A Polyhedral Approach to Online Bipartite Matching

Alfredo Torrico Shabbir Ahmed Alejandro Toriello
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
atorrico3 *at* gatech *dot* edu, {sahmed, atoriello} *at* isye *dot* gatech *dot* edu

July 12, 2017

Abstract

We study the i.i.d. online bipartite matching problem, a dynamic version of the classical model where one side of the bipartition is fixed and known in advance, while nodes from the other side appear one at a time as i.i.d. realizations of a uniform distribution, and must immediately be matched or discarded. We consider various relaxations of the polyhedral set of *achievable* matching probabilities, introduce valid inequalities, and discuss when they are facet-defining. We also show how several of these relaxations correspond to ranking policies and their time-dependent generalizations. We finally present a computational study of these relaxations and policies to determine their empirical performance.

1 Introduction

Bipartite matching is one of the fundamental combinatorial optimization models, with a century-long history of research and applications in many areas. In online or dynamic optimization, a widely studied variant has the right side of the bipartition fixed and known to the decision maker ahead of time, while the nodes in the left side appear one after another dynamically, and must immediately be matched to a remaining compatible right-hand node or discarded. This model has applications in a variety of resource allocation and revenue management areas, particularly in online search advertisement, where right-hand nodes represent ads and left-hand nodes are search terms that the search engine wants to show compatible ads to. Partly because of the connection to online search, the computer science community has been interested in this model for several years beginning in [13], which showed that for a maximum cardinality objective, a randomized ranking policy achieves the best possible competitive ratio $1 - 1/e$, assuming an adversary chooses which left-hand nodes appear; see e.g. [6] for a corrected and simplified proof.

The adversarial model of node arrival is relatively pessimistic, and more recent research has focused on models where arrivals are at least partly governed by a distribution. In the simplest case, each arriving node is an i.i.d. sample from a known distribution. Work on this model and its variants began in [9], which presented two simple algorithms. The first one is a *suggested matching* algorithm, and has a tight competitive ratio of $1-1/e$, matching the best possible performance from the adversarial case. Moreover, [9] also showed that in the i.i.d. model it is possible to improve this ratio, by presenting a *two suggested matchings* algorithm; [4] then further improved the performance analysis of this algorithm and studied its generalizations. Subsequently [16] used Monte Carlo sampling to construct an adaptive algorithm that yields a better performance guarantee. The best currently known analysis is from [12], which use a simple max-flow relaxation to generate a randomized adaptive algorithm. Other papers that study more general online bipartite matching models include [10, 11, 15]. For more references and a complete description of related problems, we refer to the reader to [17].

While these works sometimes employ simple relaxations to design policies, to our knowledge no researchers have specifically looked at the generation of good upper bounds for the problem.

- Our **first contribution** is to study the set of matching probabilities achievable by some feasible policy, which is implicitly encoded as the projection of a doubly exponential polyhedron, and to derive relaxations by identifying various classes of valid inequalities. This focus on the *achievable region* is used in applied probability, for example to study models in queueing and multi-armed bandits (e.g. [5, 7]), but to our knowledge it has not been applied to online matching.

- Our **second contribution** is then to show that optimal dual solutions of our relaxations imply specific policies, including simple ranking or time-dependent ranking policies in certain cases. This connection is established by enforcing intuitive value function approximations on the linear programming formulation of the problem's dynamic program.

The general technique of approximating a dynamic programming value function to obtain relaxations and policies is applicable in many sequential decision making problems, and dates back to [1, 8, 19, 22]. It has been used to derive relaxations and policies in many discrete dynamic optimization settings, such as economic lot scheduling [3], inventory routing [1, 2] and the traveling salesman problem [20, 21].

In the remainder of the paper, Section 2 states the problem and its formulations, and Section 3 discusses our proposed inequalities. Section 4 then shows how the relaxations imply policies, and Section 5 summarizes the results of our computational study. Section 6 then concludes and outlines future research avenues.

2 Model Description and Formulation

We are concerned with *online bipartite matching* (OBM) between two node sets, N and V, with edge set $E \subseteq N \times V$. This process occurs dynamically in the following way. The *right-hand* set V, with $|V| = m$, is known and given ahead of time. The *left-hand* set N with $|N| = n$ represents node types that may appear, but we do not know which ones will appear and how often. We know only that T elements in total will appear sequentially, each one drawn independently from the (stationary) uniform distribution over N. In other words, at each decision epoch a new node appears, and the probability that it belongs to any of the types in N is $1/n$. A left-hand node's type indicates which elements of V it is connected to, and if the same type appears more than once, each node is a separate copy. The assumption that the distribution is uniform is without loss of generality as long as the original distribution has rational probabilities, since we can duplicate nodes to put the problem in this form. Several past works on this OBM model (e.g. [9]) require $T = n$ so that the expected number of appearances of any left-hand type is one, but we do not need this assumption. Each time a left-hand node appears, it must be immediately (and irrevocably) matched to an available compatible node in V or discarded. The objective is to maximize the expected number of matches. Following convention from previous literature and the motivating application of search engine advertisement, we sometimes call $i \in N$ an *impression*, and each $j \in V$ an *ad*.

We can more generally consider weights on edges, with the objective of maximizing the expected total weight of matched edges. Our upper bound results in Section 3 extend to this case. However, for the sake of simplicity, in the remainder of the paper we focus exclusively on the cardinality case.

For any impression $i \in N$, let $\Gamma(i) \subseteq V$ denote i's neighbors, and define $\Gamma(j)$ for ad $j \in V$ analogously; also, let η be the random variable with uniform distribution over N. We can now give a dynamic programming (DP) formulation for this OBM model. Let $v^*_t(i,S)$ denote the optimal expected value given that $i \in N$ appears when the set of ads $S \subseteq V$ is available and $t-1$ draws from N still remain. Then, for all $t = 1, \dots, T$, $i \in N$ and $S \subseteq V$,

$$v^*_t(i,S) = \max \begin{cases} \max_{j \in S \cap \Gamma(i)} \{1 + \mathbb{E}_\eta[v^*_{t-1}(\eta, S \backslash j)]\} \\ \mathbb{E}_\eta[v^*_{t-1}(\eta, S)], \end{cases} \tag{1}$$

where $v^*_0(\cdot,\cdot)$ is identically zero, and the model's optimal expected value is given by $\mathbb{E}_\eta[v^*_T(\eta, V)]$. The first term in this recursion corresponds to matching i with one of its remaining neighbors $j \in S \cap \Gamma(i)$; the second corresponds to discarding i. As with any DP, the optimal value function v^* induces an optimal policy: At any state (t,i,S), we choose an action that attains the maximum in (1). It is easy to see that an optimal policy always matches an impression whenever possible, so that the discarding action is only taken when no compatible ad remains.

The recursion (1) can be equivalently captured with the linear program that we denote $\mathcal{D}$:

$$\min_v \ \mathbb{E}_\eta[v_T(\eta, V)] \tag{2a}$$
$$\text{s.t. } v_t(i, S \cup j) - \mathbb{E}_\eta[v_{t-1}(\eta, S)] \geq 1, \qquad t \in [T],\ i \in N,\ j \in \Gamma(i),\ S \subseteq V \backslash j \tag{2b}$$
$$v_t(i,S) - \mathbb{E}_\eta[v_{t-1}(\eta, S)] \geq 0, \qquad t \in [T],\ i \in N,\ S \subseteq V \tag{2c}$$
$$v \geq 0. \tag{2d}$$

The value function v^* defined in (1) is optimal for (2). Moreover, this LP is a strong dual for OBM, in the sense that any feasible v has an objective greater than or equal to $\mathbb{E}_\eta[v^*_T(\eta, V)]$.

