Effective Team Formation in Expert Networks

Morteza Zihayat*, Aijun An$, Lukasz Golab†,
Mehdi Kargar*, and Jaroslaw Szlichta#

*Ryerson University, Toronto, Canada, $York University, Toronto, Canada, †University of Waterloo, Waterloo, Canada, #University of Ontario Institute of Technology, Oshawa, Canada
mzihayat@ryerson.ca, aan@cse.yorku.ca, lgolab@uwaterloo.ca,
kargar@ryerson.ca, jaroslaw.szlichta@uoit.ca

Abstract. Given a project whose completion requires a set of skills, *team formation* is the problem of finding a set of experts who collectively cover all the required skills in a way that optimizes one or more business objectives. In this paper, we present a new framework for finding an effective team from a network of experts. The proposed framework considers different business objectives to find the best team to perform the given tasks. Experimental results on a real dataset show the effectiveness of the proposed framework.

1 Introduction

With the exponential growth of the Internet and Web 2.0 services, there are many expert network providers (e.g., *LinkedIn*, *DBLP* and *GitHub*) that connect professionals having specialized skills and experience. Such networks are one of the most popular tools used by businesses seeking subject matter experts to complete a project.

In recent years, there has been interest in finding teams of experts from such networks [1, 4]. Given a project, the goal is to find teams of experts who cover all the required skills and also to optimize the communication cost among the team members [4]. The expert network is modeled as a graph where nodes represent experts and each edge indicates prior collaboration between two experts. In [4], two functions are proposed to compute communication costs. The first function is the sum of the shortest paths among experts in a team while the second function defines the communication cost as the diameter of the subgraph (team), where the diameter of a graph is the largest shortest path between any two nodes in the network. Then, two algorithms are proposed to discover teams minimizing the communication cost functions. In recent years, several methods have been proposed to find expert teams efficiently. However, existing approaches do not consider other business objectives such as *personnel cost*, *experts' authority*, etc. Therefore, they fail to discover effective teams when there are other important requirements.

To motivate our approach and illustrate the shortcomings of existing methods, assume that all the feasible teams of experts for a project are presented in Figure 1. Each team is represented as a subgraph whose nodes are either *skill holders* (team members who have the desired skills) or *connectors*. Existing methods discover teams with minimal communication costs. Thus, among the four feasible teams presented in Figure 1, T_a is selected since its communication cost is the lowest (5). However, other teams can be more desirable if we consider other objective functions. For example, if we want to find a team with the minimum personnel cost, T_b is best since the budget required to hire the team members is $62. Furthermore, experts may be associated with authority metric such as the *h-index* or the number of publications. In this case, we may want to discover a team with the maximum authority. Even if all the skill

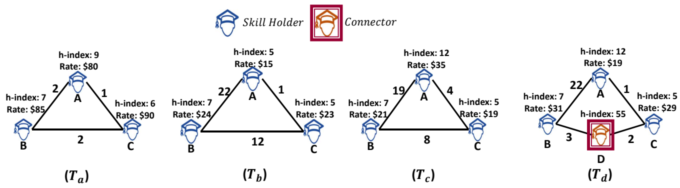

Fig. 1. An example of all feasible teams.

holders have the same authority (e.g., T_c and T_d), T_d may be preferable because its connector (e.g., expert D) has a higher authority. More importantly, if we want a team in which more than one of these objectives are optimized at the same time, there is not an obvious best choice.

We propose an effective framework to solve the problem of team formation from an expert network. Our framework considering objective functions which have not been considered in previous studies. Particularly, we consider *the personnel cost of team members*, *the authority of skill holders* and *the authority of connectors*. Then, we discover teams of experts optimizing the above objective functions. We note that this short paper is a summary of our published results in [2, 3, 5, 6].

2 Team Formation Framework

Let $C = \{c_1, c_2, \ldots, c_m\}$ be a set of m experts, and $S = \{s_1, s_2, \ldots, s_r\}$ be a set of r skills. An expert c_i has a set of skills, denoted as $S(c_i)$, where $S(c_i) \subseteq S$. If $s_j \in S(c_i)$, expert c_i has skill s_j. Furthermore, a subset of experts $C' \subseteq C$ have skill s_j if at least one of them has s_j. For each skill s_j, the set of all experts having skill s_j is denoted as $C(s_j) = \{c_i | s_j \in S(c_i)\}$. A project $P \subseteq S$ is a set of required skills. A subset of experts $C' \subseteq C$ *covers* a project P if $\forall s_j \in P \; \exists \; c_i \in C', s_j \in S(c_i)$.

