Diverse Weighted Bipartite b-Matching

Faez Ahmed
Dept. of Mechanical Engg.
University of Maryland
faez00@umd.edu

John P. Dickerson
Dept. of Computer Sci.
University of Maryland
john@cs.umd.edu

Mark Fuge
Dept. of Mechanical Engg.
University of Maryland
fuge@umd.edu

Abstract

Bipartite matching, where agents on one side of a market are matched to agents or items on the other, is a classical problem in computer science and economics, with widespread application in healthcare, education, advertising, and general resource allocation. A practitioner's goal is typically to maximize a matching market's economic efficiency, possibly subject to some fairness requirements that promote equal access to resources. A natural balancing act exists between fairness and efficiency in matching markets, and has been the subject of much research. In this paper, we study a complementary goal—balancing *diversity* and efficiency—in a generalization of bipartite matching where agents on one side of the market can be matched to *sets* of agents on the other. Adapting a classical definition of the diversity of a set, we propose a quadratic programming-based approach to solving a supermodular minimization problem that balances diversity and total weight of the solution. We also provide a scalable greedy algorithm with theoretical performance bounds. We then define the *price of diversity*, a measure of the efficiency loss due to enforcing diversity, and give a worst-case theoretical bound. Finally, we demonstrate the efficacy of our methods on three real-world datasets, and show that the price of diversity is not bad in practice.

1 Introduction

Bipartite matching problems pair an agent or item on one side of a market to an agent or item on the other. Weighted bipartite b-matching generalizes this problem to the setting where matches have a real-valued quality, and agents on one side of the market can be matched to a cardinality-constrained *set* of items or agents on the other side; real-world examples include matching children to schools [Kurata *et al.*, 2015; Drummond *et al.*, 2015], reviewers to manuscripts [Charlin and Zemel, 2013; Liu *et al.*, 2014], donor organs to patients [Bertsimas *et al.*, 2013; Dickerson and Sandholm, 2015], and workers to firms [Horton, 2017].

Often, a matching market's central goal is to maximize economic efficiency subject to some fairness constraints, such as ensuring equal opportunity amongst participants. For example, a firm might wish to maximize the number of open positions filled subject to a fairness constraint: a firm must interview a representative number of workers from marginalized backgrounds. Yet, the firm also cares about the entire cohort's ability: the workers it hires should hold high quality, yet complementary skill sets. This paper studies the tradeoff between economic efficiency and *diversity*, where matchings provide good coverage over different *classes* of item or agent.

A representative example this paper considers is matching academic papers to possible reviewers. A paper might have highest relevance to three reviewers who all come from the same lab group, perhaps because they all published heavily in a similar area. Existing weighted bipartite b-matching (WBM) algorithms would likely assign those three reviewers to the same paper. Is this outcome desirable? On the one hand, yes, because they have expertise related to the paper. On the other hand, those reviewers would stress similar points, given their common background. So the paper may only improve in a narrow (albeit important) direction. What if we wanted to diversify the reviewer backgrounds—to find reviewers well-suited to the paper *and* complementary to each other? Ideally, the reviews would remain high quality, but would cover different, complementary aspects of the paper.

This paper addresses how to compute diverse matchings under various constraints, bounds the loss on economic efficiency due to using a diverse matching, and shows in simulation and on data from three real-world bipartite matching problems that it is possible to achieve diverse matchings with limited cost to economic efficiency.

1.1 Related Work

In practice, the weighted bipartite b-matching (WBM) problem—find the feasible matching with maximum weight—has arisen naturally as a problem in many fields, such as: protein structure alignment [Krissinel and Henrick, 2004]; object recognition, and computer vision [Belongie *et al.*, 2002]; estimating text similarity [Pang *et al.*, 2016]; processor multithreading; VLSI design [Huang *et al.*, 1991]; and matching reviewers to papers in peer-review systems [Rodriguez and Bollen, 2008; Charlin and Zemel, 2013; Liu *et al.*, 2014; Tang *et al.*, 2010]. Driven by practical application, such previous work aims to maximize economic efficiency. We will compare against this objective in the present work.