The dual of (2) is a primal formulation; any feasible solution encodes a feasible policy and its probability of reaching any state in the DP. That formulation is the LP

$$\max_{x,y} \ \sum_{i \in N} \sum_{t \in [T]} \sum_{j \in \Gamma(i)} \sum_{S \subseteq V \backslash j} x^{t,S}_{i,j} \tag{3a}$$
$$\text{s.t. } \sum_{j \in \Gamma(i)} x^{T, V \backslash j}_{i,j} + y^{T,V}_i \leq \frac{1}{n}, \quad i \in N, \tag{3b}$$
$$\sum_{j \in \Gamma(i) \cap S} x^{t, S \backslash j}_{i,j} + y^{t,S}_i - \frac{1}{n} \sum_{k \in N} y^{t+1,S}_k - \frac{1}{n} \sum_{k \in N} \sum_{j \in \Gamma(k) \cap S} x^{t+1,S}_{k,j} \leq 0, \quad t \in [1, T-1],\ i \in N,\ \varnothing \neq S \subseteq V, \tag{3c}$$
$$\sum_{j \in \Gamma(i) \cap S} x^{T, S \backslash j}_{i,j} + y^{T,S}_i \leq 0, \qquad i \in N,\ S \subsetneq V, \tag{3d}$$
$$x, y \geq 0. \tag{3e}$$

Here, $x^{t,S}_{i,j}$, the variable corresponding to dual constraint (2b), represents the probability that the policy chooses to match impression i to ad j in state $(t, i, S \cup j)$, and $y^{t,S}_i$, which corresponds to (2c), similarly represents a discarding action. As with its dual, the LP (3) has exponentially many variables and constraints, and is therefore difficult to work with directly. However, we can equivalently consider optimizing over the matching probabilities achieved by a feasible policy; this corresponds to optimizing over a projection of the feasible region of (3),

$$\max\left\{\sum_{i\in N}\sum_{j\in\Gamma(i)} z_{ij} : \exists\ (x,y)\in \text{(3b)–(3e) with } z_{ij} = \sum_{t\in[T]}\sum_{S\subseteq V\setminus j} x^{t,S}_{i,j}\right\}, \tag{4}$$

where z_{ij} is the probability that impression i is *ever* matched to ad j. Any such z is a vector of matching probabilities that is *achievable* by at least one feasible policy. Let Q denote this projected polyhedron in the space of z variables, and note that Q is full-dimensional in $\mathbb{R}^{|E|}$.

3 Projected Relaxations

The polyhedron Q captures the matching probabilities achievable by any feasible policy, and optimizing over it would yield the expected value of an optimal policy. Although this optimization is computationally intractable, optimizing over any relaxation of Q yields a valid dual upper bound. In this section, we study various relaxations by presenting several classes of inequalities that are valid for Q. We begin by presenting the simplest relaxation.

Recall that each ad j can be matched at most once, so this constrains all probabilities involving j to not exceed one in total. Similarly, each impression type i appears in each epoch with probability $1/n$, and there are T stages, so the expected number of matches for i cannot exceed T/n. This gives us the LP

$$\max \sum_{(i,j)\in E} z_{ij} \tag{5a}$$

$$\text{s.t.} \sum_{j\in\Gamma(i)} z_{ij} \le \frac{T}{n}, \qquad i\in N \tag{5b}$$

$$\sum_{i\in\Gamma(j)} z_{ij} \le 1, \qquad j\in V \tag{5c}$$

$$z \ge 0. \tag{5d}$$

In particular, when $T = n$, (5) gives the deterministic bipartite matching formulation over $(N\cup V, E)$, and more generally it encodes a simple max-flow model; see e.g. [9].

We can use similar probabilistic ideas to strengthen the relaxation. An impression $i \in N$ will not appear at all with probability $(1-1/n)^T$, and thus

$$z_{ij} \le 1-(1-1/n)^T, \quad i\in N,\ j\in\Gamma(i) \tag{6}$$

is valid for Q; see e.g. [11]. Though these inequalities were already known, the following result is new to our knowledge.

Proposition 3.1. *Constraints* (6) *are facet-defining for the polyhedron of achievable probabilities* Q*.*

Proof. Let $\alpha := 1-(1-1/n)^T$ denote the inequality's right-hand side, and define also the numbers $\beta := 1-(1-1/n)^T - \frac{T}{n}(1-1/n)^{T-1}$, $\gamma := \frac{1}{n}(1-1/n)^{T-1}$. Here α represents the probability that $i\in N$ appears at least once, β corresponds to the probability that $i\in N$ appears at least twice, while γ is the probability that $i\in N$ does not appear during the first $T-1$ epochs times the probability it appears in the last epoch. We denote the canonical vector by $e_{ij}\in\mathbb{R}^{|E|}$, i.e. a vector with a one in the coordinate (i,j) and zero elsewhere. We can construct the following $|E|$ affinely independent points corresponding to policies that satisfy (6) with equality:

- The policy that simply matches (i,j) when possible and ignores other edges corresponds to αe_{ij}.
- For any edge (i',j') that does not share an endpoint with (i,j), the policy that matches either edge when possible corresponds to $\alpha(e_{ij}+e_{i'j'})$.
- For any $j' \in \Gamma(i)\backslash j$, the policy that matches (i,j) the first time i appears and then matches (i,j') the second time corresponds to $\alpha e_{ij}+\beta e_{ij'}$.
- For any $i' \in \Gamma(j)\backslash i$, the policy that matches (i,j) when possible but in the last epoch matches (i',j) if i' appears and i hasn't appeared corresponds to $\alpha e_{ij}+\gamma e_{i'j}$. □

We can generalize the previous concept to any set of impressions incident to an ad $j \in V$. Let $I \subseteq \Gamma(j)$; no nodes from this set will appear at all with probability $(1-|I|/n)^T$, and hence the set of *right-star* inequalities

$$\sum_{i \in I} z_{ij} \leq 1-(1-|I|/n)^T, \qquad j \in V,\ I \subseteq \Gamma(j) \tag{7}$$

is valid. Moreover, their separation can be achieved in polynomial time by sorting the z_{ij} in non-increasing order of i, and testing each successive sum against the corresponding right-hand side. Inequality (7) also shows that (5c) can never be tight for a feasible policy unless $|\Gamma(j)|=n$; in this case (5c) coincides with (7) for $I=\Gamma(j)=N$, and we get the following result.

Proposition 3.2. *If $j \in V$ and $\Gamma(j)=N$, constraint* (5c) *is facet-defining for the polyhedron of achievable probabilities Q.*

Proof. Let $j \in V$ be any ad. We can construct the following $|E|$ affinely independent points corresponding to policies that satisfy (5c) with equality:

- For any $i \in \Gamma(j)$, α is the probability that this impression appears at least once. Take the policy that matches (i,j) whenever possible (with probability α), and if i never appears matches the impression $k \in \Gamma(j)\setminus i$ that appears in the last draw, with probability $\frac{1}{n-1}(1-\alpha)$ respectively for each $k \neq i$. Therefore, we have n points of the form $\alpha e_{ij}+\frac{1}{n-1}(1-\alpha)\sum_{k\in\Gamma(j)\setminus i} e_{kj}$ for each edge (i,j).
- The remaining policies are constructed similarly to the proof of Proposition 3.1.

Let us denote $\epsilon=\frac{1}{n-1}(1-\alpha)$. So, we have a block matrix of size $|E|\times|E|$, in which the first $n\times n$ block corresponds to the first group of policies, and has the following form

$$C=\begin{pmatrix} \alpha & \epsilon & \epsilon & \cdots & \epsilon \\ \epsilon & \alpha & \epsilon & \cdots & \epsilon \\ \vdots & & \ddots & & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & \alpha \end{pmatrix}.$$

Let us compute the determinant of C. Add rows 2 through n to row 1 to get

$$\det(C)=\det\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ \epsilon & \alpha & \epsilon & \cdots & \epsilon \\ \vdots & & \ddots & & \vdots \\ \epsilon & \epsilon & \epsilon & \cdots & \alpha \end{pmatrix}.$$

Multiply the first row by $-\epsilon$ and add it to every other row. Then,

$$\det(C) = \det \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & \alpha-\epsilon & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & \alpha-\epsilon \end{pmatrix} = (\alpha-\epsilon)^{n-1},$$

which is greater than zero for all $T > 0$. Therefore, this block matrix is invertible, which implies the points corresponding to the first group of policies are affinely independent. The remaining points are also affinely independent, as they are the same as in the previous proof. □

In general, inequalities of the form (7) are not facet-defining, except for those cases presented in Proposition 3.1 and 3.2.

