An Algorithm for Multi-Attribute Diverse Matching

Saba Ahmadi[1] , **Faez Ahmed**[2] , **John P. Dickerson**[1] , **Mark Fuge**[3] and **Samir Khuller**[4]
[1] Department of Computer Science, University of Maryland
[2] Department of Mechanical Engineering, Northwestern University
[3] Department of Mechanical Engineering, University of Maryland
[4] Department of Computer Science, Northwestern University
saba@cs.umd.edu, faez@northwestern.edu, john@cs.umd.edu, fuge@umd.edu,
samir.khuller@northwestern.edu

Abstract

Bipartite b-matching, where agents on one side of a market are matched to one or more agents or items on the other, is a classical model that is used in myriad application areas such as healthcare, advertising, education, and general resource allocation. Traditionally, the primary goal of such models is to maximize a linear function of the constituent matches (e.g., linear social welfare maximization) subject to some constraints. Recent work has studied a new goal of balancing whole-match *diversity* and economic efficiency, where the objective is instead a monotone submodular function over the matching. Basic versions of this problem are solvable in polynomial time. In this work, we prove that the problem of simultaneously maximizing diversity along several features (e.g., country of citizenship, gender, skills) is NP-hard. To address this problem, we develop the first combinatorial algorithm that constructs provably-optimal diverse b-matchings in pseudo-polynomial time. We also provide a Mixed-Integer Quadratic formulation for the same problem and show that our method guarantees optimal solutions and takes less computation time for a reviewer assignment application. The source code is made available at https://github.com/faezahmed/diverse_matching.

1 Introduction

The bipartite matching problem occurs in many applications such as healthcare, advertising, and general resource allocation. Weighted bipartite b-matching is a generalization of this problem where each node on one side of the market can be matched to many items from the other side, and where edges may also have associated real-valued weights. Examples of weighted bipartite b-matching include assigning children to schools [Drummond *et al.*, 2015; Kurata *et al.*, 2017], reviewers to manuscripts [Charlin and Zemel, 2013; Liu *et al.*, 2014], and donor organs to patients [Dickerson and Sandholm, 2015; Bertsimas *et al.*, 2019].

Ahmed *et al.* [2017] introduced the notion of *diverse* bipartite b-matching, where the goal was to simultaneously maximize the "efficiency" of an assignment along with its "diversity." For example, a firm might want to hire several highly-skilled workers, but if that firm also cares about diversity it may want to ensure that some of those hires occur across marginalized categories of employees. They proposed an objective which combined economic efficiency and diversity demonstrating that, in practice, reducing the efficiency of a matching by small amounts can often lead to significant gains in diversity across a matching. However, their formulation was limited to diversity for a single feature. It also relied on solving a general Mixed-Integer Quadratic Program (MIQP), which is flexible but computationally intractable.

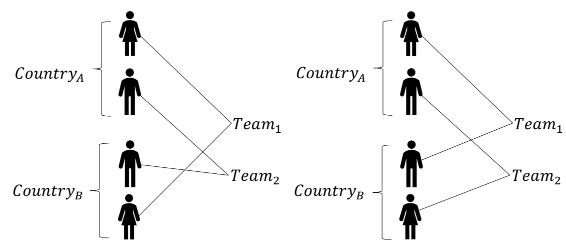

Figure 1: An illustrative example of single feature diverse matching (left) versus multi-feature diverse matching (right); here, the matching creates teams with workers from each country and gender.

In this work, we generalize the diverse matching problem and introduce matchings where each worker has *multiple* features (e.g., country of origin, gender) and our goal is to form diverse teams with respect to all these features. We found that the problem with a single feature, studied by Ahmed *et al.* [2017], can be reduced to a minimum quadratic cost maximum flow formulation and solved in polynomial time by an existing algorithm [Minoux, 1986]. In contrast, we provide NP-hardness results for the general case of multiple features.

Our contributions. The paper's main contributions follow:

- We provide the first pseudo-polynomial time algorithm for the diverse bipartite b-matching w.r.t. multiple features problem with class-specific weights.[1] The key insight lies in detecting *negative cycles* in an auxiliary graph repre-

[1] That is, under conditions when the cost of assigning all items from one feature set to an item on the other side of the graph is the same. This holds when, e.g., one is matching academic papers to reviewers where each reviewer can specify exactly one field of expertise and the cost of assigning a paper to any of the reviewers *within* the same field is the same but differs *across* fields.

sentation, which we use to either provide incremental improvements to the incumbent diverse matching or prove that our negative-cycle-detection algorithms have found a globally-optimal matching. We also provide a general MIQP formulation for this problem.
- We then extend the algorithm to the diverse bipartite b-matching problems with *general* edge weights, where edge weights of nodes within a category can be different.
- Lastly, we demonstrate our algorithm's applicability to paper-reviewer matching. Our algorithm takes less time to converge to an optimal solution than the proposed MIQP approach (using a state-of-the-art commercial solver).

2 Related Work

Matching people to form diverse teams leverages the intersection of two past areas of research: the role of team diversity in collaborative work and how diversity among groups of resources is measured and used to form/match teams. Compared to related work, this paper provides a practical, high-performing method to perform diverse b-matching that can enable applications like diverse team formation or diverse resource allocation. Below we will use the example of diverse team formation (for example, in project teams within a company) to provide a concrete example to place prior work in context; however, our proposed approach is generally applicable to any diverse matching problem.

