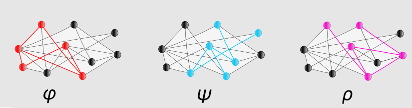
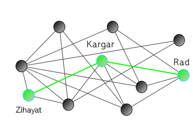
2 Problem Formalization

The graph-based modeling approaches tackle the Team Formation problem through defining subgraph optimization problems on a graph where the different aspects of real-world teams and the Team Formation process are explored. In this section, we formalize subgraph optimization problems in a unified framework with integrated notations for better readability and fostering conventions in this realm. Table 1 shows the summary of the notations.

2.1 Notations and Definitions

Definition 2.1. **Expert Graph.** In the Team Formation problem, a skillful individual who can be a candidate for a team is referred to as an *expert*. In graph-based approaches, an expert graph, also known as an expert collaboration network, is an undirected weighted and attributed graph denoted by $G = (V, A, E, w)$, where V represents a set of experts as nodes and A is the set of attributes associated to each expert as node attributes.

Table 1. Summary of the notations used in this survey. Examples are fictional.

Notation	Description	Example
$G = (V, A, E, w)$	Expert Graph, an undirected weighted attributed graph of experts (nodes) V, attributes A, edges E, and weight function w.	A co-authorship graph where two authors are connected if they publish at least one paper together.
$V = \{v_i\}$	The set of all experts (nodes) enumerated by i.	Computer science researchers with at least one paper.
v	An expert (node) of the expert graph.	*'Lappas'*.
$A_v = (S_v, a_v, b_v, c_v)$	An attribute for an expert (node) (v) in the expert graph that represents the expert's skills along with the level of expertise in each skill ($S_v \times \mathcal{R}^+$), the level of authority (a_v), budget (b_v) and capacity (c_v).	$A_{v:\text{'Lappas'}} = (S_v : \{(s:$ *'team formation'*$, x: 10), (s:$ *'recommender systems'*$, x{:}5)\}, a_v{:}24, b_v{:}\$150,000, c_v{:}5)$.
b_v	Required budget to hire expert v or her requested salary.	$b_{v:\text{'Lappas'}} = \$150,000$.
c_v	Maximum capacity in terms of the number of responsibilities or tasks that could be handled by expert v.	$c_{v:\text{'Lappas'}} = 5$.
a_v	The level of authority of expert v.	Researcher's h-index.
$S = \{s_j\}$	The set of all skills enumerated by j.	Research domains in computer science as in the ACM/IEEE computing classification system like {*'graph theory'*, *'social network analysis'*, *'operations research'*, ...}.
$S_v \times \mathcal{R}^+ = \{(s, x_{v_s}) : s \in S_v, x_{v_s} \in \mathcal{R}^+\}$	The subset of skills the expert v has along with the level of expertise for each skill (x_{v_s}).	$S_{v:\text{'Lappas'}} = \{(s :$ *'team formation'*$, x : 10), (s :$ *'recommender systems'*$, x{:}5)\}$.
x_{v_s}	The level of expertise in skill s for an expert.	Number of citations to the publications related to the skill s.
$V_s \subseteq V$	The subset of experts that are skillful for s.	$V_{s:\text{'team formation'}} = \{$*'Lappas'*, *'Kargar'*$\}$.
$s^\dagger$	The rarest skill with the fewest number of skillful experts, i.e., $\lvert V_{s^\dagger}\rvert = min_{s \in S} \lvert V_s \rvert$.	$s^\dagger$: *'theory of fairness'*.
E	The set of all edges of the expert graph.	The set of all collaborations.
$e_{v,v'}$	A weighted edge of a graph that represents a connection between a pair of experts v and v'.	Co-authorship in a publication.
$w_\uparrow(e_{v,v'})$	The weight of the edge e, desired to be high.	The number of successful joint publications between experts v and v'.
$w_\downarrow(e_{v,v'})$	The weight of the edge e, desired to be low.	Geographical distance between experts v and v'.
p	A project yet to be successfully accomplished.	Research project on *'fairness-aware team formation'*.
$S_p = \{(s, x_{v_s}, l_s, u_s)\} \subseteq S \times \mathcal{R}^+ \times \mathcal{N} \times \mathcal{N}$	The predefined subset of skills required by the project p. For each skill $s \in S_p$, l_s and u_s denote the lower and upper bound on the number of required experts, respectively, and x_{v_s} is its required level of expertise.	$\{(s :$ *'graph theory'*$, x_{v_s} : 100, l_s : 1, u_s : 2), (s :$ *'theory of fairness'*$, x_{v_s} : 1, l_s : 2, u_s : 3)\}$.
l_s	The lower bound on the number of required experts for skill s.	$l_s = 3$ for a research project that needs at least 3 postdoctoral fellows.
u_s	The upper bound on the number of required experts for skill s.	$u_s = 6$ for a research project that needs at most 6 postdoctoral fellows.
$V_p \subseteq V$	The subset of experts in the expert graph whose attributes include at least one of required skills S_p for project p. It represents the subset of experts that collectively cover required skills for p.	$V_{p:\text{'fairness-aware team formation'}} = \{$*'Lappas'*, *'Schwiegelshohn'*$\}$.
$G[V_p] = (V_p, A_p, E_p, w)$	An induced subgraph with the experts V_p that represents the prediction for an optimal subset of experts who will almost surely accomplish the project p successfully.	
$\mathcal{P}(G)$	Power set of a graph including all subgraphs of a graph	All possible teams that can be formed.
$\varphi, \phi, \rho, \chi, \psi(G[V_p])$	Optimization functions that seek to optimize overall communication cost (φ), proficiency (ϕ), personnel cost (ρ), geographical distance (χ) or density (ψ) within a subgraph $G[V_p]$ of the expert graph G.	
$d(v, v')$	Distance between experts (nodes) v and v' in the expert graph G defined as sum of edge weights of the shortest path between v and v'.	$d($*'Zihayat'*, *'Rad'*$) = 2$, both of them have joint papers with *'Kargar'*, but never with each other. 

For instance, the nodes V could be researchers in an academic network (e.g., DBLP), software developers in an open-source network (e.g., GitHub), inventors in patents (e.g., USPTO) or cast and crew in moving picture industry (e.g., IMDB). The node's attributes generally express the capabilities of an expert from different perspectives, including:

- Expert's skillset S_v, which is a subset of the entire set of skills S, i.e., $S_v \subseteq S$, that represents the expertise of the expert v such as research domains for a researcher like {*'graph theory'*, *'social network analysis'*, ...}, programming languages for a developer like {*'python'*, *'java'*, ...}, patent classes for an inventor like {*'textile'*, *'engineering'*, ...}, or movie genres for an actress like {*'sci-fi'*, *'comedy'*,...};
- Expert's level of expertise per skill (also referred to as *authority* per skill) $x_{v_s} \in \mathcal{R}^+$ given a skill $s \in S$, that indicates the level of competence or excellence of the expert v in the skill s among her peers like *'the number of citations'* to the publications related to the skill s;
- Expert's level of authority a_v that indicates the *overall* competence or excellence of the expert among her peers, like the *'h-index'* of a researcher in an academic network. This is different from level of expertise, which refers to competence per skill;
- Expert's hiring budget b_v, that can be the requested or allocated salary for an expert position in a project; and,
- Expert's capacity c_v, the maximum number of responsibilities that an expert can handle.

The subset of skills an expert v has along with her level of expertise in the skills is denoted by $S_v \times \mathcal{R}^+$. Inversely, for a given skill $s \in S$, $V_s = \{v \in V : (s, x_{v_s}) \in S_v \times \mathcal{R}^+\}$ denotes the set of all experts owning it. A rare skill $s^\dagger$ is the one possessed by the fewest number of experts, i.e., $|V_{s^\dagger}| = min_{s \in S}|V_s|$. In total, given an expert v, her attribute is a tuple $A_v = (S_v, a_v, b_v, c_v)$. For example, the attribute for the expert v = *'Lappas'* would be $A_v = (S_v : \{(s :$ *'team formation'*$, x_{v_s} : 10), (s :$ *'recommender systems'*$, x_{v_s} : 5)\}, a_v : 24, b_v : \$150,000, c_v : 5)$ whose levels of expertise in *'team formation'* and *'recommender systems'* are 10 and 5, respectively, based on *'the number of citations'* to the expert's publications in each of the research domains, and 24 being the expert's level of authority based on *'h-index'*. As will be explained in Section 3.1 (Expert Graph Construction) and Section 4.1 (Datasets), an expert's skills S_v may not be self-declared and should be inferred via an assumption based on an underlying benchmark dataset. For instance, in the academic network, an expert is an author of one or more publications and her skills can be set to the words in the titles of her publications after removing the stopwords [17, 72, 81, 83].

Next, E represents the connection between the experts as graph edges and $w(e_{v,v'}) : E \longrightarrow \mathcal{R}$, is the weight function that maps an edge $e_{v,v'} \in E$ between two experts v and v' to its weight. Edges can represent experts' previous collaborations, followership in a social network, or spatial proximity, and their weights $w(e_{v,v'})$ can be based on the number of joint projects, number of recent communications, or the geographical distances in kilometres between experts v and v'. In graph-based methods, the weight function is the primary element of the optimization process based on which an optimization function is to be maximized or minimized. For instance, experts who have collaborated the most in the past can communicate the best (technically, have the lowest communication cost) [1, 12, 17, 25, 40, 41, 72–74, 81, 111]; therefore, they are the optimum choice for an *almost surely* successful team. In this case, higher edge weights between experts are desired and can be indicated by $w_\uparrow(e_{v,v'})$. In contrast, or rather meanwhile, the less geographical distance, the better/easier communication can be observed [34] such that lower edge weights are desired and can be shown as $w_\downarrow(e_{v,v'})$.

Definition 2.2. **Experts Distance.** The distance between a pair of experts v and v' in the expert graph G, $d(v, v')$, is defined as the sum of edge weights of the shortest path between v and v'. Given the weight function $w(e_{v,v'})$, the shortest path from v to v' is the path $(v_1 \rightarrow v_2 \rightarrow \ldots \rightarrow v_n)$ including n intermediary experts, where $v_1 = v$ and $v_n = v'$,

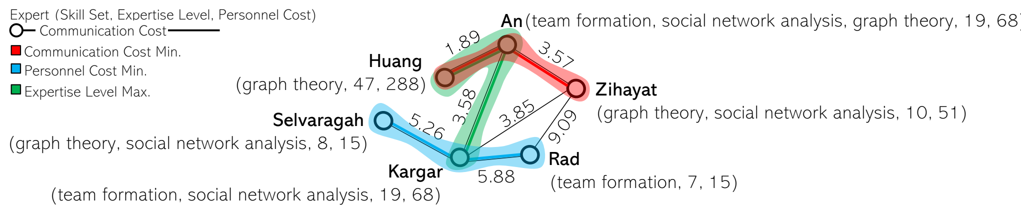


Fig. 3. An example of subgraph optimization using different objectives for the Team Formation problem.

that minimizes the sum in $\sum_1^{n-1} w(e_{v_i,v_{i+1}})$ over all possible n. While Floyd-Warshall algorithm [56] solves all pairs shortest paths in $\Theta(V^3)$, a shortest path, for efficiency, can be calculated stochastically based on a random walk [32], that is, random selection of experts (nodes) on a path (single steps of a walk) from v to v' based on a probability distribution, e.g., uniformly where the adjacent experts are selected with the same probability, or non-uniformly such as in relation to the edge weight $\text{P}(v_{i+1}|v_i) \propto w(e_{v_i,v_{i+1}})$ assuming Markov property. Further, the distance between an expert v and a subgraph $G[V_p], V_p \subseteq V$ is defined as:

$$d(v, V_p) = \min_{v' \in V_p} d(v, v') \tag{1}$$

Definition 2.3. **Project Team.** From an organizational perspective, the search for a team is determined for a given project p with a predefined required subset of skills $S_p \subseteq S$. In graph-based methods, a team is estimated by an induced subgraph of the expert graph G, denoted by $G[V_p]$, with nodes $V_p \subseteq V$ representing the team's members such that for each required skill $s \in S_p$ there should be *at least one* expert v whose skill subset S_v includes s. Therefore, the team's members are required to *collectively* cover the project's required subset of skills S_p, that is, $\forall s \in S_p, \exists v \in V_p : s \in S_v$ and so $S_p \subseteq \bigcup_{v \in V_p} S_v$. Further, an organization may limit the number of experts in a team per each required skill of a project along with a predefined level of expertise. For skill $s \in S_p$, we denote l_s and u_s for lower and upper bound on the number of required experts, respectively, and x_{v_s} is its required level of expertise. Hence, the predefined required subset of skills $S_p \subseteq S$ for a project p becomes a set of tuples $(s, x_{v_s}, l_s, u_s) \in S_p$ where $S_p \subseteq S \times \mathcal{R}^+ \times \mathcal{N} \times \mathcal{N}$.

For example, to form a research team for a project on '*fairness-aware team formation*' with a required subset of skills (research domains) $S_p = \{$'*graph theory*', '*theory of fairness*'$\}$ with one or two senior researchers having more than 100 publications in '*graph theory*', and 2 to 3 junior researchers with at least one publication in '*social network analysis*', S_p would become $S_p = \{(s\text{:}$'*graph theory*'$, x_{v_s}\text{:}100, l_s\text{:}1, u_s\text{:}2), (s\text{:}$'*social network analysis*'$, x_{v_s}\text{:}1, l_s\text{:}2, u_s\text{:}3)\}$.

Once the expert graph G is created and the project p is defined, the next step involves the search for a subgraph $G[V_p]$ whose experts (nodes) V_p would form a team. This step faces two important challenges: *i*) there may be more than one solution, that is, two or more subgraphs whose experts collectively cover the required skill subset of the project p, and *ii*) the search over all possible subgraphs of G is computationally prohibitive in medium to large-scale expert graphs. A common greedy method would involve selecting those subgraphs whose experts have had many previous successful collaborations. However, this is not always possible due to, for example, the unavailability of such experts. Also, there are many instances of successful collaborations in teams whose members meet for the first time (no prior collaborations). For instance, '*Barolli*' and '*Xhafa*''s first collaboration in 2006 [109] ignited their long-lasting successful collaborations *thereafter* on 220+ research publications until 2019. To find an optimum subgraph (team) of experts efficiently (fast), optimization objectives have been proposed based on varied assumptions to reduce the search space. In optimization, subgraphs that maximize or minimize a selected objective or a combination of objectives in tandem are of interest. In the following, we formalize such optimization objectives and lay out their details.

Table 2. An overview of objectives employed in subgraph optimizations for Team Formation problem.

			Communication Cost						Proficiency						
	Year	Hybrid	Diameter	MST	Leader Distance	Sum of Edge Weights	Random Walk	Sum of Distance	Trust Score	Expertise Level	Connector Authority	Skill Holder Authority	Personnel Cost	Geographical Distance	Density
Lappas et al. [81]	2009		✓	✓											
Li et al. [83]	2010			✓											
Farhadi et al. [25]	2011	■	✓							✓					
Kargar et al. [72]	2011				✓			✓							
Datta et al. [17]	2012		✓	✓											
Kargar et al. [73]	2012	■						✓					✓		
Gajewar et al. [57]	2012														✓
Kargar et al. [74]	2013	■	✓					✓					✓		
Kargar et al. [1]	2013	■	✓					✓					✓		
Rangapuram et al. [96]	2013	■												✓	✓
Zihayat et al. [111]	2014	■						✓		✓			✓		
Huang et al. [30]	2016				✓			✓							
Zihayat et al. [40]	2017	■				✓					✓	✓			
Zihayat et al. [41]	2018	■				✓					✓	✓	✓		
Juarez et al. [70]	2018	■								✓					✓
Nemec et al. [32]	2021						✓								
Selvarajah et al. [34]	2021	■						✓	✓					✓	

2.2 Subgraph Optimization Objectives

Given a graph G and a project p, the task is to find an *optimal* subgraph $G[V_p]$ as the recommended team, given an objective function with a set of constraints. Examples of such objectives are a team of the highest expertise, minimum geographical distance or communication cost, or all in tandem summing over *all* its members. As seen in Figure 3, given an organization aims to hire a team of experts for a project p requiring a subset of skills S_p={*'team formation', 'social network analysis', 'graph theory'*}, with candidate experts connected in an expert graph based on their joint project teams in the past and the edges being weighted based on '*the more joint project teams the less communication cost*' (Equations 20 or 19), different optimum subgraphs are selected depending on an objective we seek to optimize; (red) communication cost (Equation 3): {*'Huang', 'An', 'Zihayat'*}, (cyan) personnel cost (Equation 14): {*'Rad', 'Kargar', 'Selvarajah'*}, or (green) expertise level (Equation 10): {*'Huang', 'An', 'Kargar'*}. Table 2 shows an overview of the objectives studied in the literature. In the following, we elaborate on these objectives and their backgrounds.

2.2.1 Minimizing Communication Cost (φ)

A key team performance indicator is how effectively experts communicate, and is measured by a metric well-known as communication cost; a lower communication cost in a team indicates easier communication, better understanding and collaboration among team members, and hence, more likely for that team to succeed. For instance, a friend or a long-life colleague can be reached out to quickly and effortlessly through various media like messaging services or direct phone calls, whereas a more formal request or an appointment through administrative steps is needed to meet a manager at a higher level of an organization's hierarchical network structure like a dean or chair of a department. Hence, the communication cost of the first case is much less than the second one.

Communication cost has been the prominent objective to be minimized among the graph-based methods and is quantified based on *i*) previous joint (successful) collaborations [1, 17, 30, 40, 41, 72, 81, 111], *ii*) the time of the last collaboration [34], as well as *iii*) the similarity between cultural backgrounds and languages [34]. Though, communication cost has primarily been estimated based on the previous joint collaborations assuming an equality condition (if and only if) between communication cost and team performance (success) where low (minimum) communication cost is the *necessary* and *sufficient* condition for team performance. In other words, when low communication costs in the past lead to past successful collaborations, past successful collaborations would also indicate low communication costs in a yet-to-be-formed team in the future and would yield the team's success. Looking closely at this assumption, communication cost objective becomes a mediator from past teams' successes to signal new successful ones.

Once the edges of the expert graph G are weighed based on communication costs, the optimum team is a subgraph within which the overall communication cost is minimized based on, e.g., the summation of edge weights in the subgraph's diameter [1, 17, 81], its spanning tree [81, 83], or its entirety [40, 41], among others, as detailed below. We identify three types of search algorithms based on communication cost minimization in the literature ~~depending on the temporality of the communications and graph structures~~: non-temporal (static), hierarchical, and temporal (dynamic).