This paper incorporates diversity objectives into the WBM problem. The closest related paper, due to Liu *et al.* [2014], performs a node-specific diversity-inspired preprocess before solving a related matching problem; in our work, we consider the "global" diversity of the full matching, a function of the diversity of *sets* of vertices. Other papers have addressed diversity in ranking problems (*e.g.*, diverse recommendations [Adomavicius and Kwon, 2012]), but not for matching. Past approaches mathematically represent *coverage* of a set of items, such that a diverse set better *covers* the space of items. This coverage is often defined via diminishing marginal utility over a space, such that adding more items to nearby areas of space is less useful. There are many application-dependent choices for what such a space entails including vector spaces such as text vectors [Puthiya Parambath *et al.*, 2016] or metrics over graphs [Zhang *et al.*, 2005], among others. To represent diminishing marginal utility, families of *submodular functions* are natural candidates that have shown promise in diversity tasks like document summarization [Lin and Bilmes, 2011; Zhu *et al.*, 2007]. We will use similar reasoning when defining our objectives.

The most similar work to ours is due to Chen *et al.* [2016], who propose Conflict-Aware WBM (CA-WBM). They consider conflict constraints between vertices on the same side of a bipartite graph. In CA-WBM, if two vertices are in conflict, they may not both be matched to a vertex on the other side of the graph. CA-WBM enforces a kind of binary diversity by manually defining conflicts between specific nodes. In contrast, this paper treats diversity as an objective, not a constraint, allowing us to flexibly control the degree to which a matching algorithm encourages or discourages diverse solutions to the standard WBM problem. This is useful when one wants conflicts or diversity to vary in degree, or trade off diversity with other measures of match quality.

1.2 Our Contributions

This work studies the tradeoff between diversity and efficiency in matching markets. This is different from earlier work as the diversity measure is modeled as an objective and not as constraints, and diversity is defined over *sets* of items.

Our main contributions are as follows.

1. We formulate the diverse weighted bipartite b-matching optimization problem.
2. We propose a polynomial-time greedy algorithm for constrained b-matching, and prove performance bounds on that relative to the NP-hard main problem.
3. We show via simulation and data from three large real-world bipartite matching problems that our method produces matchings with much higher diversity than standard efficient matchings, at little overall cost to economic efficiency.

In the following Section 2, we formalize the weighed b-matching optimization problem; then, in Section 3, we define our diversity-promoting objective and present the *price of diversity*, a measure of the tradeoff in economic efficiency under a diverse matching objective. Section 4 presents an optimal method for solving our problem, a scalable polynomial-time greedy algorithm with performance bounds, and a worst-case bound on the price of diversity. Section 5 shows via sim-
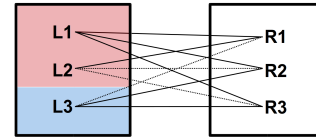

Figure 1: Bipartite b-matching problem where the left side nodes are divided into two clusters.

ulation and on real data from three matching problems that (i) our method promotes diversity in matching, (iii) the greedy approximate algorithm is both scalable and performs comparable to optimal, (ii) both algorithms retain dramatically more efficiency than our worst-case bounds implied; that is, the price of diversity in practice is quite good.

2 Weighed Bipartite Matching

Weighted bipartite b-matching is a combinatorial optimization problem formulated as follows. Given a weighted bipartite graph $G = (U, V, E)$ with weights $W : E \rightarrow R^+$, where U, V and E represent left vertices, right vertices and edges, the weighted bipartite b-matching problem is to find a subgraph $T \subset G$ such that each vertex i in T has at most b edges (*i.e.*, a degree constraint). WBM maximizes or minimizes the objective depending on the application.

We use similar notation to Chen *et al.* [2016] to define a weighed bipartite b-matching problem, with two notable differences. First, we define it as a minimization problem, and second, we define a harder problem which has both node-specific upper- and lower-cardinality constraints. The constrained weighed bipartite b-matching (WBM) problem can be expressed as follows.

$$
\begin{aligned}
\min_{X} \quad & f_1 = WX \\
\text{s.t.} \quad & L^- \leq AX_i \leq L^+ && \forall i \in \{1, \dots, M\} \\
& R^- \leq AX_i \leq R^+ && \forall i \in \{M+1, \dots, M+N\} \\
& x_{ij} \in \{0, 1\} && \forall i, j, 1 \leq i \leq M, 1 \leq j \leq N
\end{aligned}
\tag{1}
$$

We have N items on the right side with R^- and R^+ integral lower and upper cardinality constraints, respectively, and M items on the left side with L^- and L^+ as integral cardinality bounds. Here, X is an $M \times N$ matrix of binary variables, with $x_{ij} = 1$ if left item i is matched to right item j, and $x_{ij} = 0$ otherwise. W is a matrix of weights w_{ij} representing the local quality of matching items i and j.