In the example of forming teams, the traditional approach is to use weighted bipartite b-matching (WBM) methods [Basu Roy *et al.*, 2015]. These methods maximize the total weight of the matching while satisfying some constraints. However, there are two major issues with these approaches. First, it assumes that the value provided by a person in a team is always fixed and independent of who else is in the team. This assumption may not hold in many cases. A new team member may provide more added value to the team if she is added to a smaller team compared to the case if she is added to a larger team. This property of diminishing marginal utility can be mathematically captured by a family of functions called submodular functions, which we define later. Second, existing approaches do not account for diversity *within* a team, where teams with workers from different backgrounds may be desirable. For example, different types of worker diversity have a direct impact on the success rate of tasks [Ross *et al.*, 2010]. Likewise, firms with a higher number of employees with higher education and diversity in the types of educations have a higher likelihood of innovating [Østergaard *et al.*, 2011] and increasing revenue for firms [Hunt *et al.*, 2015]. In this paper, we address both these issues.

Past researchers have generally measured diversity by defining some notion of *coverage*—that is, a diverse set is one that covers the space of available variation. Mathematically, researchers have done so via the use of *submodular functions*, which encode the notion of diminishing returns [Lin and Bilmes, 2012]; that is, as one adds items to a set that are similar to previous items, one gains less utility if the existing items in the set already "cover" the characteristics added by that new item. For example, many previous diversity metrics used in the information retrieval or search communities—including Maximum Marginal Relevance (MMR) [Carbonell and Goldstein, 1998] and Determinantal Point Processes [Kulesza *et al.*, 2012]—are instances of submodular functions. These functions can model notions of coverage, representation, and diversity [Ahmed and Fuge, 2018] and they have been shown to achieve top results on common automatic document summarization benchmarks—*e.g.*, at the Document Understanding Conference [Lin and Bilmes, 2012].

Within matching, our work is closest to that of Ahmed *et al.* [2017], which used a supermodular function to propose a diverse matching optimization method. Other researchers have also approached similar problems, with diversity either as an objective or as a constraint. For instance, Gölz and Procaccia [2019] match migrants to localities in a way that maximizes the expected number of migrants who find employment. Benabbou *et al.* [2018] study the trade-off between diversity and social welfare for the Singapore housing allocation. They model the problem as an extension of the classic assignment problem, with additional diversity constraints. Lian *et al.* [2018] solve the assignment problem when preferences from one side over the other side are given and both sides have capacity constraints. They use order weighted averages to propose a polynomial-time algorithm which leads to high quality and more fair assignments. Agrawal *et al.* [2018] show that a simple iterative proportional allocation algorithm can be tuned to produce maximum matching with high entropy. Finally, Kobren *et al.* [2019] proposed two fairness-promoting algorithms for the paper-reviewer matching problem. They demonstrate that their algorithm achieves higher utility compared to state of the art matching algorithms that optimize for fairness only. In contrast, our goal is to develop an algorithm for finding the optimal assignment which maximizes utility as well as diversity along multiple features as an objective—along with having constraints on workload.

We define a utility function that can be tuned to balance the diversity and total weight of matching. The diversity function is inspired by the Herfindahl index [Hirschman, 1964], which is a statistical measure of concentration and commonly used in economics. We provide a new algorithm that models the problem using an auxiliary graph and uses a heuristic improvement of the negative cycle detection of Bellman-Ford by Goldberg and Radzik [1993][2] to find negative cycles and cancel them on a new graph to obtain an optimal solution for the original problem.

3 Preliminaries

In this section, we first define the preliminaries for a diverse matching problem, where workers are to be matched to teams and each team wants workers belonging to a diverse set of features. In our problem, we are given a set of features for the workers. Let $\mathcal{F} = \{f_1, \cdots, f_{|\mathcal{F}|}\}$ denote the feature set for the workers. An example of a feature set could be {country of citizenship, gender}. Each feature $f_k \in \mathcal{F}$ has one of

[2]We used the negative cycle detection algorithm by Goldberg and Radzik [1993]. Cherkassky *et al.* [1993] compared the performance of multiple negative cycle detection algorithms, and the algorithm by Goldberg and Radzik [1993] was one of the fastest.

the values $\mathcal{F}_k = \{f_{k,1}, \cdots, f_{k,|\mathcal{F}_k|}\}$. Let $|\mathcal{F}_{k,k'}|$ denote the number of workers having value $f_{k,k'}$ for feature f_k. The set of workers is denoted by $X = \{x_1, \ldots, x_n\}$. X is partitioned into $|\mathcal{V}|$ subsets $V_1, \cdots, V_{|\mathcal{V}|}$, where each subset V_j corresponds to a feature set $v_j = \{v_{j,1}, \cdots, v_{j,|\mathcal{F}|}\}$, where $\forall 1 \leq k \leq |\mathcal{F}|, v_{j,k} \in \mathcal{F}_k$.

We wish to form a set of teams $\{T_1, \ldots, T_t\}$ of the workers where each team T_i has a demand of d_i, specifying the number of workers that need to be assigned to it. Each worker can be assigned to at most one team.