Non-Temporal (Static): Several *static* optimization functions $\varphi : \mathcal{P}(G) \rightarrow \mathcal{R}^{+}; \varphi(G[V_p])$ have been defined to measure the overall communication cost within a subgraph of G among all possible ones $\mathcal{P}(G)$, overlooking the time dimension and the fact that experts' skills and social attributes are constantly changing over time:

- *Sum of distances* (φ_D) [32, 34, 72, 73, 111]:

$$\min_{V_p \in \mathcal{P}(G)} \varphi_D(G[V_p]) = \sum_{v,v' \in V_p} d(v, v') \tag{2}$$

  where $d(v, v')$ is the distance between two experts and should be calculated between all pairs of experts, each contributing to the subgraph's overall communication cost. Using random walks, should $\varphi_D(G[V_p])$ be the lowest, it is expected that a *random* communication between a random pair of v and v' in V_p is also the lowest, and hence, V_p would be an optimum team for the project p.
- *Sum of edge weights* (φ_E) [40, 41] which is defined as:

$$\min_{V_p \in \mathcal{P}(G)} \varphi_E(G[V_p]) = \sum_{e \in E_p} w_{\downarrow}(e) \tag{3}$$

  where $w_{\downarrow}(e)$ is the communication cost based on a pair of experts. Herein, the summation is over all edges of a subgraph, in contrast to Equation 2 where only edges in shortest paths are considered.
- *Diameter* (φ_R) is the largest path among shortest paths between pairs of experts in the subgraph [1, 17, 81]

$$\min_{V_p \in \mathcal{P}(G)} \varphi_R(G[V_p]) = \max_{v,v' \in V_P} d(v, v') \tag{4}$$

  Intuitively, as shown in Figure 4 (left)(a), the diameter signifies the longest distance, and hence, the highest communication cost possible between a pair of experts as a worst-case scenario in a team. Equation 4 is used to search for an optimum team whose worst communication cost is minimal.
- *Cost of the spanning tree* (φ_{MST}) is the sum of the weights of the minimum spanning tree (MST) on the subgraph ($G[V_p]$) [81, 83]:

$$\min_{V_p \in \mathcal{P}(G)} \varphi_{MST}(G[V_p]) = \sum_{e \in MST(G[V_p])} w_{\downarrow}(e) \tag{5}$$

  A minimum spanning tree (MST) is a tree subgraph including all the nodes of a graph (spans all nodes) in such a way that the total edge weights is minimum among all possible tree subgraphs of the graph, as shown in Figure 4 (left)(b). Herein, the subgraph's MST represents the best-case scenario where a team of connected experts has the lowest possible communication cost overall. Equation 5 is used to find the optimal team with the minimum overall communication cost across all subgraphs, represented by their MSTs in the best-case scenarios. Sometimes, as seen

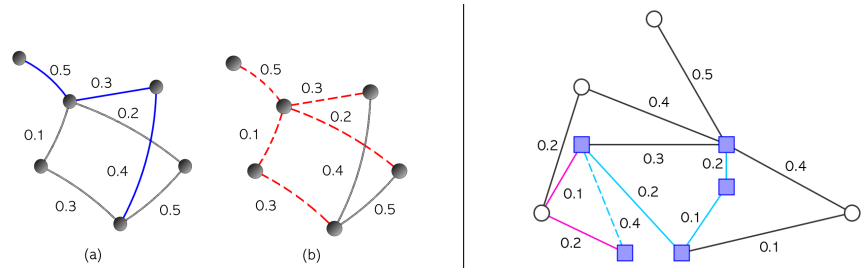


Fig. 4. Left: Minimum subgraph based on (a) diameter (solid blue edges) vs. (b) spanning tree (dashed red edges). Right: Minimum spanning tree vs. Steiner tree. Given a subgraph with square blue nodes, a minimum spanning tree has the minimum cost 0.9. However, adding extra nodes from outside (the circle white nodes) forms a Steiner tree and reduces the minimum cost to 0.8.

in Figure 4 (right), adding extra nodes to a subgraph from outside can lead to forming a tree with a smaller cost. The resulting tree is referred to as the Steiner tree of a subgraph. While a minimum spanning tree spans all nodes of a given subgraph, a Steiner tree spans extra nodes in an effort to reduce the cost.

Hierarchical: On some occasions, a team pertaining to a hierarchical structure is desired, like having a leader who is responsible for intra- and inter-team communications. Leader distance $\varphi_L(G[V_p])$ in such teams is defined as the sum of all distances between the leader L and other members as follows:

$$\min_{V_p \in \mathcal{P}(G)} \varphi_L(G[V_p]) = \sum_{v \in V_p} d(v, L) \tag{6}$$

In a more general case, levels of hierarchy can be increased to have a team of *sub*teams, in which the team's leader L is in contact with each subteam's leader L_i. Accordingly, hierarchical communication cost [30, 34, 72] is defined as:

$$\min_{V_p \in \mathcal{P}(G)} \varphi_H(G[V_p]) = \sum_{v \in V_i} d(v, L_i) + \sum_{i} d(L_i, L) \tag{7}$$

where V_i is the set of experts in the i-th subteam and $V_p = \bigcup_i V_i$.

In a cross-functional team, where experts from various seemingly unrelated disciplines are needed and having experts directly work together on a joint task could be challenging, subteams are formed to include experts with related skill sets. For example, a city council may have the mayor as the leader of a team with a lead urban architect, a chief environmental planner, and a chief police officer, each of whom is a leader of subteams of architects, environmental planners, and police officers, respectively. As another example, a research laboratory can have the principal investigator as the lead for a team of postdocs, each leading a subteam of graduate students.

Temporal (Dynamic): Most Team Formation methods face a significant challenge when *currency* (timeliness) is of the prime concern. They overlook that experts' skills and social attributes constantly change over time. A successful collaboration of experts in a team *years ago* does not tailor a successful team *now*. Also, experts enjoy varying behaviour patterns and propensities, leading to distinct temporal behaviours in response to popular skill sets. While temporality has been well-explored in other disciplines like social network analysis [72, 81], recommender systems [46], and information retrieval [89], the positive impacts of considering the time have been scarcely studied in Team Formation.

Temporal (dynamic) communication cost is based on the fact that the *least* communication cost exists between experts who could maintain many successful collaborations over time until recently or currently; the factor that might have caused the termination of collaboration in the *past* may be an obstacle to re-cooperation *now* or, the experts' interests toward skills might have changed. Further, recent few successful collaborations should outweigh successful ones in the far past. Accordingly, temporal communication cost is defined as the sum of distances with temporal regularization [34]:

$$\min_{V_p \in \mathcal{P}(G)} \varphi_{DT}(G[V_p]) = \sum_{v,v' \in V_p} d(v,v') + \alpha(t-t') \tag{8}$$

where $d(v,v')$ is based on Definition 2.2, $(t-t')$ denotes the time gap since the last (most recent) collaboration for a pair of experts; the longer the time gap the more costly the communication, and α is an attenuation factor.

2.2.2 Maximizing Proficiency (ϕ)

Proficiency of experts indicates the level of expertise in a particular profession or a skill. *Mastery level* [34], *expertise level* [70], *authority* [40, 41] and *trust score* are proposed to measure proficiency based on some criteria, for example, *'h-index'* or *'number of citations'* in academic networks. Given a graph G, the optimization task aims to find the optimum subgraph $G[V_p]$ that maximizes the proficiency objective function ϕ, defined based on the sum of node weights:

- The *mastery level* [34] of a team $G[V_p]$ is the sum over the team members' expertise among all of their skills:

$$\max_{V_p \in \mathcal{P}(G)} \phi_M(G[V_p]) = \sum_{v \in V_p, s \in S_p} x_{v_s} \tag{9}$$

  Note that x_{v_s} can be dynamic [34]; an expert becomes more professional in a skill as she gains more experience.
- The *total expertise level* [70] of a team $G[V_p]$ is the value of mastery level normalized by the size of the team:

$$\max_{V_p \in \mathcal{P}(G)} \phi_E(G[V_p]) = \frac{\phi_M(G[V_p])}{|V_p|} \tag{10}$$

  where $|V_p|$ is the size of the subgraph $G[V_p]$ that collectively covers the required skills for p.
- The *skill holder authority* is defined in conjunction with *connector authority* where experts holding the required skills may not be directly connected. Therefore, a subgraph $G[V_p]$ representing a team has two types of nodes: skill holders, H, who have the required skills, and connectors, C, who lack the skills, yet connect skill holders. For instance, in real-life scenarios, a reputable project manager (e.g., a professor) may lack the required skills (e.g., a programming language) but can facilitate communication between experts who hold the skills (e.g., programmers). Accordingly, the proficiency of a team is defined through the help of skill holders and connectors separately as [40, 41]:

$$\min_{V_p \in \mathcal{P}(G)} \phi_A(G[V_p]) = \sum_{v \in A} \frac{1}{a_v}; A \in \{H, C\} \tag{11}$$

  where the sum of the inverse of authority is minimized instead of maximizing the sum of authorities since it can be jointly employed with other objectives whose minimizations are required, like with communication cost objective under a single optimization function, as explained further in Section 2.3 (Hybrid (Multi-Objective) Optimization).
- The trust score measures the trust between a pair of experts v and v', and is of special importance when two experts have never interacted with one another in the past. It can be calculated based on [34]:
  i) *explicit trust score* ($\varkappa$) valued between −1 to +1 based on the number of collaborations (edges) between a pair of experts. The trust score is 0 if there has been no collaboration between two experts in the past.
  ii) *profile similarity score* (F) based on the similarity of skills or number of common skills between a pair of experts. This score is rooted in social psychology studies [90], which have shown that people with similar interests tend to trust each other more.
  iii) *emotional intelligence index* (ϖ), which refers to the ability of an expert to perceive, control and evaluate emotions in oneself and others to make decisions, solve problems, and communicate.

The final trust score between a pair of experts is denoted by $\tau(v, v')$ and is defined as the weighted sum of explicit trust, profile similarity and emotional intelligence index:

$$\tau(v, v') = \alpha_1 \varkappa(v, v') + \alpha_2 F(v, v') + \alpha_3 \varpi(v) \tag{12}$$

where $\sum \alpha_i = 1$. The *collective* trust score of a subgraph $G[V_p]$ is the sum of trust scores over all team members:

$$\max_{V_p \in \mathcal{P}(G)} \phi_\tau(G[V_p]) = \sum_{v,v' \in V_p} (\tau(v, v') + \tau(v', v)) \tag{13}$$

Note that the trust score is not commutative [34]; the trust score of a team is the sum of the trust scores of expert v to v' *and* v' to v for all pairs of experts v and v' in the team.

2.2.3 Minimizing Personnel Cost (ρ)

In the real world, experts are compensated financially for their efforts on a given project. The amount of compensation is relative to their expertise level as well as the available budget for the project. Thus, it is crucial to form a team with a reasonable amount of financial compensation based on the project funds. To this end, the personnel cost of a team, denoted by $\rho(G[V_p])$, is the sum of the required budgets to hire team members working on project p [41, 73]:

$$\min_{V_p \in \mathcal{P}(G)} \rho(G[V_p]) = \sum_{v \in V_p} b_v \tag{14}$$

which is to be minimized as the less money spent on a project, the better, given the project is accomplished successfully.

2.2.4 Minimizing Geographical Distance (χ)

In the search for an optimal team, companies also look for experts in the geographical region where the organization is based. This leads to new challenges as it requires drilling down on the skills of experts while maintaining the condition of a given geolocation, which is not necessarily met under a skill-driven selection of candidates. Although remote work via online platforms has facilitated today's globalized work environment, geographical proximity remains important for different reasons, including enabling face-to-face interactions, promoting cultural understanding, time zone differences, and the necessity of access to such local resources as certain region-locked services by companies [34], and so forth. These factors can impact team dynamics, coordination, and effectiveness [112]. For example, forming a team of experts from different time zones, e.g., from GMT and EDT time zones, where the business hours/days of one expert are non-working/resting periods for another, heavily discounts the efficiency of communication and accrue more costs associated with time, effort, and resources. Out of the few works that have considered geographical proximity, Selvarajah et al. [34] defined an objective function as the sum of geographical distances between members as follows:

$$\min_{V_p \in \mathcal{P}(G)} \chi(G[V_p]) = \sum_{v,v' \in V_p} d_{geo}(v, v') \tag{15}$$

where $d_{geo}(v, v')$ is the geographical distance between v and v' in G.

2.2.5 Maximizing Density(ψ)

Assuming that a greater number of collaborations between team members in the past increases the likelihood of success for the team, a *dense* subgraph with many internal edges has been sought to represent an optimum team. It is expected that increasing the density of a subgraph yields a team with stronger collaboration and less communication cost [57, 96]. The density of a subgraph is defined as the ratio of the total number of edges within the subgraph to the number of its nodes or, in weighted graphs, the sum of edge weights in a subgraph to the sum of the node weights; intuitively, the greater the number of edges compared to nodes in a graph, the denser the subgraph. Formally, the density of a

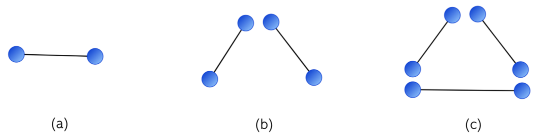


Fig. 5. Based on Equation 18, the density of graph (a), (b), and (c) are 1.0, $\frac{1}{3}$, and $\frac{1}{5}$, respectively, but by Equation 17 and 16, the density of all these graphs are equal to $\frac{1}{2}$. Here, all edge weights and node weights are assumed to be 1.0.

subgraph is defined as [96]:

$$\max_{V_p \in \mathcal{P}(G)} \psi(G[V_p]) = \frac{\sum_{v,v' \in V_p} w_{\uparrow}(e_{v,v'})}{\sum_{v \in V_p} w(v)} \tag{16}$$

where $w(v)$ is the weight of the node v. In the case of unweighted nodes, $w(v) = 1.0$ and Equation 16 is simplified to:

$$\max_{V_p \in \mathcal{P}(G)} \psi(G[V_p]) = \frac{\sum_{v,v' \in V_p} w_{\uparrow}(e_{v,v'})}{|V_p|} \tag{17}$$

Alternatively, the density of a subgraph is defined by Juarez et al. [70] as:

$$\max_{V_p \in \mathcal{P}(G)} \psi(G[V_p]) = 2 \times \frac{\sum_{v,v' \in V_p} w_{\uparrow}(e_{v,v'})}{|V_p|(|V_p| - 1)} \tag{18}$$

where density decreases quadratically by increasing the number of team members, favoring teams of small sizes, as opposed to Equation 16 and 17. Figure 5 compares the density formulas in sample graphs. As seen, while the density based on Equation 16 or 17 is equal to $\frac{1}{2}$ for all graphs, the graph in (a) has the highest density based on Equation 18.

2.3 Hybrid (Multi-Objective) Optimization

The previous sections introduced and formalized various objectives crucial in Team Formation literature. The interplay between different objectives in forming a team is also important to consider so as to model real-world circumstances more accurately. For example, companies may demand experts possessing the skills required to accomplish the tasks, and yet such experts must be in a region where the organization is geographically located while limiting the total personnel cost to the project budget. In the literature, the optimization of the following combinations of objectives is studied in a *multi-objective* optimization framework:

- Communication cost (§2.2.1) and proficiency (§2.2.2) where teams with minimum communication costs and maximum expert proficiencies are desired. This is motivated by social science studies showing that collecting a group of the best experts in each required skill does *not* necessarily ensure success as they may lack proper communication, among other reasons [77]. Zihayat et al. [40, 41] studied variations of proficiency in terms of authority metrics for skill holder experts and connector experts (Equation 11) in conjunction with communication cost.
- Communication cost and personnel cost (§2.2.3) [1, 41, 73, 74] where finding a team of skillful experts with minimum communication cost and financial compensation (salary) is desired. This is motivated by real-world scenarios where a project is constrained by a limited amount of money while different experts incur different compensations for performing the tasks of the project.
- Communication cost, personnel cost and proficiency (§2.2.2) [111] where team members can smoothly communicate with each other *and* hold the highest level of expertise (Equation 9) in the required skills while the total amount of money paid to them is minimized. Therefore, communication and personnel costs are to be minimized, yet proficiency is to be maximized jointly.
- Dynamic communication cost, geographical proximity (§2.2.4), and proficiency where communication cost and geographical distances between experts of a team are to be minimized while experts' proficiency is to be maximized.

Selvarajah et al. [34] proposed to maximize mastery level and trust score (§2.2.2) as representations for experts' proficiencies, and to minimize communication cost and geographic distances to find an optimum team.

- Density (§2.2.5) and proficiency [70] where the mostly connected experts with the highest proficiencies within a subgraph are desired as an optimum team. As explained earlier, this is motivated by the fact that highly skillful experts who have collaborated frequently in the past are likely to succeed if they group together in a new team. Maximizing density and proficiency disfavour collaboration with unknown, new or early career experts.

2.4 Optimization Constraints

As explained, subgraph optimization objectives on the expert graph G aim to identify an *optimal* subgraph $G[V_p]$ as the recommended team for a project p. This process is typically performed within a set of *predefined* constraints, also referred to as boundaries, restrictions or criteria, that any feasible subgraph $G[V_p]$ for an optimum team must satisfy. The most common constraint has been skill coverage [1, 25, 40, 57, 72, 73, 81, 83, 96], that is, the expert nodes of the optimum team must collectively cover the required skills for the project p, i.e, $\forall s \in S_p, \exists v \in V_p : s \in S_v$ and $S_p \subseteq \bigcup_{v \in V_p} S_v$. Other constraints include team size (cardinality) [30, 96], availability of experts [34], capacity (c_v) of experts [17], required budget for the project [96], i.e, $\sum_{v \in V_p} b_v$, and a minimum number of experts with specific skills per team [57], i.e., l_s. Table 9 in Appendix E provides a summary of the constraints studied in this domain along with the approaches that incorporate them. While constraints are domain-dependent and vary based on the underlying organizational requirements, they have been addressed during the optimization process by the optimization algorithms. For example, in the case of skill coverage, an algorithm foremost identifies expert nodes that possess at least one of the required skills, disregarding the rest of expert nodes [30, 40]. Similarly, when constraints such as an upper limit on team size are present, an algorithm can check the number of experts with the required skills at each iteration and continue searching neighboring nodes until the constraint is satisfied [57].

Thus far, we have formalized Team Formation problems and the optimization objectives that are defined to find optimum teams, mostly related to the question of *'what are the optimization objectives?'*. Next, we lay out details of the proposed techniques developed to provide solutions to the optimization problem, i.e., *'how are the objectives optimized?'*.

3 Optimization Techniques

Subgraph optimization problems are proven to be NP-hard [76], and so finding an optimum team via subgraphs of the expert graph optimizing an objective function is computationally prohibitive. Therefore, different heuristics have been developed to solve this problem in polynomial time using greedy and approximation algorithms. In this section, we first describe proposed methods for modelling teams with collaborative ties within their members as an expert graph (Definition 2.1). Next, we explain seminal works followed by the majority of researchers. As will be explained, most of the graph-based works have either targeted minimizing the communication cost only, or have considered hybrid optimization of several other objectives on top of the communication cost. In light of that, we categorize graph-based Team Formation studies into three groups, those that consider: *i*) minimizing communication cost only [17, 25, 30, 72, 81, 83]; *ii*) additional objectives like personnel cost, expertise level and geographical proximity jointly with communication cost [1, 34, 40, 41, 73, 74, 111]; and, *iii*) maximizing the teams' density only [57, 70]. ~~It is noteworthy that there are works that define their problem as a subgraph optimization and propose an objective function, yet attempt to solve it through methods other than subgraph optimization techniques [96] such as linear programming in operations research. We do not consider these works as part of the graph-based category of approaches.~~

3.1 Expert Graph Construction

Central to all the works in graph-based Team Formation is building the expert graph G (Definition 2.2) that represents the experts and their past collaborations. All works in this realm consider experts as nodes, ~~if they have participated in a minimum number of teams~~ and assign experts' skills as node attributes. As we will explain in Section 4.1 (Datasets), skills might not be predefined by nature and should be inferred based on what makes intuitive sense depending on the underlying benchmark dataset. For instance, authors of a research paper can be the experts and the skills of all of the experts can be determined by the keywords in the paper title [17, 72, 81, 83], even though the authors of the paper may be skillful in a few, not all, of the skills. As to the edges of the expert graph, two experts can be connected if they have collaborated in a given minimum number of teams. In the same example, co-authors are supposed to have collaborated as part of a team. Edges can be weighted by a number between zero and one according to pairwise communication costs between experts (nodes). In academic networks of the scientific research community, for instance, the *power* of collaboration between a pair of experts v and v' is defined based on the Jaccard Index as the number of joint publications over the sum of publications by each, and the communication cost will be its complement (one minus the power) [1, 17, 25, 40, 41, 72–74, 81, 111]:

$$w_{\downarrow}(e) = w(u, v) = 1 - |\frac{P_u \cap P_v}{P_u \cup P_v}| \in \mathcal{R}^{[0,1]} \tag{19}$$

where $w_{\downarrow}$ shows the desired less value of the weights, and $P_{.}$ is the set of teams that an expert has participated in. However, there are concerns about this equation:

- The expert graph G becomes a fully dense graph with edge weights in $\mathcal{R}^{[0,1]}$, including edges of weight 1.0 for all pairs of experts with no previous collaboration, which is infeasible in large-scale graphs. To relieve this, a threshold is usually used to filter out edges with a weight above that threshold, making the graph a sparse graph [17, 25, 40, 70, 72];
- The equation is invariant to the number of past collaborations. A pair of experts with only one past joint collaboration receive zero (best) communication cost;
- The expert graph G has edges of weight 0.0 for the minimum communication cost, which is counterintuitive in graph theory where non-existent edges are commonly represented with this value. Further, in terms of implementing the expert graph with an adjacency matrix, it hinders the usage of efficient sparse matrix data structures.