Proposition 3.3. *If $I \neq N$, the face of Q induced by constraint* (7) *has dimension $|E| - |I|$.*

Proof. Let $j \in V$ and $I \subseteq \Gamma(j)$ with $I \neq N$. Using the same notation as in previous propositions, construct the following points that satisfy (7) with equality:

- The policy that matches the first $i \in I$ that appears to j produces the point $\frac{1}{|I|}(1 - (1 - |I|/n)^T) \sum_{i \in I} e_{ij}$.
- For any other edge we can construct points as we did in Proposition 3.1.

This gives us $|E| - |I| + 1$ affinely independent points. Furthermore, the only way to achieve the probability $1 - (1 - |I|/n)^T$ from the edges between I and j is to follow the outlined policy: The first time any node from I appears, it must be matched to j. Conditioning on an arrival from I, any of its members are equally likely to appear; so any point on this face must have $z_{ij} = \frac{1}{|I|}(1 - (1 - |I|/n)^T)$ for every $i \in I$. This implies we cannot produce more affinely independent points. □

Let us examine the analogous situation on the other side. For an impression $i \in N$, consider a set $J \subseteq \Gamma(i)$ of adjacent ads. Since i may appear more than once, the previous argument does not apply. However, we can still upper bound the corresponding probabilities by considering the expected number of matches we can hope to make with i in J. As before, i will never appear with probability $(1 - 1/n)^T$. Similarly, i will appear exactly once (and can thus only be matched once) with probability $\frac{T}{n}(1 - 1/n)^{T-1}$. This continues until we consider the event that i appears $|J|$ or more times, because we cannot match i more than these many times in J. Let $B(T, 1/n)$ denote a binomial random variable with T trials and probability of success $1/n$. The preceding argument shows that the *left-star* inequalities

$$\sum_{j \in J} z_{ij} \leq \mathbb{E}\left[\min\left\{|J|, B(T, 1/n)\right\}\right], \qquad i \in N,\ J \subseteq \Gamma(i) \tag{8}$$

are valid. In addition, the same greedy algorithm used for (7) applies to separate them, this time sorting in non-decreasing order of j.

Theorem 3.4. *Constraints* (8) *are facet-defining for Q when $|J| < T$.*

Proof. Consider $i \in N$ and $J = \{j_1, \ldots, j_r\} \subseteq \Gamma(i)$, where $r := |J| < T$. Denote by α_k the probability that i appears at least k times, i.e.

$$\alpha_1 = 1 - \left(1 - \frac{1}{n}\right)^T, \quad \alpha_2 = 1 - \left(1 - \frac{1}{n}\right)^T - \frac{T}{n}\left(1 - \frac{1}{n}\right)^{T-1}, \text{ etc.}$$

So, it is clear that

$$\mathbb{E}\left[\min\left\{r, B(T, 1/n)\right\}\right] = \sum_{k=1}^{r} \alpha_k.$$

Therefore, we can consider the following $|E|$ affinely independent points:

- Take the $|J|$ policies that match just node i, each time it appears, with nodes $j \in J$ following each of the permutations given by the ordering; that is, the first permutation is $(j_1, \ldots, j_r)$, the second is $(j_2, \ldots, j_r, j_1)$, and so forth until the r-th permutation, $(j_r, j_1, \ldots, j_{r-1})$. So, point $\sum_{k=1}^{r} \alpha_k e_{ij_k}$ corresponds to the first policy, $\alpha_r e_{ij_1} + \sum_{k=1}^{r-1} \alpha_k e_{ij_{k+1}}$ to the second, and so on. Thus, we need to prove that the matrix

$$C = \begin{pmatrix} \alpha_1 & \alpha_r & \alpha_{r-1} & \ldots & \alpha_2 \\ \alpha_2 & \alpha_1 & \alpha_r & \ldots & \alpha_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_r & \alpha_{r-1} & \alpha_{r-2} & \ldots & \alpha_1 \end{pmatrix}$$

  is invertible, where $\alpha_1 > \alpha_2 > \ldots > \alpha_r$. This type of matrix is called a *circulant*, and non-singularity follows from Proposition 24 in [14].

- For all $j' \in \Gamma(i) \backslash J$, we simply match (i, j') after matching all nodes in J; this corresponds for instance to $\sum_{k=1}^{r} \alpha_k e_{ij_k} + \alpha_{r+1} e_{ij'}$. Note that $|J| < T$, so $\alpha_{r+1} > 0$.

- The points for the remaining edges follow an analogous construction to Proposition 3.1. $\square$

When $|J| \geq T$, we have $\mathbb{E}\left[\min\left\{|J|, B(T, 1/n)\right\}\right] = T/n$, and therefore (8) for $J \subsetneq \Gamma(i)$ is dominated by (5b). We can, however, use a similar proof to determine when this inequality is also facet-defining.

Corollary 3.5. *Constraints* (8) *for* $J = \Gamma(i)$ *are facet-defining for* Q *regardless of* T*, and thus* (5b) *is facet-defining when* $|\Gamma(i)| \geq T$*.*

The proof for this corollary does not need an affinely independent point for (i, j) with $j \notin J$ (since $J = \Gamma(i)$), and hence does not require $|J| < T$.

Finally, consider two sets $I \subseteq N$ and $J \subseteq V$. In the best case, they induce a complete bipartite subgraph, and we can proceed as before. No edges within the two sets will be matched at all with probability $(1 - |I|/n)^T$, exactly one will be matched with probability $\frac{|I|}{n}(1 - |I|/n)^{T-1}$, and so forth. Generalizing, let $B(T, |I|/n)$ denote a binomial random variable with T trials and probability $|I|/n$ of success. Then

$$\sum_{(i,j) \in E \cap (I \times J)} z_{ij} \leq \mathbb{E}\left[\min\left\{|J|, B(T, |I|/n)\right\}\right], \qquad I \subseteq N,\ J \subseteq V, \tag{9}$$

are valid. This general set of inequalities contains both (7) and (8) as special cases, by respectively taking $J = \{j\}$ and $I = \{i\}$. Moreover, for any fixed I or J they can be separated using the same greedy algorithm, now applied to sums of the z variables; more generally, they can be separated with an integer program that maximizes the left-hand side for every fixed value of $|I|$ and $|J|$. These inequalities are not necessarily facet-defining, except for the cases we have already pointed out.

All the inequalities described so far have 0-1 coefficients; a natural question is whether all of Q's facets can be written with 0-1 coefficients. However, this is not true even for very small instances. We have constructed Q in PORTA for an instance with $T = 3$, $N = \{1, 2, 3\}$, $V = \{a, b\}$ and $E = \{1a, 1b, 2b\}$. Not counting the non-negativity constraints, Q has 13 facets, but only the four identified in (8) have 0-1 coefficients.

4 Policies Derived from Bounds

Any value function approximation v implies a policy by substituting it into (1): At any encountered state (t, i, S), choose an ad that maximizes the right-hand side,

$$\arg\max_{j \in S \cap \Gamma(i)} \{1 + \mathbb{E}_\eta[v_{t-1}(\eta, S \setminus \eta)]\} = \arg\max_{j \in S \cap \Gamma(i)} \mathbb{E}_\eta[v_{t-1}(\eta, S \setminus \eta)],$$

or discard the impression if $S \cap \Gamma(i) = \varnothing$; recall that an optimal policy only discards an impression if no match is possible. We next focus on policies based on the relaxations in the previous section, by generating approximations of the true value function that are feasible in the dynamic programming LP (2) but efficient to compute. We also show that several of these approximations lead to ranking policies (cf. [13]) and their generalizations. A *ranking policy* is specified by an ordering or permutation of V: Assuming we label ads in the permutation's order as $V = \{1, \dots, m\}$, at any decision epoch (t, i, S) we match the appearing impression i to $\min\{j : j \in S \cap \Gamma(i)\}$, the lowest-indexed compatible ad that is available. Such policies are appealing from a practical perspective, as they are completely specified by a permutation and can be implemented efficiently.