The diversity loss of an assignment is denoted by D and is equal to $\sum_{k=1}^{|\mathcal{F}|} \lambda_k D_k$, where D_k shows the diversity loss w.r.t. feature f_k, and $\lambda_k \in \mathcal{Z}^{\geq 0}$ is a constant. Let $c_{i,k,k'}$ denote the number of workers in T_i having value $f_{k,k'} \in \mathcal{F}_k$ for feature f_k. Then, $D_k = \sum_{i=1}^{t} \sum_{k'=1}^{|\mathcal{F}_k|} c_{i,k,k'}^2$. The cost of assigning each worker having value $f_{k,k'} \in \mathcal{F}_k$ for feature f_k to team T_i is denoted by $u_{i,k,k'} \in \mathcal{Z}^{\geq 0}$. We assume all costs are integers. The total cost of an assignment is $TU = \sum_{k=1}^{|\mathcal{F}|} TU_k$ where $TU_k = \sum_{i=1}^{t} \sum_{k'=1}^{|\mathcal{F}_k|} c_{i,k,k'} \cdot u_{i,k,k'}$.

Our goal is to minimize the objective function which is equal to $D + \lambda_0 TU$, where $\lambda_0 \in \mathcal{Z}^{\geq 0}$. To understand why minimizing D_k makes an assignment more diverse w.r.t f_k, consider Figure 1. In the left assignment, $D_2 = 8$, and in the right assignment $D_2 = 4$, and the right assignment is more diverse w.r.t f_2, i.e. gender. By setting λ parameters, we assume that the relative importance between factors is not qualitative and can be quantified. Next, we provide Theorem 1, which shows that this problem is NP-hard.

Theorem 1. *Minimizing the supermodular diversity loss function w.r.t multiple features is NP-hard.*

Proof. We show a reduction from a variation of 3-COLOR problem which is as follows: given a graph $G = (V, E)$ with n vertices, is there a coloring with n_1 vertices of color c_1, n_2 vertices of color c_2, and n_3 vertices of color c_3, such that no two adjacent vertices receive the same color, and all the vertices are colored? It could be shown this variation is NP-hard by a reduction from classic 3-COLOR problem.

The reduction from 3-COLOR to multiple-attribute diverse matching is as following: In 3-COLOR, assign a feature f_k to each edge $e_k = (v_{k_1}, v_{k_2}) \in E$, and a worker to each vertex. Let $f_{k,i}$ denote the value of f_k for the worker corresponding to $v_i \in V$. Let $f_{k,i} = i$ if $i \neq k_1, k_2$. Otherwise, let $f_{k,i} = 0$. The goal is to form three teams T_1, T_2, T_3 with demands $d_1 = n_1, d_2 = n_2, d_3 = n_3$, respectively. We assume that all the costs of assigning workers to the teams are zero, therefore the objective function is to minimize the total diversity loss. Consider an arbitrary edge $e_k = (v_{k_1}, v_{k_2})$. If the endpoints of e_k belong to different teams, f_k contributes $n_1 + n_2 + n_3$ to the objective function since all the workers inside a team have different values for f_k. Otherwise, it contributes $n_1 + n_2 + n_3 - 2 + 2^2$ since workers corresponding to v_{k_1}, v_{k_2} are the only workers having the same value for f_k inside a team. If the cost of the optimal solution for the diverse matching problem is $(n_1 + n_2 + n_3) \cdot |E|$, there does not exist a pair of workers in a team where the vertices corresponding to them are neighbouring in G. Otherwise if the cost of the optimal solution is more than $(n_1 + n_2 + n_3) \cdot |E|$, the 3-COLOR instance is infeasible. □

	$country_1, g_1$	$country_1, g_2$	$country_2, g_1$	$country_2, g_2$
T_0	$w_{0,1}$	$w_{0,2}$	$w_{0,3}$	$w_{0,4}$
T_1	$w_{1,1}$	$w_{1,2}$	$w_{1,3}$	$w_{1,4}$
T_2	$w_{2,1}$	$w_{2,2}$	$w_{2,3}$	$w_{2,4}$
T_3	$w_{3,1}$	$w_{3,2}$	$w_{3,3}$	$w_{3,4}$

Table 1: Matrix representation of three teams and workers from two countries and two genders. Dummy team T_0 accommodates unassigned workers. Arrows represent a local exchange.

We are interested in solving this NP-hard problem. We begin by presenting two different representations of instances of the problem: one in matrix form (used for expositional ease), and the other in graph form (used to build our optimal diverse matching algorithm in Section 4).

Matrix Representation. An example of matrix representation with three teams and two countries and two genders is shown in Table 1. Each column corresponds to a feature set and each row corresponds to a team. Entry $w_{i,j}$ shows the number of workers with feature set v_j assigned to T_i. We introduce a *dummy team* T_0, and $w_{0,j}$ shows the number of workers with feature set v_j who are not assigned to any team.

Matching Representation. In this representation, a bipartite graph $G = (\mathcal{X} \cup \mathcal{T}, E)$ is given. The nodes in $\mathcal{X}$ correspond to the workers, and are partitioned into $|\mathcal{V}|$ subsets where each subset corresponds to a feature set. Each vertex in $\mathcal{T}$ corresponds to a team in $\{T_0, T_1, T_2, \cdots, T_t\}$. The assignment of workers to teams forms a b-matching, where the degree of each node T_i for $1 \leq i \leq t$ is d_i. All workers who are not assigned to any team $T_1, \cdots, T_t$ get assigned to the dummy team T_0. Degree of each node $x \in \mathcal{X}$ is exactly one.