To alleviate the problem of zero edge weights, Li et al. [83] have proposed :

$$w_{\downarrow}(e) = |\frac{1}{P_u \cap P_v}| \in \mathcal{R}^{(0,1] \cup \{\infty\}} \tag{20}$$

Equation 20, however, overlooks the total number of collaborations in teams for each of the experts, as a result of which early career experts and long-standing well-experienced experts are treated equally, which may render it unfair for either of the groups in different scenarios. For instance, Equation 19 favours early career experts as their total number of participation in teams would be low, hence, an instance of joint collaboration would weigh more toward minimum communication cost compared to experienced experts (e.g., $1 - \frac{1}{10} = 0.900$ better communication/lower cost vs. $1 - \frac{1}{1000} = 0.999$ worse communication/higher cost).

Nonetheless, Equations 19 and 20 have been employed in scenarios where lower weights signal the communication cost based on the number of joint collaborations between two experts and subgraph minimization has been desired. In some works, however, a maximization optimization is involved, where the edge weights are desired to be maximized.

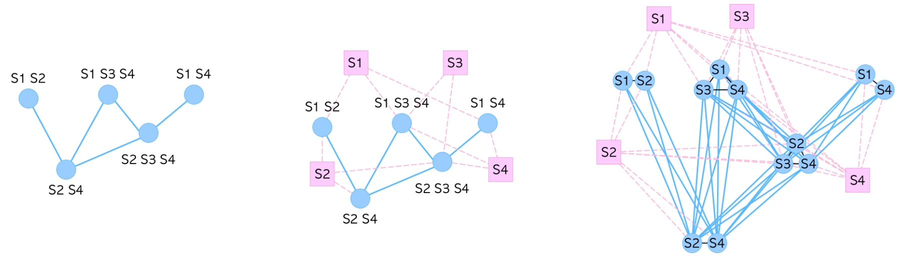


Fig. 6. Graph extension of the Steiner tree algorithm where, in the first step, the graph extends by adding additional skill nodes and connecting them to all of the experts associated with that skill, and in the second step, original nodes are replaced with a complete graph of the size $|S_v|$ in such a way that each node in the complete graph is assigned a specific skill.

Hence, the edge weight between two experts can be directly set by the number of joint collaborations as in [57, 70, 96]:

$$w_{\uparrow}(e) = |P_u \cap P_v| \in \mathcal{N}^{[0,\infty)} \tag{21}$$

3.2 Communication Cost Minimization Algorithms

The work of Lappas et al. [81] is the first attempt to form a team based on the subgraph optimization on the expert graph. A major component of Lappas et al.'s approach is minimizing the communication cost among experts based on two alternatives over possible subgraphs: *i*) diameter, and *ii*) spanning trees. For the former alternative, they proposed *RarestFirst* wherein a subgraph with the minimum diameter is to be found. For each required skill s, it finds a subset of experts that have the skill (V_s), where the skill $s^{\dagger}$ with the minimum number of experts is called a *rare* skill (§2.1). To form a team, an expert who holds the rare skill is selected as the seed (rarest first), and other experts who own the other required skills and have the minimum distance from the seed expert are selected as the team members. Since there may exist more than one expert who holds the rare skill (more than one seed expert), there can be more than one possible team. The optimum team is the one with the minimum diameter in its respective subgraph.

For the latter alternative, i.e., spanning tree-based optimization of communication cost, Lappas et al. [81] proposed *CoverSteiner* and *EnhancedSteiner* algorithms. *CoverSteiner* is a heuristic greedy two-step algorithm: in the first step, the set of candidate experts covering all of the project's required skills are selected by solving a classic *set cover* problem [107]. The set of chosen experts is then expanded to include other experts in order to build a minimum spanning tree via the Steiner tree algorithm [106]. *CoverSteiner* algorithm, however, overlooks the underlying graph structure of experts in the set cover algorithm in the first step, which may lead to possible teams with high communication costs, like when a pair of selected experts is disconnected in the expert graph. To avoid such expert selections, Lappas et al. [81] proposed *EnhancedSteiner* algorithm. *EnhancedSteiner* is also a two-step algorithm. In the first, as shown in Figure 6, it forms an extended *heterogeneous* graph, H, of the expert graph as follows: for each required skill s, an additional node (an expert with that skill, which they refer to as skill node) is added and connected to all of the expert nodes that have the skill s. Then, each node v, excluding the added nodes, is replaced by a complete graph of the size $|S_v|$ in a way that each node of the mentioned complete graph is equipped with one of the skills of S_v. The weights of edges between skill nodes and expert nodes are set to a large number, and the edge weights of the mentioned complete graphs are set to zero. In the second step, like *CoverSteiner*, the same Steiner tree algorithm is applied to the newly formed extended graph where the set of required skills is the input subset for *CoverSteiner* algorithm. Finally, skill nodes are omitted from the output set to find the optimum subgraph.

To improve the time complexity of the proposed diameter-based and spanning tree-based algorithms, Lappas et al. [81] further proposed greedy variations, referred to as *GreedyDiameter* and *GreedyMST*. Lappas et al. [81] evaluated their proposed algorithms based on minimizing the communication cost, team size and connectivity of the respective subgraphs for the output recommended teams on DBLP dataset. In terms of communication cost, the greedy versions of the proposed algorithms performed poorer than the exact algorithms, as expected. Surprisingly, no comparative analysis has been performed between the diameter-based algorithm, i.e., *RarestFirst*, and the spanning tree-based algorithms, i.e., *CoverSteiner* and *EnhancedSteiner*. Therefore, it is unclear which one is a better estimator for communication cost. In terms of team size, the *RarestFirst* algorithm results in larger teams, which is usually undesirable due to other factors like personnel cost, while the *EnhancedSteiner* algorithm generally forms smaller teams, owing to their different minimization methods where they consider the diameter and spanning tree of the subgraphs, respectively. In terms of team connectivity, *CoverSteiner* often fails to find connected subgraphs as an optimum team, whereas *RarestFirst* (diameter-based) and *EnhancedSteiner* (spanning tree-based) can recommend more connected subgraphs, as they consider the graph structure of the expert graph.

Following Lappas et al. [81], a considerable number of studies propose a wide variety of algorithms that minimize the communication cost incurred by the subgraphs in an expert graph through minimization of the diameter, Steiner tree, or sum of distances in the subgraphs [17, 25, 30, 32, 72, 81, 83]. Given an additional constraint on expert capacity, Datta et al. [17] propose *MinDiamSol* aiming to minimize the diameter and *MinAggrSol* aiming to minimize the spanning tree of the team's subgraph (§2.2.1) in their definition of the Team Formation problem. Both algorithms first transform the expert graph G to a simpler graph g by excluding expert nodes whose distance from the seed node is more than a desired (arbitrary) number. Then, *MinDiamSol* attempts to minimize the distance between the seed node and subgraph's nodes, while *MinAggrSol* greedily picks those nodes of g whose addition to the subgraph leads to maximizing the skill coverage. They also extended Lappas et al.'s *EnhancedSteiner* and *RarestFirst* algorithms to consider the capacity of experts as an additional constraint, referred to as *GreedySteiner* and *GreedyDiam*, respectively. Regarding efficacy, Datta et al.'s algorithms empirically outperform Lappas et al.'s counterparts on GitHub and DBLP datasets. Specifically, on GitHub, *MinDiamSol* and *MinAggrSol* can reduce the communication costs by up to 40%. In terms of efficiency, the diameter-based algorithms, that is, *GreedyDiam* and *MinDiamSol*, were faster than others in both datasets.

Another extension to Lappas et al.'s algorithms has been made by Li et al. [83] who proposed *GeneralizedEnhanced-Steiner* to consider an arbitrary number of experts for each required skill ($l_s = u_s > 1$) to form a team; Lappas et al.'s algorithms were limited to *one* expert per skill ($l_s = u_s = 1$). Similar to Lappas et al.'s *EnhancedSteiner*, Li et al.'s *GeneralizedEnhancedSteiner* extends the graph with nodes for skills foremost (see Figure 6). Next, the algorithm selects an expert with a higher density in the neighborhood as a seed node. Li et al. presumed that the higher the neighborhood density of a skill node is, the more likely it is for that skill node to have a smaller distance from other nodes. Given a pair of expert nodes v and v', the density of the expert v is more than v' should more experts be closer to v, and intuitively, the density increases as more experts come closer to the node. This approach is, however, computationally intractable when many skills are required, or the expert graph is large. To reduce the time complexity, Li et al. proposed *GroupingDensity* and *GroupingRandom* algorithms. These algorithms firstly form a modified expert graph, referred to as a *'group graph'*, by merging expert nodes who have a required skill into a single *super* node followed by modified edge weights based on communication cost between such super nodes, each of which representing a subset of skill holder experts. Next, the *GeneralizedEnhancedSteiner* algorithm is applied but only on the *'group graph'* to find an optimum Steiner tree whose super nodes are expanded to the original expert nodes as the optimum team.

Farhadi et al. [25] build upon Li et al.'s algorithms by considering those experts who have at least $x_{v_s} \in \mathcal{R}^+$ level of expertise for each required skill $s \in S$ only (§2.1). They consider the level of expertise of each expert in each skill as a node attribute, and experts whose expertise are higher than or equal to the minimum required expertise level are selected to cover a required skill. Farhadi et al. compare their algorithm with Lappas et al. [81]'s *RarestFirst* and Li et al. [83]'s *GeneralizedEnhanceSteiner* on DBLP dataset and show improvements in both optimum team size and communication cost via taking expertise levels into account.

Diameter-based or minimum spanning tree-based optimization methods are, however, *unstable* or *sensitive* to changes, even minor, in the expert graph; adding or removing an edge leads to major changes in the solution [57, 72, 83]. Towards more stable or robust optimization methods, Kargar et al. [72] introduce the *sum of distances* (Equation 2) in a subgraph for minimizing the communication cost, yet proving that finding a team that minimizes this distance is NP-hard. They proposed an approximation algorithm, *BestSumDistance*, that finds the optimum team with 2−approximation. This algorithm works as follows: given a required skill s and each of the experts v that own it, i.e., $v \in V_s$ (§2.1), the algorithm computes the distance of v to each of the experts v' in the complement set of V_s, i.e, $v' \in V \setminus V_s$. The expert v will be added to the subgraph if it has the least sum of distance over all such v'. In the end, *BestSumDistance* returns a subgraph with the minimum sum of distances as the optimum team.

To reduce the time complexity, even more yet address a real-world scenario, Kargar et al. [72] introduced forming a team with a leader (Equation 6). They propose an exact algorithm, *BestLeader*, that tries to minimize the sum of distances between the team's leader and all team members. Initially, this algorithm assumes each expert L as a leader. Then, for each required skill s, the expert v of V_s who has the least distance to L is added to the subgraph. *BestLeader* computes the distance of each expert node to the leader to determine its addition to the subgraph, as opposed to the *BestSumDistance* that computes the distance of each expert to all other experts. In *BestLeader*, hence, the algorithm forms a subgraph per each assumed leader L. In the end, a subgraph with the least leader distance and its leader are returned by the algorithm as the optimum team. They evaluate their algorithms by experimenting on DBLP and show that *BestLeader* and *BestSumDistance* outperform the baselines, *RarestFirst* and *EnhancedSteiner* of Lappas et al. [81].

Like Kargar et al. [72], to form a team with a leader, Huang et al. [30] argue that in real-life scenarios, especially when teams' sizes are large, a team should be considered as a collection of subteams each of which is equipped with a leader. The effectiveness of the team as a whole, thus, depends on inter-subteam collaborations through subteams' leaders; there is no need for all experts in different subteams to collaborate with each other. Therefore, Huang et al. aim to minimize the communication cost between subteam leaders with subteam members and the main leader. To minimize leaders' communication cost (Equation 7), they propose the *Assignment and Pruning (AP)* algorithm grounded on the Assignment Problem [6] and a pruning framework. The Assignment Problem involves finding the optimal way to match a set of tasks or jobs (herein, required skills of a project) to a set of workers or resources (herein, experts), while optimizing an objective function related to cost, time, or efficiency. In the Assignment Problem, however, no graph structure has been utilized. Huang et al. use *AP* algorithm to find all possible teams that cover all of the required skills foremost, followed by pruning all but those that have a minimum communication cost in their respective subgraphs. According to their experiments on the GitHub dataset and a small-scale synthetic dataset, Huang et al.'s method based on leaders' distances yields competitive performance in comparison with solutions based on diameter minimization, minimum spanning trees, and the sum of distances between team members.

More recently, Nemec et al. [32] have proposed to employ random walks between a pair of experts in the expert graph as opposed to the shortest path when minimizing communication cost. They argue that the shortest path overlooks the structure of the expert graph and falls short when the majority of experts share the same or similar shortest paths.

Table 3. Summary of multi-objective algorithms and the method that has been used for optimizing objectives in tandem.

		Multi-Objective Optimization Method		
	Algorithms	Integration of objectives	Trade-off	Pareto search
Kargar et al. (2012) [73]	Approx	✓	✓	
	Replace	✓	✓	
	MCC	✓	✓	
	MCC-Rare	✓	✓	
Kargar et al. (2013) [1, 74]	Approx{diameter, PCost}		✓	
	Approx{sumDistance, PCost}		✓	
	Approx{PCost, diameter}		✓	
	Approx-Pareto			✓
Zihayat et al. (2014) [111]	TwoPhase			✓
	FirstPhase			✓
	PLS			✓
Zihayat et al. (2017) [40]	CA-CC	✓	✓	
	SA-CA-CC	✓	✓	
Zihayat et al. (2018) [41]	PC-CC	✓	✓	
Selvarajah et al. (2021) [34]	MOCA			✓

They present *Random Walk with Restart (RWR)* algorithm where the communication cost between a pair of expert nodes v and v' has been defined in inverse relation to the estimated probability (likelihood) of reaching from v' to v, via a uniform random selection of edges in a walk; connected sparse expert nodes with few edges would have more probabilities as opposed to connected expert nodes that are located in a dense subgraph. Higher probability indicates stronger connections, hence, less communication cost between pairs of expert nodes. Nemec et al. [32] show that the Team Formation problem still remains NP-hard when the distances between expert nodes are computed by the *RWR* algorithm. To relieve this, they propose a greedy algorithm to reduce the search space by choosing an expert v with the rarest skill $s^\dagger$ as a seed node. In each step, an expert that covers one of the remaining skills while having the minimum communication cost with the seed node based on *RWR* is added to the subgraph. In the end, among all of the formed subgraphs around the seed node, the algorithm returns the one with the minimum communication cost.

3.3 Hybrid Optimization Algorithms

Forming a team based on a single objective like minimizing communication cost only, forgo other human and non-human factors affecting the team's success, such as project budget, personnel cost, proficiency and geographical proximity. While there may be no solution that optimizes all criteria simultaneously, many works [1, 34, 40, 41, 70, 73, 74, 111] target multi-objective optimization in the Team Formation problem considering a combination of crucial factors such as the ones listed above. Approaches to tackling the multi-objective optimization problems in graph-based Team Formation studies can be categorized as follows, also summarized in Table 3:

(1) Initially, integration of objectives was proposed where multiple objectives are merged into a modified expert graph and proxied by a single synthetic objective whose optimum is also optimum for other constituent objectives [40, 41, 73]. However, such methods depend heavily on the merging protocol to build the modified expert graph.
(2) Next, linear interpolation of conflicting objectives with trade-off parameters (coefficient) was proposed, mainly when optimizing multiple objectives simultaneously is impossible, and a balance should be considered among them [1, 74] like communication cost and geographical distance joint minimization.
(3) Finding Pareto solutions has been the state of the art in hybrid subgraph optimization for Team Formation wherein multiple objectives reach a win-win state such that any attempt to favour one objective makes at least one other objective worse [1, 74, 111]. In other words, a subgraph (team) $G[V_p]$ dominates another subgraph $G[V'_p]$ with

respect to two objectives if $G[V_p]$ is better than $G[V'_p]$ in one objective and not worse than $G[V'_p]$ (i.e., equal or better) in the other objective. A subgraph is a Pareto optimum team if it is *not* dominated by any other subgraphs.

Kargar et al. [73] is the pioneering graph-based approach that considers two objectives jointly for minimizing communication cost and personnel cost, a *bi*-objective optimization function that has been shown to be NP-hard. Following (1) integration of an objective approach to solve the problem, they present an approximation algorithm called *Approx*. In an expert graph, *Approx* initially spreads node weights, representing experts' personnel cost, and interpolates them with the existing edge weights, representing communication cost, such that minimizing the sum of distances in the modified graph is equal to minimizing the *bi*-objective function in the original graph. *Approx* then selects an expert holding at least one required skill as the seed node and incrementally adds the nearest expert nodes to the seed expert node to cover all required skills for a team. Hence, a candidate subgraph is formed based on the nearest distances of experts to the seed expert node only and inter-distances between selected experts are ignored. After forming all such candidate subgraphs, a subgraph with the smallest sum of distances to the seed is selected as the optimum team. To improve the efficiency of *Approx*, Kargar et al. [73] further propose three heuristic algorithms, *Replace*, *Minimal Cost Contribution (MCC)* and *MCC-Rare*. *Replace* selects an expert with the lowest personnel cost holding at least one of the required skills and replaces her with another expert holding the same skills but with a lower communication cost. *MCC* is similar to *Approx*, but it examines the communication cost of the yet-to-be-selected expert with not only the seed expert but also with other selected experts. In *MCC*, new team members are added incrementally, and each new member is chosen by comparing its communication cost with that of all the current members (not only the seed member) of the team in addition to the personnel cost of the new member. Finally, to reduce the run time of the *MCC* algorithm, they propose *MCC-Rare*, which is a variation of *MCC* where an expert with the rarest required skill $s^\dagger$ is used as a seed node. Kargar et al.'s experiments on DBLP and IMDB datasets show that *MCC* outperforms other algorithms regarding combined communication and personnel costs, and *MCC-Rare* and *Approx* are the runners-up. However, in terms of time complexity, *Replace* was the fastest, yet the poorest algorithm with the highest combined communication and personnel costs, *MCC-Rare* and *Approx* are the second and third fastest algorithms, and *MCC* is the slowest one. Further, Kargar et al. [73] benchmark their algorithms on each objective separately. To minimize communication cost, *Approx* yields the best results, while *MCC-Rare* could form a team with the lowest personnel cost.

Following the (2) integration of objectives with trade-off parameters, Kargar et al. [1, 74] propose an $(\alpha, \beta)-$approximation algorithm to find teams where the value of the first objective (e.g., personnel cost) is at most α times a given fixed amount (e.g., overall allocated budget) while the value of the second objective (e.g., communication cost) is at most β times the minimum value of any solution that meets the first objective. Their first algorithm *Approx{diameter, PCost}* computes the diameter-based communication cost (Equation 4) considering a threshold on diameter, then attempts to minimize the personnel cost. It is a $(2, log|S_p|)-$approximation algorithm where S_p is the set of required skills (§2.1). This algorithm works as follows: for each node v equipped with the rarest skill $s^\dagger$, it computes the distance of v from all nodes of $V \setminus V_{s^\dagger}$. Then, for each required skill, the node v' will be added to the subgraph if its distance from v is less than the mentioned threshold on diameter. If all of the expert nodes' distances from v exceed this threshold, then there is no possible solution. In the end, among all the formed subgraphs, the algorithm returns the subgraph with the minimum personnel cost. *Approx{sumDistance, PCost}* is the second algorithm by Kargar et al. [1, 74] which is similar to the previous one, only it uses the sum of distances based communication cost (Equation 2) instead of diameter. This is a $(|S_p|, log\ |S_p|)-$approximation algorithm and considers each expert node with one of the required skills as a seed node, unlike the previous algorithm that only uses rare skill holders as seed.

The third algorithm by Kargar et al. [1, 74] is *Approx{PCost, diameter}*. This algorithm is a $(log\ |S_p|, 2)-$approximation; it aims to minimize diameter-based communication cost given a desired threshold for personnel cost. It first starts by computing the diameter of the input graph $\varphi_R(G[V])$ and stores it. Then, *Approx{PCost, diameter}* calls *Approx{diameter, PCost}* and considers $\varphi_R(G[V])$ as a threshold for diameter to find a team with the minimum personnel cost and a diameter less than $\varphi_R(G[V])$. If the personnel cost of the formed team is more than the required threshold on personnel cost, then there is no possible team since personnel cost and communication cost are conflicting objectives. However, if the personnel cost of the formed team is less than the required threshold for personnel cost, the algorithm decreases the diameter threshold $\varphi_R(G[V])$ to a smaller value (e.g., $\frac{\varphi_R(G[V])}{2}$) and calls the *Approx{diameter, PCost}* again to find a team with a diameter less than the new threshold and the minimum personnel cost. This process ends when the diameter becomes less than a threshold.