Items on the same side of bipartite graph cannot be matched; thus, we use A as a linking matrix, such that any row i indicates which nodes are allowed to be connected to item i. We index edges (i, j) uniquely using a function $\ell : E \rightarrow \{1, \dots, MN\}$. Then, A is an $(M + N) \times MN$ matrix, where $a_{i\ell((i,j))} = 1$ if edge (i, j) exists; otherwise, $a_{i\ell((i,j))} = 0$. The degree constraints for left nodes are given by $L^- \leq Ax_i \leq L^+$, where AX_i denotes the ith element in (vector) Ax. L^+ is the upper bound on cardinality of ith node and L^- is the lower bound.

The above formulation shows a constrained matching problem, where nodes on both sides have capacity constraints. This discrete linear optimization problem is NP-hard [Chen *et al.*, 2016]. Its optimal solution will minimize the weights, emphasizing on efficiency and neglecting diversity.

3 Diversity in Matching

Diversity in matching can be defined as the need to match a node with other nodes from different groups. To add diversity, we consider a scenario where left-side nodes are divided into K groups. Let us say that we want a matching which matches each node on the right side to nodes from different clusters. The diversity is calculated using supermodular functions. These functions have been widely used in extractive document summarization [Lin and Bilmes, 2011] to get a diverse high quality summary of documents. We use a quadratic function which can incorporate diversity by balancing the number of nodes (e.g., items or agents) selected from different clusters.

Let $E_l = \{(1, l), \ldots, (M, l)\}$ be the set of all M edges from a node $l \in \{1, \ldots, N\}$ on the right side of the graph. Let the subset $S_l \subseteq E_l = \{(1, l), \ldots, (m, l)\}$ be the matched m edges for node l. Let $(P_i)^l, i \in \{1, \ldots, K\}$ is a partition of the ground set E_l into K separate clusters (*i.e.*, $\cup_i P_i^l = E_l$, and $\cap_i P_i^l = \varnothing$). That is, a left item can only belong to one cluster. The weight of an edge from left node n to right node l is $w_{n,l}$. We define the quality of match for node l on the right side as:

$$f(l) = \sum_{k=1}^{K} \left\{ \sum_{j \in S_l \cap P_k^l} w_{j,l} \right\}^2 \quad (2)$$

This quadratic function gives lower cost to solutions with even coverage over all clusters. As an example, Figure 1 shows three nodes on either side, each requiring two edges. If all edge weights w are one, the node-specific utility of a matching $\{(L1, R1), (L2, R1)\}$ is 4, while the utility of alternate matching $\{(L1, R1), (L3, R1)\}$ is 2. Hence, by minimizing the function in Equation 2, diversity is encouraged. By transforming the matching problem to quadratic minimization, the resultant objective function simultaneously optimizes quality and diversity. In the next section, we provide a formal framework to generalize this function to constrained b-matching problems.

To the best of our knowledge, no known general measure exists to measure the performance of diverse b-matching methods. Even verification of diverse matching is difficult, due to different definitions of diversity in the literature. One way of comparing our diversity results is to look at the Shannon entropy of a match for each item, with and without our method. Shannon entropy has been widely used in the ecological literature as a diversity index and quantifies the uncertainty in predicting the cluster label of an individual that is taken at random from the dataset. Here entropy of a node is given by: $-\sum_{i=1}^{K} (p_k \log p_k)$, where p_k is the proportion of selected edges in cluster K.

Entropy for an item is maximized if it is matched to other items with even coverage of different clusters; it is zero when all such items are from the same cluster. Hence, the impact of diverse matching can be measured as improvement in average entropy. We define the *entropy gain* (EG) as:

$$EG = \frac{\text{Average entropy using a diverse matching rule}}{\text{Average entropy using WBM}} \quad (3)$$

We also propose a new metric to measure the efficiency lost due to diversity. We define the *price of diversity* (PoD) as:

$$PoD = \frac{\text{Utility using WBM}}{\text{Utility using a diverse matching rule}}. \quad (4)$$

Later in the paper, we will show in simulation and on real data that the entropy gain under our proposed diverse matching method is high, at very little cost to overall efficiency.

4 Exact and Approximate Algorithms

In our bipartite matching formulation with utility minimization, the degree constraints L^- and R^- can be interpreted as setting the demand. The short side of market determines the number of edges in the matching, which is $\min\{ML^-, NR^-\}$. If the right side is short, the maximum capacity on the left should be more than the demand: $NR^- \leq ML^+$ for any matching to be feasible. For the purpose of this paper, we always assume that right side is the short side of market and the left side is clustered into groups. The cardinality constraints make the problem more difficult than what has usually been solved for matching as nodes cannot be matched independent of each other.

