Online Stochastic Matching: New Algorithms with Better Bounds*

Patrick Jaillet‡ Xin Lu§

May 2012; revised January 2013, June 2013

Abstract

We consider variants of the online stochastic bipartite matching problem motivated by Internet advertising display applications, as introduced in Feldman et al. [6]. In this setting, advertisers express specific interests into requests for impressions of different types. Advertisers are fixed and known in advance while requests for impressions come online. The task is to assign each request to an interested advertiser (or to discard it) immediately upon its arrival.

In the adversarial online model, the ranking algorithm of Karp et al. [11] provides a best possible randomized algorithm with competitive ratio $1-1/e \approx 0.632$. In the stochastic i.i.d. model, when requests are drawn repeatedly and independently from a known probability distribution over the different impression types, Feldman et al. [6] prove that one can do better than $1-1/e$. Under the restriction that the expected number of request of each impression type is an integer, they provide a 0.670-competitive algorithm, later improved by Bahmani and Kapralov [3] to 0.699, and by Manshadi et al. [13] to 0.705. Without this integrality restriction, Manshadi et al. [13] are able to provide a 0.702-competitive algorithm.

In this paper we consider a general class of online algorithms for the i.i.d. model which improve on all these bounds and which use computationally efficient offline procedures (based on the solution of simple linear programs of maximum flow types). Under the integrality restriction on the expected number of impression types, we get a $1-2e^{-2}(\approx 0.729)$-competitive algorithm. Without this restriction, we get a 0.706-competitive algorithm.

Our techniques can also be applied to other related problems such as the online stochastic vertex-weighted bipartite matching problem as defined in Aggarwal et al. [1]. For this problem, we obtain a 0.725-competitive algorithm under the stochastic i.i.d. model with integral arrival rate.

Finally we show the validity of all our results under a Poisson arrival model, removing the need to assume that the total number of arrivals is fixed and known in advance, as is required for the analysis of the stochastic i.i.d. models described above.

1 Introduction

Bipartite matching problems have been studied extensively in the operations research and computer science literature. We consider in this paper variants of the online stochastic bipartite matching problem motivated by Internet advertising display applications, as introduced in Feldman et al. [6].

We are given a bipartite graph $G = \{A \cup I, E\}$, where A is a set of advertisers and I is a set of impression types. An edge $(a, i) \in E$ if and only if advertiser $a \in A$ is interested in impressions of

*To appear in *Mathematics of Operations Research.* Research funded in part by NSF (grant 1029603) and ONR (grants N00014-09-1-0326 and N00014-12-1-0033).

‡Department of Electrical Engineering and Computer Science, Laboratory for Information and Decision Systems, Operations Research Center, MIT, Cambridge, MA, 02139; jaillet@mit.edu

§Operations Research Center, MIT, Cambridge, MA, 02139; luxin@mit.edu

type $i \in I$. The set of advertisers and their preferences are fixed and known in advance. Requests for impression come online one at a time at periods $t = 1, 2, \cdots, n$ (n being fixed and known in advance), and the impression type of each request is chosen randomly and independently from a given probability distribution over the set I.

Upon the arrival of a request, an online algorithm must irrevocably assign it to one of the interested advertisers or drop it. Overall, every request cannot be assigned to more than one advertiser, and every advertiser can be assigned at most one request. The goal is to maximize the expected number of assigned requests over the random sequence of impressions.

Given that there is a total of n requests, the probability that a request is for an impression of type i can be written as r_i/n, where r_i is the expected number of requests of type i among the random sequence. Without loss of generality, we assume that $r_i \leq 1$ for all type i; note that if a type i were to be such that $r_i > 1$, we could duplicate node in $i \in I$ into a set of identical nodes, each with the same adjacent edge structure as the original node, and each with expected number of arrival no more than one.

In this paper, we consider two variants of this online stochastic i.i.d. model: a special case for which $r_i = 1$ for all i, which we refer to as the case with integral arrival rates; and the unrestricted case with general arrival rates.

We also consider a Poisson arrival model, removing the need to assume that the total number of arrivals is fixed and known in advance, as is required for the analysis of the stochastic i.i.d. models.

1.1 Our results and techniques

In Feldman et al. [6], the authors provide a 0.670-competitive algorithm for the online stochastic bipartite matching problem with integral arrival rates, the first result to show that stochastic information on request arrivals could provably improve upon the $1 - 1/e$ competitive ratio of Karp et. al. [11]. Removing this integrality restriction, Manshadi et al. [13] show it is still possible to do better than $1 - 1/e$ and propose a 0.702-competitive algorithm, using offline statistics drawn from Monte Carlo sampling. The authors further prove that the algorithm has a better competitive ratio of 0.705 when the arrival rates are integral. More recently, Mahdian and Yan [12] and Karande et al. [10] study a much less restrictive version of the problem where not only the arrival rates are arbitrary, they are not known to the algorithm a priori.

In this paper we consider a general class of online algorithms for the i.i.d. model which improve on [6][13] and which use computationally efficient offline procedures (based on the solution of simple linear programs of maximum flow types). Under the integrality restriction on the expected number of impressions of each types, we get a $(1 - 2e^{-2})$-competitive algorithm. Without this restriction, we get a 0.706-competitive algorithm. Although the model we consider is more restrictive than the one in [12][10], we obtain better competitive ratio.

Our techniques can be applied to other related problems such as the online stochastic b-matching problem (quite trivially) and to the online stochastic vertex-weighted bipartite matching problem as defined in Aggarwal et al. [1]. For that problem, we obtain a 0.725-competitive algorithm under the stochastic i.i.d. model with integral arrival rates. Our vertex-weighted model is a special case of the edge-weighted model considered by Haeupler et al. [8], who propose a 0.667-competitive algorithm for the edge-weighted case. Finally we show the validity of all our results under a Poisson arrival model, removing the need to assume that the total number of arrivals is fixed and known in advance, as is required for the analysis of the stochastic i.i.d. models.

In order to introduce the main general ideas behind our techniques, let us first define some basic concepts about optimal offline solutions. From the available information about the problem (the initial graph $G = \{A \cup I, E\}$, the probability distribution over the set of impression types I, and the number n of i.i.d. draws from that distribution), one can solve an optimal maximum cardinality matching for each possible realization of the n i.i.d. draws. Let OPT be the random variable corresponding to the values obtained by these offline matchings. The expected value $\mathbb{E}[\mathrm{OPT}]$ can be written as $\sum_{e \in E} f_e^*$, where f_e^* is the probability that edge e is part of an optimal solution in a given realization. Note that, as in [13], $\mathbf{f}^* = (f_e^*)_{e \in E}$ can also be defined as a so-called optimal offline fractional matching.

Instead of computing $\mathbf{f}^*$ (or estimating it as in [13]) and using the information for guiding online strategies, our strategy is to formulate special maximum flow problems whose optimal solutions provide the input for the design of good online algorithms. Moreover these maximum flow problems are defined in such a way that $\mathbf{f}^*$ corresponds to feasible flows, allowing us to derive upper bounds on $\sum_{e \in E} f_e^*$, and get valid bounds for the competitive ratios of the related online algorithms.

We now provide more details. Consider an instance of a single-source single-destination node-capacitated maximum flow problem on G, with a source s connected to all elements of A, a destination t connected to all elements of I, a unit capacity on all $a \in A$, and a capacity r_i (expected number of arrivals) on each $i \in I$. Define $f_e = f_{a,i}$ to be the flow on $e = (a, i)$ for all $e \in E$. This problem can equivalently be formulated as a linear program (LP):

$$\begin{array}{lll} \max & \sum\limits_{e} f_e & \\ & \sum\limits_{i \sim a} f_{a,i} \leq 1 & \forall a \in A \\ & \sum\limits_{a \sim i} f_{a,i} \leq r_i & \forall i \in I \\ & f_e \geq 0 & \forall e \in E \end{array} \tag{1}$$

where $i \sim a$ and $a \sim i$ are shortcuts for $i : (a, i) \in E$ and $a : (a, i) \in E$, respectively.

One of the key steps behind our approach is to find appropriate additional constraints on the flows (to add to (1)) so that the resulting optimal solutions of the constrained LP lead to improved guidance for online strategies, while keeping the optimal offline fractional matching $\mathbf{f}^*$ feasible with respect to the constrained LP.

Let us now formally introduce the concept of a "list of interested advertisers for an impression of type i". Consider the set $A_i = \{a \in A : (a, i) \in E\}$ and let Ω_i be the set of all possible non-empty ordered subsets of elements of A_i. An element of Ω_i will be called a list of interested advertisers for impression of type i. We are ready to describe our class of online algorithms:

Random Lists Algorithms (RLA)

1. Add appropriate constraints to (1) to get a new constrained LP. Let $\mathbf{f}$ be an optimal solution to this LP.
2. Using $\mathbf{f}$, construct a probability distribution $\mathcal{D}_i$ over the set Ω_i for each impression type i.
3. When a request for an impression of type i arrives, select a list from Ω_i using the probability distribution $\mathcal{D}_i$:
   - if all the advertisers in the list are already matched, then drop the request;
   - otherwise, assign the request to the first unmatched advertiser in the list.

Steps 1 and 2 are problem-specific. Different solutions $\mathbf{f}$ and different construction of distributions $\mathcal{D}_i$ will lead to online algorithms that may have different properties and competitive ratios. However, these algorithms all share one common and important property: with high probability, they are robust with respect to different realizations of the n i.i.d. sequence of impression types. This property will be useful for the rigorous analysis of competitive ratios. Random lists used in this paper are extensions of ideas given in [13], but unlike [13], where the length of lists is at most 2, we consider lists of length 3 in this paper.

Paper outline: In the remainder of this section we provide an overview of related work. In Section 2 we justify the choice of looking at ratios of expected values for evaluating our class of online algorithms under the i.i.d. stochastic model. The first two main sections follow: In Section 3 we analyze the online stochastic bipartite matching problem under integral arrival rates, and in Section 4 we extend the results to the online stochastic vertex-weighted bipartite matching problem. The next major result is contained in Section 5 where we consider the online stochastic bipartite matching problem under general arrival rate. Finally we show in Section 6 the validity of our results under a Poisson arrival model, removing the need to assume that the total number of arrivals is fixed and known in advance, as is required for the analysis of the stochastic i.i.d. models. We conclude with some final remarks and open problems.

1.2 Related work

As indicated above, bipartite matching problems and related advertisement allocation problems have been studied extensively in the operations research and computer science literature.

Under an adversarial online model where no information is known about requests, Karp et al. [11] look at the bipartite matching problem and give a best possible randomized algorithm (ranking) with competitive ratio $1-1/e$. Kalyanasundaram and Pruhs [9] give a $1-1/e$-competitive algorithm for b-matching problems. Mehta et al. [15, 16] and Buchbinder et al. [4] propose two different $1-1/e$ competitive algorithms for the AdWords problem. More recently, Aggarwal et al. [1] give a $1-1/e$-competitive algorithm for the vertex-weighted bipartite matching problem.

However, adversarial models may be too conservative for some applications where worst-case scenarios are unlikely to happen. Less conservative models have been proposed. In the random permutation model, when the set of requests is unknown, but the order of the sequence is random, Goel and Mehta [7] show that a greedy algorithm is $1-1/e$ competitive. Devanur and Hayes [5] propose a near optimal algorithm for AdWords under some mild assumptions. Agrawal et al. [2] further propose a near optimal algorithm for general online linear programming problems using similar techniques. Mahdian and Yan [12] and Karande et al. [10] simultaneously show RANKING algorithm is 0.696-competitive for matching problem. Mirrokni et al. [17] propose an algorithm works well under both adversarial and random arrival model for Adwords.

The random permutation model may still be too conservative in practice, when statistics about requests may be available. In the stochastic i.i.d. model, when requests are drawn repeatedly and independently from a known probability distribution over the different impression types, Feldman et al. [6] prove that one can do better than $1-1/e$. Under the restriction that the expected number of request of each impression type is an integer, they provide a 0.670-competitive algorithm. They also show that no algorithm can achieve a competitive ratio of 0.989. Bahmani and Kapralov [3] modify the algorithm and give a competitive ratio of 0.699 under the same assumption. They also improved the upper bound to 0.902. More recently, Manshadi et al. [13] removed the assumption that the expected number of arrivals is integral, and present a 0.702-competitive algorithm (the same algorithm achieves a competitive ratio of 0.705 under the integral assumption). They also improve the upper bound to 0.86 with the integral assumption and 0.823 without the integral assumption. Finally Haeupler et al. [8] recently proposed a 0.667-competitive algorithm for the edge-weighted problem under the stochastic i.i.d. model.

2 Preliminary on competitive ratios

As a measure of performance, online algorithms are typically compared to optimum offline solutions using ratios. In this paper, an algorithm is called α-competitive if $\frac{\mathbb{E}[\mathrm{ALG}]}{\mathbb{E}[\mathrm{OPT}]} \geq \alpha$ for any given probability distributions. The goal is to find algorithms with large competitive ratios. One could use a stronger notion of competitive ratio that $\frac{\mathrm{ALG}}{\mathrm{OPT}} \geq \alpha$ would hold for most realizations as used in [6]. In this section we show that for the algorithms in our proposed class, the two concepts are in fact closely related and lead to competitive ratios that are valid under either of these measures.