Last, Kargar et al. [1, 74] propose the *Approx-Pareto* algorithm to find (3) Pareto optimal teams for the personnel and communication costs. The *Approx-Pareto* algorithm repeatedly calls *Approx{diameter, PCost}* and for several different diameter thresholds forms a team with minimum personnel cost. Thus, a set of teams is formed, each with the minimum personnel cost under the given diameter threshold. In the end, the algorithm checks if each generated team is dominated by other teams or not, and then ignores the dominated ones and returns non-dominated teams as Pareto optimal teams. In DBLP and IMDB, *Approx-Pareto* could achieve state-of-the-art performance in terms of accuracy and speed.

In another line of work, Zihayat et al. [40, 41] optimize communication cost and authority of experts (proficiency) in a team. As explained in Section 2.2.2 (Maximizing Proficiency), skill holders are team members with required skills, which may be connected via experts who cover *none* of the required skills, referred to as connectors. Zihayat et al. [40, 41] argue that the authorities of connectors have an impact on the team performance besides their connector role. They define two separate objective functions to incorporate the connectors' authorities and skill holders' authorities. Following (1) the integration of objectives, Zihayat et al. merge the expert node weights, representing the authority level of experts, with edge weights, representing communication costs. In doing so, they first transfer each node weight to all the edges that come to the node and then define a new edge weight for the graph. They show that minimizing the communication cost based on the sum of edge weights (Equation 3) for the modified expert graph results in an optimized communication cost and connector authorities. Zihayat et al. [40, 41] show that minimizing the final synthetic objective is NP-hard and propose the *Communication Cost-Connector Authority (CC-CA)* algorithm for finding the optimum team. This algorithm tries to build a tree around each expert node v as a root so that it has the minimum sum of edge weights. For each root v, the algorithm finds the nearest skill holder v' to the root and adds v' as well as all of the expert nodes between v and v' to the team. In the end, among all the formed tree subgraphs, the algorithm returns the subgraph with the lowest sum of edge weights based on Equation 3 as the optimum team. Zihayat et al. [40, 41] also propose *Skill holder Authority-Connector Authority-Communication Cost (SA-CA-CC)* algorithm to optimize communication cost, connector authority and skill holder authority by merging skill holder authority, similar to *CC-CA*, in the modified graph.

With respect to more than two objectives, Zihayat et al. [111] show that the optimization of three objectives is NP-hard and, based on (2) Pareto solutions, they propose *TwoPhase (FirstPhase and PLS)*, an approximation algorithm to optimize communication cost, personnel cost and expert proficiency. The first step of this algorithm is responsible for finding a subset of all Pareto teams. This subset is the output provided by the *lexicographic* minimization solution of a trade-off relation between communication cost, personnel cost and proficiency. In a lexicographic optimization problem, candidate optimal outputs are reached by initially optimizing the first objective. Then, if two candidate solutions are the same as per the first objective, the optimum one is selected based on the second objective. The same procedure is pursued for the subsequent objectives. It is proven by Paquete et al. [93] that the output of the mentioned optimization problem always consists of Pareto teams; however, the reverse does not hold in this first phase of the algorithm, i.e., there might be Pareto teams that are missed by the algorithm.

Zihayat et al.'s *FirstPhase* works as follows (herein, we assume the communication cost is the first objective to be optimized without loss of generality to any of the three criteria mentioned): an expert with a required skill is selected as the seed node randomly. Then, for each uncovered skill, a skill holder closest to the seed node is added to the subgraph. This continues until all of the required skills are covered. By the end of this step, the algorithm has formed subgraphs with minimum communication cost as experts are selected in a way that they have a minimum distance from the seed node. In the next step, the algorithm tries to consider other objectives as well, i.e., personnel cost and proficiency. In the final step of the first phase, the algorithm compares all the formed subgraphs based on a second objective (e.g., personnel cost), and the subgraph with the desired rate is kept as the optimum team. If they tie on the second objective, the third objective (e.g., personnel costs) is considered, and the one with the desired rate is kept as the optimum team. The aforementioned procedure is done for all experts with the required skills. Since not all Pareto teams can be found, as explained earlier, another subset of Pareto optimal teams is discovered in the second phase of *TwoPhase* algorithm via a *Pareto Local Search (PLS)* algorithm.

More recently, Selvarajah et al. [34] optimize communication cost, level of expertise, trust score and geographical proximity within the context of a multi-objective Team Formation problem. They propose a unified framework for multi-objective Team Formation problems and present a *Multi-Objective Cultural Algorithm (MOCA)*, which is an extended version of a class of evolutionary algorithms [69], that returns the highest ranked Pareto optimal teams. The crucial difference between their work and previous works is considering the dynamic nature of the mentioned objectives. For instance, an expert's proficiency can change over time through consistent practicing and improving their skills. Therefore, they attempt to include the time-related aspects of objectives in the Team Formation problem. Selvarajah et al. [34] convert static formulas to dynamic versions by adding a term which accounts for the difference between the current time and the last time that two experts have collaborated on a task.

While Pareto optimization has remained the state of the art in graph-based Team Formation, modern multi-objective optimization techniques have been developed beyond Pareto [99] such as multi-factorial [60, 84] and multi-task optimization [26, 64] to support parallel optimization of multiple functions simultaneously, each of which can be single- or multi-objectives, and can be employed for multiple subgraph optimization to form multiple optimum teams that are interrelated. We will provide a more in-depth explanation in Section 5.3. Modern Hybrid Optimization Algorithms.

3.4 Community-based Optimization Algorithms

~~While the majority of works consider communication cost as the prominent objective to optimize, either individually or in tandem with other objectives, a~~ A few works propose detecting communities as a proxy for optimum teams [57]. In communities, edges, which represent previous collaborations between the experts, are highly densely connected and indicate internal cohesion, hence minimum communication cost, while edges between experts in different communities are sparse and indicate externally loose connections, therefore, maximum communication cost.

Gajewar et al. [57] is the first to present community-based algorithms. They target forming a team with a maximal subgraph density and a desired, arbitrary number of experts required for each skill to cover the project. They prove that finding a team with maximal density is NP-hard and propose two approximation algorithms, *s-DensestAlk* and *m-DensestAlk*, to solve this problem. Gajewar et al. [57] also present three heuristics called *Enhanced-Dense*, *PartialTrimmedDense* and *CompleteTrimmedDense* which are extensions of previous approximation algorithms. These heuristics trade off between the size of the returned solution and the density while covering all required skills. Gajewar et al. [57] also extend *RarestFirst* of Lappas [81] to *MinDiameter* algorithm for projects wherein more than one expert may be needed for each skill, as happens in real-world situations. In evaluation, the density-based algorithms *s-DensestAlk* and *m-DensestAlk* outperform the diameter-based algorithm *MinDiameter* as per different metrics such as *team cardinality*, *teamPubs*, *partialTeamPubs* and *teamPubRati*, which will be explained later in the Section 4 (Evaluation Methodology). Also, formed teams will be connected if the social network is connected; however, there is no upper bound for team size.

Juarez et al. [70] also present a multi-objective formulation to maximize the density along with expert proficiency among team members. Using a different formula to compute the density of a subgraph (Equation 18), followed by a genetic algorithm for optimization, Juarez et al.'s algorithm yields smaller and more connected teams compared to Gajewar et al.'s work. In comparison with the *RarestFirst* algorithm of Lappas et al. [81] and *m-DensestAlk* of Gajewar et al. [57], Juarez et al.'s algorithm outperforms the baselines on DBLP dataset in terms of optimal team size, minimum communication cost and maximum density.

3.5 Efficiency and Scalability Enhancement

Mid- to large-scale datasets yielding large graphs pose high computational complexity on graph-based optimization techniques. For scalability, researchers have tried to reduce the computational cost, including time and space, while maintaining accuracy in two directions: (1) pruning the dataset or expert graph during preprocessing, and (2) utilizing heuristics during subgraph optimization. Pruning techniques aim to reduce the search space on the expert graph by filtering the number of teams, the skills S, expert nodes V, and/or joint collaboration among experts for edge set E. To reduce the number of teams, which also may lead to a further decrease in the expert nodes and skills, researchers filter out teams based on the team size, e.g., removing the papers or movies with a single author or cast and crew in DBLP and IMDB, respectively. This filtering has been implied by their definition of a team, as it assumes that collaboration is necessary, although may not be explicitly mentioned. Alternatively, they focus merely on a subset of projects based on the areas of specialty, e.g., only *database* and *data mining*-related papers in DBLP [17, 25, 40, 72–74, 81, 83, 111]. There have been works that select a subset of teams based on a timestamp associated with a team, e.g., movies with release dates from 2000 to 2002 [72, 73]. For skill set reduction, a minimum distribution threshold is set. For instance, in DBLP dataset, the set of terms that appear in a minimum number of paper titles are considered as the skill set [40, 74, 83]. Likewise, to reduce the expert set, a minimum threshold on the number of participations in teams has been used, e.g., minimum number of papers in DBLP [25, 74] and movies in IMDB [72, 111]. Some works apply the pruning on edge set E based on a minimum threshold for collaborations among experts like minimum number of co-authorship in papers in DBLP or co-staring in movies in IMDB [17, 25, 72–74, 111]. Rarely, researchers apply graph sampling techniques to address scalability of their proposed algorithms like Datta et al. [30]'s work where subgraphs of varying sizes were generated from the expert graph in GitHub by randomly selecting a node and performing breadth-first search around it. In Table 10 of Appendix F, we summarize the filtering methods in each of the graph-based Team Formation algorithms.

4 Evaluation Methodology

In this section, we outline the methodologies used to evaluate the performance of the surveyed Team Formation approaches. ~~Specifically, 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 output recommended teams by certain proposed approaches as compared to the gold truth.~~

Table 4. ~~Mainstream datasets in Team Formation research studies.~~

Dataset	Used by	Scale	Nodes (experts)	Edges	Skills	Projects	Public Link	Version
DBLP	Lappas et al. (2009) [81]	Mid	5,508	11,905	1,792			
	Li et al. (2010) [83]		5,482	10,339	11,905			
	Farhadi et al. (2011) [25]		8,805	5,588	1,792			April 12, 2006
	Kargar et al. (2011) [72]		5,658	8,588			`http://dblp.uni-trier.de/xml/`	
	Datta et al. (2012) [17]		7,332			19,248		
	Kargar et al. (2012) [73]		5,658	8,588			`http://dblp.uni-trier.de/xml/`	
	Gajewar et al. (2012) [57]		6,137					May 17, 2010
	Kargar et al. (2013) [1, 74]		6,229	9,400		100	`http://dblp.uni-trier.de/xml/`	April 25, 2011
	Zihayat et al. (2014) [111]					100	`http://dblp.uni-trier.de/xml/`	
	Zihayat et al. (2017) [40]	Large	40,000	125,000		50	`http://dblp.uni-trier.de/xml/`	Up to 2015
	Zihayat et al. (2018) [41]					50	`http://dblp.uni-trier.de/xml/`	Up to 2015
IMDB	Kargar et al. (2011) [72]		6,774	35,875			`http://www.imdb.com/interfaces`	
	Kargar et al. (2012) [73]		6,784	35,875			`http://www.imdb.com/interfaces`	2000-2002
	Kargar et al. (2013) [1, 74]		6,784	35,875		100	`http://www.imdb.com/interfaces`	
GitHub	Datta et al. (2012) [17]	Large	135,346		52	905,000		
	Huang et al. (2017) [30]		5,276	259,084	50	3,750		
Synthetic	Li et al. (2010) [83]					100		
	Huang et al. (2017) [30]	Small	100	1,000	50			
	Selvarajah et al. (2021) [34]		200	10,136				

4.1 Datasets

As summarized in Table 11, the major datasets used in Team Formation studies vary in terms of domain, including computer science publications (DBLP), moving pictures (IMDB), and open-source software (GitHub). As seen, employed datasets can be from the same domain but with different graph sizes, from small to large scales.

4.1.1 DBLP

Following Lappas et al. [81], the computer science bibliography dataset, known as DBLP, has become a benchmark dataset not only in the graph-based studies but also in the entire field [1, 17, 25, 40, 41, 57, 70, 72–74, 83, 96, 111]. The DBLP dataset has been generated from the DBLP website, established in 1993, resulting in a large-scale publicly available dataset of more than four million publications authored by about two million computer scientists and researchers. This dataset contains extensive information on publications, including titles, authors' names and fields of research, references, venues, publishers, publishing dates, DOI, number of citations, and so forth, as well as information pertinent to authors such as full names and affiliations. Collections of all publications are made available as daily snapshots as well as monthly releases. An advantage of DBLP is the availability of open-source libraries that can efficiently load the data, which is essential for researchers. To form the expert graph, each publication is mapped to a team whose expert members are the publication's authors and the required skills S_p are set either as the publication's fields of studies [57] or keywords in the paper's title [1, 25, 40, 72–74, 81, 83]. In this setting, thus, the graph nodes (V) and edges (E) represent authors and co-authorship in publications, respectively. Node attributes A could include the *h-index* of the authors as the proficiency level a_v, the number of citations to the expert's publications in each of the research domains as their expertise x_{v_s} in each domain, and the maximum number of projects that the experts can handle in a year as their capacity c_v. From Table 11, DBLP dataset can be obtained from its official online repository or via third-party bibliographic databases such as AMiner [33][2].

[2]https://www.aminer.org/citation

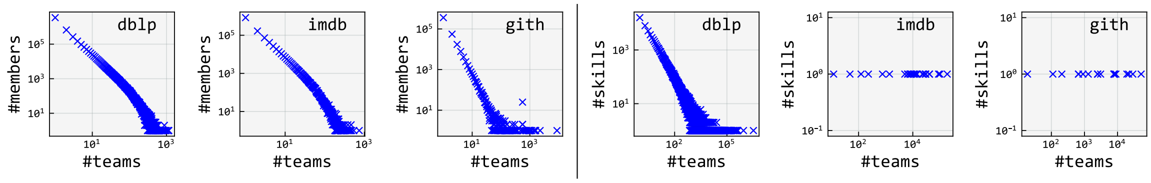


Fig. 7. Distribution of teams over members (left) and skills (right) for DBLP, IMDB, and GitHub datasets.

4.1.2 IMDB

Another popular dataset in Team Formation studies, the Internet Movie Database known as IMDB [1, 72–74] is publicly available and includes information about moving pictures such as movies, television shows, and video games, as well as information on actors, actresses, and other members of cast and crew. The choice of IMDB in Team Formation literature is not to be confused with its use cases in recommender systems or review analysis research; herein, the goal is to form a team of cast and crew for a moving picture's production, as opposed to a moving picture recommendation, for instance. To form the expert graph based on the IMDB dataset, each moving picture is considered a team whose expert members (V) are the cast and crew and the moving picture's genre/s are the team's required skills S_p. Therefore, nodes can be members of cast, edges (E) can represent co-plays in moving pictures, and node attributes A could include the personnel cost ρ as the day/week rate of the member of cast and crew acting or working in a moving picture, and the number of moving pictures in which the members of cast and crew worked in as their proficiency level a_v. As it will be described later in Section 5 (Challenges and Future Direction), as opposed to DBLP, opting for movie titles' keywords to represent skills can be naive since movies' titles rarely correspond to the movie making skills. The IMDB dataset is easily accessible for research purposes via its official repository at the mentioned link in Table 11 of Appendix G.

4.1.3 GitHub

The other dataset is GitHub, which hosts a large number of open-source software projects (repositories) and software developer profiles [17, 30, 32]. Unlike DBLP and IMDB, GitHub provides explicit edges (E) through followership connections. In GitHub, two experts are connected in the expert graph if they have contributed to the same repository. Attributes for an expert A_v can include the list of followers and the repositories created and contributed to. The skill set S_p is either derived from the repositories' titles or the programming languages employed in the repositories. The GitHub dataset can be accessed through the official APIs to stream the information on a per-repository basis[3]. It is worth noting that there are GitHub datasets that are streamed and publicly available in a single compressed file, but mostly for software engineering research purposes [50] and may miss the information required to form an expert graph for the Team Formation problem.

Figure 7 has rendered for the latest versions of DBLP, IMDB, and GitHub datasets. As seen, in left the long tail problem occurs in the distributions of teams over experts in these datasets; this is because many experts (researchers in DBLP, cast and crew in IMDB, and developers in GitHub) have participated in very few teams (papers in DBLP, movies in IMDB, and repositories in GitHub). For instance, 10^6 researchers have participated in 1 team only, while few researchers have co-authored more than 10^3 papers in DBLP. With respect to the set of skills, DBLP follows different distributions compared to IMDB and GitHub. While DBLP suffers further from the long-tailed distribution of skills in teams, IMDB and GitHub follow a more fair distribution, as shown in Figure 7 right. Specifically, IMDB and GitHub have a limited variety of skills (genres and programming languages), which are, by and large, employed by many movies and repositories, respectively. Such distinct distributions in varying domains open benchmark challenges for

[3] https://api.github.com/repos

Table 5. Evaluation methodologies adopted by proposed methods in Team Formation problem.

		Lappas et al. (2009) [81]	Li et al. (2010) [83]	Farhadi et al. (2011) [25]	Kargar et al. (2011) [72]	Datta et al. (2012) [17]	Kargar et al. (2012) [73]	Gajewar et al. (2012) [57]	Kargar et al. (2013) [74]	Kargar et al. (2013) [1]	Rangapuram et al. (2013) [96]	Zihayat et al. (2014) [111]	Huang et al. (2017) [30]	Zihayat et al. (2017) [40]	Zihayat et al. (2018) [41]	Juarez et al. (2018) [70]	Nemec et al. (2021) [32]	Selvarajah et al. (2021) [34]
Intrinsic	Precision								✓	✓		✓						
	Recall								✓	✓		✓						
	Top-cited Papers	✓																
	Highly-rated Venues													✓				
	Exact Algorithm				✓		✓		✓	✓		✓		✓				✓
Quantitative		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Qualitative	Team Size (Cardinality)	✓	✓	✓	✓													
	Team Connectivity	✓																
	Number of Liaisons		✓															
	Number of Common projects				✓													
	Skill Count		✓		✓													
	Hypervolume								✓	✓		✓						
	Average Distance								✓	✓		✓						
	Maximal Distance								✓	✓		✓						
	Number of Non-dominated Solutions																	✓
	TeamPubs							✓										
	PartialTeamPubs							✓										
	Number of Top-cited Authors							✓										

reproducibility and generalizability of Team Formation algorithms, which have received little to no attention in the literature, as discussed further in Appendix D.

4.1.4 Synthetic

Last, when obtaining real-world datasets has become challenging or restricted, researchers adopt or generate synthetic datasets to benchmark their proposed algorithm [8, 30, 34]. For instance, Selvarajah et al. [34] generated synthetic expert graphs following the Lancichinetti et al. [80]'s graph sampling method, designed to benchmark community detection methods in graphs with heterogeneity (power law) in the distributions of node degrees and community sizes. There have been other works benchmarked on synthetic datasets, yet the underlying graph sampling methods to generate the datasets have mostly remained unexplained [30, 83].

4.2 Evaluations

The proposed approaches surveyed in this paper provide different experimental analyses of their performances based on both effectiveness (accuracy) and efficiency (speed). ~~We draw upon both, starting with effectiveness measures. Then, we elaborate on how efficiency has been measured. Lastly, the baselining efforts and their limitations will be introduced.~~

4.2.1 Effectiveness (Accuracy)

Measuring the effectiveness of Team Formation approaches faces a fundamental challenge since the basic question of '*what it takes for a team to be considered successful*' has gone under examination and has remained controversial in the Team Formation literature. Finding experts who collectively cover the required skills and have optimum communication cost, personnel cost, or any other objective individually or jointly for a team might be insufficient for labelling the *truth* about its success or failure, as these criteria do not necessarily guarantee the team's success. Nonetheless, almost all graph-based works have assumed existing teams in a dataset as successful teams (positive samples). For instance, in DBLP, a publication is considered a tangible indication of a team's success. However, the rejected manuscripts, which by comparison represent teams' failures, are missing in the dataset. Similarly, in IMDB, all instances of moving pictures are considered successful since they have at least completed the production procedure and have been released for public viewing. In the following, we categorize the evaluation metrics and means used in the Team Formation realm into three types: intrinsic, quantitative, and qualitative, as summarized in Table 5, and elaborate on each.