4.1 Diverse WBM

To generalize the quadratic function (cf. Equation 2) to an optimization framework for all nodes, we define a $MN \times MN$ block-diagonal matrix $B = \text{diag}(B_1, \ldots, B_M)$ such that:

$$B_l = \begin{cases} w_i \cdot w_j & \text{if edges } i, j \in P_k^l \text{ (same cluster)} \\ 0 & \text{otherwise} \end{cases}$$

B_l is the block diagonal matrix for every right node, with K blocks on the diagonal corresponding to each cluster. Matrix B is a diagonal matrix for all B_l matrices combined. Later, we show a visualization of the symmetric B_l matrix in Fig. 3 for five clusters for reviewer matching application. Hence, the optimization problem for Diverse WBM (D-WBM) can be written as:

$$\underset{X}{\text{minimize}} \quad f_2(X) = X^T B X \quad (5)$$

The constraints and variables are the same as in WBM (cf. Equation 1). Our new model has a quadratic objective with linear constraints and integrality requirement for variables. We solve it using two different approaches, first using Gurobi's Mixed Integer Quadratic Programming (MIQP) Solver [Gurobi Optimization, 2016], and second by using a novel greedy algorithm that builds up a set by minimizing marginal gain. Next, we propose this greedy algorithm and give bounds on its performance.

4.2 Greedy Diverse WBM

The objective of the D-WBM formulation is supermodular, that is, adding new elements leads to (strictly) increasing differences. Hence a method which greedily adds edges by minimizing the marginal gain can attain reasonable performance bounds [Nemhauser *et al.*, 1978]. Secondly, solving the MIQP exactly requires storing the block diagonal matrix;

for large problems, MIQP may run out of memory as the number of non-zero terms in the matrix scales by $M^2 N$.

The incorporation of bounds L^-, L^+, R^-, and R^+ makes greedily adding edges non-trivial because the trivial matching solution may be infeasible. A naïve greedy algorithm which adds edges to minimize marginal gain and terminates when lower bounds are satisfied may fail in such cases. We will discuss two examples from Figure 1 to demonstrate that. If each node requires exactly two edges ($L^- = L^+ = R^- = R^+ = 2$) and naïve greedy adds edges in the order $(R1, L1), (R1, L2), (R2, L1), (R2, L2)$, then $R3$ will have no option to satisfy the constraint—yet clearly a feasible matching does exist. Consider another scenario, where right nodes require exactly one edge ($R^- = R^+ = 1$) while left nodes require at least one edge ($L^- = 1$). If naïve greedy selects edges $(R1, L1), (R2, L1), (R3, L1)$, then nodes $L2$ and $L3$ cannot satisfy their lower bounds.

To address these issues, we propose an algorithm which incrementally *satisfies* the lower degree constraints for all nodes. It does this by increasing the lower bound of all nodes unit step at a time and selecting edges by minimizing marginal gain in the objective f_2. In selecting edges, it gives preference to the set of nodes with unsatisfied lower bound. This ensures that the greedy selection always finds a feasible matching with good empirical performance.

Algorithm 1: GD-WBM Greedy Diverse Matching

Input: A set of $N + M$ nodes, bounds L^-, L^+, R^-, R^+ and edge weights matrix B
Output: A feasible matching
1 $C \leftarrow \emptyset$
2 **for** $i \leftarrow 1$ **to** $\max\{L^-, R^-\}$ **do**
3      $L_i^- = \min\{i, L^-\}; R_i^- = \min\{i, R^-\}$
4   **for** $j \leftarrow 1$ **to** $N + M$ **do**
5   **if** *j's lower bound is not satisfied* **then**
6   Select the feasible edge e with lowest marginal gain $f(C \cup \{e\}) - f(C)$ whose opposite node's lower bound is not satisfied
7   If no such node exists, pick the feasible edge with lowest marginal gain.

8 **return** C

For a right constrained matching, the matching for every right node is independent of others as they do not have overlapping constraints. Hence GD-WBM achieves a $(1 - 1/e)$-approximation of the optimum due to its supermodular objective function. In practice, Section 5 will show that GD-WBM performs much better than the lower bound.

4.3 Price of Diversity Bound

In this section, we propose lower bounds on the price of diversity (PoD)—the utility loss due to diverse matching—in right-constrained market. Theorem 1 gives such a bound.