Our first two policies are ranking policies, and it then follows from [10] that their competitive ratio is $1 - 1/e$ when $m = n = T$. The *competitive ratio* of a policy is the infimum, over all instances, of the policy's expected value divided by the expected value of the *offline* optimum, i.e. the maximum matching on the realized graph. The offline optimum provides an upper bound on the optimal policy's expected value, so a competitive ratio guarantee for a policy also guarantees the same multiplicative performance with respect to the optimal policy.

To begin, suppose we approximate the value at any state by considering the impression that has just appeared and the remaining available ads. Specifically, suppose $q_i \geq 0$ represents the value of having $i \in N$ appear and be available to match. Similarly, let $r_j \geq 0$ be the value we assign to each available ad $j \in V$. This leads to an approximation of the expected value of state (t, i, S) as

$$v_t(i, S) \approx q_i + (t-1)\mathbb{E}_\eta[q_\eta] + \sum_{j \in S} r_j. \tag{10}$$

In the approximation, i has just appeared, and thus the state has value q_i. In addition, there are $t - 1$ more draws remaining, so we will get the expected value of q_η that many more times. Finally, each $j \in S$ still available to match contributes its value r_j.

Proposition 4.1. *Suppose we restrict the feasible region of $\mathcal{D}$ by forcing solutions to have the form* (10)*, where the decision variables are now q_i, $i \in N$ and r_j, $j \in V$. The resulting LP is equivalent to*

$$\min_{q,r \geq 0} \left\{ \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j : \ q_i + r_j \geq 1, \ (i,j) \in E \right\}, \tag{$\mathcal{D}_1$}$$

the dual of (5)*.*

Proof. To prove the proposition, we must establish two facts. First, the objective function (2a) of $\mathcal{D}$ reduces to $(\mathcal{D}_1)$'s objective under restriction (10); and second, the feasible region of $\mathcal{D}$ collapses to the feasible region of $(\mathcal{D}_1)$ under restriction (10).

First, for the objective function (2a) we have

$$\mathbb{E}_\eta \left[q_\eta + (T-1)\mathbb{E}_\eta[q_\eta] + \sum_{j \in V} r_j \right] = \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j.$$

For the feasible region, we take each constraint class from $\mathcal{D}$ separately. For any $t \in [T]$, $i \in N$, $j \in \Gamma(i)$, and $S \subseteq V \backslash j$ in the matching constraint (2b), we get

$$q_i + (t-1)\mathbb{E}_\eta[q_\eta] + \sum_{\ell \in S} r_\ell + r_j - \mathbb{E}_\eta \left[(t-1)q_\eta + \sum_{\ell \in S} r_\ell \right] = q_i + r_j \geq 1.$$

Similarly, the discarding constraint (2c) has

$$q_i + (t-1)\mathbb{E}_\eta[q_\eta] + \sum_{\ell \in S} r_\ell - \mathbb{E}_\eta \left[(t-1)q_\eta + \sum_{\ell \in S} r_\ell \right] = q_i \geq 0.$$

We also require $r_j \geq 0$, since the value function must be non-negative. Altogether, this yields $\mathcal{D}_1$, the dual of (5). □

Let $(q^{(10)}, r^{(10)})$ be an optimal extreme point solution of $(\mathcal{D}_1)$. This feasible region is the convex hull of node covers of $(N \times V, E)$; thus $(q^{(10)}, r^{(10)})$ is the incidence vector of a cover, and when $T = n$ it is a minimum cardinality cover. Suppose we are at state (t, i, S) and employ the value function approximation (10) given by this solution in the dynamic programming recursion (1) to choose an action. Assuming $S \cap \Gamma(i) \neq \varnothing$, this yields

$$\arg\max_{j \in S \cap \Gamma(i)} \left\{ 1 + \mathbb{E}_\eta \left[(t-1)q_\eta^{(10)} + \sum_{\ell \in S \backslash j} r_\ell^{(10)} \right] \right\} = \arg\min_{j \in S \cap \Gamma(i)} r_j^{(10)},$$

where the equivalence follows simply by removing terms that do not depend on j. This corresponds to a *cover ranking* policy: Given an optimal node cover, we match an arriving impression if possible to a non-cover ad, and only match it to an ad in the cover when no remaining non-cover ad is compatible. Note that any ranking that orders ads so that non-cover ads appear before ads in the cover induces a cover ranking policy.

Corollary 4.2. *If $m = n = T$, a cover ranking policy has a competitive ratio of $1 - 1/e$.*

Proof. This follows from [10], who show that any ranking algorithm (which they term "greedy") has this competitive ratio. □

This first approximation (10) does not capture the interaction between impressions and ads. Suppose we add a value $p_{ij} \geq 0$ to a state whenever $i \in N$ appears and $j \in S$ is one of the remaining ads. The new value function approximation is

$$v_t(i, S) \approx q_i + (t-1)\mathbb{E}_\eta[q_\eta] + \sum_{j \in S} \left(r_j + p_{ij} + \left(1 - (1 - 1/n)^{t-1}\right) \sum_{k \in N \backslash i} p_{kj} \right). \tag{11}$$

Since $i \in N$ is the current impression, the approximation includes a value p_{ij} for all remaining ads j. (This value will be zero if $(i, j) \notin E$, but we include it to simplify the expressions.) Furthermore, each other impression $k \in N \setminus i$ will appear at least once in the remaining epochs with probability $1 - (1 - 1/n)^{t-1}$, so we include these values as well, discounted by that probability; we only count these values once, because an ad can only be matched once.

Proposition 4.3. *Suppose we restrict the feasible region of $\mathcal{D}$ by forcing solutions to have the form* (11)*, where the decision variables are now q_i, $i \in N$, r_j, $j \in V$ and p_{ij}, $(i, j) \in E$. The resulting LP is equivalent to*

$$\min \; \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j + \left(1 - \left(1 - \frac{1}{n}\right)^T\right) \sum_{(i,j) \in E} p_{ij}$$

$$\text{s.t. } q_i + r_j + p_{ij} \geq 1, \quad (i,j) \in E, \qquad (\mathcal{D}_2)$$
$$q, r, p \geq 0,$$

the dual of the LP obtained by adding constraints (6) *to* (5).

Proof. The proof follows in a similar fashion to the proof of Proposition 4.1. First, we have the following expectation

$$\mathbb{E}_\eta \left[p_{\eta j} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N \setminus \eta} p_{kj} \right] = \left(1 - (1-1/n)^t\right) \sum_{i \in N} p_{ij},$$

which immediately shows that for $t = T$ the term

$$\mathbb{E}_\eta[v_T(\eta, V)] = \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j + \left(1 - \left(1 - \frac{1}{n}\right)^T\right) \sum_{(i,j) \in E} p_{ij}$$

is the objective (2a). Furthermore, for any $t \in [T]$, $i \in N$, $j \in \Gamma(i)$, and $S \subseteq V \setminus j$, we have

$$\begin{aligned}
&\sum_{l \in S \cup j} \left(p_{il} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N \setminus i} p_{kl} \right) \\
&\qquad\qquad - \mathbb{E}_\eta \left[\sum_{l \in S} \left(p_{\eta l} + \left(1 - (1-1/n)^{t-2}\right) \sum_{k \in N \setminus \eta} p_{kj} \right) \right] \\
&= p_{ij} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N \setminus i} p_{kj} \\
&\quad + \sum_{l \in S} \left(p_{il} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N \setminus i} p_{kj} \right) - \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N} \sum_{l \in S} p_{kl} \\
&= p_{ij} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N \setminus i} p_{kj} + (1-1/n)^{t-1} \sum_{l \in S} p_{ij} \quad \geq \quad p_{ij},
\end{aligned}$$

with equality holding when $t = 1$ and $S = \varnothing$. It follows that the restriction of the feasible region of (2) with this approximation yields the model $(\mathcal{D}_2)$, the dual of (5) with the additional constraints (6). $\square$

Let $(q^{(11)}, r^{(11)}, p^{(11)})$ be an extreme point optimal solution for $(\mathcal{D}_2)$, and suppose we use the approximation given by this optimal solution in (1); we call this a *probability bound policy*. At state (t, i, S) with $S \cap \Gamma(i) \neq \varnothing$, after removing terms that do not depend on j, this results in the optimization problem

$$\begin{aligned}
&\arg\max_{j \in S \cap \Gamma(i)} \sum_{\ell \in S \setminus j} \left(r_\ell^{(11)} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N} p_{k\ell}^{(11)} \right) = \\
&\arg\min_{j \in S \cap \Gamma(i)} \left\{ r_j^{(11)} + \left(1 - (1-1/n)^{t-1}\right) \sum_{k \in N} p_{kj}^{(11)} \right\}.
\end{aligned}$$

Though it is not immediately clear, this policy is also a ranking policy.