Intrinsic vs. Project Generation Evaluation. An intrinsic evaluation schema, as we denote it, hinges on the idea that we can have a standalone gold standard (ground truth) against which we can measure an approach's effectiveness. Gold data is commonly created through the selection of a subsample of a given dataset, a procedure that is based on the assumption that the existing instances of the teams in a dataset are of type *successful*. In this schema, one can measure their approach's effectiveness based on the existence of their output teams in this gold set. As opposed to the intrinsic evaluation, an extrinsic evaluation schema evaluates a given method's output based on its impact on the performance of other applications [65]. In the DBLP dataset of publications, some works put a stricter constraint on a team to be a gold team, like if the team is cited more than a certain number [81] or published in a highly-rated venue [40]. A third evaluation schema, utilized by some graph-based works [17, 25, 81, 96], is performed differently than intrinsic and extrinsic schemas. In such works, the dataset is entirely used to build the expert graph, and for evaluation purposes, *synthetic* projects are generated, each of which is followed by an exhaustive search over *all* possible subgraphs of the expert graph. A subgraph that obtains the optimum value for a project is labelled as a gold team. Such evaluation schema has been also referred to as project generation [17, 25, 30, 40, 41, 72, 74, 81, 96, 111]. Based on these evaluation schemas, metrics have been employed to measure the effectiveness of an algorithm, as explained hereafter.

Quantitative Metrics. Evaluation metrics can be utilized where we have access to a gold standard, either in the context of an intrinsic evaluation methodology or a gold set acquired via an exhaustive search for a task (synthetic subset of skills). Intrinsic evaluation has been primarily measured via such classification metrics as precision and recall [1, 74, 111]. However, in project generation evaluation, the results of the methods are evaluated in comparison with the optimum subgraphs; a method that outputs a more similar subgraph or a subgraph whose objective value, e.g., communication cost, is closer to the optimum value is considered more effective [1, 34, 40, 72–74, 111]. Since an exhaustive search for optimum subgraphs might be computationally prohibitive, some works [25, 81, 96] simply omit the exhaustive search for an optimum subgraph per projects and proceed with the direct comparison between different methods based on how well they optimize objective functions; a method that yields the lowest (highest) value for an objective or a combination of objectives that are to be minimized (maximized) is considered as the most effective one on average over all of the projects. Some of the main objectives include minimizing the communication cost by searching for subgraphs with minimum diameter, sum of the edge weights, sum of distances between experts, or spanning tree. Other objectives include personnel cost and geographical distance (to be minimized), or trust score and expertise level (to be maximized), as explained in Section 2.2

Qualitative Metrics. Last, some works have adopted qualitative evaluation methodologies to demonstrate the effectiveness of the predicted teams. Qualitative metrics are a go-to choice that make intuitive sense in the absence of a gold standard. This is similar to the case of project generation-based evaluation without exhaustive search. Even with the existence of gold standard data, this schema has been utilized in some works as complementary confidence in the methods' performance. Such qualitative measures are described below:

Table 6. ~~Qualitative metrics.~~

Qualitative Metrics	Description
Team Size (Cardinality)	The number of team members.
Team Connectivity	Indicates the degree of connectedness of the team's members in their respective subgraph.
Number of Liaisons	The number of team members not covering required skills but keeping other skill-holder members connected.
Skill Count	The number of times an expert has participated in projects in the past that shared the same required skill.
Hypervolume	Measures how well a set of solutions covers the possible optimum outcomes across multiple objectives.
~~Average Distance~~	~~Measures the distance between teams formed by a definite algorithm and a set of gold truth teams.~~
Maximal Distance	Greatest distance between teams formed by a definite algorithm and a set of gold truth teams.
TeamPubs	The number of publications in the past whose authors are exactly the same as in the recommended team.
PartialTeamPubs	The number of publications where at least half of their authors belong to the recommended team.
Number of Common Projects	The number of joint projects shared by at least two experts of the recommended team.
Number of Top-cited Authors	The number of experts in the recommended team who are among the top-cited authors.
Number of Non-dominated Solutions	The number of solutions that cannot be improved in one objective without compromising another objective.

- *Team Size (Cardinality)* is the number of team members; effective teams are usually desired to be small and more cost-effective [72, 81, 83].
- *Team Connectivity* indicates the degree of connectedness of the team's members in their respective subgraph; densely connected team members show many previous collaborations, directly or via other team members, hence have lower communication costs [81].
- *The Number of Liaisons*, also referred to as intermediators [83] or connectors [40], are the number of team members that do not cover any of the required skills for the project but keep other skill-holder members connected in the respective subgraph of the team. It is desired to minimize the number of liaisons for a team.
- *Skill Count* shows how many times an expert member of a team has participated in projects in the past that shared the same required skill. This value is averaged over all required skills and all members of the team. A higher skill count indicates a greater level of experience, and hence better expertise, within the team concerning the set of skills required for the project [72].
- *Hypervolume* evaluates the quality of solutions in multi-objective optimization problems and measures how well a set of solutions covers the possible optimum outcomes across multiple objectives. A higher hypervolume value suggests a better distribution of non-dominated solutions and a more comprehensive representation of the trade-offs between different objectives. This can help in comparing and selecting solutions that provide a balanced and diverse set of outcomes in multi-objective optimization [1, 74, 111].
- *Average Distance* is a qualitative metric that measures the distance between teams formed by a definite algorithm (e.g., Pareto optimum teams) and a set of gold truth teams (e.g., outputs of a Pareto-exact algorithm). The distance is computed via the mean distance from each member of the first set and the closest member in the second set. The Maximal distance has also been used as a qualitative measure in several works [1, 74, 111], which computes the greatest distance between two mentioned sets.
- *TeamPubs* and *PartialTeamPubs* are metrics that have been used for the DBLP dataset yet can be generalized for other domains. For a recommended team, the TeamPubs is equal to the number of publications (collaboration) in the past whose authors (members) are exactly the same as in the recommended team, [57] and *PartialTeamPubs* is the number of publications where at least half of their authors belong to the recommended team [57]. A larger value of these metrics shows that the team is more likely to be successful due to its more successful collaborations in the past.
- *Number of Common Projects* (e.g., publications) is the number of joint projects shared by at *least two* experts of the recommended team [72].

Table 7. Time complexity of graph-based Team Formation methods. ($\times$) shows the authors forego calculating the time complexity.

Complexity Class	Optimization Method	Algorithm	Time Complexity	
NP-Hard / NP-Complete	Diameter	RarestFirst [81]	$O(\|V\|^2)$	
		MinDiamSol [17]	$O(\|S_p\|\|V\|^2(\log\|V\| + \log\|S_p\|)\log\|V\|)$	
		Approx{diameter, PCost} [1, 57]	$O(\|V\|^2\|S_p\|)$	
		Approx{PCost, diameter} [1, 57]	$O(\|V\|^2\|S_p\|(\log_2 \frac{MaxDiameter}{\epsilon} + 1))$	ϵ is an input precision
		Generalized Diameter [25]	$\times$	
		MinDiameter [57]	$O(\|V\|^2)$	
		Generalized Diameter with Skill Grading [25]	$\times$	
	MST	CoverSteiner [81]	$O(\|V\|^3)$	
		EnhancedSteiner [81]	$O(k\|E\|)$	k is the number of nodes added to form the extended graph.
		MinAggrSol [17]	$O(\|S_p\|\log\|S_p\|)$	
		GeneralizedEnhancedSteiner-Random [83]	$\times$	
NP-Hard	Density	m-DensestAlk [57]	$O(\alpha\|V\|^3)$	$\alpha = \sum_{s\in S_p} l_s$
		GeneralizedEnhancedSteiner-Density [83]	$\times$	
		GroupingDensity [83]	$\times$	
	Sum of Distance	Best-SumDistance [72]	$O(\|S_p\|^2\|V\|^2)$	
		FirstPhase [111]	$O(\|S_p\|^2\|V\|^2)$	
		Assignment and Pruning (AP) [30]	$\times$	
		Approx [73]	$O(\beta^2\|S_p\|^2)$	$\beta = \max_{s\in S_p}\|V_s\|$
		MCC-Rare [73]	$O(\beta^2\|S_p\|^3)$	$\beta = \max_{s\in S_p}\|V_s\|$
		MCC [73]	$O(\beta\gamma\|S_p\|^2)$	$\beta = \max_{s\in S_p}\|V_s\|$, $\gamma = \min_{s\in S_p}\|V_s\|$
		Replace [73]	$O(\beta\log\beta)$	$\beta = \max_{s\in S_p}\|V_s\|$
		MOCA [34]	$\times$	
	Sum of Edge Weights	PC-CC [41]		
		CC [40]	$O(\beta\|S_p\|\|V\|)$	$\beta = \max_{s\in S_p}\|V_s\|$
Polynomial	Leader Distance	Best-Leader [72]	$O(\beta\|S_p\|\|V\|)$	$\beta = \max_{s\in S_p}\|V_s\|$

- *Number of Top-cited Authors* is obtained by comparing the experts of the formed teams against the list of top-cited authors and is equal to the number of experts in the recommended team who are among the top cited authors [57].
- *Number of Non-dominated Solutions*, also referred to as the number of Pareto solutions, is the number of solutions that cannot be improved in one objective without compromising another objective.

4.2.2 Efficiency (Speed)

It is crucial to have an algorithm that not only forms an effective team, but also gets the answer in an acceptable time. Lappas et al. [81] is the first work to address the difficulty class of Team Formation problems. They prove that finding a team that covers all the required skills and minimizes communication costs via the diameter or spanning tree is NP-complete. In the following, the difficulty class of other Team Formation problems are also studied, such as, minimizing sum of distance [57, 72, 73, 111], minimizing sum of edge weight [40, 41] and maximizing density [57, 83]. They prove that all of these problems are NP-hard. Out of all Team Formation optimization problems, forming a team with a leader is proved to be solved in polynomial time [72]. Researchers have proposed different approximation and heuristic workarounds to address NP time complexities of such algorithms. For instance, to form a team by minimizing the diameter, some works [1, 57, 74, 81] propose algorithms that grow quadratically with the number of required skills. Table 7 is the summary of algorithms attempting to decrease time complexity categorized based on their optimization method as well as their time complexities.

As an empirical view on real data, several works [1, 17, 30, 34, 72–74, 83, 111] also studied time complexity as the elapsed run time of the algorithm (wall clock). Some works compare the run time of their algorithms with their own proposed algorithms, e.g., a greedy version or heuristic algorithms [1, 17, 30, 73, 74, 83, 111]. Another approach involves addressing the scalability of algorithms by studying their run time by increasing nodes of the graph or the number of required skills [1, 34, 72, 74]. While there is no work that compares the efficiency of their algorithms with other works, a joint analysis of efficiency (time complexity) vs. efficacy has been studied rarely [17, 30, 34, 73, 83]; the general findings in most of these works show that the more effective an algorithm, the lower its runtime.

4.3 Baselines

Surprisingly, almost all the works in graph-base Team Formation forgo benchmarking their algorithms against other baselines or at-the-time state of the art, most presumably due to: *i*) lack of a common benchmark dataset, *ii*) unavailability of source codes, and *iii*) improvements have been mostly made for an *additional* objective, as opposed to the objective optimization method, which makes the comparison inevitably unfair. For example, Zihayat et al. [111] propose considering personnel cost and expertise level in addition to communication cost and seek multi-objective optimization through minimizing the sum of distance method. A fair comparison would be with other existing methods that optimize the same objectives; this has been rare and has led to no available baseline. To fill the gap and further support the efficiency or efficacy of their proposed algorithms, researchers compare their proposed algorithms with their variations. For example, Huang et al. [30] compare their proposed method with its greedy versions, or Datta et al. [17] benchmark their proposed method by comparison with Lappas et al.'s [81] algorithms, which have not considered Capacity.

Almost all proposed works compare their algorithms with their variations and only a few have been compared with the pioneering works by Lappas et al. [81] and Kargar et al. [73]. In Appendix A, we illustrate the cross-comparing of all proposed algorithms with each other in a table where the main diameter is mostly selected. Lappas et al.'s *RarestFirst* and *EnhanceSteiner* have been the main baselines for works that propose better optimization for communication cost. Also, a few works [83] that propose an extra node attribute on top of communication cost compare their methods with *modified* versions of Lappas et al.'s methods to have a fair comparison.

Last, a common baseline to almost all proposed works has been a *random* algorithm, which randomly selects a subset of experts in the expert graph as a recommended team to provide a minimum comparison basis upon which the proposed methods desire to improve efficacy.

5 Challenges and Future Directions

With a bird's-eye view of the chronological evolution of approaches to Team Formation, we observe that graph-based algorithms represent a significant leap forward by modeling collaboration and social ties utilizing expert graphs [81]. This breakthrough not only surpassed earlier methods that relied solely on operations research [24, 35, 71] but also established graph-based approaches as an indispensable consideration for both operations research [21, 37] and the emerging learning-based paradigm [43, 45, 98] in the field. While there are some aspects of graph-based methods that remain consistent, including expert graph construction and subgraph optimization for an optimum project team, one fundamental aspect has been variable over time, the optimization objectives. Initially, communication cost was the only constraint consistently optimized by early works and treated as a universal constraint [17, 25, 30, 32, 72, 81, 83]. Gradually, other critical constraints such as personnel cost [73, 74], proficiency [111], geographical distance [96] and evolving expert collaborations over time [34] have been included to enhance the effectiveness of the formed project team for the real-world applications. Moreover, early approaches focused on single-objective optimization [17, 25, 30, 32, 72, 81, 83]. However, multi-objective optimization methods began to emerge, initially optimizing both communication cost and

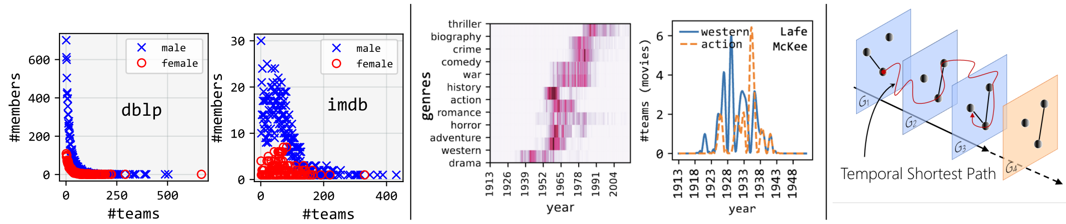


Fig. 8. Left: distribution of genders in DBLP and IMDB. Middle: temporal distribution of movies over genres, and temporal activity of an actor in two genres. Right: temporal expert graph as a stream of graph snapshots in each time interval from the past to the future.

personnel cost simultaneously [73]. As multi-objective optimization methods have expanded, different approaches have introduced new objectives to the model, such as trust score and authority, which are followed by others in different methods [34, 74, 111]. Meanwhile, graph-based methods faced significant challenges in efficiency and robustness. They are computationally intractable because subgraph optimizations are NP-hard [76]. Therefore, techniques for subgraph search space reduction and polynomial heuristics have been proposed. Also, as the expert networks are dynamic and collaborations occur in real-time, few methods have been proposed to accommodate regular changes to the expert graph while preventing a complete re-optimization computation [57, 73]. In the following, also in Appendix D, we go more in depth in some future research directions that show promise in further unleashing the potential of subgraph optimization in forming effective teams, prioritized based on their societal impact and immediate applications to real-world scenarios.

5.1 Fairness and Diversity

The primary focus of existing Team Formation methods is to optimize objectives toward successful teams while largely ignoring fairness and diversity when recommending experts. It has been well-explored that data-driven methods that produce recommendations suffer from unfair biases. They result in discrimination and reduced visibility for an already disadvantaged group [63], disproportionate selection of popular candidates [49, 110], and over/under-representation and racial/gender disparities [78] since they are trained on real-world datasets that already inherit hidden societal biases. Meanwhile, social science research provides compelling evidence about the synergistic effects of diversity on team performance [66, 82]; diversity breeds innovation and increases teams' success by enabling a stronger sense of community and support, reducing conflict, and stimulating more creative thinking. As explained in Section 4 (Evaluation Methodology), having an expert who has participated in many teams previously, i.e., densely connected with other experts, is an advantage for a team. In spite of this or rather because of it, a method that pursues optimization of communication cost based on, e.g., spanning tree or sum of distances, would select/recommend such experts more often, leading to *popularity* bias and overlooking early career experts. Furthermore, no study has been performed on the datasets from the real world in terms of the distribution of demographic attributes like gender, race, or age over teams to uncover the hidden societal biases. As shown in Figure 8 (left), DBLP and IMDB datasets are suffering from gender bias; that is, teams are dominated by the majority males while females are heavily under-represented, which is overlooked by the proposed methods. To the best of our search, there is no fairness-aware algorithmic method that mitigates societal biases in Team Formation methods except that of the recent work by Barnabò et al. [27] that proves *fair* Team Formation is NP-complete; therefore, computationally prohibitive for practical use. Therefore, the front-most future direction is to develop fairness-aware objectives and efficient (fast) optimization techniques in Team Formation methods for recommending a *diverse* list of experts in terms of *i*) popularity and *ii*) demographic attributes (e.g., gender, race, or age), given required skills while almost surely promising the success of the recommended experts considering objectives such as communication cost. It is worth noting that a pure fairness-centric Team Formation algorithm that solely overfits to satisfy fairness, neglecting the success of the team, is also undesirable for and unfair to the organizations, e.g., a team of experts from an under-represented group (e.g., all team members are early career experts or all team members are from the same cultural background) who are unable to accomplish the tasks.

In summary, existing Team Formation methods need urgent further development to counter unfairness. The concrete research path would include: *i*) quantifying notions of fairness for teams, *ii*) automating forming fair yet successful teams, *iii*) mitigating existing unfair biases in real-world datasets by training strategies, *iv*) explore the synergistic balance of fusing bias mitigation methods based on notions of fairness into the Team Formation.

5.2 Temporality

Experts' interests, skills, and levels of expertise change due to society's demands, novel technologies, and working experience. For instance, with the growth of automation, more and more experts are acquiring skills related to computer science, as seen in social science, biology, and linguistics, among other sciences [87, 88]. Figure 8 (middle) demonstrates the *non*-uniform and temporal distribution of movies over genres (skills) and casts and crews (experts) within *yearly* time intervals in IMDB dataset. Although the set of genres remains the same over 100 years, the number of movies that adopt each genre varies over time (e.g., *'drama'* vs. *'musical'*). Further, we observe that actors who were active in 1930s are not active in years after 2000 (presumably due to aging). Therefore, should a method recommend actors in a yet-to-be-made movie in the genre of *'drama'*, it should learn the actors' temporal experiences in this genre from the *past* and avoid recommending *in*active or *in*experienced ones for the *future*.

Despite a large body of graph-based methods, the positive impacts of considering temporality have been considered in no work but in the Selvarajah et al. [34]'s where communication cost has been further weighted based on recency (Section 2.2.1). Also, there has been little work in operations research that used time as a *constraint* to model experts' availability or predefined start and due dates of projects. Durfee et al. [24] considered scheduling constraints or preferences in a two-step Team Formation process. First, teams are built in the matchmaking optimization stage, taking into account the ability to be more readily (re)scheduled with respect to the timing requirement. Next, in the scheduling optimization stage, time slots are allotted to the team such that they minimize the starting times of all the members. Rahmanniyay et al. [95] studied the impact of various factors like weather conditions that can change the duration of a project or delay the delivery of material to a manufacturing company. Yang et al. [23] apply integer programming to determine the optimum team of experts available at a certain point in time. Contrary to considering time as an optimization constraint, consuming time as an *aspect* through which experts' skills and collaboration ties evolve remains an open question and is a promising research direction in graph-based Team Formation methods.

As opposed to graph-based methods that are trained on a stationary expert graph, temporal methods need to consider an evolving expert graph whose expert node attributes (e.g., experts' level of expertise or authority), and edges (collaboration within time) change over time. A commonly proposed solution is to model a temporal expert graph as a stream of snapshots of expert graphs at each time interval, i.e., $\mathcal{G} = [G_0, G_1, \cdots, G_t, \cdots, G_T], 0 \leq t \leq T$, and generalize the definitions for the distance, shortest path, or random walk between a pair of experts but in graphs of different time intervals [38, 92], as shown in Figure 8 (right). Temporal graph has been successfully applied in route planning in a road network [52] and nervous system modeling [101].