Let $L_1, L_2, \cdots, L_n$ be the sequence of random lists for the n successive requests. Every list only contains interested advertisers, and the assignment of requests only depends on the order of advertisers in the list and their current status. Thus, from a given realization of this sequence of random lists, we can construct the corresponding matching and find its cardinality. We can show that the cardinality is stable with respect to the realization in the following sense:

Claim 1. *If two realizations $(l_1, \cdots, l_t, \cdots, l_n)$ and $(l_1, \cdots, l'_t, \cdots, l_n)$ only differ by one list, then the cardinality of their resulting matchings differs at most by one.*

Proof. Let W_j and W'_j be the set of matched advertisers right after the j^{th} arrival corresponding to the two realizations above, respectively. We will show by induction that $\forall j, |W_j \backslash W'_j| \leq 1$ and $|W'_j \backslash W_j| \leq 1$. For all $j \leq t-1$, since the two realizations are identical for the first j lists, $W_j = W'_j$. Since in every period, at most one advertiser becomes matched, the claim is also true for $j = t$. Let us consider $j \geq t+1$. If $W_j \backslash W_{j-1} \subset W'_{j-1}$, then by induction, $|W_j \backslash W'_j| \leq |W_{j-1} \backslash W'_{j-1}| \leq 1$. Otherwise, let $\{k\} = W_j \backslash W_{j-1}$. Then, in the list l_j, all advertisers in front of k are in W_{j-1}. Noting that k is unmatched for ALG$'$ before the j^{th} period, we have $W'_j \backslash W'_{j-1} \subset W_{j-1} \cup \{k\}$. Therefore, $|W_j \backslash W'_j| = |W_{j-1} \backslash W'_{j-1}| \leq 1$. Similarly, we can show $|W'_j \backslash W_j| \leq 1$. Hence, $|\mathrm{ALG} - \mathrm{ALG}'| \leq ||W_n| - |W'_n|| \leq \max\{|W_n \backslash W'_n|, |W'_n \backslash W_n|\} \leq 1$. □

Note that L_j only depends on the impression type of the j^{th} request, and does not depend on types and assignments of earlier impressions. Thus, $L_1, \cdots, L_n$ are independently and identically distributed. We can then apply McDiarmid's Inequality which we recall here for convenience:

McDiarmid's Inequality [14]: Let $X_1, X_2, \ldots, X_n$ be independent random variables all taking values in the set $\mathcal{X}$. Let $f : \mathcal{X}^n \mapsto \mathbb{R}$ be a function of $X_1, X_2, \ldots, X_n$ that satisfies $\forall i$, $\forall x_1, \ldots, x_n, x'_i \in \mathcal{X}$, $|f(x_1, \ldots, x_i, \ldots, , x_n) - f(x_1, \ldots, x'_i, \ldots, x_n)| \leq c_i$. Then $\forall \epsilon > 0$,

$$\mathbb{P}(f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] > \epsilon) \leq \exp(-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2})$$

and

$$\mathbb{P}(f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] < -\epsilon) \leq \exp(-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}).$$

Combining McDiarmid's Inequality with Claim 1 we obtain:

Lemma 1. $\mathbb{P}(\mathrm{ALG} - \mathbb{E}[\mathrm{ALG}] < -n\epsilon) \leq \exp(-2n\epsilon^2)$.

Similarly, note that the offline solution only depends on the realization of the impression types. So we can show a similar result:

Lemma 2. $\mathbb{P}(\mathrm{OPT} - \mathbb{E}[\mathrm{OPT}] > n\epsilon) \leq \exp(-2n\epsilon^2)$.

From the two lemmas above, we can conclude that

$$\mathbb{P}(\frac{\text{ALG}}{\text{OPT}} \geq \frac{\mathbb{E}[\text{ALG}]}{\mathbb{E}[\text{OPT}]} - \frac{2\epsilon}{c+\epsilon}) \geq 1 - 2\exp(-2n\epsilon^2),$$

where $c = \mathbb{E}[\text{OPT}]/n$. If $\mathbb{E}[\text{OPT}] = \Theta(n)$, the inequality above indicates that the two notions of competitive ratios are closely related and essentially equivalent as far as our results are concerned.

Throughout the paper we will assume that n is large enough so that a factor of $1 + O(1/n)$ is negligible when analyzing the performance of online algorithms.

3 Stochastic matching with integral arrival rates

In this section and the next, we assume that $r_i = 1$ for all i.

3.1 Online algorithm

As we mentioned in 1.1, two steps in RLA are problem-specific: finding offline solutions and constructing random lists. In this subsection, we propose methods for these two steps.

3.1.1 Offline solution

Let us consider the following maximum flow problem on the graph G:

$$\begin{array}{rll} \max & \sum_{a,i} f_{a,i} & \\ s.t. & \sum_{i\sim a} f_{a,i} \leq 1 & \forall a \in A \\ & \sum_{a\sim i} f_{a,i} \leq 1 & \forall i \in I \\ & f_e \in [0, 2/3] & \forall e \in E \end{array} \tag{2}$$

Note that compared to (1) as introduced in 1.1, the additional constraints on the flows are very simple. Since the set of vertices of the feasible polytope of (2) is a subset of $\{0, 1/3, 2/3\}^E$, there exists an optimal solution to (2) in $\{0, 1/3, 2/3\}^E$.

To ease the construction of random lists and the analysis, based on the optimal solution $\mathbf{f}$, we first construct a resulting graph $G_\mathbf{f} = \{A \cup I, E_\mathbf{f}\}$, where $E_\mathbf{f} = \{e \in E : f_e > 0\}$. For simplicity, we try to make $E_\mathbf{f}$ as sparse as we can by doing the following two types of transformations. As argued below, there exists an optimal solution $\mathbf{f}$ such that its resulting graph $G_\mathbf{f}$ does not contain cycles of length 4, unless the four nodes in such a cycle do not have any other neighbors; and such a solution can be found in polynomial time. Cycles in Figure 1 are the only three possible cycles of length 4. The four nodes in the left cycle cannot have any other neighbor outside the cycle; the middle and right cycle can be transformed into a non-cycle with the same objective value. Furthermore, if there exists impression i that has two neighbors a_1 and a_2 with $f_{i,a_1} = f_{i,a_2} = 1/3$ and $f_{a_1} + f_{a_2} < 2$, without loss of generality, we assume $f_{a_2} < 1$. Another solution $\mathbf{f}'$ with $f'_{i,a_1} = 0$, $f'_{i,a_2} = 2/3$, and everything else unchanged has the same objective value, and less edges in its resulting graph. We transform $\mathbf{f}$ to $\mathbf{f}'$. Note that each time, transformations mentioned above remove one or two edges and does not introduce any new edge. Thus, given any initial optimal solution, after at most $|E|$ transformations, the optimal solution cannot be transformed further in the above two ways.

The extra constraint $f_e \leq 2/3$ is added for two reasons: LP (2) provides a upper bound on the offline solution; the resulting graph is sparse. In fact, as showed in Section 3.2, any constraint $f_e \leq c$ with $c \geq 1 - e^{-1}$ provides an upper bound on the offline solution; however, only $c = 2/3$

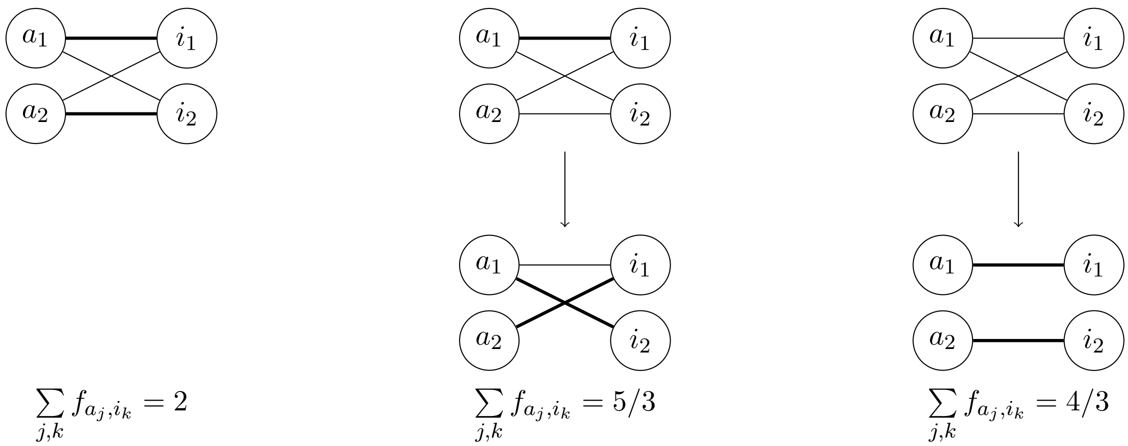

Figure 1: Cycles of length 4. Thin edges carry 1/3 flow; and thick edges carry 2/3 flow.

makes the resulting graph sparse. The sparsity not only helps the construction of random lists as described in Section 3.1.2, but also eases the analysis of the algorithm.

3.1.2 Generation of the random lists

In order to simplify the description of the specific probability distribution used to generate the random lists, and the analysis of the corresponding online algorithm, let us first add dummy advertisers a_d^i and dummy edges (a_d^i, i) with $f_{a_d^i,i} = r_i - \sum_{a \in A} f_{a,i}$ for all $i \in I$ with $\sum_{a \sim i} f_{a,i} < 1$. Dummy advertisers are flagged as matched from the start, so no impression are ever assigned to them. Since every edge in the graph has value 1/3 or 2/3, every node has two or three neighbors.

The construction of the random lists goes as follows. Given an impression type i, if it has two neighbors a_1 and a_2 in the resulting graph, the list is $\langle a_1, a_2 \rangle$ with probability $f_{a_1,i}$; the list is $\langle a_2, a_1 \rangle$ with probability $f_{a_2,i}$. Otherwise, if i has three neighbors a_1, a_2, and a_3 (in this case, $f_{a_1,i} = f_{a_2,i} = f_{a_3,i} = 1/3$), the list is a uniformly random permutation of $\langle a_1, a_2, a_3 \rangle$.

3.2 Upper bound on the offline algorithm

In order to show that LP (2) provides an upper bound on the offline solution, we prove that the offline solution is feasible to the LP. The feasibility of the first two constraints is obvious. The feasibility of the last constraint is a consequence of the following simple lemma:

Lemma 3 (Manshadi et al.[13])**.** $\forall e \in E, f_e^* \le 1 - e^{-1} < 2/3$.

From Lemma 3, the expected optimal offline solution is thus a feasible to LP (2). Therefore, the optimal solution $\mathbf{f}^T \cdot \mathbf{1}$ is an upper bound on the offline optimal $\mathbf{f}^{*T} \cdot \mathbf{1}$. From now on, we will compare the online algorithm with the optimal solution of LP (2) instead of with the offline solution, because the former one is much easier to find than the latter one.

3.3 Certificate events

One difficulty encountered in previous papers is that an advertiser being matched is highly dependent on other advertisers. The strong dependence is difficult to deal with, and difficult to be decoupled. In this paper, we use a local approach to avoid this issue. To be more specific, we compute a lower bound on p_a, the probability that advertiser a is matched, using only knowledge of a's neighborhood.

To ease the analysis, we consider lists associated with all arriving impressions rather than the types of impressions, because online matching results are a deterministic function of the former one. As mentioned in Section 2, all lists are i.i.d. distributed. It is not difficult to see that the distribution can be easily inferred from the resulting graph $G_{\mathbf{f}}$. For example, a local structure as showed in Figure 2 implies that with probability $1/n$, a list starts with $\langle a_1, ...\rangle$; and with probability $1/6n$, a list is $\langle a_1, a, a_2\rangle$.

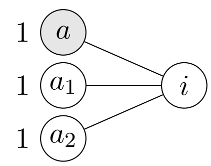

Figure 2: Possible configurations of i's neighborhood in the graph. All edges carry 1/3 flow. The number next to advertiser a indicates $f_a = \sum_{i\sim a} f_{a,i}$.

Assume a is the advertiser we are considering, and i is a neighbor of a. Let us consider the following two types of events: $B_a = \{$among the n lists, there exists a list starting with $\langle a, ...\rangle\}$ and $G_a^i = \{$among the n lists, there exist successive lists starting with advertisers different from a but which are neighbors of i, and ensuring that a is matched$\}$. For example, in a local structure as showed in Figure 2, if three lists appear in order: $\langle a_1, ...\rangle$, $\langle a_2, ...\rangle$, and $\langle a_1, a_2, a\rangle$, then advertiser a is matched; and hence G_a^i happens. B_a and G_a^i (for any i) will be called "certificate events", in the sense that if any of these events happen, they provide a certificate that advertiser a is matched.

We now show that these certificate events have some good properties and their probabilities are easy to find. In this section and the next, we will use these certificate events to lower bound the probability that an advertiser is matched; and further lower bound the competitive ratios of our algorithms.

3.3.1 Asymptotic independence

For notation simplicity, we define supporting set $S(G_a^i)$ as the set of lists that start with advertisers that are neighbors of i but not a; $S(B_a)$ as the set of lists that start with a. The supporting set of the intersection of two certificate events is defined as the union of the supporting sets of the two certificate events.