Local Exchange. A local exchange happens when a group of teams decides to transfer one or more workers between each other while maintaining the total number of workers in each of them. The exchange is done in a way that the initial demands of all the teams are fulfilled. Arrows in Table 1 show a local exchange in a matrix representation.

In this exchange, one worker from V_2 is moved from T_3 to T_1. Two workers from V_1 are moved. One is moved from T_1 to T_2, and the other one is moved from T_2 to T_3. The set of edges of local exchange in a matrix representation is called a cycle. The source-transitions of a cycle are the cells without any input edges, and the sink-transitions are the cells without any output edges. In Table 1, the cells corresponding to $w_{3,2}$ and $w_{1,1}$ are source-transitions, and the cells corresponding to $w_{1,2}$ and $w_{3,1}$ are sink-transitions.

Figure 2 shows the same local exchange operation using a matching representation. In this figure, the black matching shows the initial assignment, and the dotted matching shows the assignment after the exchange operation is done.

Gain of a local exchange. Our goal is to minimize the objective function f, by doing some local exchanges. To find out, we first calculate the marginal gain from a given exchange operation which is the difference between the objective values before and after a local exchange. In order to simplify this concept, we use the following definition:

	$country_1$	$country_2$
T_0	$c_{0,1,1} = w_{0,1} + w_{0,2}$	$c_{0,1,2} = w_{0,3} + w_{0,4}$
T_1	$c_{1,1,1} = w_{1,1} + w_{1,2}$	$c_{1,1,2} = w_{1,3} + w_{1,4}$
T_2	$c_{2,1,1} = w_{2,1} + w_{2,2}$	$c_{2,1,2} = w_{2,3} + w_{2,4}$
T_3	$c_{3,1,1} = w_{3,1} + w_{3,2}$	$c_{3,1,2} = w_{3,3} + w_{3,4}$

Table 2: Matrix representation embedding w.r.t country.

	g_1	g_2
T_0	$c_{0,2,1} = w_{0,1} + w_{0,3}$	$c_{0,2,2} = w_{0,2} + w_{0,4}$
T_1	$c_{1,2,1} = w_{1,1} + w_{1,3}$	$c_{1,2,2} = w_{1,2} + w_{1,4}$
T_2	$c_{2,2,1} = w_{2,1} + w_{2,3}$	$c_{2,2,2} = w_{2,2} + w_{2,4}$
T_3	$c_{3,2,1} = w_{3,1} + w_{3,3}$	$c_{3,2,2} = w_{3,2} + w_{3,4}$

Table 3: Matrix Representation Embedding w.r.t to gender.

Embedding of Matrix Representation. Consider a given matrix representation M, it can be embedded into a matrix M_k for a fixed feature f_k in the following way: all the columns in M corresponding to the same value $f_{k,k'}$ of f_k, are combined into a single column in M_k. For example, embedding of the matrix representation in Table 1 into M_1, M_2 w.r.t. the features country and gender are shown in Tables 2 and 3. Since in M_1, the number of people assigned from each country to each team is not changed, $\Delta_1 = \Delta(\lambda_0 \cdot TU_1 + \lambda_1 D_1) = 0$. According to M_2, $\Delta_2 = \Delta(\lambda_0 \cdot TU_2 + \lambda_2 D_2) = \lambda_0\big(-u_{3,2,2} + u_{1,2,2} - u_{1,2,1} + u_{3,2,1}\big) + \lambda_2\big((c_{3,2,2}-1)^2 - (c_{3,2,2})^2 + (c_{1,2,2}+1)^2 - (c_{1,2,2})^2 + (c_{1,2,1}-1)^2 - c_{1,2,1}^2 + (c_{3,2,1}+1)^2 - c_{3,2,1}^2\big)$.

It can be seen that the contribution of the cells which are not source-transition or sink-transition to the gain of a local exchange is zero (all the cells in the local exchange in Table 2, and the node corresponding to $c_{2,2,1}$ in M_2). If the net gain, i.e. $\Delta_1 + \Delta_2$, is negative, then the local exchange can be considered beneficial and we can transfer the workers.

4 Negative-Cycle-Detection-based Algorithms

In this section, we explain our algorithm for finding the optimum assignment. First, we build an auxiliary graph G'. For each team $T_i \in \{T_0, \cdots, T_t\}$, there is a switch in G' with $|\mathcal{V}|$ input ports, and $|\mathcal{V}|$ output ports. Each port is a node in G', and each switch is a directed bipartite graph, with edges going from its input ports (nodes) to its output ports. In Figure 3, each box is a switch. Inside a switch T_i, there is a directed edge from each input port to each output port. If the directed edge is connecting two ports such that their corresponding combinations of features do not have the same value for any features, the weight of this edge is equal to zero. Otherwise, per each feature f_k that has the same value, $-2\lambda_k$ is added to the weight of this edge.

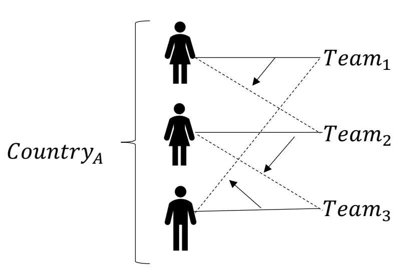

Figure 2: Local exchange operation (in matching representation).