Temporal study of Team Formation problem finds immediate application for online, on-demand, and large-scale crowd work in crowdsourcing platforms [29, 48] where experts hold portfolios and timelines, showcasing their successes and additional evidence of their expertise through previously established credentials within time. In a crowdsourcing platform, candidate experts advertise their certified skills and bid prices for their participation while continuously adapting to changing task demands by upskilling or shifting their skills to remain relevant. While there has been Team Formation research on crowdsourcing platforms using social networks [29, 48], no work has considered the temporal aspect of skill set and experts' skills.

5.3 Modern Hybrid Optimization Algorithms

While Pareto optimum is based on trade-offs among multiple, mostly conflicting, objectives, the search space is a single vector space, each dimension of which is one objective, and the target is to find a *single* optimum team among the set of feasible solutions. However, organizations, more often than not, aim to form *multiple* project teams from a shared pool of experts based on different sets of objectives, which may be interrelated. For instance, forming two project teams, one based on maximizing proficiency and minimizing personnel cost, and the other based only on communication cost. Available experts may have a limited capacity of one project team at a time, and hence, forming an optimum project in the first one influences the available pool of experts for the second one, and vice versa. As such, we face two optimization tasks: one bi-objective and one single objective, each of which has its own vector spaces that can be searched *simultaneously*, yet solutions from one space impact the solutions in the other one. To fill the gap, multi-task optimization techniques like multi-factorial optimization can be leveraged [26, 60, 64, 64, 84]. Particularly, multi-factorial optimization employs the implicit parallelism between multiple orthogonal or inter-related search spaces, each of which can be a single- or multi-objective search space individually, for simultaneous concurrent search while seamlessly transferring knowledge between different inter-related search spaces. Multi-factorial or other multi-task optimization algorithms to form inter-related project teams have remained unexplored in subgraph optimization algorithms despite their widespread success in operations research for other problems like engineering structural design [79], supply chain and logistics [97], and algorithmic trading [20], to name a few.

6 Concluding Remarks

We presented a survey of the graph-based category of algorithmic approaches to the Team Formation problem, as the mainstream body of research in this domain. A comprehensive overview of seminal graph-based solutions, including proposed optimization objectives, has been presented. We have identified a shortfall in this field to be the lack of standards and conventions, which we addressed in this survey by proposing and leveraging a unified set of notations to formalize the problem and its different aspects and sub-tasks. Adopting a chronological approach that reflects the advancements over time, we provided a detailed study of the graph-based optimization techniques in the Team Formation. We introduced the benchmark datasets in this field along with their unique characteristics as well as methodologies and metrics for intrinsic, quantitative, and qualitative evaluations of proposed graph-based algorithms in terms of efficacy and efficiency. We identified five major lines of research as future directions based on the pinpointed open issues and challenges in the literature, including fairness concerns, temporality aspects of Team Formation, modern hybrid optimization algorithms, pitfalls to avoid in evaluation (Appendix C), and open-source implementations for proposed algorithms (Appendix D). ~~Within the first, we identified that graph-based Team Formation algorithms have foregone fairness when recommending optimum teams of experts, urging for fairness-aware optimization objectives and algorithms. Regarding temporality aspects, we elaborated on the ever-evolving essence of expert graphs over time and explained how and why it makes intuitive sense to consider time in Team Formation solutions. Considering the very little body of work on the time aspect, we laid out how taking temporality into consideration can help improve the algorithms. In the following, we present modern optimization techniques adapted from operations research, with~~

~~a proven track record in other domains to apply to Team Formation algorithms. Further, we discussed the pitfalls that we might fall into while evaluating approaches. For one, inferring the skill sets for defined teams based on different subsets faces several challenges, such as the unintuitive mapping from words to actual skills in the real world, and these challenges affect the distribution of teams over skills. For another, we argued the strong assumption Team Formation studies make, considering all instances of teams in the datasets are successful. A shortcoming this involves is the lack of leveraging negative instances that, in turn, leads to the inability to utilize algorithmic variations for better-improved performances. Last, despite many graph-based Team Formation algorithms, codebases and details are scarcely publicly available, which resulted in a lack of cross-algorithm benchmarking results, threatening the veracity of outcomes.~~ We believe that these directions considerably help improve state-of-the-art not only in the graph-based within the search-based category but also can act as guidelines to consider in other novel learning-based works.

References

[1] A. An, M. Kargar, and M. Zihayat. Finding affordable and collaborative teams from a network of experts. In *Proceedings of the 13th SIAM International Conference on Data Mining,*, pages 587–595, 2013.

[2] A. Anagnostopoulos et al. Online team formation in social networks. In *Proceedings of the 21st international conference on World Wide Web*, pages 839–848, 2012.

[3] B. K. Baiden and A. D. Price. The effect of integration on project delivery team effectiveness. *International Journal of Project Management*, 29(2):129–136, 2011.

[4] R. B. Bernabé, I. Navia, and F. J. García-Peñalvo. Faat: freelance as a team. In *Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM*, pages 687–694, 2015.

[5] M. T. Brannick et al. *Team performance assessment and measurement: Theory, methods, and applications*. Psychology Press, 1997.

[6] R. Burkard, M. Dell'Amico, and S. Martello. *Assignment problems: revised reprint*. 2012.

[7] K. M. Bursic. Strategies and benefits of the successful use of teams in manufacturing organizations. *IEEE transactions on engineering management*, 39(3):277–289, 1992.

[8] Mi. Cheatham and K. Cleereman. Application of social network analysis to collaborative team formation. In *2006 International Symposium on Collaborative Technologies and Systems, CTS*, pages 306–311, 2006.

[9] O. Chen, F. Paas, and J. Sweller. A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review*, 35(2):63, 2023.

[10] S. Chen. An integrated methodological framework for project task coordination and team organization in concurrent engineering. *Concurr. Eng. Res. Appl.*, 13(3):185–197, 2005.

[11] S. Chen and L. Lin. Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. *IEEE Trans. Engineering Management*, 51(2):111–124, 2004.

[12] P. R. Cohen, H. J. Levesque, and I. A. Smith. On team formation. *Synthese Library*, pages 87–114, 1997.

[13] M. Craig and D. McKeown. How to build effective teams in healthcare. *Nursing times*, 111(14):16–18, 2015.

[14] K. Daniel. *Thinking, fast and slow*. 2011.

[15] A. Dashti, S. Samet, and H. Fani. Effective neural team formation via negative samples. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3908–3912, 2022.

[16] A. Dashti, K. Saxena, D. Patel, and H. Fani. Opentf: A benchmark library for neural team formation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3913–3917, 2022.

[17] S. Datta, A. Majumder, and K. V. M. Naidu. Capacitated team formation problem on social networks. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, pages 1005–1013, 2012.

[18] L. A. DeChurch and J. R Mesmer-Magnus. The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of applied psychology*, 95(1):32–53, 2010.

[19] A. Anagnostopoulos et al. Power in unity: forming teams in large-scale community systems. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM*, pages 599–608, 2010.

[20] A. Gupta et al. Multiobjective multifactorial optimization in evolutionary multitasking. *IEEE Trans. Cybern.*, 47(7):1652–1665, 2017.

[21] C. Ding et al. Teamgen: An interactive team formation system based on professional social network. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 195–199, 2017.

[22] C. Lin et al. Smallblue: Social network analysis for expertise search and collective intelligence. In *Proceedings of the 25th International Conference on Data Engineering, ICDE*, pages 1483–1486, 2009.

[23] D. Yang et al. On social-temporal group query with acquaintance constraint. *Proc. VLDB Endow.*, 4(6):397–408, 2011.

[24] Durfee et al. Using hybrid scheduling for the semi-autonomous formation of expert teams. *Future Gener. Comput. Syst.*, 31:200–212, 2014.

[25] F. Farhadi et al. An effective expert team formation in social networks based on skill grading. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference*, pages 366–372, 2011.

[26] F. Ming et al. Constrained multiobjective optimization via multitasking and knowledge transfer. *IEEE Trans. Evol. Comput.*, 28(1):77–89, 2024.

[27] G. Barnabò et al. Algorithms for fair team formation in online labour marketplaces10033. In *Companion of The 2019 World Wide Web Conference WWW*, pages 484–490, 2019.

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

[29] I. Lykourentzou et al. Team dating: A self-organized team formation strategy for collaborative crowdsourcing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1243–1249, 2016.

[30] J. Huang et al. Forming grouped teams with efficient collaboration in social networks. *Comput. J.*, 60(11):1545–1560, 2017.

[31] J. Juárez et al. A comprehensive review and a taxonomy proposal of team formation problems. *ACM Comput. Surv.*, 54(7):153:1–153:33, 2022.

[32] J. Nemec et al. Rw-team: Robust team formation using random walk. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management*, pages 4759–4763, 2021.

[33] J Tang et al. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 990–998, 2008.

[34] K. Selvarajah et al. A unified framework for effective team formation in social networks. *Expert Syst. Appl.*, 177:114886, 2021.

[35] L. Wang et al. Team recommendation using order-based fuzzy integral and NSGA-II in starcraft. *IEEE Access*, 8:59559–59570, 2020.

[36] L. Wu et al. Graph neural networks for natural language processing: A survey. *Found. Trends Mach. Learn.*, 16(2):119–328, 2023.

[37] M. B. Campêlo et al. The sociotechnical teams formation problem: a mathematical optimization approach. *Ann. Oper. Res.*, 286(1):201–216, 2020.

[38] M. Latapy et al. Stream graphs and link streams for the modeling of interactions over time. *Soc. Netw. Anal. Min.*, 8(1):61:1–61:29, 2018.

[39] M. Twyman et al. Teammate invitation networks: The roles of recommender systems and prior collaboration in team assembly. *Soc. Networks*, 68:84–96, 2022.

[40] M. Zihayat et al. Authority-based team discovery in social networks. In *Proceedings of the 20th International Conference on Extending Database Technology, EDBT*, pages 498–501, 2017.

[41] M. Zihayat et al. Effective team formation in expert networks. In *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management*, volume 2100 of *CEUR Workshop Proceedings*, 2018.

[42] R. Bing et al. Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications. *Artif. Intell. Rev.*, 56(8):8003–8042, 2023.

[43] R. Hamidi Rad et al. Learning to form skill-based teams of experts. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management*, pages 2049–2052, 2020.

[44] R. Hamidi Rad et al. Pytfl: A python-based neural team formation toolkit. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management*, pages 4716–4720, 2021.

[45] R. Hamidi Rad et al. A variational neural architecture for skill-based team formation. *ACM Trans. Inf. Syst.*, 42(1):7:1–7:28, 2024.

[46] S. Wu et al. Graph neural networks in recommender systems: A survey. *ACM Comput. Surv.*, 55(5):97:1–97:37, 2023.

[47] T. Y. Lim et al. An information entropy-based evolutionary computation for multi-factorial optimization. *Appl. Soft Comput.*, 114:108071, 2022.

[48] W. Wang et al. Toward efficient team formation for crowdsourcing in noncooperative social networks. *IEEE Trans. Cybern.*, 47(12):4208–4222, 2017.

[49] Y. Zhang et al. Causal intervention for leveraging popularity bias in recommendation. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 11–20, 2021.

[50] Z. Alshara et al. Pi-link: A ground-truth dataset of links between pull-requests and issues in github. *IEEE Access*, 11:697–710, 2023.

[51] Z. Liang et al. Multi-factorial optimization for large-scale virtual machine placement in cloud computing. *CoRR*, abs/2001.06585, 2020.

[52] Z. Luo et al. Diversified top-k route planning in road network. *Proc. VLDB Endow.*, 15(11):3199–3212, 2022.

[53] Z. Wu et al. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Networks Learn. Syst.*, 32(1):4–24, 2021.

[54] Z. Ye et al. A comprehensive survey of graph neural networks for knowledge graphs. *IEEE Access*, 10:75729–75741, 2022.

[55] E. Fitzpatrick and R. G. Askin. Forming effective worker teams with multi-functional skill requirements. *Comput. Ind. Eng.*, 48(3):593–608, 2005.

[56] R. W. Floyd. Algorithm 97: Shortest path. *Commun. ACM*, 5(6):345, 1962.

[57] A. Gajewar and A. Das Sarma. Multi-skill collaborative teams based on densest subgraphs. In *Proceedings of the Twelfth SIAM International Conference on Data Mining*, pages 165–176, 2012.

[58] M. E. Gaston, J. Simmons, and M. desJardins. Adapting network structure for efficient team formation. In *Artificial Multiagent Learning, Papers from the 2004 AAAI*, volume FS-04-02, pages 1–8, 2004.

[59] G. Gigerenzer and R. Selten. *Bounded rationality: The adaptive toolbox*. The MIT press, 2002.

[60] A. Gupta, Y. Ong, and L. Feng. Multifactorial evolution: Toward evolutionary multitasking. *IEEE Trans. Evol. Comput.*, 20(3):343–357, 2016.

[61] R. Gyanchandani, B. Nathani, and D. Jaroliya. Factors affecting team performance in it sector: An exploratory analysis. *J-GIBS*, 11(1), 2019.

[62] J. R Hackman. *Leading teams: Setting the stage for great performances*. Harvard Business School Press, 2002.

[63] S. Hajian, F. Bonchi, and C. Castillo. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2125–2126, 2016.

[64] W. Han, H. Li, and M. Gong. Multi-regularization sparse reconstruction based on multifactorial multiobjective optimization. *Appl. Soft Comput.*, 136:110122, 2023.

[65] G. Hirst. The handbook of computational linguistics and natural language processing. *Language*, 87(4):897–899, 2011.

[66] B. Hofstra et al. The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, 117(17):9284–9291, 2020.

[67] D. Horowitz and S. D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of the 19th International Conference on World Wide Web, WWW*, pages 431–440, 2010.

[68] D. Horowitz and S. D. Kamvar. Searching the village: models and methods for social search. *Commun. ACM*, 55(4):111–118, 2012.

[69] T. Jansen. *Analyzing Evolutionary Algorithms - The Computer Science Perspective*. Natural Computing Series. 2013.

[70] J. Juárez and C. A. Brizuela. A multi-objective formulation of the team formation problem in social networks: preliminary results. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO*, pages 261–268, 2018.

[71] S. J. Kalayathankal et al. A modified fuzzy approach to project team selection. *Soft Computing Letters*, 3:100012, 2021.

[72] M. Kargar and A. An. 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*, pages 985–994, 2011.

[73] M. Kargar, A. An, and M. Zihayat. Efficient bi-objective team formation in social networks. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD*, volume 7524 of *Lecture Notes in Computer Science*, pages 483–498, 2012.

[74] M. Kargar, M. Zihayat, and A. An. Affordable and collaborative team formation in an expert network. Technical report, Department of Computer Science and Engineering, York University, 2013.

[75] M. Kargaret al. Effective keyword search over weighted graphs. *IEEE Transactions on Knowledge and Data Engineering*, 34(2):601–616, 2020.

[76] R. M. Karp. Reducibility among combinatorial problems. In *Proceedings of a symposium on the Complexity of Computer Computations,*, The IBM Research Symposia Series, pages 85–103, 1972.

[77] J. R. Katzenbach and D. K. Smith. *The wisdom of teams: Creating the high-performance organization*. Harvard Business Review Press, 2015.

[78] M. Kay, C. Matuszek, and S. A. Munson. Unequal representation and gender stereotypes in image search results for occupationsCHI. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3819–3828, 2015.

[79] B. Kodess et al. Reference data and multi-factorial methods for development of modern materials. *Key Engineering Materials*, 910:966–975, 2022.

[80] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78:046110, Oct 2008.

[81] T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 467–476. ACM, 2009.

[82] J. Lauring and F. Villesèche. The performance of gender diverse teams: what is the relation between diversity attitudes and degree of diversity? *European Management Review*, 16(2):243–254, 2019.

[83] C. Li and M. Shan. Team formation for generalized tasks in expertise social networks. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom / IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT*, pages 9–16, 2010.

[84] G. Li, Q. Lin, and W. Gao. Multifactorial optimization via explicit multipopulation evolutionary framework. *Inf. Sci.*, 512:1555–1570, 2020.

[85] P. Liu and Z. Li. Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, 42(6):553–568, 2012.

[86] J. Magill-Evans, M. Hodge, and J. Darrah. Establishing a transdisciplinary research team in academia. *Journal of allied health*, 31(4):222–226, 2002.

[87] J. Mahdavimoghaddam et al. Exploring the utility of social content for understanding future in-demand skills. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), nov 2022.

[88] T. H. McCormick et al. Using twitter for demographic and social science research: Tools for data collection and processing. *Sociological methods & research*, 46(3):390–421, 2017.

[89] M. McGregor. Defining and measuring cost, effort, and load in information retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3490–3490, 2023.

[90] D. H. McKnight, V. Choudhury, and C. J. Kacmar. The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *J. Strateg. Inf. Syst.*, 11(3-4):297–323, 2002.

[91] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[92] L. Oettershagen, P. Mutzel, and N. M. Kriege. Temporal walk centrality: Ranking nodes in evolving networks. In *WWW '22: The ACM Web Conference 2022*, pages 1640–1650, 2022.

[93] L. Paquete, M. Chiarandini, and T. Stützle. Pareto local optimum sets in the biobjective traveling salesman problem: An experimental study. In *Metaheuristics for Multiobjective Optimisation*, volume 535, pages 177–199. 2004.

[94] C. R. Paris, E. Salas, and J. A. Cannon-Bowers. Teamwork in multi-person systems: a review and analysis. *Ergonomics*, 43(8):1052–1075, 2000.

[95] F. Rahmanniyay, A. J. Yu, and J. Seif. A multi-objective multi-stage stochastic model for project team formation under uncertainty in time requirements. *Comput. Ind. Eng.*, 132:153–165, 2019.

[96] S. S. Rangapuram, T. Bühler, and M.s Hein. Towards realistic team formation in social networks based on densest subgraphs. In *22nd International World Wide Web Conference, WWW*, pages 1077–1088, 2013.

[97] A. Rauniyar, R. Nath, and P. K. Muhuri. Multi-factorial evolutionary algorithm based novel solution approach for multi-objective pollution-routing problem. *Comput. Ind. Eng.*, 130:757–771, 2019.

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

[99] S. Sharma and V. Kumar. A comprehensive review on multi-objective optimization techniques: Past, present and future. *Archives of Computational Methods in Engineering*, 29(7):5605–5633, 2022.

[100] P. D. Sherer. Leveraging human assets in law firms: Human capital structures and organizational capabilities. *ILR Review*, 48(4):671–691, 1995.

[101] A. Sizemore and D. S. Bassett. Dynamic graph metrics: Tutorial, toolbox, and tale. *NeuroImage*, 180(Part):417–427, 2018.

[102] M. Sozio and A. Gionis. 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*, pages 939–948, 2010.

[103] D. Stokols et al. The science of team science: overview of the field and introduction to the supplement. *American journal of preventive medicine*, 35(2):S77–S89, 2008.

[104] D. Strnad and N. Guid. A fuzzy-genetic decision support system for project team formation. *Appl. Soft Comput.*, 10(4):1178–1187, 2010.

[105] R. I. Swaab et al. The too-much-talent effect: Team interdependence determines when more talent is too much or not enough. *Psychological Science*, 25(8):1581–1591, 2014.

[106] H. Takahashi. An approximate solution for the steiner problem in graphs. *Math. Japonica*, pages 573–577, 1980.

[107] V. Vazirani. *Approximation algorithms*. 1st edition, 2001.

[108] X. Wang, Z. Zhao, and W. Ng. A comparative study of team formation in social networks. In *Database Systems for Advanced Applications - 20th International Conference, DASFAA*, volume 9049 of *Lecture Notes in Computer Science*, pages 389–404, 2015.

[109] F. Xhafa, L. Barolli, and A. Durresi. Batch mode scheduling in grid systems. *International Journal of Web and Grid Services*, 3(1):19–37, 2007.

[110] E. Yalcin and A. Bilge. Investigating and counteracting popularity bias in group recommendations. *Inf. Process. Manag.*, 58(5):102608, 2021.

[111] M. Zihayat, M. Kargar, and A. An. Two-phase pareto set discovery for team formation in social networks. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pages 304–311, 2014.