Lemma 4. *Let E_1 and E_2 be certificate events or intersections of two certificate events. If their supporting sets $S(E_1) \cap S(E_2) = \emptyset$, then E_1 and E_2 are asymptotically independent, i.e. $|\mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1)\mathbb{P}(E_2)| < O(1/n)$.*

Proof. Let M_1 be the number of lists among all n lists in $S(E_1)$; M_2 be the number of lists among all n lists in $S(E_2)$. The proof consists of three key parts: with high probability M_1 and M_2 are small; when M_1 and M_2 are small, they are asymptotically independent; given M_1 and M_2, E_1 and E_2 are independent.

According to the construction of our algorithm, we can show that a given list belongs to $S(E_1)$ (or $S(E_2)$) with probability less than $6/n$. From the Chernoff bound, with high probability M_1 and M_2 are close to their mean: $\mathbb{P}(M_1 \geq 6\mu) \leq \exp(-\frac{6\mu^2}{2+\mu}) \leq O(1/n)$ and $\mathbb{P}(M_2 \geq 6\mu) \leq \exp(-\frac{6\mu^2}{2+\mu}) \leq O(1/n)$, where $\mu = n^{1/3}$.

Assuming $\mathbb{E}[M_1] = n_1$ and $\mathbb{E}[M_2] = n_2$, for all $m_1 < 6\mu$ and $m_2 < 6\mu$, we have

$$\frac{\mathbb{P}(M_1 = m_1, M_2 = m_2)}{\mathbb{P}(M_1 = m_1)\mathbb{P}(M_2 = m_2)} = \frac{(n-m_1)!(n-m_2)!}{n!(n-m_1-m_2)!} \frac{(1-(n_1+n_2)/n)^{n-m_1-m_2}}{(1-n_1/n)^{n-m_1}(1-n_2/n)^{n-m_2}} = 1 + O(1/n),$$

where the last inequality is due to $m_1 m_2 = o(n)$.

Since all advertisers other than neighbors of i are assumed to have infinite capacities, all the lists that are not in $S(E_1)$ do not affect E_1. Thus, given $M_1 = m_1$, E_1 is independent of $n - m_1$ lists that are not in $S(E_1)$. Because of the assumption $S(E_1) \cap S(E_2) = \emptyset$, E_1 is independent of E_2 given M_1 and M_2.

From the three facts above, we have

$$\begin{aligned}
\mathbb{P}(E_1 \cap E_2) &= \sum_{m_1, m_2} \mathbb{P}(M_1 = m_1, M_2 = m_2)\mathbb{P}(E_1, E_2 | M_1 = m_1, M_2 = m_2) \\
&= \sum_{m_1, m_2 < 6\mu} \mathbb{P}(M_1 = m_1, M_2 = m_2)\mathbb{P}(E_1, E_2 | M_1 = m_1, M_2 = m_2) + O(1/n) \\
&= \sum_{m_1, m_2 < 6\mu} \mathbb{P}(M_1 = m_1, M_2 = m_2)\mathbb{P}(E_1 | M_1 = m_1)\mathbb{P}(E_2 | M_2 = m_2) + O(1/n) \\
&= \sum_{m_1, m_2 < 6\mu} \mathbb{P}(M_1 = m_1)\mathbb{P}(M_2 = m_2)\mathbb{P}(E_1 | M_1 = m_1)\mathbb{P}(E_2 | M_2 = m_2) + O(1/n) \\
&= \sum_{m_1, m_2} \mathbb{P}(M_1 = m_1)\mathbb{P}(M_2 = m_2)\mathbb{P}(E_1 | M_1 = m_1)\mathbb{P}(E_2 | M_2 = m_2) + O(1/n) \\
&= \mathbb{P}(E_1)\mathbb{P}(E_2) + O(1/n)
\end{aligned}$$

□

By applying Lemma 4 twice, we can show that any four (or less than four) certificate events are asymptotic independent, as long as their supporting sets do not intersect:

Corollary 1. *Consider a set of at most four certificate events* $\{C_j\}_{j \in J}$ *(*$|J| \leq 4$*). If* $\cap_{j \in J} S(C_j) = \emptyset$*, then* $\mathbb{P}(\cap_{j \in J} C_j) = \prod_{j \in J} \mathbb{P}(C_j) + o(1/n)$*.*

3.3.2 Computing probabilities

In this section and the next, supporting sets of certificate events are of small sizes because of the construction of the distribution. In such cases, the probabilities of certificate events can be calculated via double summation, which is doable even by hand, though time-consuming.

On the other hand, it also can be done via a dynamic programming approach. Given n, the probability of an advertiser being matched at the end given the current state can be computed backward. As we can easily check, the probability converges to the limit with an error term of $O(1/n)$. In fact, when $n = 10^4$, the computed probability is within 10^{-5} accuracy.

In this paper, we simply provide the probabilities of certificate events and omit the process of finding them due to the following reasons. First, the computation of probabilities is not the key to our approach, though the actual numbers matter. Second, it is just simple algebra and too long to present in the paper.

3.4 Lower bound on the online algorithm

For notational simplicity, define $f_a \triangleq \sum_{i \sim a} f_{a,i}$, and let p_a be the probability that an advertiser a is matched by the online algorithm. Since every edge in the graph G with a non-zero flow will carry a flow of 1/3 or 2/3, there are very few different local configurations in the graph $G_f = \{A \cup I, E_f\}$, where $E_f = \{e \in E | f_e > 0\}$. For example, for an edge $e = (a, i)$ such that $f_a = 1$ and $f_{a,i} = 2/3$, the only four possibilities for i's neighborhood are $\alpha 1$, $\alpha 2$, $\alpha 3$, and $\alpha 4$ in Figure 3; for an edge $e = (a, i)$ such that $f_a = 1$ and $f_{a,i} = 2/3$, the only five possibilities for i's neighborhood are $\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$, and $\beta 5$ in Figure 3.

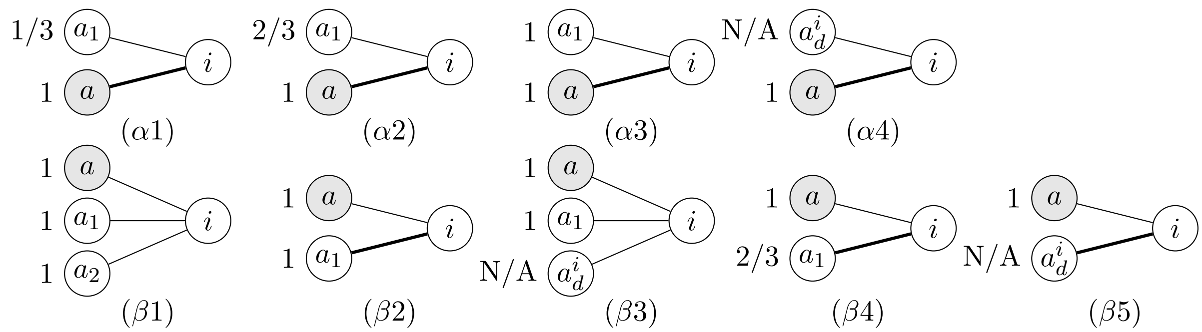

Figure 3: Possible configurations of i's neighborhood in the graph. Thin edges carry 1/3 flow, and thick edges carry 2/3 flow. The number next to advertiser a indicates $f_a = \sum_{i \sim a} f_{a,i}$.

For each configuration, because a has at most three neighbors, we can easily compute a lower bound on the probability of its being matched. For example, assume a has two neighbors i_1 and i_2, and they are not part of a cycle of length 4. a is matched if one of the three certificate events happens: B_a, $G_a^{i_1}$ or $G_a^{i_2}$. Since those three events are asymptotically independent, $p_a \geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1-p)(1-p_1)(1-p_2)$, where $p = \mathbb{P}(B_a)$, $p_1 = \mathbb{P}(G_a^{i_1})$, and $p_2 = \mathbb{P}(G_a^{i_2})$ are easy to find. Using such an idea, we can show case by case that:

Lemma 5. *$\forall a \in A$, let N_a be the set of advertisers who are at an edge-distance no more than 4 from a in G_f. Then, there exists $\mu_{a,a'} \in [0,1]$ for all $a' \in N_a$, such that*

$$\sum_{a' \in N_a} \mu_{a,a'} p_{a'} \geq (1 - 2e^{-2}) \sum_{a' \in N_a} \mu_{a,a'} f_{a'}.$$

Proof. The detailed proof that goes through all cases can be found in the appendix. □

Lemma 6. *$\exists \{\lambda_a \geq 0\}_{a \in A}$ such that $\sum_a \lambda_a \mu_{a,a'} = 1, \forall a'$.*

Proof. The proof can be found in the appendix. □

Combining the two lemmas above, a conical combination of inequalities leads to our main result:

Theorem 1. $\mathbb{E}[ALG] = \sum\limits_{a \in A} p_a \geq (1 - 2e^{-2}) \sum\limits_{a \in A} f_a \geq (1 - 2e^{-2}) \mathbb{E}[OPT]$.

3.5 Tight example

It is worth mentioning that the ratio of $1 - 2e^{-2}$ is tight for this algorithm. The ratio can be achieved with the following example: Consider the case of the complete bipartite graph $K_{n,n}$, where n is an even number. One optimal solution to LP (2) consists of a disjoint union of $n/2$ cycles of length 4; within each cycle, two edges carry 1/3 flow, and two carry 2/3 flow. Since the underlying graph is $K_{n,n}$, the optimal offline solution is n. On the other hand, for any cycle in the offline optimal solution, the expected number of matches is $2(1 - e^{-2})$. Therefore, the competitive ratio in this instance is $1 - 2e^{-2} \approx 0.729$.

4 Extension to vertex-weighted stochastic matching

In this section, we consider the online stochastic vertex-weighted matching problem as defined in Aggarwal et al. [1]. The problem is exactly the same as the online stochastic matching problem introduced in Section 1 except for the objective function. In the weighted problem, every advertiser a has a nonnegative weight w_a, indicating his/her importance or value. The objective is to maximize the sum of weights of matched advertisers rather than the number of matched advertisers as in the unweighted problem.

The techniques used in Section 3.4 are based on local properties of graphs and thus also work for the vertex-weighted case.

4.1 Original algorithm

Let us consider the maximum flow problem on the graph G:

$$\begin{array}{rlr} \max & \sum\limits_{a,i} w_a f_{a,i} & \\ s.t. & \sum\limits_{i} f_{a,i} \leq 1 & \forall a \in A \\ & \sum\limits_{a} f_{a,i} \leq 1 & \forall i \in I \\ & f_e \in [0, 2/3] & \forall e \in E \end{array} \tag{3}$$

Again, since the set of vertices of the feasible polytope of (3) is a subset of $\{0, 1/3, 2/3\}^E$, there exists an optimal solution to (3) in $\{0, 1/3, 2/3\}^E$, and let $\mathbf{f}$ be such an optimal solution that satisfies requirements in Section 3.2. To ease the analysis, we try to make $E_{\mathbf{f}}$ as sparse as we can by doing the following two types of transformations as we did in Section 3.2. As argued before, there exists an optimal solution $\mathbf{f}$ such that its resulting graph $G_{\mathbf{f}}$ does not contain cycles of length 4, unless the four nodes in such a cycle do not have any other neighbors. Furthermore, if there exists impression i that has two neighbors a_1 and a_2 with $f_{i,a_1} = f_{i,a_2} = 1/3$, $f_{a_1} < 1$, and $f_{a_2} < 1$, without loss of generality, we assume $w_{a_1} < w_{a_2}$. Another solution $\mathbf{f}'$ with $f'_{i,a_1} = 0$, $f'_{i,a_2} = 2/3$, and everything else unchanged has a larger or equal objective value, and less edges in its resulting graph. We transform $\mathbf{f}$ to $\mathbf{f}'$. Note that each time, transformations mentioned above remove one or two edges and does not introduce any new edge. Thus, given any initial optimal solution, after at most $|E|$ transformations, the optimal solution cannot be transformed further in the above two ways.

Based on $\mathbf{f}$, the probability distributions over lists can be constructed as in 3.1.2, and the same idea as in 3.4 leads to the proof that $p_a \geq 0.682 f_a$ for all $a \in A$. Summing up these inequalities, we have $\sum_{a \in A} w_a p_a \geq 0.682 \sum_{a \in A} w_a f_a$, which implies that the algorithm is 0.682-competitive.

It is worth noting that, although Lemma 5 and Lemma 6 still hold, they are of little value to weighted problems because of different weights associated with different advertisers. For the same reason, results and techniques proposed in previous papers dealing with unweighted stochastic matching problems are unlikely to be adapted for weighted problems.

4.2 Modification

However, modifying $\mathbf{f}$ and the construction of random lists can lead to a better algorithm. If i has neighbors with $f = 1$ and $f < 1$, as showed in Figure 4, $\mathbf{f}$ will be modified as follows: in (1), $\tilde{f}_{a_1,i} = 0.1$ and $\tilde{f}_{a_2,i} = 0.9$; in (2), $\tilde{f}_{a_1,i} = 0.15$ and $\tilde{f}_{a_2,i} = 0.85$; in (3), $\tilde{f}_{a_1,i} = 0.6$ and $\tilde{f}_{a_2,i} = 0.4$; in (4), $\tilde{f}_{a_1,i} = 0.1$, $\tilde{f}_{a_2,i} = 0.45$ and $\tilde{f}_{a_3,i} = 0.45$; in (5), $\tilde{f}_{a_1,i} = 0.15$, $\tilde{f}_{a_2,i} = 0.425$ and $\tilde{f}_{a_3,i} = 0.425$. For all the other edges e, $\tilde{f}_e = f_e$.