Algorithm 1: Find optimal diverse b-matching

Input : Directed weighted graph G', initial feasible b-matching Q which satisfies team demands.
Output: Optimal diverse b-matching
while $\exists$ *a negative cycle* $C \in G'$ **do**
    // Perform a local exchange operation along C;
    for $e \in C$ **do**
        // Assume edge e is from output port $O_j^{i_1}$ of team T_{i_1} to input port $I_j^{i_2}$ of another team T_{i_2};
        // Move one worker with feature set $v_j = \{f_{1,k'_1}, \cdots, f_{|\mathcal{F}|,k'_{|\mathcal{F}|}}\}$ from team T_{i_1} to team T_{i_2}:
        $\forall k \in \{1, \cdots, |\mathcal{F}|\}$:
        $c_{i_1,k,k'_k} -= 1, c_{i_2,k,k'_k} += 1$;
        Update weight of edges of G' w.r.t to the new values of c_{i_1,k,k'_k}, and c_{i_2,k,k'_k};

The reason behind assigning these weights to the edges is to make sure in a local exchange, considering a fixed feature f_k, the cells which are not a source-transition or a sink-transition w.r.t. M_k, have zero contribution to $\Delta(D_k)$.

For each pair of teams T_{i_1} and T_{i_2} where $i_1 \neq i_2$, and for each feature combination v_j, there is a directed edge from output port $O_j^{i_1}$ of switch T_{i_1} to the input port $I_j^{i_2}$ of switch T_{i_2}, and weight of this edge captures the difference in the objective function when in the matrix representation a person in column V_j (with feature set v_j) is moved from T_{i_1} to $T_{i,2}$.

Each cycle in this graph corresponds to a cycle in a matrix representation and local exchanges along them have the same gain. Figure 3 shows a cycle which is corresponding to the cycles in Table 1 and Figure 2.

After constructing the auxiliary graph, we run Algorithm 1. Algorithm 1 moves workers from one team to another if it detects a negative cycle.

Algorithm 1 takes as input an initial feasible solution Q as input. To find Q, we first find a feasible solution, which satisfies all the demand constraints. In order to find an initial feasible solution, in each iteration, consider the first subset of workers in the the bipartite graph G (V_j) with at least one unassigned worker, and the first team (T_i) such that the number of workers assigned to it is less than its demand (In the first iteration, we start with V_1, T_1, and all the workers are unassigned). Assign un-assigned workers from V_j to T_i, until either demand of T_i is fully satisfied, in this case, move to the next team ($i = i+1$), or all the workers from V_j are assigned, then let $j = j+1$. Repeat this procedure until all the demand constraints are satisfied. Time complexity of this procedure is $\mathcal{O}(|\mathcal{V}| + t)$.

In Algorithm 1, any negative cycle detection algorithm can be used to detect negative cycles in G'. We use a heuristic improvement of Bellman-Ford proposed by Goldberg and Radzik [Goldberg and Radzik, 1993] in our experiments.

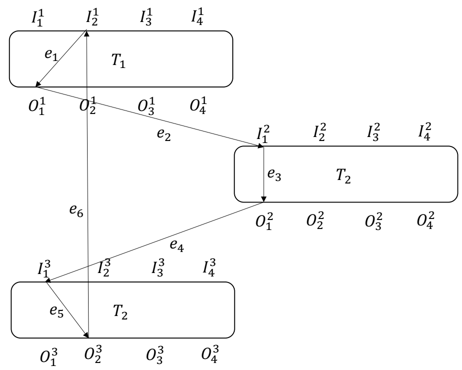

Figure 3: A local Exchange in graph representation.

	$country_1, g_1$	$country_1, g_2$	$country_2, g_1$	$country_2, g_2$
T_0	$w_{0,1}$	$w_{0,2}$	$w_{0,3}$	$w_{0,4}$
T_1	$w_{1,1}$	$w_{1,2}$	$w_{1,3}$	$w_{1,4}$
T_2	$w_{2,1}$	$w_{2,2}$	$w_{2,3}$	$w_{2,4}$
T_3	$w_{3,1}$	$w_{3,2}$	$w_{3,3}$	$w_{3,4}$

Table 4: Maximal cycle decomposition

5 Proof of Optimality

In this section, we prove that Algorithm 1 gives the optimum solution for diverse bipartite b-matching problem.

Assume after the algorithm ends, the final assignment is a local optimum P, and the optimum solution is P^*. Consider the matching representations of P and P^*. Since all the nodes in P and P^* are matched, the symmetric difference of P and P^* ($P \oplus P^*$) can be decomposed into a set of alternating even cycles. Each local exchange along an alternating cycle corresponds to a cycle in the matrix representation.

Before proving Thm. 2, we need the following definitions:

Maximal Cycle. A cycle y in a matrix representation M is maximal if its source-transitions and sink-transitions are source-transition and sink-transition w.r.t all the edges in M as well. For example, consider Table 4. Let's call the dotted cycle y_g, the dashed cycle y_r, and the solid-line cycle y_b. y_g has two source-transitions $w_{1,1}, w_{0,3}$, and it has two sink-transitions $w_{0,1}, w_{1,3}$. Since there are no edges going out of $w_{1,3}, w_{0,1}$, and no edges going into $w_{0,3}, w_{1,1}$, y_g is a maximal cycle. Cycles y_r, y_b are maximal cycles as well. Therefore, $\{y_g, y_b, y_r\}$ gives a maximal cycle decomposition for M. However, if we consider embedding of M w.r.t gender (M_2), then y_r is not a maximal cycle anymore, and $\{y_g, y_r \cup y_b\}$ gives a maximal cycle decomposition w.r.t M_2 and M_1(embedding w.r.t countries). A cycle is called all-maximal cycle if it is maximal w.r.t all the matrix representations $M_1, \cdots, M_{|\mathcal{F}|}$. In this example, $\{y_g, y_r \cup y_b\}$ gives an all-maximal cycle decomposition.