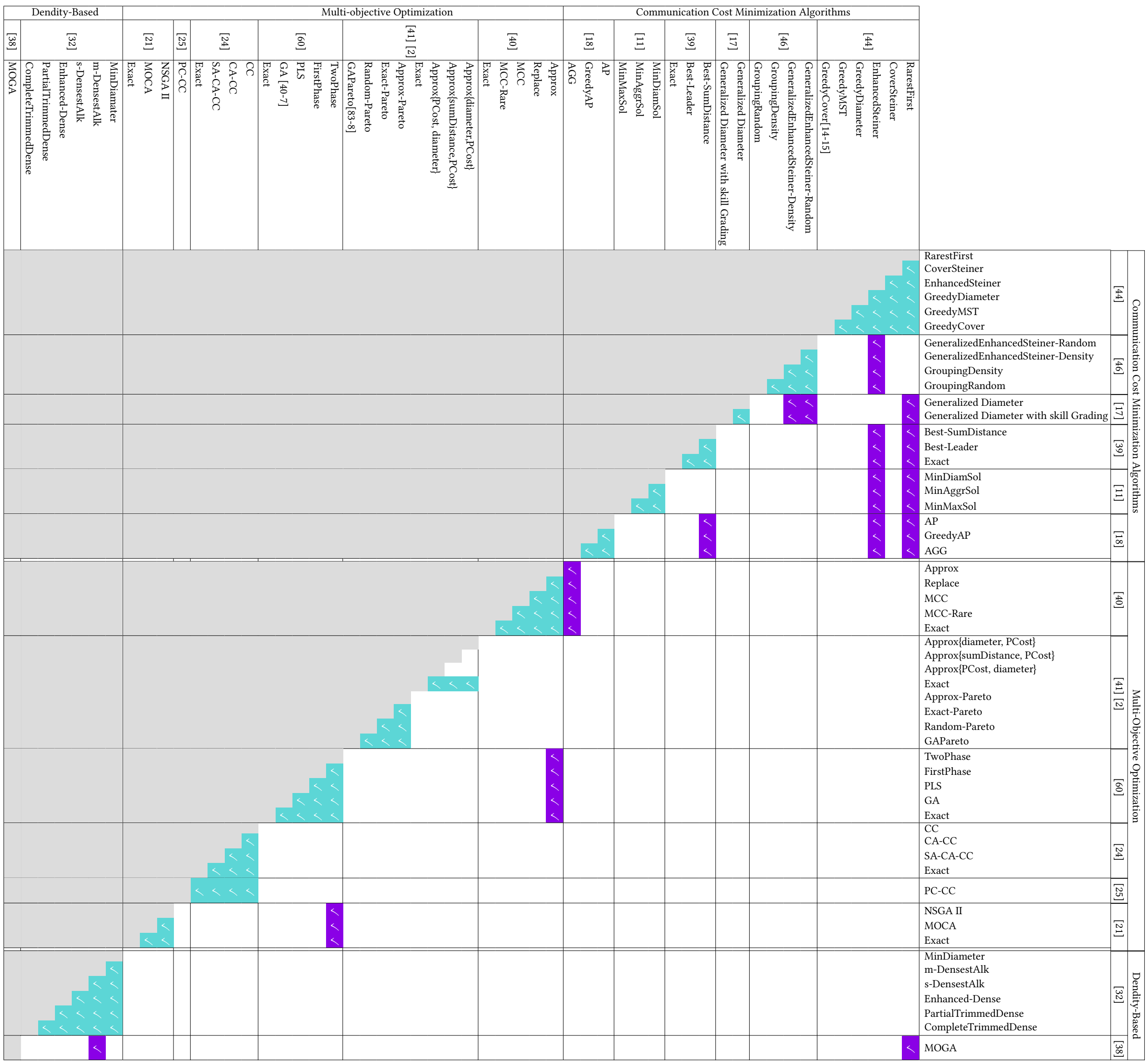
[112] A. Zzkarian and A. Kusiak. Forming teams: an analytical approach. *IIE transactions*, 31(1):85–97, 1999.

Appendices

Appendix A Baselines

In this section, we sketch a who-to-who comparison map for the proposed approaches in the literature to demonstrate how they build upon each other. As indicated in the table, a notable shortcoming in the literature is that most works have not compared their algorithms with other baselines, leaving room for more comprehensive evaluations.

Table 8. Cross-comparison of graph-based Team Formation methods primarily against their variations, with few benchmarked against pioneering works.



Appendix B Background on Team Formation Problem

In this section, we introduce notions around the topic of the team, including itself, aiming at bridging the gap between audiences with different backgrounds. We leverage relevant literature across several disciplines, followed by a brief history of Team Formation studies and their evolution through time. This will facilitate an understanding of the formalizations offered in this survey.

B.1 Team vs. Group

There are varying definitions for the term *'team'* in the literature. Brannik et al. [3] define a team as a group of experts who collaborate together with a common purpose in order to accomplish the requirements of a task. This way, coordination is emphasized as a central characteristic of a team. Zzkarian and Kusiak [61] view teams as a group of experts who independently endeavor to accomplish their individual tasks to reach a shared goal or value, while actively interacting and adapting. Here, we see an emphasis on the roles or functions that each individual has to carry out for the sake of collective gain. In a similar view, Katzenbach and Smith [42] refer to experts as key team members, each of whom must not only have the expertise and be skillful in a needed functional area, but also such skills need to be complementary in order to create a cross-functional team like a city planning team which include *'lead urban architect'*, *'chief environmental planner'*, *'chief police officer'*, and the *'mayor'*. As clear from the aforementioned, a recurring term when discussing teams is *'group'*, which can, and should, be distinguished from *'team'* by several factors. Most prominently, a major team behavior is coordination, which is missing (or sometimes is missed) in groups where several people are actively trying to reach their personal goals although sharing similarities like being like-minded. Further, in teams, experts are responsible for the success of their portion as well as the whole team's mission [3].

B.2 A Successful Team

As the scope and complexity of tasks exceed the capability of each individual, teams are formed to shoulder the burgeoning demand and accomplish the tasks in various functional areas. The way team members are selected is extremely important as it directly determines the *team performance* (throughput), also referred to as team effectiveness, which is a team's potential to accomplish its tasks *successfully* and reach its goals [8]. Team performance can be quantified by indicators like the number of impactful publications for a research team (h-index), or the number of issued patents for a team of inventors. In some domains, however, it remains controversial what constitutes a performance indicator. For example, in the entertainment industry, the performance of the cast and crew for a movie can be measured based on the movie's immediate reception by the people (box office) or critical reviews (ratings) within a long span of time. Accordingly, the best or *optimum* team is the one that achieves or almost surely achieves the highest performance. The team performance depends on how a team is developed. Tuckman [57] introduces four distinct phases of development for a team, including *i*) *forming*: human resource allocation by a leader, *ii*) *storming*: developing shared goals and values resulting in conflicts, *iii*) *norming*: conflict resolution between team members, and *iv*) *performing*: focusing on individual tasks to reach the shared goal or value.

The forming stage, known as Team Formation, also variously referred to as team selection, team allocation, team composition, and team configuration, is the focus of algorithmic studies, also this survey, and involves the selection of team members from a pool of expert candidates in a way that all the required skills are covered while trying to optimize one or several objectives that directly or indirectly maximizes the team performance. Prior to algorithmic approaches, the teams used to be formed manually. This process can be very tedious and time-consuming for a human team selector as well as considerably error-prone. Additionally, the dependence on human judgment can result in

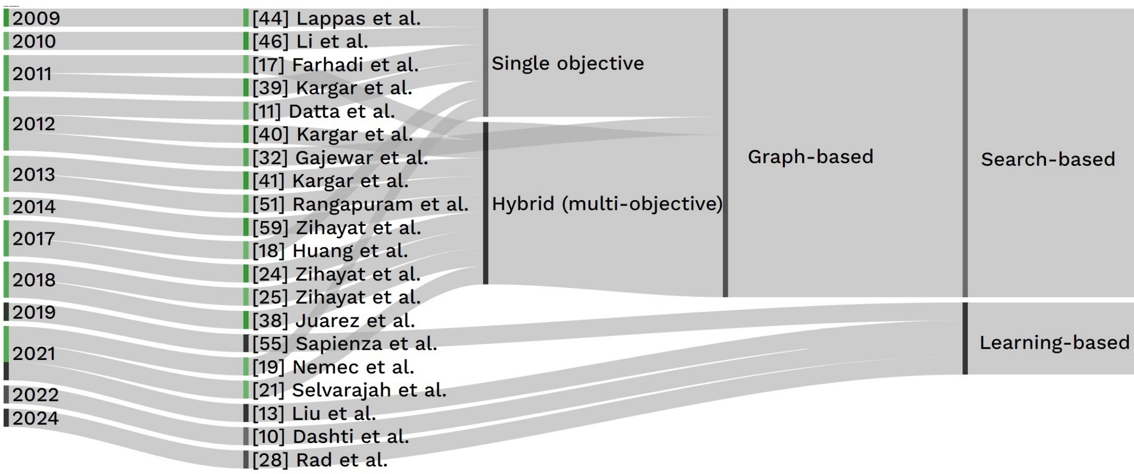


Fig. 9. Team Formation methods in time; a substantial number of methods are graph-based.

subjectivity and *bias* arising from the selector's personal experiences and perceptions. The decision-making process for the selector is even more complex due to several other related factors, including the task importance, budget, time constraints, team size, and geographical proximity, among others. On a personal level, the characteristics of candidates for forming the team have a considerable impact on the Team Formation procedure. Such elements include technical abilities, availability, individual cost, productivity, behavior, personality, knowledge, negotiation skills, proactivity, and intra-team communication. The tremendous difficulty of this process has led to an active field of research where researchers from several disciplines endeavor to propose algorithmic ways for Team Formation.

B.3 History of Computational Team Formation

The problem of Team Formation has been approached through *i*) search-based and, subsequently, *ii*) learning-based methods. Both of these paradigms aim at forming the *best* or *optimum* team, that is, a team with maximum team performance with respect to human and non-human constraints. Within the context of search-based works, solutions were generally based on operation research and graph-based modeling to optimize multiple objective functions.

The operation research approaches produced some forms of mathematical models [9, 14, 15, 36, 59] such as fuzzy model [22] for Team Formation problem. These works mainly performed the optimization either through genetic algorithms [1, 4, 5, 22, 38] or via an integer linear and/or nonlinear programming on a large search space containing all the possible subsets of the set of skilled candidates [12, 16, 23, 31, 31]. Operation research work, however, was premised on the mutually independent selection of candidates and overlooked the organizational and/or social ties among them.

The graph-based modeling solutions, the subject of this survey, include a significantly larger number of studies where the problem of Team Formation has been translated into graph mining. Graph has been employed to fill the gap by incorporating social and collaborative ties, and other interpersonal features using measures such as density, degree centrality and closeness centrality. In Team Formation, the network of experts is considered as a graph with experts being represented as its nodes and the edges representing the ties between them, e.g., collaborative ties [25], hence providing a modeling convenience for this problem. Meanwhile, the exponential growth of the Internet has led to the advent of services such as LinkedIn[4] and GitHub[5] where experts connect with each other and create social ties (e.g., followership). In graph-based methods, the search for an optimum team occurs over all possible subgraphs and the most optimized subgraph is sought to be the optimum team. A seminal work in this category is by Lappas et al. [44] who defined an optimum team as a group of experts with minimum communication cost and reduced it into a subgraph with the minimum diameter or minimum spanning tree. We lay out more details on why graph-based methods build up a following and become well-established in Team Formation literature in the next section.

From Figure 9, recently, a paradigm shift to machine learning-based methods, including artificial neural networks and graph neural networks, has been observed due to advanced hardware computing power, especially graphic processing units (GPUs) and tensor processing units (TPUs), that reduced elapsed time from months to days and/or hours opening doors to the analysis of massive collections of candidates coming from different fields. These methods are different from search-based solutions in that they *learn* the inherent structure of the ties among candidates and their skills. Wherein, all past successful and *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. This line of research started with Sapienza et al. [55] who employed a non-variational autoencoder neural network architecture and is being followed by researchers through other neural-based architectures such as variational Bayesian neural network [10, 13, 26–28]. Although learning-based literature is novel, it is still in its early stages. We chose to survey the graph-based approaches as they comprise the mainstream and a relatively larger body of research to be investigated in a survey such as the present one. Further, this body of work serves as a prelude to the learning-based works, especially the ones based on graph neural network in terms of the Team Formation problem definition as well as experimental and evaluation settings. We are however tracing the learning-based paradigm's literature and its progress for a future potential survey.

B.4 Why Graph-based Team Formation

In summary, due to *i*) organizations' inherent hierarchical network structure, *ii*) experts' social and collaborative ties, and *iii*) the synergistic interdisciplinary discoveries from social network analysis and graph theory, graphs have become a natural choice and pervasive in Team Formation literature and have been firmly established not only in search-based methods but also in the newly emerged learning-based methods [26]. Therefore, a survey of the graph-based literature is both beneficial for the community and quite timely.

Originally, organizations fostered the formation of teams to accomplish complicated tasks. Organizations are designed based on hierarchical *network* structures that impact the Team Formation process. Initial studies of Team Formation in organizational dynamics suggested the importance of network structure on organizational performance [33]. Huberman and Hogg [37] use a probabilistic modeling approach to study organizational change in large-scale organizations, finding that clustered organizations are more stable than organizations with dispersed connectivities. Similarly, Glance and Huberman [35] show that collaboration is more likely in hierarchically structured organizations with fluid (changeable) groupings, and Miller [48] demonstrates the importance of network structures on information processing.

Later findings suggest that not only organizations but also real-world *social* networks have rich, purposeful, and meaningful structures, which have significant effects on the overall team performance. In other words, a key performance indicator of a team is the underlying network topology, which determines the direct interactions among the individuals

[4] `https://linkedin.com/`
[5] `https://github.com/`

of the organization [7, 33, 34]. Via simulation on different synthetic network structures, Gaston et al. [33, 34] were among the first who demonstrated the impact of the network structure on the number of possible teams for an optimum team given a set of tasks. Given a synthetic set of experts (agents), they form an attributed network whose nodes are experts associated with a single skill as their attributes and edges represent the organizational structure. They define a *valid* team as a connected sub-network (subgraph) of agents whose collective skills cover the skill requirements for a given task. They showed that *scale-free* networks wherein connections (nodes degree) follow a power law distribution with a long tail, that is few agents have many connections while the majority has scarcely connected, can cover more valid teams and, hence are more efficient for the set of tasks compared to networks of types lattice and small-world in which most agents are not neighbors of one another yet can be reached from every other by a path.

On the other hand, social network analysis (SNA), the process of studying interactions among individuals or groups in a network where the data is inherently relational, has been employed to incorporate interpersonal attributes such as communication, collaboration attributes such as the number of projects and the amount of time the team members worked together, and social attributes such as their level of friendship and the number of co-worker friends they have in common [34, 44, 56]. For instance, coordination, communication and cooperation can be analyzed using measures of social network analysis, such as density, degree centrality and closeness centrality [49]. Cheatham et al. [6] used techniques from social network analysis to help identify colleagues that would be helpful in a newly formed team [6]. An attributed weighted network whose nodes are experts with attributes (e.g., skills) was formed, and weighted links are established either *explicitly* or are *inferred* based on interpersonal attributes between experts. Team Formation then translates into finding communities and *densely* connecting subgraphs.

Affiliation networks wherein experts are related by membership in a community (club), co-authorship on scholarly publication, and hierarchical organization of employees in a company are samples of social networks with explicit links that are readily available (e.g., LinkedIn[6]) and encode the fact that experts in the same community or department, or experts that have had successful collaborations in the past can communicate easier leading to a more effective team formation.

Appendix C Pitfalls in Evaluation

C.1 Inferring Skills

As explained in Section 4.1 (Datasets), proposed Team Formation methods have been benchmarked on different datasets from varying domains where skill sets are either a predefined set or have to be inferred during the experimental setup while the choice of the expert set has been straightforward. In DBLP, while a given publication already includes keywords or fields of study (fos) representing the scope of the paper and could be adopted as the required skills for that publication [32, 38], most researchers have adopted customized policies to infer the skill set from words in the titles of publications [17, 24, 39–41, 44, 60]; this heavily impacts the distribution of teams over skills. For instance, Kargar et al. [2, 39–41] rank the words based on the number of occurrences in publications' titles and filter out sparse words as non-informative and only keep frequent ones, as a result of which the distribution of teams over the inferred skill set becomes more concentrated around commonly occurring skills, that is, each skill has been adopted in many teams. Contrary to DBLP, choosing keywords in a moving picture's title in IMDB to represent the required skills for the moving picture's production [39] could be misleading. For example, for the movie *'Rosemary's Baby'*[7], the keywords *'Rosemary'* or *'Baby'* fall short of representing the skills required for making such a movie. A promising approach, which has

[6]https://www.linkedin.com/
[7]https://en.wikipedia.org/wiki/Rosemary's_Baby_(film)

received no attention yet, can be considering keywords in the moving picture's plot as the required skills since they describe the narrative sequence of events and characters in the moving picture. A possible alternative can be to infer the skill set based on the genres and subgenres of the moving pictures. For instance, our example movie, *'Rosemary's Baby'* has been made in the genres of {*'horror'*, *'mystery'*, *'drama'*}, which is a better representative for the required skills. The choice of genres in the IMDB domain would, however, pose challenges for the proposed methods when performing a comparative analysis of results across different domains, e.g., DBLP vs. IMDB, since there are a limited number of genres and subgenres, and this yields a different distribution of teams over skills, compared to DBLP [39]. Similar challenges can be pointed out in the GitHub dataset, where programming languages are assumed to be the set of skills. Therefore, it is crucial to consider data distribution in different domains for a fair comparison of algorithms, which is overlooked by existing graph-based Team Formation methods; they ill-assumed that the distributions of teams over skills are similar or they made them similar.

C.2 Unsuccessful Teams

Oddly, graph-based Team Formation methods have assumed all existing instances of teams in a dataset are successful (positive samples), and they filter out team instances which are not complying with their definition of success, like having a given minimum number of publications for a research group in DBLP domain, and performed solely on remaining teams as successful ones thereafter. This is because most real-world datasets in the Team Formation literature do not have explicit unsuccessful teams as negative samples (e.g., collections of rejected papers in DBLP), or what constitutes a failure remains controversial. For example, in the movie industry, a movie's success can be measured based on its immediate reception by the people (box office) or critical reviews (ratings) within a long span of time. While literature in other disciplines has shown that leveraging not only positive samples (e.g., friendship in social networks) but also negative samples (e.g., distrust) convey complementary signals to an algorithm and improve accuracy in various tasks [20, 29, 30, 43, 45, 50, 52], no work in graph-based Team Formation, however, has benchmarked the proposed algorithms with *un*successful teams or proposed an optimization objective that leverages both successful and unsuccessful teams and investigated the synergistic effect of utilizing unsuccessful teams during optimization.

Appendix D Open-Source Implementation

As shown in Appendix A (Baselines), the cross-comparison of the results among proposed algorithms is limited mainly because the proposed algorithms suffer from the lack of public availability in terms of implementation and experimental setup, limiting them being a comparative baseline. We investigated online public web pages associated with the authors of the graph-based papers; our search yielded *no* public code repositories, except that of Wang et al. [58] who *re*-implemented a limited number of Team Formation algorithms, signifying that accessibility and reproducibility of proposed algorithms were *not* a concern among researchers in this area. Below, we present proposals that we believe the community, including researchers, conference organizers, and institutions, should collectively pursue to foster better benchmarking practices and, subsequently, easier and more unified comparative analyses.

D.1 Proposing and Standardizing Benchmark Libraries for Team Formation

An interesting avenue of research is to solicit benchmark libraries that repeat, reproduce, generalize, and analyze prior work with a focus on generating new findings from the re-application of established approaches akin to a test of time and investigating the extent to which assumptions of the original work hold up. Particularly, *reproducibility* efforts by different teams under different experimental setups, rather than *replicability* efforts, i.e., with different teams but the same experimental setup, could be of more interest for generating new research insights with the existing approaches.

D.2 Organizing Task-Specific Tutorials and Workshops at Cross-Disciplinary Forums and Conferences.