Theorem 4.4. *The probability bound policy is a ranking policy. Therefore, when $m = n = T$ it has a competitive ratio of $1 - 1/e$.*

Proof. Let $(q^{(11)}, r^{(11)}, p^{(11)})$ be an extreme point optimal solution for $(\mathcal{D}_2)$; then it is binary because $(\mathcal{D}_2)$ is the dual of a network flow. Moreover, by optimality it follows that if $p^{(11)}_{ij} = 1$, then $q^{(11)}_i = r^{(11)}_j = 0$. In particular, for any $j \in V$, at most one of $r^{(11)}_j$ and $\sum_k p^{(11)}_{kj}$ can be positive, and

$$(1 - (1 - 1/n)^{T-1}) \sum_k p^{(11)}_{kj} < 1,$$

because otherwise we can set $r^{(11)}_j = 1$ and $p^{(11)}_{ij} = 0$ for all $i \in N$ and obtain a new solution with a better objective.

Define a ranking of ads that orders them in non-decreasing order of

$$\epsilon_j(T-1) := r^{(11)}_j + (1 - (1 - 1/n)^{T-1}) \sum_k p^{(11)}_{kj}.$$

We claim the probability bound policy chooses nodes to match based on this ranking. To see this, observe that because the solution is binary, $\epsilon_j(T-1)$ can take at most $n+2$ values: It can be zero or one, or $(1 - (1-1/n)^{T-1}) \sum_k p^{(11)}_{kj}$, where $\sum_k p^{(11)}_{kj} \in \{0, \dots, n\}$. The proof then follows by noting that the ordering of the $\epsilon_j(t)$ values is the same for all $t = 2, \dots, T-1$.

□

Note that a probability bound policy does not necessarily perform strictly better than a cover ranking policy. For instance, suppose $m = n = T$ and the graph is an even cycle. In this case, it is easy to see that $p^{(11)}_{ij} = 0$ for all edges (i, j), and thus the two solutions coincide.

We next generalize this approach to include all right-star inequalities.

Theorem 4.5. *Consider the value function approximation*

$$\begin{aligned} v_t(i, S) \approx \;& q_i + (t-1)\mathbb{E}_\eta[q_\eta] \\ &+ \sum_{j \in S} \Bigg(r_j + \sum_{\substack{I \subseteq \Gamma(j) \\ I \ni i}} p_{Ij} + \sum_{\substack{I \subseteq \Gamma(j) \\ I \not\ni i}} \left(1 - (1 - |I|/n)^{t-1}\right) p_{Ij} \Bigg), \end{aligned} \tag{12}$$

where $q \in \mathbb{R}^N_+$, $r \in \mathbb{R}^V_+$, and $p_{Ij} \in \mathbb{R}_+$ for $j \in V$ and $I \subseteq \Gamma(j)$. Restricting the feasible region of $\mathcal{D}$ with this approximation yields

$$\begin{aligned} \min \;& \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j + \sum_{j \in V} \sum_{I \subseteq \Gamma(j)} (1 - (1 - |I|/n)^T) p_{Ij} \\ \text{s.t. } & q_i + r_j + \sum_{\substack{I \subseteq \Gamma(j), \\ I \ni i}} p_{Ij} \geq 1, \;\; (i, j) \in E, \\ & q, r, p \geq 0, \end{aligned} \tag{$\mathcal{D}_3$}$$

the dual of the LP obtained by adding constraints (7) *to* (5).

Proof. The proof follows the same structure as the proofs of Propositions 4.1 and 4.3. First, we have the following expectation

$$\mathbb{E}_\eta \left[\sum_{\substack{I \subseteq \Gamma(j), \\ I \ni \eta}} p_{Ij} + \sum_{\substack{I \subseteq \Gamma(j), \\ I \not\ni \eta}} \left(1 - (1 - |I|/n)^{t-1}\right) p_{Ij} \right] = \sum_{I \subseteq \Gamma(j)} (1 - (1 - |I|/n)^t) p_{Ij}$$

which immediately shows that for $t = T$ the term

$$\mathbb{E}_\eta[v_T(\eta, V)] = \frac{T}{n}\sum_{i\in N} q_i + \sum_{j\in V} r_j + \sum_{j\in V}\sum_{I\subseteq\Gamma(j)} (1-(1-|I|/n)^T)p_{Ij},$$

is the objective (2a). Furthermore, for any $t \in [T]$, $i \in N$, $j \in \Gamma(i)$, and $S \subseteq V\backslash j$, we have

$$\begin{aligned}
&\sum_{l\in S\cup j}\left(\sum_{\substack{I\subseteq\Gamma(l),\\ I\ni i}} p_{Il} + \sum_{\substack{I\subseteq\Gamma(l),\\ I\not\ni i}} \left(1-(1-|I|/n)^{t-1}\right)p_{Il}\right)\\
&\qquad\qquad - \mathbb{E}_\eta\left[\sum_{l\in S}\left(\sum_{\substack{I\subseteq\Gamma(l),\\ I\ni\eta}} p_{Il} + \sum_{\substack{I\subseteq\Gamma(l),\\ I\not\ni\eta}} \left(1-(1-|I|/n)^{t-2}\right)p_{Il}\right)\right]\\
&= \sum_{\substack{I\subseteq\Gamma(j),\\ I\ni i}} p_{Ij} + \sum_{\substack{I\subseteq\Gamma(l),\\ I\not\ni i}} \left(1-(1-|I|/n)^{t-1}\right)p_{Ij} + \sum_{l\in S}\sum_{\substack{I\subseteq\Gamma(l),\\ I\ni i}} (1-|I|/n)^{t-1}p_{Il}\\
&\geq \sum_{\substack{I\subseteq\Gamma(j),\\ I\ni i}} p_{Ij},
\end{aligned}$$

with equality holding when $t = 1$ and $S = \varnothing$. It follows that the restriction of the feasible region of (2) with this approximation yields the model $(\mathcal{D}_3)$, the dual of (5) with the additional constraints (7). $\square$

Let $(q^{(12)}, r^{(12)}, p^{(12)})$ be optimal for $(\mathcal{D}_3)$. At state (t, i, S) with $S \cap \Gamma(i) \neq \varnothing$, using the value function approximation (12) we obtain the optimization problem

$$\begin{aligned}
&\underset{j\in S\cap\Gamma(i)}{\arg\max}\sum_{\ell\in S\backslash j}\left(r_\ell^{(12)} + \sum_{I\subseteq\Gamma(\ell)} (1-(1-|I|/n)^{t-1})p_{I\ell}^{(12)}\right)\\
&= \underset{j\in S\cap\Gamma(i)}{\arg\min}\left\{r_j^{(12)} + \sum_{I\subseteq\Gamma(j)} (1-(1-|I|/n)^{t-1})p_{Ij}^{(12)}\right\}.
\end{aligned}$$

The proof of Theorem 4.4 does not apply here, because the coefficients multiplying the p_{Ij} values may decay at different rates with respect to t, depending on the cardinality of $|I|$. Nevertheless, the policy is a *time-dependent* ranking policy: At any epoch t, the policy's ranking is given by a linear combination of the $r_j^{(12)}$ and $p_{Ij}^{(12)}$ values; the influence of the p values in the ranking is highest in the first epoch, and decays until vanishing in the last one. As with a (static) ranking policy, we can pre-compute the T rankings and implement the policy efficiently.