Lemma 1. *The set of all the edges of $P \oplus P^*$ can be decomposed into a set of all-maximal cycles.* [3]

Theorem 2. *Algorithm 1 finds the global optimum for the diverse b-matching problem.*

Proof. Let $f(P)$ show the value of the objective function for the assignment P. $f(P^*) - f(P) < 0$ therefore:

$$f(P^*) - f(P) = gain(y'_{1,1}) + gain(y'_{2,2}) + \cdots + gain(y'_{\ell',\ell'}) < 0$$

Where y'_k ($1 \leq k \leq \ell'$) is the k^{th} cycle in the all-maximal cycle decomposition, and $y'_{k,k}$ is applying the local exchange of the cycle y'_k at step k. The initial step is the assignment P. Since $f(P^*) - f(P) < 0$, there must be a maximal cycle y'_g such that $gain(y'_{g,g}) < 0$. We wish to show $gain(y'_{g,1}) < 0$, which implies starting from the initial assignment P, a local exchange can be done with a negative gain, and P is not a local optimum which is a contradiction.

Let $D(y'_{g,g}), U(y'_{g,g})$ denote respectively the change in the diversity loss, and the change in the utility when applying a local exchange y'_g in step g. Let $D_{f_k}(y'_{g,g}), U_{f_k}(y'_{g,g})$ denote the change in the diversity loss w.r.t the feature f_k, when applying $y'_{g,g}$. Therefore:

$$gain(y'_{g,g}) = \sum_{k \in |\mathcal{F}|} \Big(D_{f_k}(y'_{g,g}) + U_{f_k}(y'_{g,g}) \Big)$$

Lemma 2 shows if $D_{f_k}(y'_{g,g}) < 0$, then $D_{f_k}(y'_{g,1}) < 0$. As a result, $D(y'_{g,g}) < 0$ implies $D(y'_{g,1}) < 0$. It is easy to see that $U(y'_{g,g}) = U(y'_{g,1})$. Therefore, $gain(y'_{g,g}) < 0$ implies $gain(y'_{g,1}) < 0$, and the proof is complete. □

Lemma 2. *If $D_{f_k}(y'_{g,g}) < 0$, then $D_{f_k}(y'_{g,1}) < 0$.* [3]

Theorem 3. *The running time of the algorithm is $\mathcal{O}((\lambda_{\max} \cdot |\mathcal{F}| \cdot n^2 + \lambda_0 U) \cdot |\mathcal{V}|^2 \cdot t^2(|\mathcal{V}| + t))$, where U is the maximum cost of an initial feasible b-matching and $\lambda_{max} = \max\{\lambda_1, \cdots, \lambda_{|\mathcal{F}|}\}$.*

In order to prove this theorem, first we show the following lemmas hold.

Lemma 3. *The number of iterations of our algorithm is at most $\lambda_{max} \cdot |\mathcal{F}| \cdot n^2 + \lambda_0 U$.* [3]

Lemma 4. *The complexity of each iteration of the algorithm is $\mathcal{O}(|\mathcal{V}|^2 \cdot t^2(|\mathcal{V}| + t))$.* [3]

Combining Lemma 3 with Lemma 4, and considering $\mathcal{O}(|\mathcal{V}| + t)$ time complexity for finding an initial feasible solution, yields Theorem 3.

6 Diverse Weighted Bipartite b-Matching

In this section, we extend our algorithm to solve the case where the cost of assigning workers from the *same* feature set to a team can be different. First, in each switch we put input and output ports for each worker. Inside each switch, there is a complete bipartite graph from input ports to the output ports. Consider an edge between an input port to an output port corresponding to workers x_i and x_j. Per each feature f_k where x_i, x_j have the same values for f_k, $-2\lambda_k$ is added to the weight of the edge between x_i, x_j.

Consider an edge from output port $x_k^{i_1}$ of switch T_{i_1} to input port $x_k^{i_2}$ of switch T_{i_2}, where $x_k \in V_j$. The weight of this edge is equal to the change in the objective function by moving one worker from V_j out of T_{i_1}, and adding that worker to T_{i_2}. The proof of the following theorem is similar to Thm. 3.

Theorem 4. *The running time of the algorithm for general weights is $\mathcal{O}((\lambda_{max} \cdot |\mathcal{F}| \cdot n^2 + \lambda_0 U) \cdot n^2 \cdot t^2(n + t))$, where U is the maximum cost of any feasible b-matching.*

[3] Due to space constraints, we omit this proof to the full version.

7 Experimental Validation & Discussion

To demonstrate the effectiveness of the proposed method, we apply it to a dataset of reviewer paper matching. First, we find the optimal solution for multi-feature reviewer paper matching and compare it to the single feature diverse matching method. We also provide the MIQP formulation of the same problem based on literature and show how our algorithm is faster to the Gurobi based MIQP solver.