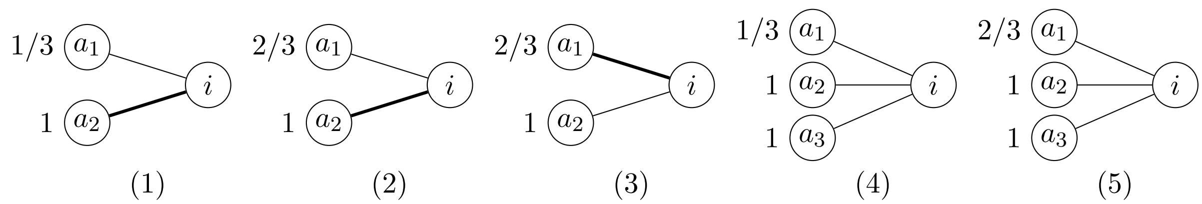

Figure 4: Modification of $\mathbf{f}$. Thin edges carry 1/3 flow, and thick edges carry 2/3 flow. The number next to advertiser a indicates f_a.

Now use $\tilde{\mathbf{f}}$ instead of $\mathbf{f}$ for the construction of the probability distributions over lists in a way similar to the one described in Section 3.1.2 as follows. Given an impression type i, if it has two neighbors a_1 and a_2 in the resulting graph, the list is $\langle a_1, a_2\rangle$ with probability $\tilde{f}_{a_1,i}$; the list is $\langle a_2, a_1\rangle$ with probability $\tilde{f}_{a_1,i}$. If i has three neighbors a_1, a_2, and a_3; the list is $\langle a_j, a_k, a_l\rangle$ with probability $\tilde{f}_{a_j,i}\tilde{f}_{a_k,i}/(1-\tilde{f}_{a_j,i})$.

Let $\tilde{p}_a$ be the probability that advertiser a is matched in the modified algorithm. Using the same idea as in Section 3.4, we can then show that:

Lemma 7. $\tilde{p}_a \geq 0.725 f_a, \forall a \in A$.

Proof. The detailed proof which goes through all cases can be found in the appendix. □

Summing these inequalities up, we have:

Theorem 2. $\mathbb{E}[ALG] = \sum_{a\in A} w_a \tilde{p}_a \geq 0.725 \sum_{a\in A} w_a f_a \geq 0.725\mathbb{E}[OPT]$.

5 Stochastic matching with general arrival rates

In this section, we assume that $r_i \leq 1$ for all i. The algorithm and basic ideas here are very similar to [13]: in the offline stage, we approximate the expected offline optimal solution; then, in the online stage, we use the approximation solution to generate lists of length two. However, our algorithm is different in two aspects. First, we use a max flow problem instead of Monte Carlo methods to approximate the offline solution; second, the way lists are generated is different. The first difference leads to much less computation in the offline stage, while the second difference results in a slightly better competitive ratio.

5.1 Offline solution

One possible approach to find useful offline information in the general case is to use sampling methods to estimate the optimal offline solution, as described in Manshadi et al. [13]. However, some properties that hold for the optimal offline solution may not hold for the estimated one. Furthermore, a large number of samples may be needed in order to estimate the offline optimal solution within a desirable accuracy, which takes a long time. Therefore, we consider the following

LP instead:

$$
\begin{array}{llll}
\max & \sum\limits_{a,i} f_{a,i} & & \\
s.t. & \sum\limits_{i\sim a} f_{a,i} \le 1 & \forall a \in A & \\
 & \sum\limits_{a\sim i} f_{a,i} \le r_i & \forall i \in I & (4)\\
 & \sum\limits_{i\sim a} (2f_{a,i} - r_i)^+ \le 1 - \ln 2 + \frac{1}{n} & \forall a \in A & \\
 & f_e \ge 0 & \forall e \in E &
\end{array}
$$

Note that that LP(4) is equivalent to a single-source s single-destination t maximum flow problem on a directed network $\hat{G} = \{\hat{V}, \hat{E}\}$ with $|A| + 2|I| + 2$ vertices and $2|E| + |A| + 2|I|$ arcs. The vertex set $\hat{V} = \{s,t\} \cup A \cup I \cup I'$, where I' is a duplicate of I, and the arc set $\hat{E} = \{(s,a),(a,i),(i',i),(a,i'),(i,t)|a \in A, i \in I, i \text{ is a duplicate copy of } i\}$. The capacity of (s,a) is 1; the capacity of (a,i) is $r_i/2$; the capacity of (i',i) is $1 - \ln 2 + 1/n$; the capacity of (i,t) is r_i; (a,i') have infinite capacities.

5.2 Upper bound on the optimal offline solution

Let $\mathbf{f}^*$ be an optimal offline solution. All but the third constraints in (4) are trivially valid for $\mathbf{f}^*$. The third constraint has been proven in [13]:

Lemma 8. *[[13], Lemma 5]* $\sum\limits_{i\sim a}(2f^*_{a,i} - r_i)^+ \le 1 - \ln 2 + \frac{1}{n}, \forall a \in A.$

5.3 Randomized algorithm

For simplicity, let us again first add a dummy advertiser a_d with $f_{a_d} \triangleq 1$, and dummy edges (a_d, i) for all i with $f_{a_d,i} \triangleq r_i - \sum_{a\in A} f_{a,i}$. The dummy advertiser is full at the very beginning. Every time an impression of type i arrives, a random list consisting of two advertisers will be generated as follows. Assume $a_1, ..., a_k$ are the advertisers interested in i. Choose a random number x uniformly over $[0, r_i]$. If $x \in [\sum_{l=1}^{j-1} f_{a_l,i}, \sum_{l=1}^{j} f_{a_l,i}]$, then a_j is the first advertiser in the list to be considered; if $x \pm r_i/2 \in [\sum_{l=1}^{k-1} f_{a_l,i}, \sum_{l=1}^{k} f_{a_l,i}]$ then a_k is the second in the list to be considered. Worth noting is the possibility that a_j and a_k correspond to the same advertiser; in that case, the list degenerates to a singleton.

Let $m^i_{a_j,a_k}$ be the expected number of requests for impressions of type i and corresponding lists given by $\langle a_j, a_k\rangle$. Since all lists are i.i.d., the probability that an impression is of type i and its corresponding list is $\langle a_j, a_k\rangle$ is $m^i_{a_j,a_k}/n$. From the construction of the lists, we have $m^i_{a_j,a_k} = m^i_{a_k,a_j}$. As we mentioned in Section 2, from a given realization of the sequence of random lists, we can find the cardinality of the corresponding online matching. Since the random list associated with the j^{th} request only depends on the impression type of that request, and not on types and assignments of earlier requests, these random lists are all i.i.d.. Thus, we can focus on the lists themselves, rather than on the impression types that they are associated with. Then, $m_{a_j,a_k} \triangleq \sum_{i\in I} m^i_{a_j,a_k}$ is the expected number of lists that are $\langle a_j, a_k\rangle$, irrespective of the impression types which they are associated with. Furthermore, because $m^i_{a_j,a_k} = m^i_{a_k,a_j}$, we have $m_{a_j,a_k} = m_{a_k,a_j}$. Since all lists are i.i.d., the probability that a list is $\langle a_j, a_k\rangle$ is $m_{a_j,a_k}/n$.

5.4 Lower bound on the online algorithm

The analysis here is almost the same as in [13] except for some minor changes due to the different ways we generate random lists, e.g. $m_{a_j,a_k} = m_{a_k,a_j}$. To help better understand the arguments, we present the full proof in this section. The following main result is proved by way of successive claims.

Theorem 3. $\sum_{a \in A} p_a \geq 0.706 \sum_{a \in A} f_a$.

Let $A_a = A \backslash \{a\}$, $A^* = A \cup \{a_d\}$, and $A^*_a = A^* \backslash \{a\}$. $\forall a \in A^*, a_1 \in A^*_a$, define events $B_a = \{\exists j$, such that $L_j = \langle a, . \rangle\}$, $E_{a_1,a_2} = \{\exists j < k$ such that $L_j = \langle a_1, . \rangle, L_k = \langle a_1, a_2 \rangle\}$, and $E_{a_d,a} = \{\exists j$, such that $L_j = \langle a_d, a \rangle\}$. If any of B_a, $E_{a_1,a}$, and $E_{a_d,a}$ happens, then advertiser a is matched. Thus, the probability p_a that advertiser a is matched is at least:

$$
\begin{aligned}
p_a &\geq \mathbb{P}(B_a) + \mathbb{P}(\bar{B}_a)\mathbb{P}(\bigcup_{a_1 \in A^*_a} E_{a_1,a} | \bar{B}_a) \\
&\geq 1 - e^{-f_a} + e^{-1}\Big(\sum_{a_1 \in A^*_a} \mathbb{P}(E_{a_1,a}) - \tfrac{1}{2} \sum_{a_1 \neq a_2 \in A^*_a} \mathbb{P}(E_{a_1,a}, E_{a_2,a}) \Big) \\
&\geq 1 - e^{-f_a} + e^{-1}\Big(\sum_{a_1 \in A^*_a} \mathbb{P}(E_{a_1,a}) - \tfrac{1}{2} \sum_{a_1 \neq a_2 \in A^*_a} \mathbb{P}(E_{a_1,a})\mathbb{P}(E_{a_2,a}) \Big),
\end{aligned}
$$

where the last two inequalities are due to asymptotic independence. The proof of asymptotic independence is similar to the proof of Lemma 4, and is omitted in the paper.

Let us now provide a way to compute $\mathbb{P}(E_{a_1,a})$.

Claim 2. *We have* $\mathbb{P}(E_{a_1,a}) = g(f_{a_1}, m_{a_1,a})$ *for all* $a_1 \in A$ *and* $a \in A^*_{a_1}$*, and* $\mathbb{P}(E_{a_d,a}) \geq g(f_{a_d}, m_{a_d,a})$ *for all* $a \in A$*, where:*

$$
g(y,x) = h(y,0) - h(y,x), \quad \textit{and} \quad h(y,x) = \begin{cases} \dfrac{y}{y-x}(e^{-x} - e^{-y}), & \text{if } x \neq y \\ ye^{-y}, & \text{if } x = y \end{cases}
$$

Proof. Define $F^j_{a_1} = \{$the j^{th} list is $\langle a_1, . \rangle\}$ and $G^j_{a_1,a} = \{$there exists $k \geq j$ such that the k^{th} list is $\langle a_j, a \rangle\}$. Then,

$$
\begin{aligned}
\mathbb{P}(E_{a_1,a}) &= \sum_j \mathbb{P}(F^j_{a_1})\mathbb{P}(G^{j+1}_{a_1,a}) \\
&= \sum_j \big(1 - \frac{f_{a_1}}{n}\big)^{j-1} \frac{f_{a_1}}{n} \Big(1 - \big(1 - \frac{m_{a_1,a}}{n}\big)^{n-j}\Big) \\
&\approx \sum_j \frac{f_{a_1}}{n} e^{-\frac{j}{n} f_{a_1}} \big(1 - e^{-\frac{n-j}{n} m_{a_1,a}}\big).
\end{aligned}
$$

If $m_{a_1,a} \neq f_{a_1}$,

$$
\begin{aligned}
\mathbb{P}(E_{a_1,a} | \bar{B}_a) &= \sum_j \frac{f_{a_1}}{n} \big(e^{-\frac{j}{n} f_{a_1}} - e^{-m_{a_1,a}} e^{-\frac{j}{n}(f_{a_1} - m_{a_1,a})}\big) \\
&= 1 - e^{-f_{a_1}} - \frac{f_{a_1}}{f_{a_1} - f_{a_1,a}} e^{-f_{a_1,a}} \big(1 - e^{-(f_{a_1} - f_{a_1,a})}\big) = g(f_{a_1}, f_{a_1,a}).
\end{aligned}
$$

If $m_{a_1,a} = f_{a_1}$,

$$
\begin{aligned}
\mathbb{P}(E_{a_1,a} | \bar{B}_a) &= \sum_j \frac{f_{a_1}}{n} \big(e^{-\frac{j}{n} f_{a_1}} - e^{-f_{a_1}}\big) \\
&= 1 - e^{-f_{a_1}} - f_{a_1} e^{-f_{a_1}} = g(f_{a_1}, f_{a_1}).
\end{aligned}
$$

$\square$

We have transformed a probabilistic problem into an algebraic problem. In the remaining part of the section, we only use algebraic manipulations and the following properties of functions g and h to find a lower bound of the competitive ratio.