We can state a similar correspondence between constraints (8) and a value function approximation.

Theorem 4.6. *Consider the value function approximation*

$$\begin{aligned}
v_t(i,S) \approx\ & q_i + (t-1)\mathbb{E}_\eta[q_\eta] + \sum_{j\in S} r_j\\
&+ \sum_{J\subseteq\Gamma(i)} p_{iJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)+1\}]
\end{aligned}$$

$$+ \sum_{k \in N \setminus i} \sum_{J \subseteq \Gamma(k)} p_{kJ} \mathbb{E}[\min\{|J \cap S|, B(t-1, 1/n)\}], \tag{13}$$

where $q \in \mathbb{R}_+^N$, $r \in \mathbb{R}_+^V$, $p_{iJ} \in \mathbb{R}_+$ *for* $i \in N$ *and* $J \subseteq \Gamma(i)$. *Restricting the feasible region of* $\mathcal{D}$ *with this approximation yields*

$$\begin{aligned} \min\ & \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j + \sum_{i \in N} \sum_{J \subseteq \Gamma(i)} p_{iJ} \mathbb{E}[\min\{|J|, B(T, 1/n)\}] \\ \text{s.t. } & q_i + r_j + \sum_{\substack{J \subseteq \Gamma(i), \\ J \ni j}} p_{iJ} \geq 1,\ (i,j) \in E, \\ & q, r, p \geq 0, \end{aligned} \tag{$\mathcal{D}_4$}$$

the dual of the LP obtained by adding constraints (8) *to* (5).

Proof. First, we have the following expectation

$$\begin{aligned} \mathbb{E}_\eta \Big[& \sum_{J \subseteq \Gamma(\eta)} p_{\eta J} \mathbb{E}[\min\{|J \cap S|, B(t-1, 1/n) + 1\}] \\ & \qquad + \sum_{k \in N \setminus \eta} \sum_{J \subseteq \Gamma(k)} p_{kJ} \mathbb{E}[\min\{|J \cap S|, B(t-1, 1/n)\}] \Big] \\ & \qquad = \sum_{i \in N} \sum_{J \subseteq \Gamma(i)} p_{iJ} \mathbb{E}[\min\{|J \cap S|, B(t, 1/n)\}], \end{aligned}$$

which immediately shows that for $t = T$ the term

$$\mathbb{E}_\eta[v_T(\eta, V)] = \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j + \sum_{i \in N} \sum_{J \subseteq \Gamma(i)} p_{iJ} \mathbb{E}[\min\{|J|, B(T, 1/n)\}]$$

is the objective (2a) by taking $t = T$ and $S = V$. Furthermore, for any $t \in [T]$, $i \in N$, $j \in \Gamma(i)$, and $S \subseteq V \setminus j$, we have

$$\begin{aligned} v_t(i, S \cup j) = \ldots + & \sum_{J \subseteq \Gamma(i)} p_{iJ} \mathbb{E}[\min\{|J \cap (S \cup j)|, B(t-1, 1/n) + 1\}] \\ + \sum_{k \in N \setminus i} & \sum_{J \subseteq \Gamma(k)} p_{kJ} \mathbb{E}[\min\{|J \cap (S \cup j)|, B(t-1, 1/n)\}]. \end{aligned}$$

First, notice that if $J \ni j$ then $|J \cap (S \cup j)| = 1 + |J \cap S|$, otherwise if $J \not\ni j$ then $|J \cap (S \cup j)| = |J \cap S|$, so

$$\begin{aligned} \sum_{J \subseteq \Gamma(i)} p_{iJ} \mathbb{E}[\min\{|J \cap (S \cup j)|, & B(t-1, 1/n) + 1\}] \\ & = \sum_{\substack{J \subseteq \Gamma(i), \\ J \ni j}} p_{iJ} \mathbb{E}[\min\{1 + |J \cap S|, B(t-1, 1/n) + 1\}] \\ & + \sum_{\substack{J \subseteq \Gamma(i), \\ J \not\ni j}} p_{iJ} \mathbb{E}[\min\{|J \cap S|, B(t-1, 1/n) + 1\}]. \end{aligned}$$

In the same way we have

$$\sum_{k\in N\setminus i}\sum_{J\subseteq\Gamma(k)} p_{kJ}\mathbb{E}[\min\{|J\cap(S\cup j)|, B(t-1,1/n)\}]$$
$$=\sum_{k\in N\setminus i}\sum_{\substack{J\subseteq\Gamma(k),\\ J\ni j}} p_{kJ}\mathbb{E}[\min\{1+|J\cap S|, B(t-1,1/n)\}]$$
$$+\sum_{k\in N\setminus i}\sum_{\substack{J\subseteq\Gamma(k),\\ J\not\ni j}} p_{kJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}].$$

Hence, for $v_t(i,S\cup j)$ we obtain

$$v_t(i,S\cup j)=\ldots+\sum_{\substack{J\subseteq\Gamma(i),\\ J\ni j}} p_{iJ}\mathbb{E}[\min\{1+|J\cap S|, B(t-1,1/n)+1\}] \tag{A}$$
$$+\sum_{\substack{J\subseteq\Gamma(i),\\ J\not\ni j}} p_{iJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)+1\}] \tag{B}$$
$$+\sum_{k\in N\setminus i}\sum_{\substack{J\subseteq\Gamma(k),\\ J\ni j}} p_{kJ}\mathbb{E}[\min\{1+|J\cap S|, B(t-1,1/n)\}] \tag{C}$$
$$+\sum_{k\in N\setminus i}\sum_{\substack{J\subseteq\Gamma(k),\\ J\not\ni j}} p_{kJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}], \tag{D}$$

and from previous observations we know

$$\mathbb{E}_\eta\left[v_{t-1}(\eta,S)\right]=\sum_{i\in N}\sum_{J\subseteq\Gamma(i)} p_{iJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}]$$
$$=\sum_{\substack{J\subseteq\Gamma(i)\\ J\ni j}} p_{iJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}] \tag{A$'$}$$
$$+\sum_{\substack{J\subseteq\Gamma(i)\\ J\not\ni j}} p_{iJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}] \tag{B$'$}$$
$$+\sum_{k\in N\setminus i}\sum_{\substack{J\subseteq\Gamma(k)\\ J\ni j}} p_{kJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}] \tag{C$'$}$$
$$+\sum_{k\in N\setminus i}\sum_{\substack{J\subseteq\Gamma(k)\\ J\not\ni j}} p_{kJ}\mathbb{E}[\min\{|J\cap S|, B(t-1,1/n)\}]. \tag{D$'$}$$

First, we clearly have $(D)-(D')=0$, and $(B)-(B'),(C)-(C')\geq 0$. Also, we obtain

$$(A)-(A')=\sum_{\substack{J\subseteq\Gamma(i),\\ J\ni j}} p_{iJ}.$$

In other words we have

$$\sum_{J\subseteq\Gamma(i)} p_{iJ}\mathbb{E}[\min\{|J\cap(S\cup j)|, B(t-1,1/n)+1\}]$$

$$
\begin{aligned}
&\qquad + \sum_{k \in N \setminus i} \sum_{J \subseteq \Gamma(k)} p_{kJ} \mathbb{E}[\min\{|J \cap (S \cup j)|, B(t-1, 1/n)\}] \\
&\qquad - \sum_{i \in N} \sum_{J \subseteq \Gamma(i)} p_{iJ} \mathbb{E}[\min\{|J \cap S|, B(t-1, 1/n)\}] \\
\geq \quad & \sum_{\substack{J \subseteq \Gamma(i), \\ J \ni j}} p_{iJ},
\end{aligned}
$$

with equality holding when $t = 1$ and $S = \varnothing$. So, it follows that

$$
v_t(i, S \cup j) - \mathbb{E}_\eta[v_{t-1}(\eta, S)] \geq q_i + r_j + \sum_{\substack{J \subseteq \Gamma(i), \\ J \ni j}} p_{iJ} \geq 1.
$$

Therefore, the restriction of the feasible region of (2) with this approximation yields ($\mathcal{D}_4$), the dual of (5) with the additional constraints (8). □

This approximation generalizes the intuition behind approximation (11) to subsets of ads. Suppose we model the value p_{iJ} of impression i interacting with a set of compatible ads $J \subseteq \Gamma(i)$. When i appears in epoch t, we expect no more than $\mathbb{E}[\min\{|J \cap S|, B(t-1, 1/n) + 1\}]$ matches between i and J from that point forward: The number of matches cannot exceed the number of remaining ads in the set, $|J \cap S|$, but it also cannot exceed the number of times we expect i to appear, the current appearance plus $B(t-1, 1/n)$ more. A similar argument applies to any impression that did not appear in this epoch.