Theorem 1. *The worst-case Price of Diversity (PoD) for a*

right-constrained diverse matching is:

$$\frac{1}{N} \sum_{l=1}^{N} \frac{z_l}{1 + \sqrt{R^- - 1}\sqrt{z_l^2 - 1}}, \ where \ z_l = \frac{\sum w_{jl}}{\min w_{jl}} \quad (6)$$

Proof sketch. We wish to find a problem instance where the best diverse matching has high weight under the utilitarian objective. Such PoD for a matching will be minimized when diversity leads to all low-weight weight edges being replaced by higher weight edges. Such a situation can occur when WBM provides a match for a right node with all m edges going to left nodes in the same cluster 1. Let $\{w_1, \ldots, w_m\}$ be such edge weights. In this instance, let D-WBM select edges going to m unique clusters. Here, D-WBM will select the single edge with least weight from each cluster, including cluster 1. Call those edges $\{a_1, \ldots, a_m\}$ be the selected edge weights by D-WBM. The edge weights will satisfy the following constraints:

$$\sum_{i=1}^{m} w_i \leq \sum_{i=1}^{m} a_i \quad \text{and} \quad (\sum_{i=1}^{m} w_i)^2 \geq \sum_{i=1}^{m} a_i^2 \quad (7)$$

Both WBM and D-WBM select edge $a_1 = \min_{i \in [m]} w_i$ from cluster 1. To minimize PoD, the denominator—f_1 of the diverse matching—$\sum_{i=1}^{n} a_i$ should be maximized. Using the Lagrangian method, this constrained maximization problem is solved, with optima occurring at the surface of a hypersphere and $a_2 = a_3 \ldots = a_m$. $\square$

If the minimum weight for every right node is 0, then the PoD becomes $\frac{1}{\sqrt{R^- - 1}}$. In the succeeding section, we will show that Theorem 1 is quite conservative.

5 Results and Discussion

We begin by comparing the two methods' PoD to its theoretical bound on a synthetic dataset. Next, we solve the b-matching problem on three datasets, one for movie recommendation and two for papers–reviewers matching. We analyze the effect of problem size by increasing the number of nodes and the cardinality requirements. Finally, we discuss the trade-off between diversity and utility.

5.1 Artificial Dataset

In this section, we simulate matching 10 nodes with another 10 nodes; by Theorem 1, the worst-case PoD is 0.5. Weights are selected randomly from a uniform distribution between 0 to 1 and the cluster labels of the left-side nodes are selected randomly. For right-constrained matching, we use $R^- = 5$, implying that each right side node will be matched to at least five items. The number of clusters are varied from 2 to 10, and 100 trials are done to compare D-WBM with WBM.

Figure 2 shows that in practice, PoD is never below 0.9 despite random clusters and weights. Also, EG generally decreases when PoD increases. The greedy approximation GD-WBM finds the same matches as D-WBM for all cases.
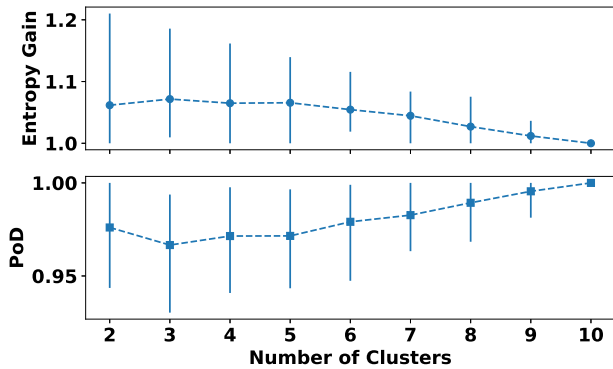
Figure 2: PoD and EG for a simulated dataset showing the average PoD with 5th and 95th percentile values. D-WBM unilaterally outperforms the worst-case PoD of 0.5.

5.2 Application to MovieLens dataset

This example considers matching movies to users, while ensuring that the movies contain diverse genres. We use a subset of the MovieLens 1M dataset [Harper and Konstan, 2016], which includes one million ratings by 6,040 users for 3,900 items. This dataset contains both users' movie ratings between 1 and 5 and genre categories for each movie (*e.g.*, comedy, romance, and action).

We first train a standard collaborative recommender system [Bradley, 2016] to obtain ratings for all movies by every user. We cluster the movies into 5 clusters using their vector of 18 genres using spectral clustering,[1] so that each movie gets a unique cluster label. We solve the right constrained matching problem for 500 movies and make recommendations for 500 users with at least 10 recommendations for every user. Table 1 shows that with average EG of 1.45, D-WBM finds a more diverse matching for all users and on average a user loses only 1% utility for this gain. To save computational time, in all our simulations we terminate D-WBM after 1 hour and take the best solution, while WBM converges to true optima. Hence the results are conservative estimates.