Claim 3. *For $y \in [0,1]$, $h(y,x)$ is convex and decreasing in $x \in [0,y]$; $g(y,x)$ is concave and increasing in $x \in [0,y]$; $g(y,x)$ is increasing in $y \in [x,\infty)$.*

Proof. The claim can be easily verified by taking first and second order partial derivatives. $\square$

Because of the convexity of h in the second argument, we have

$$\mathbb{P}(E_{a_1,a}) = h(f_{a_1},0) - h(f_{a_1},m_{a_1,a}) \le -m_{a_1,a} \cdot \frac{\partial h}{\partial y}(f_{a_1},0) \le e^{-1} m_{a_1,a}, \tag{5}$$

implying that $\sum_{a_1 \in A_a} \mathbb{P}(E_{a_1,a}|\bar{B}_a) \le e^{-1}$. Combined with $\mathbb{P}(E_{a_1,a}|\bar{B}_a) \ge g(f_{a_1},m_{a_1,a})$ for all $a_1 \in A_a^*$, we have

$$p_a \ge 1 - e^{-f_a} + e^{-1}\Big(\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a}) - \frac{1}{2}\big(\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})\big)^2 + \frac{1}{2}\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})^2\Big).$$

Since g is increasing in the first argument, we have $g(f_{a_d},m_{a_d,a}) = g(f_{a_d},m_{a,a_d}) \ge g(f_a,m_{a,a_d})$ for all $a \in A$. Thus,

$$\sum_{a \in A}\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a}) \quad \ge \quad \sum_{a \in A}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1})$$

and

$$\sum_{a \in A}\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})^2 \quad \ge \quad \sum_{a \in A}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1})^2 \quad .$$

Therefore, by switching the order of summation, we have

$$\begin{aligned}
\frac{\sum_{a\in A} p_a}{\sum_{a \in A} f_a} &\ge \frac{\sum_{a\in A}\Big(1 - e^{-f_a} + \frac{1}{e}\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a}) - \frac{1}{2e}\big(\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})\big)^2 + \frac{1}{2e}\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})^2\Big)}{\sum_{a \in A} f_a} \\
&\ge \frac{\sum_{a\in A}\Big(1 - e^{-f_a} + \frac{1}{e}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1}) - \frac{1}{2e}\big(\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})\big)^2 + \frac{1}{2e}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1})^2\Big)}{\sum_{a \in A} f_a} \\
&\ge \min_{a \in A} \frac{1 - e^{-f_a} + \frac{1}{e}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1}) - \frac{1}{2e}\big(\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})\big)^2 + \frac{1}{2e}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1})^2}{f_a}
\end{aligned}$$

Let $\beta_a = \max_{a_1 \in A_a^*} m_{a,a_1} \triangleq m_{a,a_1^*}$, $s_a = \sum_{a_1 \in A_a^*} m_{a,a_1}$. Then, we have $\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1})^2 \ge g(f_a,\beta_a)^2$. Furthermore, because g is concave in the second argument and $g(f_a,0)=0$,

$$\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1}) \ge \sum_{a_1 \in A_a^*} \frac{m_{a,a_1}}{\beta_a} g(f_a,\beta_a) = \frac{s_a}{\beta_a} g(f_a,\beta_a).$$

On the other hand, from inequality (5),

$$\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a}) = \sum_{a_1 \in A_a^* \setminus \{a_1^*\}} g(f_{a_1},m_{a_1,a}) + g(f_{a_1},\beta_a) \le e^{-1}(s_a - \beta_a) + g(1,\beta_a).$$

Therefore,

$$\begin{aligned}
&\tfrac{1}{f_a}\big(1 - e^{-f_a} + \tfrac{1}{e}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1}) - \tfrac{1}{2e}\big(\sum_{a_1 \in A_a^*} g(f_{a_1},m_{a_1,a})\big)^2 + \tfrac{1}{2e}\sum_{a_1 \in A_a^*} g(f_a,m_{a,a_1})^2\big) \\
\ge\ &\tfrac{1}{f_a}\big(1 - e^{-f_a} + \tfrac{1}{e}\tfrac{s_a}{\beta_a} g(f_a,\beta_a) - \tfrac{1}{2e}\big(e^{-1}(s_a - \beta_a) + g(1,\beta_a)\big)^2 + \tfrac{1}{2e} g(f_a,\beta_a)^2\big) \triangleq R(f_a,\beta_a,s_a).
\end{aligned}$$

From the definitions of f_a, β_a, and s_a, we have $f_a \geq s_a \geq \beta_a$, and $f_a - s_a$ is the expected number of lists that are singletons $\langle a \rangle$. From the construction of lists, the expected number of singletons $\langle a \rangle$ associated with impressions of types i is $(2f_{a,i} - r_i)^+$. Thus, $f_a - s_a = \sum_{i \sim a}(2f_{a,i} - r_i)^+ \leq (1 - \ln 2) + 1/n$. We can numerically show that, for $n \geq 100$:

Claim 4. *Subject to* $1 \geq f_a \geq s_a \geq \beta_a \geq 0$ *and* $f_a - s_a \leq (1 - \ln 2) + 1/n$, $R(f_a, \beta_a, s_a) \geq 0.706$.

Proof. We divide the feasible region into cubes of side length 0.001. In each small region S, define $f_a^{\max} \triangleq \sup f_a$ and $f_a^{\min} \triangleq \inf f_a$. Define $s_a^{\max}$, $s_a^{\min}$, $\beta_a^{\max}$, and $\beta_a^{\min}$ similarly. Then, from Claim 3, we can show that $\forall (f_a, s_a, \beta_a) \in S$, $R(f_a, s_a, \beta_a)$ is bounded from below by

$$\frac{1}{f_a^{\max}}\big(1 - e^{-f_a^{\min}} + \frac{1}{e}\frac{s_a^{\min}}{\beta_a^{\max}} g(f_a^{\min}, \beta_a^{\min}) - \frac{1}{2e}\big(e^{-1}(s_a^{\max} - \beta_a^{\min}) + g(1, \beta_a^{\max})\big)^2 + \frac{1}{2e} g(f_a^{\min}, \beta_a^{\min})^2\big).$$

We can numerically verify $R(f_a, s_a, \beta_a) \geq 0.706$ in each region. The lower bound is achieved when $f_a \in [0.999, 1]$, $s_a \in [0.692, 0.693]$, and $\beta_a \in [0.564, 0.565]$. Hence, $R(f_a, s_a, \beta_a) \geq 0.706$ is a valid inequality for the whole feasible region. $\square$

Theorem 3 follows from Claim 4.

6 Poisson arrivals

In the preceding sections, the number of arriving requests is assumed to be fixed and known in advance. However, in most applications, such an assumption is too strong. Thus, in this section, we attempt to relax this assumption.

In this section, we consider the following scenario. A set of advertisers express their interests in impressions of different types. Advertisers are fixed and known ahead of time while requests for impressions come online. Impression types are i.i.d., and the distribution may be known or unknown. The arrival of impressions is a Poisson Process with arrival rate $\lambda = n$. The task is to maximize the cardinality of matching by the end of a given fixed period $T = 1$.

6.1 Algorithms

The expected number of arrivals is $\lambda T = n$. We show that, greedy algorithms designed for stochastic matching with given number of arrivals works well for the one with Poisson arrivals (e.g. the ranking algorithm for problems with unknown distribution, our proposed algorithms in the previous sections for problems with known distribution). More specifically, we will show that a c-competitive "greedy-type" algorithm (where c is the ratio of expectation) for fixed arrivals is $c - \epsilon$ competitive for Poisson arrivals.

Because the number of Poisson arrivals concentrates around its mean, we expect both online and offline objective to concentrate around their means.

Lemma 9. *Let* N *be the number of arrivals within* $[0, T]$. *Then,* $\mathbb{P}((1-\epsilon)\lambda T < N < (1+\epsilon)\lambda T) \to 1$ *as* $\lambda T \to \infty$ *for any* $\epsilon > 0$.

Let OPT_m be the expected offline optimal solution given $N = m$ and OPT be the expected offline optimal solution.

Lemma 10. $\forall (1 - \epsilon)n < m \leq n$, $OPT_m \leq OPT_n$; $\forall n \leq m < (1 + \epsilon)n$, $OPT_m \leq (1 + \epsilon)OPT_n$.

Proof. $\forall (1-\epsilon)n < m \leq n$, an instance τ_m of m arrivals can be generated in the following way: generate an instance τ_n of n arrivals first, and then remove $n-m$ arrivals uniformly at random. Since τ_m is a subset of τ_n, $OPT(\tau_m) \leq OPT(\tau_n)$. By taking expectation, we have $OPT_m \leq OPT_n$.

$\forall n \leq m < (1+\epsilon)n$, an instance τ_n of n arrivals can be generated in the following way: generate an instance τ_m of m arrivals first, and then remove $m-n$ arrivals uniformly at random. A feasible solution of τ_n can be induced by the optimal solution of τ_m, by removing pairs corresponding to removed arrivals and not adding any other pairs. The feasible solution of τ_n has expected value of $\frac{n}{m}OPT(\tau_m)$. Thus, $OPT(\tau_m) \leq \frac{m}{n}OPT(\tau_n) \leq (1+\epsilon)OPT(\tau_n)$. □

As a consequence we have:

Corollary 2. $OPT \leq (1+\epsilon)OPT_n$.

6.1.1 The unweighted case

Let ALG_m be the expected online solution given $N = m$ and ALG be the expected online solution.

Lemma 11. $\forall (1-\epsilon)n < m \leq n, ALG_m \geq (1-\epsilon)ALG_n$; $\forall n \leq m < (1+\epsilon)n, ALG_m \geq ALG_n$.

Proof. $\forall n \leq m < (1+\epsilon)n$, an instance τ_n of n arrivals can be generated in the following way: generate an instance τ_m of m arrivals first, and then remove the last $m-n$ arrivals. Because of the greediness of the algorithm, $ALG_m \geq ALG_n$.

$\forall (1-\epsilon)n < m \leq n$, let r_i be the probability that the i^{th} arrival is matched. Because of the greediness of the algorithm and the fact that less and less bins are unmatched, r_i is non-increasing. Since $ALG_m = \sum_{i=1}^{m} r_i$ and $ALG_n = \sum_{i=1}^{n} r_i$, we have $ALG_m \geq \frac{m}{n}ALG_n \geq (1-\epsilon)ALG_n$. □

As a consequence we have:

Corollary 3. $ALG \geq (1-2\epsilon)ALG_n$.

Because of the assumption of c-competitiveness, $ALG_n \geq c \cdot OPT_n$. Therefore Corollaries 2 and 3 imply that $ALG \geq (1-2\epsilon)c \cdot OPT$.

6.1.2 The weighted case

Let ALG_m be the expected online solution given $N = m$ and ALG be the expected online solution. Let R_i be the marginal revenue in the i^{th} step. For the algorithm proposed in Section 4.1 and 4.2, we can show that although $\mathbb{E}[R_i]$ is not non-increasing as in the unweighted case, R_j cannot be too large compared to $\mathbb{E}[R_i]$ for $i < j$. Specifically:

Lemma 12. $\mathbb{E}[R_j] \leq 9\mathbb{E}[R_i], \forall i < j$.

Proof. Let I be the indicator vector of availability of advertisers right after step $i-1$. Given I, if an advertiser has zero probability to be matched to a query at step i, he has zero probability to be matched to a query at step j. Given I, if he has non-zero probability to be matched at step i, then the probability is at least $1/3n$; on the other hand, with probability at most $3/n$, he is matched at step j. From the discussion above, we have $\mathbb{E}[R_j|I] \leq 9\mathbb{E}[R_j|I]$. By taking expectation over I, we have our lemma. □

Lemma 13. $\forall (1-\epsilon)n < m \leq n, ALG_m \geq (1-9\epsilon)ALG_n$; $\forall n \leq m < (1+\epsilon)n, ALG_m \geq ALG_n$.

Proof. $\forall n \leq m < (1+\epsilon)n$, an instance τ_n of n arrivals can be generated in the following way: generate an instance τ_m of m arrivals first, and then remove the last $m-n$ arrivals. Because of the greediness of the algorithm, $ALG_m \geq ALG_n$.

$\forall (1-\epsilon)n < m \leq n$, $ALG_m = \sum_{i=1}^{m} \mathbb{E}[R_i]$ and $ALG_n = \sum_{i=1}^{n} \mathbb{E}[R_i]$. From the above lema, $ALG_m \geq \frac{m}{m+9(n-m)} ALG_n \geq (1-9\epsilon) ALG_n$. □

As a consequence:

Corollary 4. $ALG \geq (1-9\epsilon) ALG_n$.

Because of the assumption of c-competitiveness, $ALG_n \geq c \cdot OPT_n$. Therefore Corollaries 2 and 4 imply that $ALG \geq (1-10\epsilon) c \cdot OPT$.

6.1.3 Remarks

The only property of the Poisson distributed random variables we have used is that they concentrate around their means. Hence, if the number of arriving queries are different random variables also concentrating around their means, the results in this section would still apply.

7 Concluding remarks

In this paper, we have proposed new algorithms for online stochastic matching problems which led to improved competitive ratios under either integral or general arrival rates. We have also showed that our techniques can be applied to other related problems. In particular we have showed that one can do better than $1-1/e$ for the online vertex-weighted bipartite matching problem under the stochastic i.i.d. model with integral arrival rate. Finally we have showed the validity of all our results under a Poisson arrival model, removing the need to assume that the total number of arrivals is fixed and known in advance, as is required for the analysis of the stochastic i.i.d. models.