Given an expert network G and a project P that requires the set of skills $\{s_1, s_2, \ldots, s_n\}$, a **feasible expert team (FET)** T is a connected subgraph of G whose nodes cover P. With each team, we associate a set of n skill-expert pairs: $\{\langle s_1, c_{s_1}\rangle, \langle s_2, c_{s_2}\rangle, \ldots, \langle s_n, c_{s_n}\rangle\}$, where c_{s_j} is an expert in T that has skill s_j for $j = 1, \ldots, n$. Since there may be many teams covering the required skills and some teams may not be interesting, teams are ranked by their **communication cost** [4]. Suppose the edges of a team T are denoted as $\{e_1, e_2, \ldots, e_t\}$. The communication cost of T is defined as $\mathrm{CC}(T) = \sum_{i=1}^{t} w(e_i)$, where $w(e_i)$ is the weight of edge e_i. We proved that minimizing the communication cost is an NP-hard problem by a reduction from 3-SAT.

Search Algorithm. Given a project P and an expert network G, our framework returns a subtree of G corresponding to a team with the lowest sum of edge weights. It first considers each expert c_r as a potential root node for the subtree. Then, to build a tree around c_r, for each required skill s_i, the nearest skill holder is selected (i.e., *nearestExpert*), that contains s_i. The *nearestExpert* is connected to the current team, meaning that any additional nodes along the path from the root to *nearestExpert* are also added. The tree with the lowest sum of edge weights is the best team. Note that, when finding a team with the minimum communication cost, edge weights in the input graph represent the shortest path among experts.

Objective Functions. We want to find a team whose members collaborate effectively and where another objective (e.g., the personnel cost of the team) is optimized.

In this situation, there is not an obvious best choice since there is a trade-off between objectives (e.g., the personnel cost and the communication cost). Moreover, it is possible that an objective function is defined based on node weights (e.g., experts' cost). To find the best team, we transform the expert network G to G' by moving all values associated with experts (node weights) onto the edge weights and then running the aforementioned method on the transformed graph to find the best team of experts. Below, we introduce new objective functions and we discuss how we build G' based on these objective functions.

a) Experts' Authority. Suppose that the **connectors of a team** T (all nodes excluding skill holders) are denoted as $\{c_1, c_2, \ldots, c_q\}$. The connector authority of T is defined as $\mathrm{CA}(T) = \sum_{i=1}^{q} a'(c_i)$ [5]. We are also interested in optimizing the authority of skill holders. Suppose that the **skill holders of a team** T are denoted as $\{c_1, c_2, \ldots, c_n\}$. The skill holder authority of T is defined as $SA(T) = \sum_{i=1}^{n} a'(c_i)$. To build G', we use a hybrid function as follows: $\text{CA-CC}(T) = \gamma \times \mathrm{CA}(T) + (1-\gamma) \times \mathrm{CC}(T)$. In order to consider the authority of skill holders, we use the following hybrid function, $\text{SA-CA-CC}(T) = \lambda \times \mathrm{SA}(T) + (1-\lambda) \times \text{CA-CC}(T)$ [5].

b) Personnel Cost. Let the set of experts in a team T be $\{c_1, c_2, \ldots, c_q\}$. The personnel cost of T is defined as [3]: $PC(T) = \sum_{i=1}^{q} t(c_i)$ where $t(c_i)$ is the required budget to hire expert c_i. Given a team T of experts from graph G for a project and a trade-off λ between the communication and personnel costs, the combined cost of T with respect to G is defined as [2]: $\text{PC-CC}(T) = (p-1)(1-\lambda) \times PC(T) + 2\lambda \times CC(T)$, where p is the number of required skills.

We also propose two other approaches to take personnel cost into account [3]. The first approach is to consider a limit on one of the objectives and then find the best team based on the other objective. The second approach is to discover a set of teams that are not worse than any other teams based on the objectives. These teams are called *Pareto-optimal teams*. All these solutions are approximation algorithms and have provable bounds (recall that the problem we solve is NP-hard).