To further understand the matching result, we compare the movie recommendations by D-WBM and WBM for User ID 423. WBM matches her to movies that are all either Comedies or Dramas, with an average predicted rating of 4.07. In contrast, D-WBM matches the user with movies from five different clusters, with an average movie rating of 4.05, showing negligible loss in efficiency. Table 2 compares the recommended genres. While we chose to promote diversity in genre, the MovieLens dataset also provides information about the user's gender, age group, and occupation; D-WBM can encourage other types of diversity in movie recommendation.

5.3 Application to Reviewer Assignment

In this section, we present an application of diverse matching to automatically determine the most appropriate reviewers

Dataset	D-WBM		GD-WBM	
	PoD	EG	PoD	EG
Movie-Lens	0.99	1.45	0.99	1.45
UIUC Reviewer	0.92	1.63	0.83	1.60
Sugiyama	0.94	4.28	0.93	4.28

Table 1: Price of Diversity and Entropy Gain results for three real world datasets.

D-WBM		WBM	
Cluster	Genres	Cluster	Genres
2	Drama, Thriller	2	Drama,Thriller
1	Adventure,Sci-Fi	2	Crime,Drama,Thriller
3	Documentary	0	Comedy
3	Documentary	0	Comedy,Drama
0	Comedy,Romance	0	Comedy,Romance
4	Drama,Horror	2	Drama
4	Horror,Sci-Fi,Thriller	2	Drama
2	Drama	2	Drama
0	Comedy,Mystery	0	Comedy,Mystery
1	Action,Thriller	2	Action,Crime,Drama

Table 2: Genres and Cluster labels of ten movies recommended to a user by WBM and D-WBM. The movies by D-WBM provide a broader genre coverage.

for a manuscript by also ensuring that reviewers are different from each other.

UIUC Multi-Aspect Review Assignment Dataset

We use the multi-aspect review assignment evaluation dataset [Karimzadehgan and Zhai, 2009] which is 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.

For the reviewer-paper bipartite graph, edge weights between each test paper and reviewer are set as the cosine distance of their label vectors. We cluster the reviewers into 5 clusters based on their topic vectors using spectral clustering. We set the constraints such that each paper matches with at least 3 reviewers and every reviewer is allocated at least 1 paper, while no reviewer is allocated more than 10 papers.

Despite the constraints, D-WBM finds a diverse matching with PoD of 0.92. DG-WBM provides an average entropy gain of 1.60 but pays a higher price of diversity as shown in Table 1. To delve deeper into the results, we take the example of 48th paper titled "Towards musical query-by-semantic-description using the CAL500 data set" from the dataset. This paper is labeled as related to Topics T8 (Multimedia IR), T16 (Language models) and T3 (Other machine learning). WBM matches it to three reviewers with IDs 43, 34, and 158[2]. If we analyze their topic vectors, we find that all of them have the Language Models (T16) topic common between them. Not surprisingly, they are all found by our clustering method to be in the same cluster, as shown in Fig. 3.

---

[1]The exact choice of recommender system and clustering algorithm is not central to paper; it just helps set up the graph.

[2]More information on the reviewers is available here: `http://sifaka.cs.uiuc.edu/ir/data/review.html`
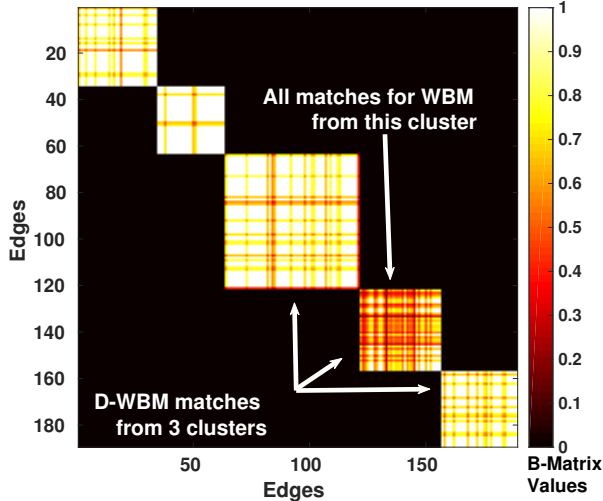
Figure 3: Block Diagonal B-Matrix for Paper 48; WBM selects all matches from a single cluster

Figure 4: Change oF PoD and EG with increasing R^-.

Figure 5: Runtime comparison as problem size increases

On the other hand, diverse matching provides a match with three reviewers (IDs 131, 153 and 158) from three different clusters. Reviewer 131 provides a balance between IR and Language Model topics but also works on User Studies. Reviewer 153 works on many topics relevant to the paper —Multimedia IR (T8), ML (T2, T3) and Text Categorization (T1). In D-WBM's reviewer set, Reviewers 153 and 131 have no common aspect between them while Reviewer 158 shares interests with both. Having a set of reviewers who are similar to the query paper but complementary in skillsets may provide a well-rounded review with good coverage of different viewpoints. GD-WBM also finds a match which has higher entropy (three different clusters) than WBM.