Some questions remain open. Gaps between 0.706 and 0.823 for problems with general arrival rates, and between 0.729 and 0.86 for problems with integral arrivals rates are yet to be closed. Note that for unweighted problems with integral rates, the bottleneck of the analysis is a 2-by-2 complete bipartite graph. The bottleneck remains even if one is using $b_1 \neq 2/3$ in the constraints $f_e \leq b_1$, because the 2-by-2 complete bipartite graphs could be part of feasible solutions. One possible approach would be to add another set of constraints $f_{a,i_1} + f_{a,i_2} \leq b_2$ with $b_2 \leq 1 - e^{-2}$. Since the offline optimal solution $\mathbf{f}^*$ satisfies $f^*_{a,i_1} + f^*_{a,i_2} \leq 1 - e^{-2}$, it would be feasible to the new LP. The techniques used in this paper may then be applied to the new LP in order to derive a better competitive ratio. The same idea could also be applied to weighted problems with integral rates.

Acknowledgements

We thank an anonymous referee for providing thoughtful and detailed comments which improved the quality and clarity of our paper.

References

[1] G. Aggarwal, G. Goel, C. Karande, and A. Mehta. Online vertex-weighted bipartite matching and single-bid budgeted allocations. In *SODA '11: Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1253–1264, 2011.

[2] S. Agrawal, Z. Wang, and Y. Ye. A dynamic near-optimal algorithm for online linear programming. Working paper, Stanford University, 2009.

[3] B. Bahmani and M. Kapralov. Improved bounds for online stochastic matching. In *ESA '10: Proceedings of the 22nd Annual European Symposium on Algorithms*, pages 170–181, 2010.

[4] N. Buchbinder, K. Jain, and S. Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *ESA '07: Proceedings of the 15th Annual European Symposium on Algorithms*, pages 253–264, 2007.

[5] N. R. Devanur and T. P. Hayes. The adwords problem: Online keyword matching with budgeted bidders under random permutations. In *EC '09: Proceedings of the 10th ACM Conference on Electronic Commerce*, pages 71–78, 2009.

[6] J. Feldman, A. Mehta, V. S. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *FOCS '09: Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 117–126, 2009.

[7] G. Goel and A. Mehta. Online budgeted matching in random input models with applications to adwords. In *SODA '08: Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 982–991, 2008.

[8] B. Haeupler, V. Mirrokni, and M. Zadimoghaddam. Online stochastic weighted matching: Improved approximation algorithms. In *WINE '11: Workshop of Network and Internet Economics*, pages 170–181, 2011.

[9] B. Kalyanasundaram and K. R. Pruhs. An optimal deterministic algorithm for online b-matching. *Theoretical Computer Science*, 233(1-2):319–325, 2000.

[10] C. Karande, A. Mehta, and P. Tripathi. Online bipartite matching with unknown distributions. In *STOC '11: Proceedings of the 43nd Annual ACM Symposium on Theory of computing*, pages 587–596, 2011.

[11] R. M. Karp, U. V. Vazirani, and V. V. Varirani. An optimal algorithm for online bipartite matching. In *STOC '90: Proceedings of the 22nd Annual ACM Symposium on Theory of computing*, pages 352–358, 1990.

[12] M. Mahdian and Q. Yan. Online bipartite matching with random arrivals: An approach based on strongly factor-revealing lps. In *STOC '11: Proceedings of the 43nd Annual ACM Symposium on Theory of computing*, pages 597–606, 2011.

[13] V. Manshadi, S. Oveis Gharan, and A. Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012.

[14] C. McDiarmid. On the method of bounded differences. *Surveys in Combinatorics, London Math. Soc. Lect. Note Series 141*, pages 148–188, 1989.

[15] A. Mehta, A. Saberi, U. Vazirani, and V. Varirani. Adwords and generalized online matching. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 264–273, 2005.

[16] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized on-line matching. *Journal of the ACM*, 54(5):Article 22 (19 pages), 2007.

[17] V. Mirrokni, S. Oveis Gharan, and M. Zadimoghaddam. Simultaneous approximations for adversarial and stochastic online budgeted allocation. In *SODA '12: Proceedings of the 23nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1690–1701, 2012.

A Complete Proofs of Lemma 5, Lemma 6, and Lemma 7

Lemma 5. *$\forall a \in A$, let N_a be the set of advertisers who are at an edge-distance no more than 4 from a in G_f. Then, there exists $\mu_{a,a'} \in [0,1]$ for all $a' \in N_a$, such that*

$$\sum_{a' \in N_a} \mu_{a,a'} p_{a'} \geq (1 - 2e^{-2}) \sum_{a' \in N_a} \mu_{a,a'} f_{a'}.$$

Proof. For advertiser a with $f_a = 1/3$, $p_a \geq \mathbb{P}(B_a) = 1 - e^{-1/3} \geq 0.850 f_a$. For advertiser a with $f_a = 2/3$, $p_a \geq \mathbb{P}(B_a) = 1 - e^{-2/3} \geq (1 - 2e^{-2}) f_a$. Thus, we only need to prove the lemma for a with $f_a = 1$.

Before doing so, let us first find probabilities of events B_a and G_a^i exactly.

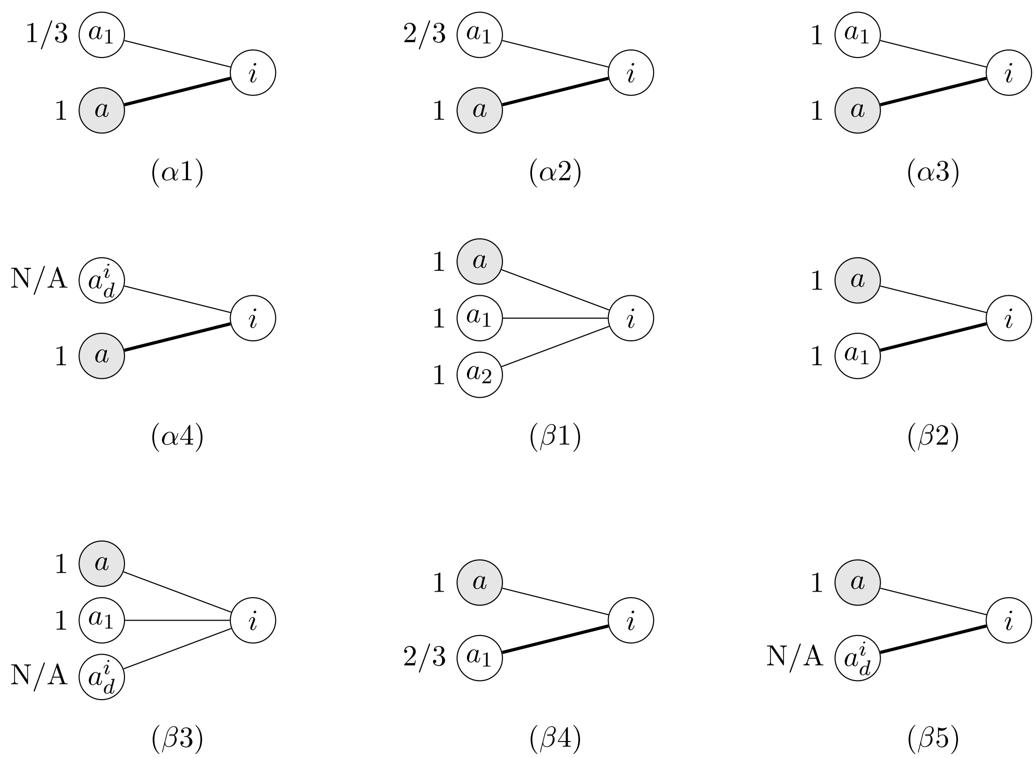

Figure 5: Possible configurations of i's neighborhood in the graph. Thin edges carry 1/3 flow, and thick edges carry 2/3 flow. The number next to advertiser a indicates f_a.

α) $f_{a,i} = 2/3$.

If i has 2 neighbors a and a_1, then $f_{a_1,i} = 1/3$:

α1. $f_{a_1} = 1/3$.

$$\begin{aligned} \mathbb{P}(G_a^i) &\geq \sum_{j=1}^{n} \tfrac{1}{3n} e^{-\frac{j}{3n}} \left(1 - e^{-\frac{n-j}{3n}}\right) \\ &\approx 1 - \tfrac{4}{3} e^{-\frac{1}{3}} \triangleq p_1 (\geq 0.044). \end{aligned}$$

α2. $f_{a_1} = 2/3$.

$$\begin{aligned}\mathbb{P}(G_a^i) &\geq \sum_{i=1}^{n} \frac{2}{3n} e^{-\frac{2j}{3n}} \left(1 - e^{-\frac{n-j}{3n}}\right) \\ &= 1 - e^{-\frac{2}{3}} - e^{-\frac{1}{3}} \cdot 2(1 - e^{-\frac{1}{3}}) \triangleq p_2 (\geq 0.080).\end{aligned}$$

α3. $f_{a_1} = 1$.

$$\begin{aligned}\mathbb{P}(G_a^i) &\geq \sum_{j=1}^{n} \frac{1}{n} e^{-\frac{j}{n}} \left(1 - e^{-\frac{n-j}{3n}}\right) \\ &= 1 - e^{-1} - e^{-\frac{1}{3}} \cdot \frac{3}{2}(1 - e^{-\frac{2}{3}}) \triangleq p_3 (\geq 0.109).\end{aligned}$$

If i has only one neighbor:

α4.

$$\mathbb{P}(G_a^i) \geq 1 - e^{-\frac{1}{3}} \triangleq p_4 (\geq 0.283).$$

β) $f_{a,i} = 1/3$.

If i_1 has 3 neighbors a, a_1, and a_2:

β1. We have $f_{a_1} = f_{a_2} = 1$; otherwise, we can find another optimal solution to LP (2) with less non-zero flow edges. Therefore,

$$\begin{aligned}\mathbb{P}(G_a^i) &\geq \sum_{k>j} \frac{2}{n} e^{-\frac{2j}{n}} \cdot \frac{4}{3n} e^{-\frac{4(k-j)}{3n}} \cdot (1 - \frac{7}{8} e^{-\frac{2(n-k)}{3n}}) \\ &= 1 - e^{-2} - \frac{21}{8} e^{-\frac{2}{3}}(1 - e^{-\frac{4}{3}}) + \frac{9}{4} e^{-\frac{4}{3}}(1 - e^{-\frac{2}{3}}) \triangleq p_5 (\geq 0.160).\end{aligned}$$

If i has 2 neighbors a and a_1. Note that $f_{a_1,i} = 1/3$ and $f_{a_1} < 1$ cannot happen together; otherwise, $\mathbf{f}$ cannot be a maximum flow:

β2. $f_{a_1,i} = 2/3$ and $f_{a_1} = 1$.

$$\begin{aligned}\mathbb{P}(G_a^i) &\geq \sum_{j=1}^{n} \frac{1}{n} e^{-\frac{j}{n}} \left(1 - e^{-\frac{2(n-j)}{3n}}\right) \\ &= 1 - e^{-1} - e^{-\frac{2}{3}} \cdot 3(1 - e^{-\frac{1}{3}}) \triangleq p_6 (\geq 0.195).\end{aligned}$$

β3. $f_{a_1,i} = 1/3$ and $f_{a_1} = 1$.

$$\begin{aligned}\mathbb{P}(G_a^i) &\geq \sum_{j=1}^{n} \frac{4}{3n} e^{-\frac{4j}{3n}} (1 - \frac{7}{8} e^{-\frac{2(n-j)}{3n}}) \\ &= 1 - e^{-\frac{4}{3}} - \frac{7}{4} e^{-\frac{2}{3}}(1 - e^{-\frac{2}{3}}) \triangleq p_7 (\geq 0.299).\end{aligned}$$

β4. $f_{a_1,i} = 2/3$ and $f_{a_1} = 2/3$.

$$\begin{aligned}\mathbb{P}(G_a^i) &\geq \sum_{j=1}^{n} \frac{2}{3n} e^{-\frac{2j}{3n}} \left(1 - e^{-\frac{2(n-j)}{3n}}\right) \\ &= 1 - \frac{5}{3} e^{-\frac{2}{3}} \triangleq p_8 (\geq 0.144).\end{aligned}$$

If i has only one neighbor:

β5.

$$\mathbb{P}(G_a^i) \geq 1 - e^{-\frac{2}{3}} \triangleq p_9 (\geq 0.486).$$

We say that "i is in $\alpha 1$ with respect to a" if (a,i) has the neighborhood structure shown in $\alpha 1$ in Figure 5. The same for $\alpha 2, \alpha 3, \alpha 4, \beta 1, \ldots, \beta 5$. "With respect to a" will be omitted unless otherwise specified. We are now ready to compute lower bounds on p_a when $f_a = 1$. We have two cases:

Case 1: a is contained in a cycle of length 4 in G_f. Let $[a_1(=a), i_1, a_2, i_2]$ be the cycle. According to the choice of the offline solution (see Section 3.1.1), $\sum_{j,k} f_{a_j,i_k} = 2$ as showed in Figure 6. Let N be the number of impressions of type i_1 or i_2. Then,

$$p_{a_1} + p_{a_2} = \mathbb{P}(N = 1) + 2\mathbb{P}(N \geq 2) = 2 - 4e^{-2} = (1 - 2e^{-2})(f_{a_1} + f_{a_2}) \approx 0.729(f_{a_1} + f_{a_2}).$$

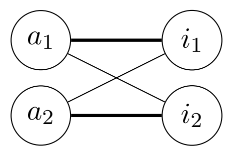

Figure 6: Cycle of length 4. Thin edges carry 1/3 flow; and thick edges carry 2/3 flow.