For the reviewer assignment problem, where each reviewer has multiple features, we want to match each paper with reviewers who are not only from different expertise areas (clusters), but also belong to different genders. We use the multi-aspect review assignment evaluation dataset [Karimzadehgan and Zhai, 2009], a benchmark dataset from UIUC. It contains 73 papers accepted by SIGIR 2007, and 189 prospective reviewers who had published in the main information retrieval conferences. The dataset provides 25 major topics and for each paper in the set, an expert provided 25-dimensional label on that paper based on a set of defined topics. Similarly for the 189 reviewers, a 25-dimensional expertise representation is provided.

To compare our method (Algorithm 1) with a baseline, we formulate a multi-feature MIQP variant of our problem, which is an extension of the single-feature formulation provided in [Ahmed *et al.*, 2017] and is given by:

$$\min \lambda_0 \sum_{k=1}^{|\mathcal{F}|} \sum_{i=1}^{t} \sum_{k'=1}^{|\mathcal{F}_k|} u_{i,k,k'} \cdot c_{i,k,k'} + \sum_{k=1}^{|\mathcal{F}|} \lambda_k \sum_{i=1}^{t} \sum_{k'=1}^{|\mathcal{F}_k|} c_{i,k,k'}^2$$

$$\sum_{k=1}^{|\mathcal{F}|} \sum_{k'=1}^{|\mathcal{F}_k|} c_{i,k,k'} = d_i, \forall 1 \leq i \leq t$$

$$\sum_{i=0}^{t} c_{i,k,k'} = |\mathcal{F}_{k,k'}|, 1 \leq k \leq |\mathcal{F}|, 1 \leq k' \leq |\mathcal{F}_k|$$

To set up the graph for our method, we first cluster the reviewers into 5 clusters based on their topic vectors using spectral clustering. To calculate the relevance of each cluster for any paper, we take the average cosine similarity of label vectors of reviewers in that cluster and the paper. We set the constraints such that each paper matches with exactly 4 reviewers, and no reviewer is allocated more than 1 paper. To increase dataset size, we double the number of reviewers by creating a copy of each reviewer. As the original dataset lacks gender information, we added a new feature to each reviewer in this dataset by randomly adding one of two gender labels (Male or Female) to each reviewer. We set $\lambda_0 = \lambda_1 = \lambda_2 = 1$ for our experiments. Note that by varying these parameters, one can create the Pareto optimal frontier too.

We run the negative cycle detection algorithm, and the MIQP solver using Gurobi to find the optimum solution. On converging to the optimal solution, we find that all 73 papers receive two male reviewers and two female reviewers, which shows that the method was capable of balancing gender diversity. Each paper receives reviewers from four different clusters. If we only optimize for cluster diversity, it is possible that the gender ratio for individual paper gets skewed. When we run the same model with $\lambda_g = 0$ (no weight to gender diversity), we find that out of 73 papers, 12 papers receive all four reviewers of the same gender and 41 papers receive three reviewers of the same gender. Hence, only 27.3% teams of reviewers are gender balanced. However, one should note that when we do not keep gender as an objective, the resultant allocation is random and different skewness can be observed in different runs based on the initial solution.

# Papers	# Reviewers	MIQP Time (s)	Our Method Time (s)
03	378	24.68	0.18
13	378	3979.90	14.84
23	378	14400.00	122.96
33	378	14400.00	400.56
43	378	14400.00	825.95
53	378	14400.00	2837.15
63	378	14400.00	5453.58
73	378	14400.00	11040.55

Table 5: Comparison of MIQP and our method for UIUC reviewer dataset with each paper needing 4 reviewers.

Finally, we compare the timing performance of our algorithm with MIQP by changing the number of papers that need to be reviewed on a Dell XPS 13 laptop with i7 processor. For MIQP, we set a maximum run time of four hours (14400 seconds) for Gurobi solver, at which we report the current best MIQP solution. Table 5 shows that for all cases with the number of papers greater than 13, MIQP does not converge within four hours, while our method finds the optimum solution in lesser time. Interestingly, MIQP current solutions are found to be the same as the optimum solution found by our method, which shows that for this application, MIQP was able to search the solution but it was not able to prove that the solution is optimum. In contrast, our method finds the solution faster as well as guarantees that it is optimum.

8 Conclusion & Future Research

In this paper, we proposed the first pseudo-polynomial time algorithms for multi-feature diverse weighted bipartite b-matching—a problem that we also showed is NP-hard. We propose an algorithm that not only guarantees an optimal solution but also converges faster than a proposed approach using a black-box industrial MIQP solver. We demonstrated our results on a dataset for paper reviewer matching. Future work could explore the extension of this method to online diverse matching [Dickerson *et al.*, 2019], where vertices arrive sequentially and must match immediately; this has direct application in advertising, where one could balance notions of reach, frequency, and immediate monetary return. Exploring connections to fairness in machine learning [Grgić-Hlača *et al.*, 2018] and hiring [Schumann *et al.*, 2019] by way of diversity are also of immediate interest.

Acknowledgements

We would like to thank anonymous reviewers for their helpful comments. Dickerson was supported by NSF CAREER Award IIS-1846237, DARPA #S4761, NIH NLM-013039-01 and a generous gift from Google. Khuller was supported by research awards from Adobe and Amazon. Ahmadi was supported in part by a research award from Amazon. Ahmed and Fuge were supported in part by NSF CMMI-1727849.