3 Experimental Evaluation

In this section, we use the proposed algorithm and various objective functions explained above to implement our framework for team discovery which optimizes CC, CA-CC, SA-CA-CC and PC-CC. We use various datasets including the DBLP XML dataset[1] to build an expert graph [3]. The algorithms are implemented in Java and the experiments are conducted on an Intel(R) Core(TM) i7 2.80 GHz computer with 4 GB of RAM.

For comparison, we also implemented *Random*, which randomly builds 10,000 teams and selects the one with the lowest value of SA-CA-CC (in Figure 3 (a)) or PC-CC (in Figure 3 (b)), and *Exact* which performs an exhaustive search to find an (SA-CA-CC)-optimal or (PC-CC)-optimal solution. Figure 3 (a) illustrates the SA-CA-CC scores of different objective functions for different values of λ (lower is better). The projects used in the experiments are generated as follows. We set the number of skills in a project to 4, 6, 8 or 10. For each number of skills, 50 sets of skills are generated randomly, corresponding to 50 random projects. The average scores over the 50 projects are computed for each objective function. According to Figure 3, SA-CA-CC produces results that are close to those of Exact. Since all the functions use the same algorithm, CC, CA-CC and SA-CA-CC have similar runtime. Figure 3 (b)

[1] http://dblp.uni-trier.de/xml/

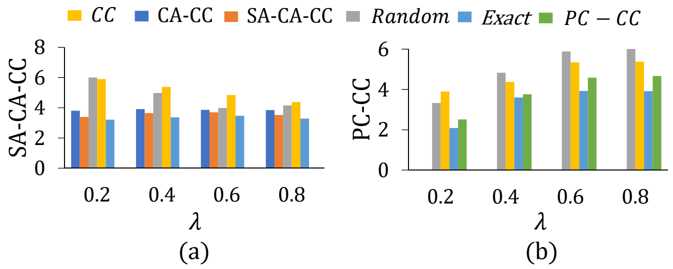

Fig. 2. SA-CA-CC and PC-CC scores of different objective functions ($\gamma = 0.6$)

shows the average PC-CC cost values of teams for different functions for the DBLP dataset. The results show that all of the algorithms outperform the Random method. The results also suggest that the PC-CC method has the lowest cost values among non-exact methods.

We check if the top-5 teams returned by CC and SA-CA-CC were successful in real life. To do so, we examined the rankings of the publication venues of these teams according to the Microsoft Academic conference ranking. We used the DBLP dataset up to 2015 for team discovery, and only consider papers published in 2016. We set γ and λ to 0.6 and generate 5 different projects with four different skills. From the teams that co-authored papers in 2016, we found that 78% of the time the teams found by SA-CA-CC published in more highly-rated venues than those found by CC.

4 Conclusions

We studied the problem of finding teams of experts from an expert network in a way that optimizes different objectives: the communication cost among team members, the authority of skill holders and connectors, and the personnel cost. We proposed a series of objective functions and methods to find the best team of experts. In future work, in order to find the distance between two experts, we plan to investigate the use of statistical methods (e.g., random walk with restart) instead of the shortest path.

References

1. Basu Roy, S., Lakshmanan, L.V., Liu, R.: From group recommendations to group formation. In: Proceedings of SIGMOD '15, pp. 1603–1616 (2015)
2. Kargar, M., An, A., Zihayat, M.: Efficient bi-objective team formation in social networks. In: Proceedings of ECML/PKDD'12, pp. 483–498. Springer (2012)
3. Kargar, M., Zihayat, M., An, A.: Finding affordable and collaborative teams from a network of experts. In: Proceedings of SDM'13, pp. 587–595. SIAM (2013)
4. Lappas, T., Liu, K., Terzi, E.: Finding a team of experts in social networks. In: Proceedings of KDD'09, KDD '09, pp. 467–476. ACM, New York, NY, USA (2009)
5. Zihayat, M., An, A., Golab, L., Kargar, M., Szlichta, J.: Authority-based team discovery in social networks. In: Proceedings of EDBT'17, pp. 498–501 (2017)
6. Zihayat, M., Kargar, M., An, A.: Two-phase pareto set discovery for team formation in social networks. In: In Proceedings of WI/IAT, vol. 2, pp. 304–311. IEEE (2014)