This value function approximation (13) seems not to define a dynamic ranking policy; let $(q^{(13)}, r^{(13)}, p^{(13)})$ be optimal for ($\mathcal{D}_4$). At state (t, i, S) with $S \cap \Gamma(i) \neq \varnothing$, employing the approximation (13) in (1) results in

$$
\arg\max_{j \in S \cap \Gamma(i)} \left\{ \sum_{\ell \in S \setminus j} r_\ell^{(13)} + \sum_{i \in N} \sum_{J \subseteq \Gamma(i)} p_{iJ}^{(13)} \mathbb{E}[\min\{|J \cap (S \setminus j)|, B(t-1, 1/n)\}] \right\}.
$$

Because the coefficients multiplying the p variables depend explicitly on the set S of remaining ads, it is impossible to compute these expressions a priori to obtain a ranking. We have nevertheless also implemented this policy in our computational experiments, outlined in the next section.

Finally, we can combine and generalize our previous approximations to derive a correspondence to (9).

Theorem 4.7. *Consider the value function approximation*

$$
\begin{aligned}
v_t(i, S) \approx \; & q_i + (t-1)\mathbb{E}_\eta[q_\eta] + \sum_{j \in S} r_j \\
& + \sum_{\substack{I \subseteq N \\ I \ni i}} \sum_{J \subseteq V} p_{IJ} \mathbb{E}[\min\{|J \cap S|, B(t-1, |I|/n) + 1\}] \\
& + \sum_{\substack{I \subseteq N \\ I \not\ni i}} \sum_{J \subseteq V} p_{IJ} \mathbb{E}[\min\{|J \cap S|, B(t-1, |I|/n)\}],
\end{aligned}
\tag{14}
$$

where $q \in \mathbb{R}_+^N$, $r \in \mathbb{R}_+^V$, $p_{IJ} \in \mathbb{R}_+$ for $I \subseteq N$ and $J \subseteq V$. Restricting the feasible region of $\mathcal{D}$ with this approximation yields

$$
\min \; \frac{T}{n} \sum_{i \in N} q_i + \sum_{j \in V} r_j + \sum_{I \subseteq N} \sum_{J \subseteq V} p_{IJ} \mathbb{E}[\min\{|J|, B(T, |I|/n)\}]
$$

$$\text{s.t. } q_i + r_j + \sum_{\substack{I \subseteq N, \\ I \ni i}} \sum_{\substack{J \subseteq V, \\ J \ni j}} p_{iJ} \geq 1, \ (i,j) \in E, \tag{$\mathcal{D}_5$}$$

$$q, r, p \geq 0,$$

the dual of (5) *with the additional constraints* (9).

The proof of Theorem 4.7 is analogous to Theorem 4.6, but also uses parts of the proof of Theorem 4.5. Finally, this approximation defines a policy similar to the one given by (13).

5 Computational Results

In this section we outline the experiments we conducted to test the proposed bounds and policies. All of the test instances we constructed have $T = n = m$, and consist of the following:

1. A cycle of size 20 ($n = 10$).
2. A cycle of size 200 ($n = 100$).
3. 20 small instances with $n = 10$, each one randomly generated by having a possible edge in $N \times V$ be present independently with a probability of 10%.
4. 20 large dense instances with $n = 100$, each one randomly generated by having a possible edge in $N \times V$ be present independently with a probability of 10%.
5. 20 large sparse instances with $n = 100$, each one randomly generated by having a possible edge in $N \times V$ be present independently with probability of 2.5%.

We use "small" instances with $n = 10$ because the exact DP recursion (1) is still computationally tractable, yet it becomes intractable for even slightly bigger instances. Furthermore, the dimension of the polyhedron Q grows as $O(n^2)$, and all of our inequality classes are separable in polynomial time, except for (9). We can solve the LP's for "large" instances with $n = 100$ in a few seconds, but as n grows the dimension of the problem itself becomes the bottleneck. For example, for our large dense instances, the expected number of variables is $1,000$; a similarly constructed instance with $n = 500$ would already have $25,000$ variables in expectation.

We tested various bounds on the instances by solving the initial relaxation (5) and then adding the inequalities we introduced in Section 3. For the policies, we simulated 20,000 realizations of the small instances and 200 realizations of the large instances, and we report the sample average of each policy. To benchmark our results, for the small instances we computed the optimal expected value given by the DP (1), and for the larger instances we calculated the sample mean of the maximum expected off-line matching, by computing the maximum cardinality matching of each simulated realization; this yields an upper bound on any policy as it affords the decision maker early access to information. As policy comparisons, we implemented two heuristics: The single-matching policy computes a maximum cardinality matching in $(N \times V, E)$, and matches only these edges, ignoring all others; this heuristic has an approximation ratio of $1 - (1 - 1/n)^T$ (approximately $1 - 1/e$ when $T = n$) [9]. The two-matching policy is a heuristic modification of the algorithm from [9] that uses a maximum cardinality 2-matching in $(N \times V, E)$.

Table 1 summarizes the results. For each instance class, in each row we present the geometric mean of each bound or policy's ratio to the best available benchmark (the DP value for small instances and the expected maximum matching for large ones). We also report the sample standard deviation of the ratios in parenthesis.

Bound/Policy	20-Cycle	200-Cycle	Small	Large Dense	Large Sparse
(5)	1.2681	1.2659	1.3151 (0.081)	1.0018 (0.0011)	1.2167 (0.010)
(5) + (6)	1.2681	1.2659	1.0886 (0.042)	1.0018 (0.0011)	1.1227 (0.011)
(5) + (7)	1.1319	1.0980	1.0536 (0.038)	1.0006 (0.0006)	1.0794 (0.009)
(5) + (8)	1.1606	1.1370	1.0570 (0.031)	1.0013 (0.0009)	1.0961 (0.011)
(5) + (7) + (8)	1.1319	1.0980	1.0536 (0.038)	1.0004 (0.0005)	1.0717 (0.009)
(5) + (9)	1.1319	1.0980	1.0288 (0.023)	-	-
Exp. Matching	1.0514	1	1.0030 (0.005)	1	1
(1)	1	-	1	-	-
Matching	0.8259	0.8025	0.8566 (0.053)	0.6351 (0.0007)	0.7768 (0.028)
2-Matching	0.9859	0.9496	0.9616 (0.037)	0.7500 (0.0023)	0.8793 (0.030)
(10)	0.9861	0.9474	0.9883 (0.020)	0.9232 (0.0048)	0.9306 (0.008)
(11)	0.9861	0.9474	0.9963 (0.005)	0.9232 (0.0048)	0.9358 (0.008)
(12)	0.9861	0.9474	0.9974 (0.005)	0.9471 (0.0091)	0.9560 (0.006)
(13)	0.9980	0.9539	0.9732 (0.032)	0.9409 (0.0036)	0.8635 (0.070)

Table 1: Summary of experiment results.