References

[Agrawal *et al.*, 2018] Shipra Agrawal, Morteza Zadimoghaddam, and Vahab Mirrokni. Proportional allocation: Simple, distributed, and diverse matching with high entropy. In *International Conference on Machine Learning (ICML)*, pages 99–108, 2018.

[Ahmed and Fuge, 2018] Faez Ahmed and Mark Fuge. Ranking ideas for diversity and quality. *Journal of Mechanical Design*, 140(1):011101, 2018.

[Ahmed *et al.*, 2017] Faez Ahmed, John P. Dickerson, and Mark Fuge. Diverse weighted bipartite b-matching. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, pages 35–41. AAAI Press, 2017.

[Basu Roy *et al.*, 2015] Senjuti Basu Roy, Ioanna Lykourentzou, Saravanan Thirumuruganathan, Sihem Amer-Yahia, and Gautam Das. Task assignment optimization in knowledge-intensive crowdsourcing. *The VLDB Journal—The International Journal on Very Large Data Bases*, 24(4):467–491, 2015.

[Benabbou *et al.*, 2018] Nawal Benabbou, Mithun Chakraborty, Xuan-Vinh Ho, Jakub Sliwinski, and Yair Zick. Diversity constraints in public housing allocation. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 973–981. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

[Bertsimas *et al.*, 2019] Dimitris Bertsimas, Theodore Papalexopoulos, Nikolaos Trichakis, Yuchen Wang, Ryutaro Hirose, and Parsia A Vagefi. Balancing efficiency and fairness in liver transplant access: tradeoff curves for the assessment of organ distribution policies. *Transplantation*, 2019.

[Carbonell and Goldstein, 1998] Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–336. ACM, 1998.

[Charlin and Zemel, 2013] Laurent Charlin and Richard Zemel. The toronto paper matching system: an automated paper-reviewer assignment system. 2013.

[Cherkassky *et al.*, 1993] Boris Cherkassky, Andrew V. Goldberg, and Tomasz Radzik. Shortest paths algorithms: Theory and experimental evaluation. *Mathematical Programming*, 73:129–174, 1993.

[Dickerson and Sandholm, 2015] John P. Dickerson and Tuomas Sandholm. Futurematch: Combining human value judgments and machine learning to match in dynamic environments. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 622–628. AAAI Press, 2015.

[Dickerson *et al.*, 2019] John P. Dickerson, Karthik Abinav Sankararaman, Aravind Srinivasan, and Pan Xu. Balancing relevance and diversity in online bipartite matching via submodularity. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

[Drummond *et al.*, 2015] Joanna Drummond, Andrew Perrault, and Fahiem Bacchus. Sat is an effective and complete method for solving stable matching problems with couples. In *IJCAI*, pages 518–525, 2015.

[Goldberg and Radzik, 1993] Andrew V. Goldberg and Tomasz Radzik. A heuristic improvement of the bellman-ford algorithm, 1993.

[Gölz and Procaccia, 2019] Paul Gölz and Ariel D Procaccia. Migration as submodular optimization. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

[Grgić-Hlača *et al.*, 2018] Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[Hirschman, 1964] Albert O Hirschman. The paternity of an index. *The American economic review*, 54(5):761–762, 1964.

[Hunt *et al.*, 2015] Vivian Hunt, Dennis Layton, and Sara Prince. Diversity matters. *McKinsey & Company*, 1:15–29, 2015.

[Karimzadehgan and Zhai, 2009] Maryam Karimzadehgan and ChengXiang Zhai. Constrained multi-aspect expertise matching for committee review assignment. In *ACM Conference on Information and Knowledge Management (CIKM)*, pages 1697–1700, 2009.

[Kobren *et al.*, 2019] Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 1247–1257, New York, NY, USA, 2019. ACM.

[Kulesza *et al.*, 2012] Alex Kulesza, Ben Taskar, et al. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012.

[Kurata *et al.*, 2017] Ryoji Kurata, Naoto Hamada, Atsushi Iwasaki, and Makoto Yokoo. Controlled school choice with soft bounds and overlapping types. *Journal of Artificial Intelligence Research*, 58:153–184, 2017.

[Lian *et al.*, 2018] Jing Wu Lian, Nicholas Mattei, Renee Noble, and Toby Walsh. The conference paper assignment problem: Using order weighted averages to assign indivisible goods. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Lin and Bilmes, 2012] Hui Lin and Jeff A Bilmes. Learning mixtures of submodular shells with application to document summarization. *arXiv preprint arXiv:1210.4871*, 2012.

[Liu *et al.*, 2014] Xiang Liu, Torsten Suel, and Nasir Memon. A robust model for paper reviewer assignment. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 25–32, New York, NY, USA, 2014. ACM.

[Minoux, 1986] M. Minoux. *Solving integer minimum cost flows with separable convex cost objective polynomially*, pages 237–239. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.

[Østergaard *et al.*, 2011] Christian R Østergaard, Bram Timmermans, and Kari Kristinsson. Does a different view create something new? the effect of employee diversity on innovation. *Research Policy*, 40(3):500–509, 2011.

[Ross *et al.*, 2010] Joel Ross, Lilly Irani, M Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 extended abstracts on Human factors in computing systems*, pages 2863–2872. ACM, 2010.

[Schumann *et al.*, 2019] Candice Schumann, Samsara N. Counts, Jeffrey Foster, and John P. Dickerson. The diverse cohort selection problem. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2019.