Case 2: a is not contained in a cycle of length 4. Then it has either three or two neighbors:

1) a has three neighbors i_1, i_2, and i_3, then

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2} \cup G_a^{i_3}) \\ &= 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2}))(1 - \mathbb{P}(G_a^{i_3})) \\ &\geq 1 - e^{-1}(1 - p_8)^3 \geq 0.769 f_a. \end{aligned}$$

Please note that the second equality is due to Corollary 1, which says that four or less certificate events are asymptotically independent if their supporting sets do not intersect. We will also use this asymptotic independence property repeatedly in the rest of the proof.

2) a has two neighbors i_1 and i_2. Without loss of generality, let us assume that $f_{a,i_1} = 1/3$ and $f_{a,i_2} = 2/3$:

2a. i_1 is in case $\beta 3$ or $\beta 5$.

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1})) \\ &\geq 1 - e^{-1}(1 - p_7) \geq 0.742 f_a. \end{aligned}$$

2b. i_1 is in case $\beta 4$. Let a_1 be the other neighbor of i_1.

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1})) \\ &\geq 1 - e^{-1}(1 - p_8). \end{aligned}$$

Similarly, we can compute

$$\begin{aligned} p_{a_1} &\geq \mathbb{P}(B_{a_1} \cup G_{a_1}^{i_1}) \\ &\geq 1 - e^{-\frac{2}{3}} + e^{-\frac{2}{3}} \sum_j \frac{1}{n} e^{-\frac{j}{n}} \left(1 - e^{-\frac{n-j}{3n}}\right) \\ &= 1 - e^{-\frac{2}{3}} + e^{-\frac{2}{3}} \left(1 - e^{-1} - e^{-\frac{1}{3}} \cdot \frac{3}{2}(1 - e^{-\frac{2}{3}})\right) \triangleq p_{10} (\geq 0.542). \end{aligned}$$

Since $f_a = 1, f_{a_1} = 2/3$, we have

$$p_a + p_{a_1} \geq 1 - e^{-1}(1 - p_8) + p_{10} \geq 0.736(f_a + f_{a_1}).$$

2c. i_1 is in case $\beta 2$.

i. i_2 is in case $\alpha 1$. Let a_1 be the other neighbor of i_2. Since

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_1)(1 - p_6) \end{aligned}$$

and

$$p_{a_1} \geq \mathbb{P}(B_{a_1}) = 1 - e^{-\frac{1}{3}} = p_4,$$

and $f_a = 1, f_{a_1} = 1/3$, we have

$$p_a + p_{a_1} \geq 1 - e^{-1}(1 - p_1)(1 - p_6) + p_4 \geq 0.750(f_a + f_{a_1}).$$

ii. i_2 is in case $\alpha 2$. Let a_1 be the other neighbor of i_2.

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_2)(1 - p_6). \end{aligned}$$

Similarly, we can compute

$$\begin{aligned} p_{a_1} &\geq \mathbb{P}(B_{a_1} \cup G_{a_1}^{i_1}) \\ &\geq 1 - e^{-\frac{2}{3}} + e^{-\frac{2}{3}} \sum_j \tfrac{1}{n} e^{-\frac{j}{n}} \left(1 - e^{-\frac{n-j}{3n}}\right) \\ &= 1 - e^{-\frac{2}{3}} + e^{-\frac{2}{3}} \left(1 - e^{-1} - e^{-\frac{1}{3}} \cdot \tfrac{3}{2}(1 - e^{-\frac{2}{3}})\right) = p_{10}. \end{aligned}$$

Since $f_a = 1, f_{a_1} = 2/3$, we have

$$p_a + 0.5 p_{a_1} \geq 1 - e^{-1}(1 - p_2)(1 - p_6) + 0.5 p_{10} \geq 0.749(f_a + 0.5 f_{a_1}).$$

iii. i_2 is in case $\alpha 3$ or $\alpha 4$.

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_3)(1 - p_6) \geq 0.736 f_a. \end{aligned}$$

2d. i_1 is in case $\beta 1$ and i_2 is not in case $\alpha 3$.

i. i_2 is in case $\alpha 1$. Let a_1 be the other neighbor of i_2. Since

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_1)(1 - p_5) \end{aligned}$$

and

$$p_{a_1} \geq \mathbb{P}(B_{a_1}) = 1 - e^{-\frac{1}{3}} = p_4,$$

and $f_a = 1, f_{a_1} = 1/3$, we have

$$p_a + p_{a_1} \geq 1 - e^{-1}(1 - p_1)(1 - p_5) + p_4 \geq 0.741(f_a + f_{a_1}).$$

ii. i_2 is in case $\alpha 2$. Let a_1 be the other neighbor of i_2.

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_2)(1 - p_5). \end{aligned}$$

Similarly, we can compute

$$\begin{aligned} p_{a_1} &\geq \mathbb{P}(B_{a_1} \cup G_{a_1}^{i_1}) \\ &\geq 1 - e^{-\frac{2}{3}} + e^{-\frac{2}{3}} \sum_j \frac{1}{n} e^{-\frac{j}{n}} \left(1 - e^{-\frac{n-j}{3n}}\right) \\ &= 1 - e^{-\frac{2}{3}} + e^{-\frac{2}{3}} \left(1 - e^{-1} - e^{-\frac{1}{3}} \cdot \frac{3}{2}(1 - e^{-\frac{2}{3}})\right) = p_{10}. \end{aligned}$$

Since $f_a = 1, f_{a_1} = 2/3$, we have

$$p_a + 0.5p_{a_1} \geq 1 - e^{-1}(1 - p_2)(1 - p_5) + 0.5p_{10} \geq 0.740(f_a + 0.5f_{a_1}).$$

iii. i_2 is in case $\alpha 4$,

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_4)(1 - p_5) \geq 0.778f_a. \end{aligned}$$

2e. i_1 is in case $\beta 1$ and i_2 is in case $\alpha 3$, then i_2 has two neighbors. Let a_1 and a_2 be the other two neighbors of i_1, and a_3 be the other neighbor of i_2.

$$\begin{aligned} p_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - p_3)(1 - p_5). \end{aligned}$$

i. a_3 has three neighbors. According to the discussion in (1),

$$p_{a_3} \geq 1 - e^{-1}(1 - p_8)^3.$$

Since $f_a = f_{a_1} = 1$, we have

$$p_a + \frac{1}{3}p_{a_3} \geq 0.736(f_a + \frac{1}{3}f_{a_3}).$$

If a_3 has two neighbors. Let the other neighbor of a_3 is i_3.

ii. i_3 is in $\alpha 1$ with respect to a_3. Let the other neighbor of i_3 be a_4. According to the discussion in (2c-i),

$$p_{a_3} + p_{a_4} \geq 1 - e^{-1}(1 - p_1)(1 - p_6) + p_4.$$

Since $f_a = f_{a_3} = 1, f_{a_4} = 1/3$, we have

$$p_a + p_{a_3} + p_{a_4} \geq 0.739(f_a + f_{a_3} + f_{a_4}).$$

iii. i_3 is in $\alpha 2$ with respect to a_3. Let the other neighbor of i_3 be a_4. According to the discussion in (2c-ii),

$$p_{a_3} + 0.5p_{a_4} \geq 1 - e^{-1}(1 - p_2)(1 - p_6) + 0.5p_{10}.$$

Since $f_a = f_{a_3} = 1, f_{a_4} = 2/3$, we have

$$p_a + p_{a_3} + 0.5p_{a_4} \geq 0.738(f_a + f_{a_3} + 0.5f_{a_4}).$$

iv. i_3 is in $\alpha 3$ or $\alpha 4$ with respect to a_3. According to the discussion in (2c-iii),

$$p_{a_3} \geq 1 - e^{-1}(1 - p_3)(1 - p_6).$$

Since $f_a = f_{a_3} = 1$, we have

$$p_a + p_{a_3} \geq 0.730(f_a + f_{a_3}).$$

□

Lemma 6. $\exists \{\lambda_a \geq 0\}_{a \in A}$ *such that* $\sum_a \lambda_a \mu_{a,a'} = 1, \forall a'$.

Proof. Let us first obtain an expression of λ_a for various types of advertisers. Consider an advertiser a such that $f_a = 1$: if a corresponds to case (1) in the proof of Lemma 5, $\lambda_a = 1 - \#(\text{nodes in (2e) that are at distance 2 from } a)/3$; if a corresponds to case (2c) and there exists a node in (2e) that is at distance 2 from a, then $\lambda_a = 0$; otherwise, $\lambda_a = 1$. For all the other advertisers a such that $f_a < 1$, we have $\lambda_a = 1 - \sum_{a':f_{a'}=1} \mu_{a',a}\lambda_{a'}$. We can now verify that for all a, $\lambda_a \geq 0$ and $\sum_{a'} \lambda_{a'}\mu_{a',a} = 1$:

- If a is in case (1), let $N'_a = \{a' : a' \text{ is a 2-neighbor of } a \text{ and } a' \text{ is in (2e)}\}$. Because $|N'_a| \leq 3$, we have $\lambda_a \geq 0$. On the other hand, from the proof of Lemma 5, $\mu_{a',a} = 1/3$ if $a' \in N'_a$, and 0, otherwise. From the construction of λ above, $\lambda_{a'} = 1$ for $a' \in N'_a$ and $\lambda_a = 1 - |N'_a|/3$. Therefore, $\sum_{a'} \lambda_{a'}\mu_{a',a} = 1$.

- If a is in (2c) and it has a 2-neighbor a_1 who is in (2e), then from the proof of Lemma 5, $\mu_{a_1,a} = 1$ and $\mu_{a',a} = 0$ for all $a' \neq a$ or a_1. From the construction of λ above, $\lambda_a = 0$ and $\lambda_{a_1} = 1$. Therefore, $\sum_{a'} \lambda_{a'}\mu_{a',a} = 1$.

- For all a with $f_a = 1$ and not in the two cases above, $\mu_{a',a} = 0$ for all $a' \neq a$. Since $\lambda_a = 1$, we have $\sum_{a'} \lambda_{a'}\mu_{a',a} = 1$.

- For all a with $f_a < 1$, $\mu_{a',a} = 0$ for all $a' \neq a$ with $f_{a'} < 1$. Because of the construction of λ, $\sum_{a'} \lambda_{a'}\mu_{a',a} = 1$ is trivially true. We will show that $\lambda_a \geq 0$.
  - $f_a = 1/3$. We can show that there is at most one advertiser a' such that $\lambda_{a'}\mu_{a',a} > 0$. Therefore, $\lambda_a \geq 0$.
  - $f_a = 2/3$, and a has only 1 neighbor. We can show that there is at most one advertiser a' such that $\lambda_{a'}\mu_{a',a} > 0$. Therefore, $\lambda_a \geq 0$.
  - $f_a = 2/3$, and a has 2 neighbors. We can show that there is at most two advertisers a' such that $\lambda_{a'}\mu_{a',a} > 0$. Furthermore, for all a', $\mu_{a',a} \leq 1/2$. Therefore, $\lambda_a \geq 0$.

□

Lemma 7. $\tilde{p}_a \geq 0.725 f_a, \forall a \in A$.

Proof. As discussed in Section 3.1.1, the left case in Figure 1 is the only possible cycle in the resulting graph. Let N be the number of impressions of type i_1 or i_2. Then, $\tilde{p}_{a_1} + \tilde{p}_{a_2} = \mathbb{P}(N = 1) + 2\mathbb{P}(N \geq 2) = 2 - 4e^{-2}$. Because of the symmetry between a_1 and a_2, $\tilde{p}_{a_1} = \tilde{p}_{a_2} = 1 - 2e^{-2} = 0.729$.

From now on, we only need to consider advertisers a who are not part of cycles of length 4. Therefore, the supporting sets of their certificate events do not intersect, thus are asymptotically independent.

We first consider the case $f_a = 1$. We can show case by case that:

Claim 5. $\forall a$ *with* $f_a = 1$, $\tilde{p}_a \geq 0.7250 f_a$.

Proof. Let us first compute probabilities of certificate events:

α) $f_{a,i} = 2/3$.

If i has 2 neighbors a and a_1, then $f_{a_1,i} = 1/3$.