Scholarly Paper Recommendation Dataset

To further test our method on matching reviewers with papers, we use the Scholarly Paper Recommendation dataset provided by Sugiyama *et al.* [Sugiyama and Kan, 2010], which contains 50 researchers and 100,351 candidate papers from proceedings in the ACM Digital Library.

We select all papers from KDD 2000 to KDD 2010 from this dataset (a total of 1184 papers) and find 3 reviewers for each of them from the given set of 50 reviewers. We calculate edge weights between papers and reviewers as the cosine distance between the tf-idf vector of the query paper and reviewer's latest paper. No limit of maximum number of papers that a reviewer can review is imposed but each reviewer must be allocated at least one paper. We divide the reviewers into 5 clusters using their tf-idf vectors with Spectral Clustering. The results show that D-WBM improves on the diversity of all 1184 papers with EG of 4.28 as shown in Table 1. This large gain in entropy is obtained at a small price of losing 6% efficiency. The larger EG in this dataset compared to UIUC is because the latter had upper bounds on reviewer capacity.
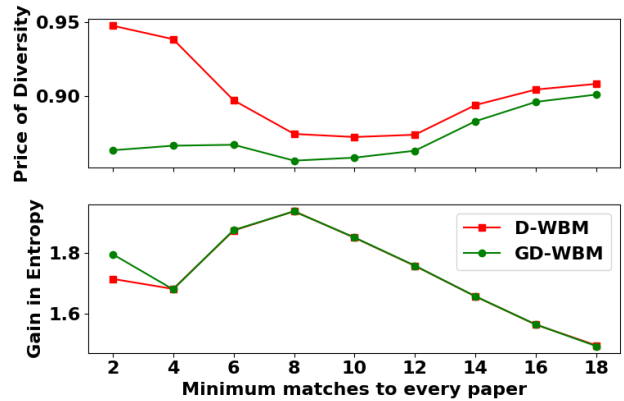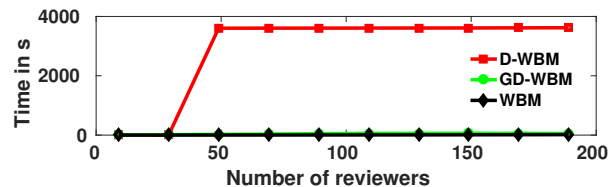
5.4 Effect of Bounds and Problem Size

So far, we have set the cardinality bounds without discussing their effect on the matching results. One would expect that as bounds become tighter, the utility of WBM and DP-WBM will suffer due to lesser search space. However, the question we answer here is how it affects the relative performance of the two methods as measured by PoD and EG.

More specifically, we study R^-, as the number of edges in the matching is determined by it. We use UIUC dataset discussed before, where each reviewer must review at least one paper and we cluster the reviewers into 10 groups. Figure 4 shows that the PoD is consistently high irrespective of the number of reviewers matched to every paper. The PoD initially decreases as more clusters contribute to diversity but then slowly increases to 1 as the problem becomes more and more constrained. Obviously, when $R^- = M$, there is only one matching possible and both WBM and D-WBM have the same utility. EG in general increases when R^- is less than the number of clusters as new clusters can contribute to diversity. Among other bounds, setting R^+ to any value has no effect on optimization. Increasing L^+ allows WBM to overuse few good nodes, who might have low edge weights with everyone. Hence, WBM's entropy suffers significantly.

Finally, we discuss the effect of problem size on the performance of WBM, D-WBM, and GD-WBM. We use the UIUC dataset with $R^- = 3, L^- = 1$ and increase the number of reviewers available to review the papers. Figure 5 shows the relative time performance of the three methods. We can see that WBM and GD-WBM take much less time than D-WBM's MIQP solver. The latter time is capped at one hour and the best solution at that point is used for analysis.

6 Conclusions & Future Research

In this paper, we presented a quantitative approach to balancing diversity and efficiency in a generalization of bipartite matching where agents on one side of the market can be matched to sets of agents or items on the other. We propose a quadratic programming-based approach to solving a supermodular minimization problem that balances diversity and total weight of the solution. The general problem is NP-hard, so we proposed a scalable greedy algorithm with theoretical performance bounds. We proposed the price of diversity (PoD), which measures efficiency loss due to enforcing diversity, and gave worst-case theoretical bounds on that metric. Finally, we validated our methods on three real-world datasets, and showed that the price of diversity is quite good in practice.