With respect to the bounds, the right-star inequalities (7) improve the basic bound (5) more than the left-star ones (8), even though the latter are facet-defining. For small instances, the complete subgraph inequalities (9) can further cut the gap to under 3%; however, we weren't able to compute this bound in a reasonable time for larger instances because of the significant additional computational burden. For the dense large instances, our bounds are all quite close to the expected maximum matching benchmark, which is unsurprising since in most realizations of these instances there is a perfect or near-perfect matching.

In terms of policies, the best performer overall is the time-dependent ranking policy corresponding to the right-star inequalities and approximation (12). However, the non-ranking policy corresponding to the left-star inequalities and approximation (13) does perform better on the two cycle instances. In contrast, the single-matching heuristic policy does not perform well, and even the 2-matching heuristic's performance significantly worsens for the large instances; this may indicate the benefit of having more than two choices per impression in larger graphs.

The next set of experiments generalizes the previous cycle instances to k-regular graphs constructed in the following way: We labeled impressions and ads from 0 to $n-1$, and for each $i \in \{0, \dots n-1\}$ we set $\Gamma(i) = \{i, i+1, \dots, i+k-1\} (\bmod n)$. We present the results in Table 2 for bipartite graphs with $n = m = 100$ and $k = 3, 4, 5, 6$. We compare bounds and policies with respect to the sample mean of the maximum expected off-line matching. As with the previous large instances, we again did not compute the exact optimal solution or the last bound corresponding to (5) + (9).

In terms of bounds, we again observe that inequality (7) improves the basic bound (5) more than the other classes of inequalities, including (8). Nonetheless, as k increases, both bounds (right-star and left-star) are much tighter. Also, we observe that (6) does not improve the bound from (5). This is expected; when $n = T$, in $(\mathcal{D}_2)$ we will have $p_{ij}^{(11)} = 0$ for all edges incident to nodes of degree 2 or greater (all nodes in these instances), since their objective coefficient satisfies $1 - (1 - 1/n)^n \geq 1 - 1/e$.

Regarding the policies, (10), (11) and (12) perform exactly in the same way. This is unsurprising, because the graphs' symmetry means none of the policies can differentiate among the ads. On the other hand, the left-star policy (13) does better than the rest of the policies, as we observed also in

Bound/Policy	3-regular	4-regular	5-regular	6-regular
(5)	1.1687	1.1202	1.0951	1.0696
(5) + (6)	1.1687	1.1202	1.0951	1.0696
(5) + (7)	1.1131	1.1013	1.0886	1.0674
(5) + (8)	1.1424	1.1157	1.0944	1.0695
(5) + (7) + (8)	1.1131	1.1013	1.0886	1.0674
Exp. Matching	1	1	1	1
Matching	0.7409	0.7102	0.6943	0.6781
2-Matching	0.8630	0.8343	0.8129	0.8030
(10)	0.9347	0.9303	0.9295	0.9254
(11)	0.9347	0.9303	0.9295	0.9254
(12)	0.9347	0.9303	0.9295	0.9254
(13)	0.9429	0.9396	0.9426	0.9454

Table 2: Ratios for k-regular bipartite graphs.

the cycle instances. Once again, the single-matching and 2-matching policies perform much worse than the ranking policies.

6 Conclusions

We have studied relaxations for the i.i.d. online bipartite matching problem, by deriving several classes of valid inequalities for the polyhedron of attainable probabilities. We have also determined which of these inequalities are facet-defining, and used them to design heuristic policies, many of which turn out to be of ranking type.

Our results motivate a variety of questions for future work. In particular, we mention at the end of Section 3 that, even in very small instances, all facets not included in our study exhibit a more complicated structure. For example, they cannot be written as 0-1 inequalities. More polyhedral results are still needed, and they may require the study of sub-structures in the underlying graph or may need to include the timing of policies' choices; our preliminary results in this vein reveal quite complex inequalities.

A more general question is to apply methods like the ones in this paper to other online matching and resource allocation problems. One prominent example is the AdWords problem [10, 18], but there are several other models in advertising and revenue management that may benefit from such an approach.

References

[1] D. Adelman, *Price-Directed Replenishment of Subsets: Methodology and its Application to Inventory Routing*, Manufacturing and Service Operations Management **5** (2003), 348–371.

[2] ———, *A Price-Directed Approach to Stochastic Inventory/Routing*, Operations Research **52** (2004), 499–514.

[3] D. Adelman and C. Barz, *A Unifying Approximate Dynamic Programming Model for the Economic Lot Scheduling Problem*, Mathematics of Operations Research **39** (2014), 374–402.

[4] B. Bahmani and M. Kapralov, *Improved Bounds for Online Stochastic Matching*, Proceedings of the 18th Annual European Conference on Algorithms: Part I, Springer-Verlag, 2010, pp. 170–181.

[5] D. Bertsimas and J. Niño-Mora, *Conservation Laws, Extended Polymatroids and Multi-Armed Bandit Problems; A Polyhedral Approach to Indexable Systems*, Mathematics of Operations Research **21** (1996), 257–306.

[6] B. Birnbaum and C. Mathieu, *On-Line Bipartite Matching Made Simple*, ACM SIGACT News **39** (2008), 80–87.

[7] E.G. Coffman, Jr. and I. Mitrani, *A Characterization of Waiting Time Performance Realizable by Single-Server Queues*, Operations Research **28** (1980), 810–821.

[8] D.P. de Farias and B. van Roy, *The Linear Programming Approach to Approximate Dynamic Programming*, Operations Research **51** (2003), 850–865.

[9] J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan, *Online Stochastic Matching: Beating* $1 - 1/e$, Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE, 2009, pp. 117–126.

[10] G. Goel and A. Mehta, *Online Budgeted Matching in Random Input Models with Applications to Adwords*, Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '08, SIAM, 2008, pp. 982–991.

[11] B. Haeupler, V.S. Mirrokni, and M. Zadimoghaddam, *Online Stochastic Weighted Matching: Improved Approximation Algorithms*, Proceedings of the 7th Workshop on Internet and Network Economics (WINE) (Berlin-Heidelberg) (N. Chen, E. Elkind, and E. Koutsoupias, eds.), Springer-Verlag, 2011, pp. 170–181.

[12] P. Jaillet and X. Lu, *Online Stochastic Matching: New Algorithms with Better Bounds*, Mathematics of Operations Research **39** (2014), 624–646.

[13] R.M. Karp, U.V. Vazirani, and V.V. Vazirani, *An Optimal Algorithm for On-line Bipartite Matching*, Proceedings of the 22nd Annual ACM Symposium on the Theory of Computing (STOC) (New York), ACM, 1990, pp. 352–358.

[14] I. Kra and S.R. Simanca, *On circulant matrices*, Notices of the AMS **59** (2012), 368–377.

[15] M. Mahdian and Q. Yan, *Online Bipartite Matching with Random Arrivals: An Approach Based on Strongly Factor-Revealing LPs*, Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing (STOC) (New York, NY, USA), ACM, 2011, pp. 597–606.

[16] V.H. Manshadi, S. Oveis Gharan, and A. Saberi, *Online Stochastic Matching: Online Actions Based on Offline Statistics*, Mathematics of Operations Research **37** (2012), 559–573.

[17] A. Mehta, *Online matching and ad allocation*, Foundations and Trends in Theoretical Computer Science **8** (2013), 265–368.

[18] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani, *Adwords and generalized online matching*, Journal of the ACM **54** (2007), 22:1–22:19.

[19] P.J. Schweitzer and A. Seidmann, *Generalized Polynomial Approximations in Markovian Decision Processes*, Journal of Mathematical Analysis and Applications **110** (1985), 568–582.

[20] A. Toriello, *Optimal Toll Design: A Lower Bound Framework for the Asymmetric Traveling Salesman Problem*, Mathematical Programming **144** (2014), 247–264.

[21] A. Toriello, W.B. Haskell, and M. Poremba, *A Dynamic Traveling Salesman Problem with Stochastic Arc Costs*, Operations Research **62** (2014), 1107–1125.

[22] M.A. Trick and S.E. Zin, *Spline Approximations to Value Functions: A Linear Programming Approach*, Macroeconomic Dynamics **1** (1997), 255–277.