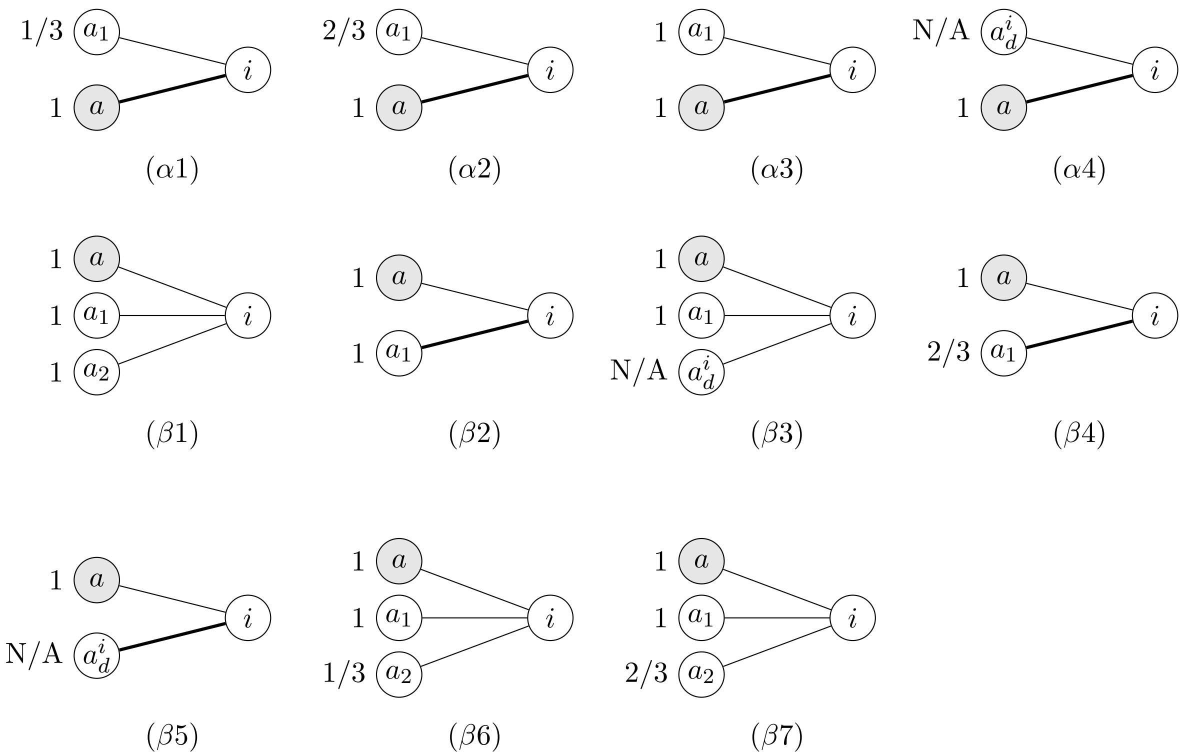

Figure 7: Possible configurations of i's neighborhood in the graph. Thin edges carry 1/3 flow, and thick edges carry 2/3 flow. The number next to advertiser a indicates f_a.

α1. $f_{a_1} = 1/3$. We use a Markov Chain approach to approximate $\mathbb{P}(G_a^i)$. The state space consists of three states: "a is full" (state 1), "a is empty and a_1 is full" (state 2), and "a is empty and a_1 is empty" (state 3). The transition probabilities are $p(3,2) = 0.1/n$, $p(3,3) = 1 - 0.1/n$, $p(2,2) = 1 - 0.1/n$, $p(2,1) = 0.1/n$, and $p(1,1) = 1$. The initial probability distribution is (0,0,1), i.e. both a and a_1 are empty. $\mathbb{P}(G_a^i)$ is the probability of state 1 after n time step. We use $n = 10^6$ here and for all other cases:

$$\mathbb{P}(G_a^i) \quad \geq \quad 0.0047 (\triangleq \tilde{p}_1)$$

The same idea can be used to compute the probability for all cases. The only difference is the size of state space, and the transition probability. Please note that we can also calculate $\mathbb{P}(G_a^i)$ exactly, as we did in the proof of Lemma 5.

α2. $f_{a_1} = 2/3$.

$$\mathbb{P}(G_a^i) \quad \geq \quad 0.0194 (\triangleq \tilde{p}_2)$$

α3. $f_{a_1} = 1$.

$$\mathbb{P}(G_a^i) \quad \geq \quad 0.1091 (\triangleq \tilde{p}_3)$$

If i has only one neighbor:

α4.

$$\mathbb{P}(G_a^i) \quad \geq \quad 0.2835 (\triangleq \tilde{p}_4)$$

β) $f_{a,i_1} = 1/3$.

If i has 3 neighbors a, a_1 and a_2. Then at least one of f_{a_1} or f_{a_2} is 1; otherwise, we can find another solution that has less non-zero flow edges and a better objective value.

β1. $f_{a_1} = f_{a_2} = 1$.

$$\mathbb{P}(G_a^i) \geq 0.1608(\triangleq \tilde{p}_5)$$

β6. $f_{a_1} = 1$ and $f_{a_2} = 1/3$.

$$\mathbb{P}(G_a^i) \geq 0.1396(\triangleq \tilde{p}_6)$$

β7. $f_{a_1} = 1$ and $f_{a_2} = 2/3$.

$$\mathbb{P}(G_a^i) \geq 0.1304(\triangleq \tilde{p}_7)$$

If i has 2 neighbors a and a_1. Note that $f_{a_1,i} = 1/3$ and $f_{a_1} < 1$ cannot happen together; otherwise, $\mathbf{f}$ cannot be a maximum flow.

β2. $f_{a_1,i} = 2/3$ and $f_{a_1} = 1$.

$$\mathbb{P}(G_a^i) \geq 0.1955(\triangleq \tilde{p}_8)$$

β3. $f_{a_1,i} = 1/3$ and $f_{a_1} = 1$.

$$\mathbb{P}(G_a^i) \geq 0.2992(\triangleq \tilde{p}_9)$$

β4. $f_{a_1,i} = 2/3$ and $f_{a_1} = 2/3$.

$$\mathbb{P}(G_a^i) \geq 0.1219(\triangleq \tilde{p}_{10})$$

If i_1 has only one neighbor:

β5.

$$\mathbb{P}(G_a^i) \geq 0.4866(\triangleq \tilde{p}_{11})$$

We are now ready to compute lower bounds on $\tilde{p}_a$ when $f_a = 1$:

1) a has 3 neighbors i_1, i_2, and i_3.

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2} \cup G_a^{i_3}) \\ &= 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2}))(1 - \mathbb{P}(G_a^{i_3})) \\ &\geq 1 - e^{-1}(1 - \tilde{p}_{10})^3 = 0.7509 f_a.\end{aligned}$$

2) a has 2 neighbors i_1 and i_2. Without loss of generality, let us assume that $f_{a,i_1} = 1/3$ and $f_{a,i_2} = 2/3$.

2a. i_2 is in case α4, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - \tilde{p}_4) = 0.7364 f_a.\end{aligned}$$

2b. i_2 is in case α1, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-0.9-1/3}(1 - \tilde{p}_1)(1 - \tilde{p}_{10}) = 0.7449 f_a.\end{aligned}$$

2c. i_2 is in case α2, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-0.85-1/3}(1 - \tilde{p}_2)(1 - \tilde{p}_{10}) = 0.7360 f_a.\end{aligned}$$

2d. i_2 is in case $\alpha 3$,

i. i_1 is in case $\beta 1, \beta 2, \beta 3$, or $\beta 5$, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-1}(1 - \tilde{p}_3)(1 - \tilde{p}_5) = 0.7250 f_a.\end{aligned}$$

ii. i_1 is in case $\beta 4$, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-2/3-0.4}(1 - \tilde{p}_3)(1 - \tilde{p}_{10}) = 0.7308 f_a.\end{aligned}$$

iii. i_1 is in case $\beta 6$, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-2/3-0.45}(1 - \tilde{p}_3)(1 - \tilde{p}_6) = 0.7491 f_a.\end{aligned}$$

iv. i_1 is in case $\beta 7$, then

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-2/3-0.425}(1 - \tilde{p}_3)(1 - \tilde{p}_7) = 0.7400 f_a.\end{aligned}$$

□

Claim 6. *$\forall a$ with $f_a = 1/3$, $\tilde{p}_a \geq 0.7622 f_a$.*

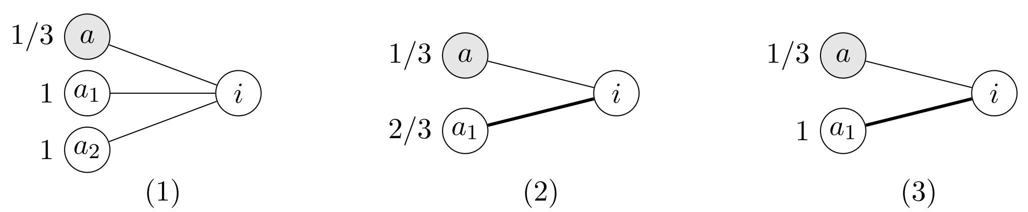

Figure 8: Possible configurations of i's neighborhood in the graph. Thin edges carry 1/3 flow, and thick edges carry 2/3 flow. The number next to advertiser a indicates f_a.

Proof. There are 3 possible local configurations:

1. The probability of certificate event $\mathbb{P}(G_a^i) \geq 0.1756$, thus,

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^i) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^i)) \\ &\geq 1 - e^{-0.1}(1 - 0.1756) = 0.2541 = 0.7622 f_a.\end{aligned}$$

2. $\tilde{p}_a \geq \mathbb{P}(B_a) = 1 - e^{-1/3} = 0.2835 = 0.8504 f_a$.

3. The probability of certificate event $\mathbb{P}(G_a^i) \geq 0.2275$, thus,

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^i) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^i)) \\ &\geq 1 - e^{-0.1}(1 - 0.2275) = 0.3010 = 0.9030 f_a.\end{aligned}$$

□

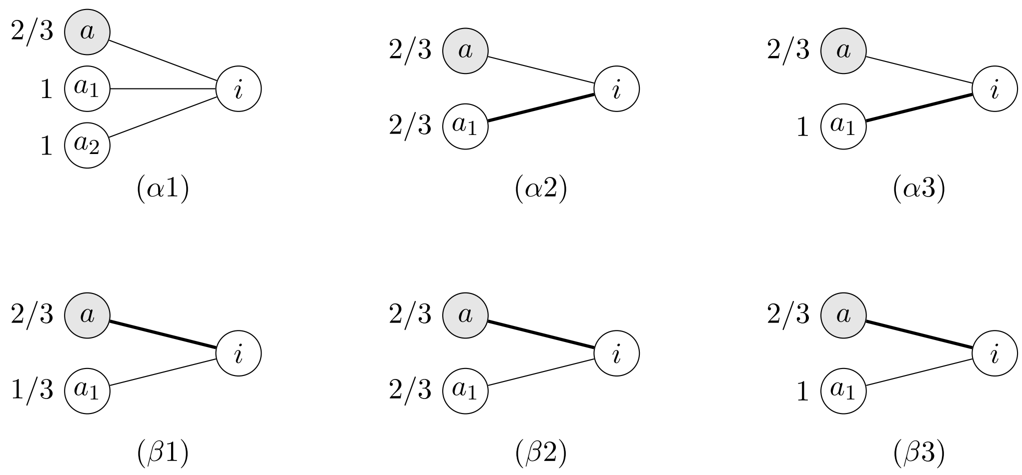

Figure 9: Possible configurations of i's neighborhood in the graph. Thin edges carry 1/3 flow, and thick edges carry 2/3 flow. The number next to advertiser a indicates f_a.

Claim 7. *$\forall a$ with $f_a = 2/3$, $\tilde{p}_a \geq 0.7299 f_a$.*

Proof. Let us first compute the probabilities of certificate events.

α). $f_{a,i} = 1/3$,

$\alpha 1$.

$$\mathbb{P}(G_a^i) \geq 0.1748$$

$\alpha 2$.

$$\mathbb{P}(G_a^i) \geq 0.1443$$

$\alpha 2$.

$$\mathbb{P}(G_a^i) \geq 0.1748$$

β). $f_{a,i} = 2/3$,

$\beta 1$.

$$\mathbb{P}(G_a^i) \geq 0$$

$\beta 2$.

$$\mathbb{P}(G_a^i) \geq 0$$

$\beta 3$.

$$\mathbb{P}(G_a^i) \geq 0.2016$$

We are now ready to bound $\tilde{p}_a$ when $f_a = 2/3$.

1). If a has only one neighbor i,

1a. i is in case $\beta 1$ or 2, then $\tilde{p}_a \geq \mathbb{P}(B_a) = 1 - e^{-2/3} = 0.4866 = 0.7299 f_a$.

1b. i is in case $\beta 3$, then,

$$\begin{aligned}\tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^i) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^i)) \\ &\geq 1 - e^{-0.6}(1 - 0.2016) = 0.5618 = 0.8427 f_a.\end{aligned}$$

2). If a has two neighbors i_1 and i_2,

2a. If neither of i_1 or i_2 is in case $\alpha 2$,

$$\begin{aligned} \tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-0.3}(1 - 0.1748)^2 = 0.4955 = 0.7433 f_a. \end{aligned}$$

2b. If at least one of i_1 or i_2 is in case $\alpha 2$,

$$\begin{aligned} \tilde{p}_a &\geq \mathbb{P}(B_a \cup G_a^{i_1} \cup G_a^{i_2}) = 1 - (1 - \mathbb{P}(B_a))(1 - \mathbb{P}(G_a^{i_1}))(1 - \mathbb{P}(G_a^{i_2})) \\ &\geq 1 - e^{-0.15-1/3}(1 - 0.1443)^2 = 0.5484 = 0.8226 f_a. \end{aligned}$$

□

In conclusion, combining the results of Claims 5, 6, and 7, we obtain the proof of Lemma 7.

□