Future research will focus on diverse matching for online problems, where edges arrive sequentially and on scaling diverse matching to larger datasets. Another area of work can be diverse matching with diversity of a set defined using Determinantal Point Process (DPP) kernels [Kulesza and Taskar, 2012], which do not require explicit clustering. Lastly, the trade-off between diversity and efficiency can be further explored using an objective function which combines efficiency maximizing WBM and entropy maximizing D-WBM.

Acknowledgements

References

[Adomavicius and Kwon, 2012] Gediminas Adomavicius and YoungOk Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 24(5):896–911, 2012.

[Belongie *et al.*, 2002] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[Bertsimas *et al.*, 2013] Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research*, 61(1):73–87, 2013.

[Bradley, 2016] P. Allen Bradley. MovieLens collaborative filtering. https://github.com/bradleypallen/keras-movielens-cf, 2016.

[Charlin and Zemel, 2013] Laurent Charlin and Richard S Zemel. The Toronto paper matching system: an automated paper-reviewer assignment system. In *International Conference on Machine Learning (ICML)*, 2013.

[Chen *et al.*, 2016] Cheng Chen, Lan Zheng, Venkatesh Srinivasan, Alex Thomo, Kui Wu, and Anthony Sukow. Conflict-aware weighted bipartite b-matching and its application to e-commerce. *IEEE Transactions on Knowledge and Data Engineering*, 28(6):1475–1488, 2016.

[Dickerson and Sandholm, 2015] John P. Dickerson and Tuomas Sandholm. FutureMatch: Combining human value judgments and machine learning to match in dynamic environments. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 622–628, 2015.

[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 *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[Gurobi Optimization, 2016] Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2016.

[Harper and Konstan, 2016] F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19, 2016.

[Horton, 2017] John Joseph Horton. The effects of algorithmic labor market recommendations: evidence from a field experiment, 2017. To appear, Journal of Labor Economics.

[Huang *et al.*, 1991] Chu-Yi Huang, Yen-Shen Chen, Youn-Long Lin, and Yu-Chin Hsu. Data path allocation based on bipartite weighted matching. In *ACM/IEEE Design Automation Conference*, 1991.

[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.

[Krissinel and Henrick, 2004] E Krissinel and K Henrick. Secondary-structure matching (ssm), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D: Biological Crystallography*, 60(12):2256–2268, 2004.

[Kulesza and Taskar, 2012] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083*, 2012.

[Kurata *et al.*, 2015] Ryoji Kurata, Masahiro Goto, Atsushi Iwasaki, and Makoto Yokoo. Controlled school choice with soft bounds and overlapping types. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2015.

[Lin and Bilmes, 2011] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL-HLT)*, 2011.

[Liu *et al.*, 2014] Xiang Liu, Torsten Suel, and Nasir Memon. A robust model for paper reviewer assignment. In *ACM Conference on Recommender Systems (RecSys)*, 2014.

[Nemhauser *et al.*, 1978] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functionsi. *Mathematical Programming*, 14(1):265–294, 1978.

[Pang *et al.*, 2016] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, and Xueqi Cheng. Text match-

ing as image recognition. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

[Puthiya Parambath *et al.*, 2016] Shameem A Puthiya Parambath, Nicolas Usunier, and Yves Grandvalet. A coverage-based approach to recommendation diversity on similarity graph. In *ACM Conference on Recommender Systems (RecSys)*, 2016.

[Rodriguez and Bollen, 2008] Marko A Rodriguez and Johan Bollen. An algorithm to determine peer-reviewers. In *ACM Conference on Information and Knowledge Management (CIKM)*, 2008.

[Sugiyama and Kan, 2010] Kazunari Sugiyama and Min-Yen Kan. Scholarly paper recommendation via user's recent research interests. In *Conference on Digital Libraries*, pages 29–38, 2010.

[Tang *et al.*, 2010] Wenbin Tang, Jie Tang, and Chenhao Tan. Expertise matching via constraint-based optimization. In *Conference on Web Intelligence and Intelligent Agent Technology (WIC)*. IEEE, 2010.

[Zhang *et al.*, 2005] Benyu Zhang, Hua Li, Yi Liu, Lei Ji, Wensi Xi, Weiguo Fan, Zheng Chen, and Wei-Ying Ma. Improving web search results using affinity graph. In *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2005.

[Zhu *et al.*, 2007] Xiaojin Zhu, Andrew B Goldberg, Jurgen Van Gael, and David Andrzejewski. Improving diversity in ranking using absorbing random walks. In *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2007.