Tutorials [53, 54] and workshops at conferences or forums with closely related problems/agenda, like Society for Industrial and Applied Mathematics (SIAM)'s Symposium on Discrete Algorithms, the annual meeting by the Institute for Operations Research and the Management Sciences (INFORMS), Special Interest Group on Information Retrieval (SIGIR), and Neural Information Processing Systems (NeurIPS), can introduce Team Formation-specific challenges. These challenges could feature realistic constraints in Team Formation from different domains and would encourage researchers to push the boundaries of current algorithms using cross-disciplinary methodologies used in Team Formation (graph-based, learning-based, and OR-based), helping to foster hybrid models that take advantage of each paradigm's strengths. Leaderboards could drive ongoing improvements while creating a reference point for comparison across different methods, fostering fair comparative analysis.

D.3 Addressing Ethical Considerations in Team Formation Benchmarks.

Ethical issues in Team Formation, such as fairness in the assignment of roles based on skills and preventing biases in the selection of team members (e.g., gender or cultural biases), must be prioritized. Ethical guidelines, potentially promoted at conferences like Fairness, Accountability, and Transparency in AI (FAccT), should ensure benchmarks promote fairness and inclusivity. The community should avoid perpetuating biases or harmful practices that could arise from poorly designed datasets or biased assumptions in the Team Formation process.

Appendix E Optimization Constraints

Optimization constraints, as discussed in Section 2.4, represent a set of predefined restrictions that vary across different domains and are determined by project requirements. Table E provides a summary of the constraints considered in Team Formation studies.

	Authority (a_v)	#Experts per Skill (l_S)	Team Size ($\|V_p\|$)	Capacity (c_v)	Budget (b_v)	Skill Coverage (S_v)	Inter-expert Distance ($d(v, v')$)
[44] Lappas et al. (2009)						✓	
[46] Li et al. (2010)						✓	✓
[17] Farhadi et al. (2011)	✓	✓				✓	
[39] Kargar et al. (2011)						✓	
[32] Gajewar et al. (2012)		✓				✓	
[40] Kargar et al. (2012)						✓	
[11] Majumder et al. (2012)				✓			
[2] Kargar et al. (2013)						✓	
[51] Rangapuram et al. (2013)			✓		✓	✓	
[47] Li et al. (2015)		✓					
[18] Huang et al. (2017)			✓	✓			
[24] Zihayat et al. (2017)						✓	

Table 9. Constraints used in graph-based Team Formation algorithms. As seen, almost all algorithms consider the Skill Coverage (S_v) as the primary constraint. The works for each dataset category have been sorted by the time of publication.

Table 10: Summary of filtering methods to address scalability and efficiency of the proposed graph-based Team Formation algorithms. The works for each dataset category have been sorted by the time of publication.

Dataset	Work	Project Team (p)	Skill Set (S)	Expert Node Set (V)	Timestamp	Edge Set (E)
DBLP	[44] Lappas et al. (2009)	Topics: Database (DB), Data Mining (DM), Artificial Intelligence (AI), Theory (T). Conferences: DB={SIGMOD, VLDB, ICDE, ICDT, EDBT, PODS}, DM={WWW, KDD, SDM, PKDD, ICDM}, AI={ICML, ECML, COLT, UAI}, T={SODA, FOCS, STOC, STACS}.		Published at least 3 papers.		
	[46] Li et al. (2010)	Topics: Data Mining. Conferences: KDD, ICDM, SDM, PAKDD, PKDD, ICML, CIKM, WWW, and SIGIR.	The set of terms occurring in at least 4 paper titles.			
	[17] Farhadi et al. (2011)	Topics: Database, Data Mining, Artificial Intelligence, Theory.	Extracted from Bibsonomy tag tools instead of paper titles.	Published at least 3 papers.		Co-authored at least 2 papers.
	[39] Kargar et al. (2011)	Topics: Database (DB), Data Mining (DM), Artificial Intelligence (AI), Theory (T). Conferences: DB={SIGMOD, VLDB, ICDE, ICDT, EDBT, PODS}, DM={WWW, KDD, SDM, PKDD, ICDM}, AI={ICML, ECML, COLT, UAI}, T={SODA, FOCS, STOC, STACS}.		Published at least 3 papers.		Co-authored at least 2 papers.
	[11] Majumder et al. (2012)	Topics: Database, Data mining, Artificial Intelligence, Theory.				Co-authored at least 2 papers.
	[32] Gajewar et al. (2012)	Topics: Database (DB), Data Mining(DM), Artificial Intelligence (AI) and Theory (T). Conferences: DB={SIGMOD, VLDB, ICDE, ICDT, EDBT, PODS}, DM={WWW, KDD, SDM, PKDD,CDM}, AI={ICML, ECML, COLT, UAI}, T={SODA, FOCS, STOC, STACS, ICALP, ESA}.		Authors with at least three papers in the domains.	May 17, 2010	
	[40] Kargar et al. (2012)	Topics: Database, Data mining, Artificial Intelligence, Theory.				Co-authored at least 2 papers.
	[41] Kargar et al. (2013)	Topics: Database, Data mining, Artificial Intelligence, Theory.	Terms in the titles of at least 2 papers of the expert.	Published at least 3 papers.		Co-authored at least 2 papers.
	[2] Kargar et al. (2013)		Published at least 3 papers.	Published at least 3 papers.		
	[51] Rangapuram et al. (2013)	Topics: Database (DB), Data Mining(DM), Artificial Intelligence (AI) and Theory (T). Conferences: DB={SIGMOD, VLDB, ICDE, ICDT, EDBT, PODS}, DM={WWW, KDD, SDM, PKDD,CDM}, AI={ICML, ECML, COLT, UAI}, T={SODA, FOCS, STOC, STACS, ICALP, ESA}.	Terms in the titles of at least 2 papers of the expert.	Published at least 3 papers.		
	[60] Zihayat et al. (2014)	Topics: Database, Data mining, Artificial Intelligence, Theory.		Published at least 3 papers.		Co-authored at least 2 papers.
	[24] Zihayat et al. (2017)		Terms in the titles of at least 2 papers of the expert.	Potential skill holders, junior researchers with fewer than 10 papers.		
	[18] Huang et al. (2017)	Topics: Database (DB), Data Mining (DM), Artificial Intelligence (AI), Theory (T). Conferences: DB={SIGMOD, VLDB, ICDE, ICDT, EDBT, PODS}, DM={WWW, KDD, SDM, PKDD, ICDM}, AI={ICML, ECML, COLT, UAI}, T={SODA, FOCS, STOC, STACS}.		Published at least 3 papers.		
	[25] Zihayat et al. (2018)				Up to 2015.	
	[38] Juarez et al. (2018)	Topics: Database (DB), Data Mining(DM), Artificial Intelligence (AI) and Theory (T). Conferences: DB={SIGMOD, VLDB, ICDE, ICDT, EDBT, PODS}, DM={WWW, KDD, SDM, PKDD,CDM}, AI={ICML, ECML, COLT, UAI}, T={SODA,FOCS, STOC, STACS, ICALP, ESA}.	Authors with at least three papers in the domains.			Co-authored at least 2 papers.
	[19] Nemec et al. (2021)		Extracted from her publications' titles.			Co-authored at least 1 paper.

Continued on next page

Table 10: Summary of filtering methods to address scalability and efficiency of the proposed graph-based Team Formation algorithms. The works for each dataset category have been sorted by the time of publication. (Continued)

Dataset	Work	Project Team (p)	Skill Set (S)	Expert Node Set (V)	Timestamp	Edge Set (E)
IMDB	[17] Farhadi et al. (2011)			Actors who have played in at least 8 movies.		
	[40] Kargar et al. (2012)			Actors who have played in at least 8 movies.	Movies from 2000 to 2002.	Co-played in at least 4 movies.
	[41] Kargar et al. (2013)			Actors who have played in at least 8 movies.		Co-played in at least 4 movies.
	[60] Zihayat et al. (2014)			Actors who have played in at least 8 movies.		Co-played in at least 4 movies.
GitHub	[32] Gajewar et al. (2012)					Co-authored at least 2 papers.
	[18] Huang et al. (2017)					Subgraphs ((10k to 135k nodes) are generated by a breadth-first search around a selected node.
	[19] Nemec et al. (2021)			IT professionals (e.g., developers) extracted from the titles of the projects they worked on.		If they work on the same project.

Appendix F Efficiency and Scalability Enhancement

Large-scale datasets lead to complex graphs, which pose significant computational challenges in subgraph optimization algorithms. Researchers employed strategies to enhance scalability and reduce computational costs through methods such as filtering datasets or using heuristics during optimization. Table 10 illustrates the filtering methods utilized within graph-based Team Formation algorithms.

Appendix G Benchmark Datasets

This section complements Section 4.1 by providing a summary of major datasets used in Team Formation approaches as summarized in Table 11 which includes computer science publications (DBLP), moving pictures (IMDB) and open-source software (GitHub).

Table 11. Mainstream datasets in Team Formation research studies.

Dataset	Used by	Scale	Nodes (experts)	Edges	Skills	Projects	Public Link	Version
	Lappas et al. (2009) [44]	Mid	5,508	11,905	1,792			
	Li et al. (2010) [46]		5,482	10,339	11,905			
	Farhadi et al. (2011) [17]		8,805	5,588	1,792			April 12, 2006
	Kargar et al. (2011) [39]		5,658	8,588			`http://dblp.uni-trier.de/xml/`	
	Datta et al. (2012) [11]		7,332			19,248		
DBLP	Kargar et al. (2012) [40]		5,658	8,588			`http://dblp.uni-trier.de/xml/`	
	Gajewar et al. (2012) [32]		6,137					May 17, 2010
	Kargar et al. (2013) [2, 41]		6,229	9,400		100	`http://dblp.uni-trier.de/xml/`	April 25, 2011
	Zihayat et al. (2014) [60]					100	`http://dblp.uni-trier.de/xml/`	
	Zihayat et al. (2017) [24]	Large	40,000	125,000		50	`http://dblp.uni-trier.de/xml/`	Up to 2015
	Zihayat et al. (2018) [25]					50	`http://dblp.uni-trier.de/xml/`	Up to 2015
	Kargar et al. (2011) [39]		6,774	35,875			`http://www.imdb.com/interfaces`	
IMDB	Kargar et al. (2012) [40]		6,784	35,875			`http://www.imdb.com/interfaces`	2000-2002
	Kargar et al. (2013) [2, 41]		6,784	35,875		100	`http://www.imdb.com/interfaces`	
GitHub	Datta et al. (2012) [11]	Large	135,346		52	905,000		
	Huang et al. (2017) [18]		5,276	259,084	50	3,750		
	Li et al. (2010) [46]					100		
Synthetic	Huang et al. (2017) [18]	Small	100	1,000	50			
	Selvarajah et al. (2021) [21]		200	10,136				

References for the Appendices

[1] M. Agarwal, N. Kumar, and L. Vig. Non-additive multi-objective robot coalition formation. *Expert Syst. Appl.*, 41(8):3736–3747, 2014.

[2] A. An, M. Kargar, and M. Zihayat. Finding affordable and collaborative teams from a network of experts. In *Proceedings of the 13th SIAM International Conference on Data Mining,*, pages 587–595, 2013.

[3] M. T. Brannick et al. *Team performance assessment and measurement: Theory, methods, and applications.* Psychology Press, 1997.

[4] J. Moreno Cadavid, D. Arturo Ovalle, and R. Maria Vicari. A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Comput. Educ.*, 58(1):560–569, 2012.

[5] Z. Che Ani et al. A method for group formation using genetic algorithm. *International Journal on Computer Science and Engineering*, 02:3060–3064, 01 2010.

[6] Mi. Cheatham and K. Cleereman. Application of social network analysis to collaborative team formation. In *2006 International Symposium on Collaborative Technologies and Systems, CTS*, pages 306–311, 2006.

[7] C. Chen. A multi-level study of free-loading in dynamic groups: The importance of initial network topology. In *2nd International Conference on Intelligent Networking and Collaborative Systems, INCoS*, pages 16–23, 2010.

[8] N. J. Cooke and M. L. Hilton. *Enhancing the effectiveness of team science.* National Academies Press, 2015.

[9] M. Craig, D. Horton, and F. Pitt. Forming reasonably optimal groups: (FROG). In *Proceedings of the 2010 International ACM SIGGROUP Conference on Supporting Group Work, GROUP*, pages 141–150, 2010.

[10] A. Dashti, S. Samet, and H. Fani. Effective neural team formation via negative samples. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3908–3912, 2022.

[11] S. Datta, A. Majumder, and K. V. M. Naidu. Capacitated team formation problem on social networks. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, pages 1005–1013, 2012.

[12] A. Anagnostopoulos et al. Power in unity: forming teams in large-scale community systems. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM*, pages 599–608, 2010.

[13] B. Liu et al. Coach-player multi-agent reinforcement learning for dynamic team composition. In *Proceedings of the 38th International Conference on Machine Learning, ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 6860–6870, 2021.

[14] C. Dorn et al. Interaction mining and skill-dependent recommendations for multi-objective team composition. *Data Knowl. Eng.*, 70(10):866–891, 2011.

[15] D. Gao et al. Top-k team recommendation in spatial crowdsourcing. In *Web-Age Information Management - 17th International Conference, WAIM*, volume 9658 of *Lecture Notes in Computer Science*, pages 191–204, 2016.

[16] Durfee et al. Using hybrid scheduling for the semi-autonomous formation of expert teams. *Future Gener. Comput. Syst.*, 31:200–212, 2014.

[17] F. Farhadi et al. An effective expert team formation in social networks based on skill grading. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference*, pages 366–372, 2011.

[18] J. Huang et al. Forming grouped teams with efficient collaboration in social networks. *Comput. J.*, 60(11):1545–1560, 2017.

[19] J. Nemec et al. Rw-team: Robust team formation using random walk. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management*, pages 4759–4763, 2021.

[20] J. Tang et al. A survey of signed network mining in social media. *ACM Comput. Surv.*, 49(3):42:1–42:37, 2016.

[21] K. Selvarajah et al. A unified framework for effective team formation in social networks. *Expert Syst. Appl.*, 177:114886, 2021.

[22] L. Wang et al. Team recommendation using order-based fuzzy integral and NSGA-II in starcraft. *IEEE Access*, 8:59559–59570, 2020.

[23] M. B. Campêlo et al. The sociotechnical teams formation problem: a mathematical optimization approach. *Ann. Oper. Res.*, 286(1):201–216, 2020.

[24] M. Zihayat et al. Authority-based team discovery in social networks. In *Proceedings of the 20th International Conference on Extending Database Technology, EDBT*, pages 498–501, 2017.

[25] M. Zihayat et al. Effective team formation in expert networks. In *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management*, volume 2100 of *CEUR Workshop Proceedings*, 2018.

[26] R. Hamidi Rad et al. Learning to form skill-based teams of experts. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management*, pages 2049–2052, 2020.

[27] R. Hamidi Rad et al. Pytfl: A python-based neural team formation toolkit. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management*, pages 4716–4720, 2021.

[28] R. Hamidi Rad et al. A variational neural architecture for skill-based team formation. *ACM Trans. Inf. Syst.*, 42(1):7:1–7:28, 2024.

[29] T. Mikolov et al. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems*, pages 3111–3119, 2013.

[30] W. Zhang et al. Optimizing top-n collaborative filtering via dynamic negative item sampling. In *The 36th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR*, pages 785–788, 2013.

[31] E. Fitzpatrick and R. G. Askin. Forming effective worker teams with multi-functional skill requirements. *Comput. Ind. Eng.*, 48(3):593–608, 2005.

[32] A. Gajewar and A. Das Sarma. Multi-skill collaborative teams based on densest subgraphs. In *Proceedings of the Twelfth SIAM International Conference on Data Mining*, pages 165–176, 2012.

[33] M. Gaston and M. desJardins. Team formation in complex networks. In *Proceedings of the 1st NAACSOS Conference*. Citeseer, 2003.

[34] M. E. Gaston, J. Simmons, and M. desJardins. Adapting network structure for efficient team formation. In *Artificial Multiagent Learning, Papers from the 2004 AAAI*, volume FS-04-02, pages 1–8, 2004.

[35] N. S. Glance and B. A. Huberman. Organizational fluidity and sustainable cooperation. In *From Reaction to Cognition, 5th European Workshop on Modelling Autonomous Agents, MAAMAW, Selected Papers*, volume 957 of *Lecture Notes in Computer Science*, pages 89–103, 1993.

[36] M. Hayano, D. Hamada, and T. Sugawara. Role and member selection in team formation using resource estimation for large-scale multi-agent systems. *Neurocomputing*, 146:164–172, 2014.

[37] B. A. Huberman and T. Hogg. Communities of practice: Performance and evolution. *Comput. Math. Organ. Theory*, 1(1):73–92, 1995.

[38] J. Juárez and C. A. Brizuela. A multi-objective formulation of the team formation problem in social networks: preliminary results. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO*, pages 261–268, 2018.

[39] M. Kargar and A. An. 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*, pages 985–994, 2011.

[40] M. Kargar, A. An, and M. Zihayat. Efficient bi-objective team formation in social networks. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD*, volume 7524 of *Lecture Notes in Computer Science*, pages 483–498, 2012.

[41] M. Kargar, M. Zihayat, and A. An. Affordable and collaborative team formation in an expert network. Technical report, Department of Computer Science and Engineering, York University, 2013.

[42] J. R. Katzenbach and D. K. Smith. *The wisdom of teams: Creating the high-performance organization*. Harvard Business Review Press, 2015.

[43] J. Kunegis, J. Preusse, and F. Schwagereit. What is the added value of negative links in online social networks? In *22nd International World Wide Web Conference, WWW*, pages 727–736, 2013.

[44] T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 467–476. ACM, 2009.

[45] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web, WWW*, pages 641–650, 2010.

[46] C. Li and M. Shan. Team formation for generalized tasks in expertise social networks. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom / IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT*, pages 9–16, 2010.

[47] C. Li, M. Shan, and S. Lin. On team formation with expertise query in collaborative social networks. *Knowl. Inf. Syst.*, 42(2):441–463, 2015.

[48] J. H. Miller. Evolving information processing organizations. Papers _001, Carnegie Mellon, Department of Decision Sciences, May 1995.

[49] C. S. Pereira and A. L. Soares. Improving the quality of collaboration requirements for information management through social networks analysis. *Int. J. Inf. Manag.*, 27(2):86–103, 2007.

[50] P. Qin, W. Xu, and J. Guo. A novel negative sampling based on TFIDF for learning word representation. *Neurocomputing*, 177:257–265, 2016.

[51] S. S. Rangapuram, T. Bühler, and M.s Hein. Towards realistic team formation in social networks based on densest subgraphs. In *22nd International World Wide Web Conference, WWW*, pages 1077–1088, 2013.

[52] S. Rendle and C. Freudenthaler. Improving pairwise learning for item recommendation from implicit feedback. In *Seventh ACM International Conference on Web Search and Data Mining, WSDM*, pages 273–282, 2014.

[53] M. Saeedi, C. Wong, and H. Fani. Collaborative team recommendation for skilled users: Objectives, techniques, and new perspectives. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, UMAP Adjunct 2024, Cagliari, Italy, July 1-4, 2024*, 2024.

[54] M. Saeedi, C. Wong, and H. Fani. Paradigm shifts in team recommendation: From historical subgraph optimization to emerging graph neural network. In *Adjunct Proceedings of the 2nd International ACM SIGIR Conference on Information Retrieval in the Asia Pacific*, 2024.

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

[56] M. Sozio and A. Gionis. 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*, pages 939–948, 2010.

[57] B. W. Tuckman. Developmental sequence in small groups. *Psychological bulletin*, 63(6):384, 1965.

[58] X. Wang, Z. Zhao, and W. Ng. A comparative study of team formation in social networks. In *Database Systems for Advanced Applications - 20th International Conference, DASFAA*, volume 9049 of *Lecture Notes in Computer Science*, pages 389–404, 2015.

[59] M. Xie, L. V. S. Lakshmanan, and P. T. Wood. Breaking out of the box of recommendations: from items to packages. In *Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys*, pages 151–158, 2010.

[60] M. Zihayat, M. Kargar, and A. An. Two-phase pareto set discovery for team formation in social networks. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pages 304–311, 2014.

[61] A. Zzkarian and A. Kusiak. Forming teams: an analytical approach. *IIE transactions*, 31(1):85–97, 1999.